



HAL
open science

L'adaptation à la sécheresse chez le pin maritime (*Pinus pinaster* Ait.): patrons de diversité et différenciation nucléotidiques de gènes candidats et variabilité de caractères phénotypiques

Emmanuelle Eveno

► **To cite this version:**

Emmanuelle Eveno. L'adaptation à la sécheresse chez le pin maritime (*Pinus pinaster* Ait.): patrons de diversité et différenciation nucléotidiques de gènes candidats et variabilité de caractères phénotypiques. Sciences du Vivant [q-bio]. Université des Sciences et Technologies (Bordeaux 1), 2008. Français. NNT: . tel-02824873

HAL Id: tel-02824873

<https://hal.inrae.fr/tel-02824873>

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.

THÈSE

Présentée à

L'UNIVERSITÉ BORDEAUX 1

ÉCOLE DOCTORALE : Sciences du Vivant, Géosciences, Sciences de l'Environnement

Par

Emmanuelle EVENO

Pour obtenir le grade de

DOCTEUR

SPÉCIALITÉ : Biologie Cellulaire et Moléculaire

**L'adaptation à la sécheresse chez le pin maritime (*Pinus pinaster* Ait.) :
 patrons de diversité et différenciation nucléotidiques de gènes candidats et
 variabilité de caractères phénotypiques.**

Soutenue le 1er Avril 2008

Devant la commission d'examen formée de :

M. Didier ALARD	Professeur, Université Bordeaux 1	Président
M. Frédéric AUSTERLITZ	Chargé de Recherche, Université Paris-Sud XI	Rapporteur
Mme. Magali SAN CRISTOBAL	Chargée de Recherche, INRA Toulouse	Rapporteur
M. Rémy PETIT	Directeur de Recherche, INRA Bordeaux I	Examineur
Mme. Dominique THIS	Maître de Conférence, Université Montpellier II	Examineur
Mme. Pauline GARNIER-GERE	Chargée de Recherche, INRA Bordeaux I	Co-directeur de thèse

☆☆ Remerciements ☆☆

Arrivée enfin au bout de cette longue aventure, l'heure est venue d'exprimer le plus simplement et le plus naturellement mes remerciements à toutes les personnes ayant contribué à ce que cette thèse soit finalisée.

Mes premières pensées vont d'abord à ma co-directrice de thèse, Pauline Garnier-Géré. Cela fait maintenant quelques années que tu m'as prise sous ton aile. Tu as été la première à me faire confiance lors de mon DEA et de ma thèse. Je retiendrai les moments d'échange et de cogitation intellectuelle que j'ai pu avoir avec toi. J'ai appris grâce à toi, que le monde de la recherche pouvait être très créatif, et que chaque idée avait sa place. Je suis loin d'avoir intégré toutes les lois de la statistique et de la génétique des populations, mais je sais maintenant qu'il ne faut pas sous-estimer le pouvoir des modèles statistiques et des échantillonnages. Sur le plan plus personnel, ton soutien moral et ton encouragement jusqu'au bout de cette thèse, m'aura permis de finaliser cette expérience. Mes discours sont courts mais tu sais combien je te suis reconnaissante de m'avoir permis de marquer un point final à cette thèse.

Je n'aurai pu être là sans mon deuxième co-directeur de thèse, Christophe Plomion. Merci pour m'avoir donné ma chance et m'avoir accueilli au sein de la Team « Maritime Pine » !

Je souhaite également remercier le grand chef Antoine Kremer pour m'avoir permis d'intégrer son laboratoire pendant ces 4 années.

Frédéric Austerlitz et Magali San Cristobal ont gentiment accepté d'être les rapporteurs de cette thèse. Merci pour leur lecture et l'intérêt porté à ce manuscrit. Je remercie également les examinateurs, Rémy Petit, Didier Alard et Dominique This. Merci aussi à Fred et Laurent pour leur relecture.

Cette thèse ayant fait parti d'un projet européen, les échanges et rencontres ont été nombreux. Je tiens à remercier tout particulièrement Santi pour ses précieux conseils en génétique des populations et sa contribution à l'article, ainsi que Ricardo pour son expertise en génétique quantitative et pour le temps qu'il a su m'accorder malgré un emploi du temps surchargé.

Cette thèse est basée sur un large jeu de données phénotypiques qui a pu être constitué grâce au travail de nombreuses personnes : d'abord tous les membres de l'Unité Expérimentale, notamment Fred Bernier, qui ont réalisé entre autre les mesures de hauteur et de biomasse. Je me souviendrai également de toutes les personnes volontaires pour désherber (à la main et à quatre pattes !!!) la parcelle expérimentale sous un soleil de plomb, ainsi que dépiotter, peser, mesurer ces trop nombreux pins maritimes ! Deux long mois de broyage et pesée d'aiguilles auraient également été beaucoup plus monotones sans l'aide précieuse de Nancy. Mimi, tu as aussi grandement contribué à la qualité des résultats phénotypiques de cette thèse. Tu as su rester calme lorsque j'ai envahi ton labo et ton étuve avec mes sacs remplis de pins. Notre prochain projet sera, je l'espère, la conquête du Vignemale avec Guy et Laurent. Pour la partie moléculaire, je remercie Delphine et Pierre pour m'avoir aidé à accumuler et analyser autant de séquences !

Il existe aussi un aparté dans ma vie de thésarde qui consista à organiser un colloque européen en Biologie Evolutive réunissant une centaine de thésards. Je dis un grand merci à tous mes collaborateurs venus des quatre coins de la France, avec qui j'ai partagé des moments inoubliables...

J'adresse également des remerciements plus personnels aux collègues mais surtout amis pierrotonnais. Par ordre d'apparition dans ma vie de stagiaire et thésarde, je citerai en premier Josquin Tibbits. Tu es celui qui m'a initié à la pipette et qui surtout a partagé mes repas durant les premiers mois de DEA à la « cantine-chambre-salle des fêtes » de Pierroton. J'ai apprécié les informations bibliographiques que tu as su me fournir même à l'autre bout du monde. Je n'oublierai pas non plus les deux David, au dimorphisme de taille. Hormis cette équipe d'anciens thésards, je ne peux oublier tous les autres dont Agnes et ses « David !!! » du fond du couloir, Patricia et Marta..

Les autres personnes que j'ai croisé tout au long de ma thèse sont nombreuses, et j'ai une pensée pour chacune d'elle: Muriel (nos bavardages plus que nos réels efforts sportifs lors des cours d'escalade), Hakim (sa vision des étoiles), Jeremy (aucun Jauzion à l'horizon !), Christophe, Philou, Erwan, Olivier, Cécile, Manon, Julie, Manuela, Florian, Eric, Tristan, Jérôme, Jean Charles, Delphine, Camille..., tous les permanents du labo..., ainsi que tous les thésards mexicains, chiliens, argentins, japonais, portugais...

Je retiendrai notamment l'accueil du gars du Nord : Laurent ! Merci pour ton hospitalité et ton grand soutien durant ces derniers temps. Les barbecues et surtout les randonnées resteront parmi mes meilleurs souvenirs. Merci aussi à Jérôme et Michela, les inséparables ! Vous aurez été là pour moi dans les moments les plus difficiles. Note personnelle pour Jérôme : aurais-je un jour le droit de savoir ce qu'est un petit géranium ????

Enfin, je ne pourrai m'arrêter là sans évoquer mes deux collègues et amis qui ont partagé avec moi la salle archive, ma deuxième maison ! Céline, j'ai le souvenir d'une épopée italienne avec deux têtes de cochon, et ton flegme à tout épreuve. Jorge, je retiendrai les franches rigolades que l'on a pu avoir, notamment au cinéma !

Enfin, je n'aurai pu réaliser cette thèse, sans le soutien indéclinable quoiqu'il arrive de mes parents. Je pense aussi à ma sœur et au petit Paul qui a eu longtemps le privilège d'avoir sa photo dans mon bureau, et a pu vérifier avec ses tout petits yeux bleus que mes alignements de séquences étaient corrects...

Mon dernier merci revient à toi Guillaume. Merci de m'avoir supporté dans mes moments de doute, et de m'avoir toujours poussé à avancer. Maintenant, une aventure d'une autre dimension nous attend, et je suis sûre qu'elle sera longue et belle. Moja bulla tlen !!! Moja połowa !!!

Table des matières

Introduction	1
Chapitre I.	
Détection de signatures de sélection au niveau moléculaire : méthodes et applications chez les arbres forestiers	
A. Les différentes formes de sélection naturelle et leurs conséquences sur la diversité et la différenciation génétiques	13
A.1. Différentes formes de sélection naturelle au niveau phénotypique	15
A.2. Différentes formes de sélection naturelle au niveau moléculaire et conséquences sur la diversité et différenciation moléculaires	15
A.2.1. <i>Sélection directionnelle</i>	16
A.2.2. <i>Sélection purifiante</i>	19
A.2.3. <i>Sélection balancée</i>	21
B. Comment détecter de la sélection naturelle au niveau moléculaire ?	22
B.1. Panorama synthétique des tests de neutralité	23
B.2. Les approches pan-génomiques	26
B.2.1. <i>L'approche pan-génomique basée sur la différenciation entre populations</i>	26
B.2.2. <i>Autres approches pan-génomiques basées sur les patrons de diversité nucléotidique et l'étendue de déséquilibre de liaison intra-population</i>	30
B.2.3. <i>Limites des approches pan-génomiques</i>	31
C. Application des tests de neutralité chez les espèces d'arbres forestiers	33
C.1. L'approche gènes-candidats	33
C.1.1 <i>Faible étendue des DL</i>	33
C.1.2 <i>Large taille du génome des conifères</i>	34
C.1.3 <i>Ressources en gènes candidats</i>	34
C.2. Tests d'écarts à la neutralité pour des approches gènes-candidats et pan-génomiques	35
Chapitre II.	
Matériel végétal, gènes candidats et marqueurs microsatellites	
A. Populations échantillonnées et données climatiques	41
B. Gènes candidats	44
B.1. Stratégie pour le choix des gènes candidats	44
B.2. Formation de la paroi cellulaire	47
B.3. Lignification	51
B.4. Métabolisme des carbohydrates	52
B.5. Protéines LEA	52
B.6. Un facteur de régulation de l'ABA	54
B.7. Des gènes aux fonctions inconnues, dont l'expression est induite par un stress hydrique	54
B.8. Obtention des séquences de gènes candidats	55
B.8.1. <i>Extraction d'ADN à partir de tissu haploïde</i>	55
B.8.2. <i>Amplification, séquençage des fragments de gènes candidats et détection des polymorphismes</i>	55

C. Choix de marqueurs microsatellites et définition d'une référence « neutre »	56
C.1. Choix des marqueurs microsatellites et génotypage	56
C.2. Estimation de la diversité moléculaire des SSRs et identification de groupes homogènes de populations	56
C.3. Résultats préliminaires sur la diversité moléculaire des SSRs et définition d'une référence neutre	58

Chapitre III.

Méthodologie d'étude des patrons de diversité et différenciation nucléotidiques

A. Echantillonnage	65
B. Estimation des diversités nucléotidique et haplotypique	66
C. Estimation de l'importance des évènements de recombinaison	68
C.1. Nombre minimum d'évènements de recombinaison	68
C.2. Estimation du taux de recombinaison d'une population	69
D. Estimation des déséquilibres de liaison entre sites	69
E. Estimation de la différenciation génétique entre populations	70
E.1. Approche basée sur les fréquences alléliques	70
E.2. Approche basée sur les distances entre allèles	71
<i>E.2.1. L'approche d'Excoffier, Smouse et Quattro (1992)</i>	71
<i>E.2.2. L'approche de Slatkin</i>	72
E.3. Estimation des paramètres	72
F. Détection d' « outliers » basée sur la différenciation génétique	72
F.1. Approche de Beaumont et Nichols implémenté dans « Fdist2 »	73
<i>F.1.1. Principe</i>	73
<i>F.1.2. Robustesse de l'approche à plusieurs modèles démographiques</i>	73
F.2. Extension de l'approche à des marqueurs bi-alléliques (méthode « FstSNP »)	76
F.3. Quelques résultats de simulations avec « Fdist2 » et « FstSNP »	79
<i>F.3.1. Effet du modèle de mutation</i>	81
<i>F.3.2. Effet de la taille d'échantillon</i>	81
F.4. Détection d' « outliers » par l'approche Bayésienne de Beaumont et Balding (2004)	83
<i>F.4.1. Particularité de l'approche bayésienne</i>	83
<i>F.4.2. Principe de l'approche de Beaumont et Balding (2004)</i>	84
<i>F.4.2.1. Modèle généalogique sous-jacent</i>	84
<i>F.4.2.2. Comptage des allèles et expression du paramètre F_{ij}</i>	84
<i>F.4.2.3. Régression du paramètre F_{ij}</i>	86
<i>F.4.2.4. Choix de la distribution a priori des paramètres α_i, β_j et γ_{ij}</i>	86
<i>F.4.2.5. Implémentation de la chaîne MCMC</i>	86
<i>F.4.2.6. Interprétation des distributions a posteriori des paramètres du modèle</i>	87
<i>F.4.3. Applications préliminaires de la méthode « BayesFst »</i>	88
<i>F.4.3.1. Déclaration du facteur d'interaction</i>	88
<i>F.4.3.2. Impact du facteur de corrélation</i>	89
<i>F.4.3.3. Regroupement des haplotypes</i>	89
G. Tests de neutralité au sein de la métapopulation atlantique	94

Chapitre IV.

Patrons de diversité et différenciation moléculaires : résultats et discussion

A. Synthèse des résultats sur l'approche de détection d' « outliers »	99
B. Résultats sur les patrons de diversité nucléotidique et tests de neutralité au sein de la métapopulation atlantique	104
B.1. Polymorphismes détectés	104
B.2. Variation des patrons de diversités nucléotidique et haplotypique entre gènes	104
B.3. Variation du déséquilibre de liaison entre sites polymorphes intra-gène	105
B.4. Des écarts à la neutralité révélant des profils hétérogènes selon les gènes	112
C. Discussion	116
C.1. Observation d'une faible diversité nucléotidique silencieuse chez le pin maritime ?	116
C.2. Pourtant, une diversité plus élevée a été observée pour d'autres marqueurs...	116
C.3. Facteurs susceptibles d'affecter les niveaux de diversité nucléotidique	120
C.3.1. <i>Le taux de mutation</i>	120
C.3.2. <i>Démographie des populations</i>	121
C.3.3. <i>Effets possibles de la sélection</i>	122
C.4. Quels scénarios évolutifs au sein des populations naturelles de pin maritime ?	123
C.4.1. <i>Signatures de sélection diversifiante entre populations, et directionnelle au sein de la métapopulation atlantique</i>	123
C.4.2. <i>Signatures de sélection homogénéisante sur les 10 populations, et balancée au sein de la métapopulation atlantique ?</i>	130
C.4.3. <i>Conclusion</i>	133

Chapitre V.

Variabilité génétique de la composition isotopique en carbone et relations avec hauteur et biomasse dans un test de provenances de pin maritime

A. Introduction	137
A.1. Résistance à la sécheresse et relation entre production de biomasse et consommation en eau	137
A.2. Calcul de l'EUE par l'approche isotopique	138
A.3. Utilisation du caractère EUE_{int} dans des programmes de sélection	141
A.4. Question de recherches	141
A.5. Objectifs	143
B. Matériels et méthodes	143
B.1. Description du dispositif expérimental	143
B.2. Modalités de mesure et stratégie d'échantillonnage de la composition isotopique du carbone	145
B.2.1. <i>Modalités de mesure</i>	145
B.2.2. <i>Stratégie d'échantillonnage</i>	147
B.3. Mesures des caractères de hauteur et de biomasse	150
B.4. Analyses statistiques	152
B.4.1. <i>Modèle d'analyse de variance et estimation des composantes de variance</i>	152
B.4.2. <i>Estimation du coefficient de variation phénotypique, de la variance génétique additive, et de l'héritabilité au sens strict</i>	153
B.4.3. <i>Estimation des corrélations génétiques et environnementales entre caractères quantitatifs</i>	153

<i>B.4.4. Etude de la structuration entre populations pour les caractères quantitatifs</i>	156
<i>B.4.5. Analyse intra population</i>	156
C. Problèmes méthodologiques et résultats préliminaires	156
C.1. Problème de l'estimation des variances entre familles pour le $\delta^{13}C$	156
C.2. Biais engendré par les analyses isotopiques	162
D. Résultats	164
D.1. Une variation phénotypique répartie différemment selon les caractères entre effets génétiques et environnementaux	164
D.2. Comparaison des moyennes phénotypiques entre populations	166
<i>D.2.1. Des populations de l'extrémité de l'aire de distribution divergentes pour l'efficience d'utilisation en eau et la hauteur</i>	166
<i>D.2.2. Des populations espagnoles moins efficaces que les populations françaises mais à biomasse similaire pour une hauteur plus faible</i>	169
D.3. Héritabilités	171
<i>D.3.1. Des caractères à héritabilités globalement élevées toutes populations confondues</i>	171
<i>D.3.2. Des valeurs d'héritabilité dans chaque population principalement liées à l'importance des variances génétiques additives</i>	171
<i>D.3.3. Variation de l'héritabilité entre populations</i>	173
D.4. Corrélations entre caractères	177
D.4.1. Des corrélations génétiques élevées entre caractères de hauteur et biomasse	177
D.4.2. Corrélations génétiques entre efficience d'utilisation en eau et hauteur ou biomasse	177
D.4.3. Des corrélations résiduelles positives et élevées entre caractères de hauteur et biomasse mais plus faibles et variables avec le $\delta^{13}C$	183
D.5. Résultats sur la différenciation entre populations	185
<i>D.5.1. Une différenciation entre populations représentative des grands groupes géographiques</i>	185
<i>D.5.2. Maintien d'une différenciation élevée entre populations espagnoles</i>	186
E. Discussion	186
E.1. Différents mécanismes de résistance à la sécheresse entre provenances de milieux humides et de milieux chauds ou secs	187
<i>E.1.1. Stratégie basée sur le contrôle stomatal</i>	187
<i>E.1.2. Stratégie basée sur le développement racinaire</i>	189
<i>E.1.3. Performances des populations locales versus non-locales</i>	190
E.2. Le cas particulier des populations espagnoles	190
E.3. Différentes intensités de corrélation génétique et physiologique entre populations	191
E.4. Conservation d'un potentiel adaptatif (variabilité génétique) plus élevé au sein de quelques populations (tel qu'il a été exprimé dans le milieu de test)	192
E.5. Conclusion et perspectives	194

Chapitre VI.

Recherche d'associations entre diversité de gènes candidats et variabilité génétique de caractères liés à la résistance au stress hydrique

A. Introduction	197
B. Matériels et méthodes	199
B.1. Choix des marqueurs SNPs	199
B.2. Méthodes et modèles d'analyses de la variation	200
<i>B.2.1. Corrélations entre fréquences alléliques et moyennes phénotypiques</i>	200
<i>B.2.2. Modèles d'analyse de la variation</i>	200
C. Résultats	206

C.1. Corrélations entre fréquences alléliques aux SNPs et moyennes phénotypiques	206
C.2 Résultats des tests d'associations entre SNPs et variation phénotypique	206
C.2.1. Effets globaux des SNPs	206
C.2.2. Effets des SNPs hiérarchisés à l'effet population	211
D. Conclusion	212
Chapitre VII.	
Discussion générale	
A. Le pin maritime face aux changements climatiques ?	215
A.1 Variation des stratégies d'adaptation à un déficit en eau	217
A.2 Niveaux de plasticité phénotypique au sein des populations ?	218
A.3 Prédiction des performances des populations dans un milieu plus aride ?	219
A.4 Perspectives pour l'amélioration du pin maritime et la conservation de ses ressources génétiques	219
B. Variation des niveaux et des patrons de diversité moléculaires	220
B.1 Inférences évolutives : scénarios de sélection ou effets de l'histoire démographique des populations?	220
B.2 Détection de signatures de sélection : quelles méthodes, quelles données et quels échantillonnages ?	222
B.2.1 Des méthodes complémentaires	222
B.2.2 Gènes candidats versus marqueurs anonymes et haplotypes versus SNPs ?	223
B.2.3. Perspectives d'amélioration des méthodes de détection d' « outliers »	223
B.2.4 Questions d'échantillonnage...	224
C. Associations entre variabilité phénotypique et diversité moléculaire	224
C.1. F_{ST} versus Q_{ST} chez le pin maritime ?	227
C.2. Bases génétiques de la variation pour la tolérance au stress hydrique en populations naturelles : utopie ou réalité proche ?	228
C.2.1 Premières estimations sur l'organisation des DL chez le pin maritime	229
C.2.2. Interprétation biologique de la fonction des gènes candidats	230
C.2.3 Importance des effets des SNPs ?	231
C.3. Quelles approches et quels échantillons pour les études d'associations chez des espèces structurées en populations?	232
C.3.1. Etudes d'association au sein de populations non structurée ou prise en compte de la structuration en populations ?	232
C.3.2. Fréquence minimum des allèles aux gènes candidats et taille des échantillons	232
Conclusion et perspectives	235
Références	241

Index des figures

<u>Figure 1.</u> Exemple de scénarios proposés par l'IPCC sur le changement des températures et précipitations au niveau mondial d'ici 2100.	2
<u>Figure 2.</u> Approche gènes-candidats adoptée au sein du projet TREESNIPS.	8
<u>Figure 3.</u> Illustration des différentes formes de sélection au niveau phénotypique.	14
<u>Figure 4.</u> Illustration des différentes formes de sélection au niveau moléculaire au sein d'une population.	18
<u>Figure 5.</u> Cas d'un balayage sélectif incomplet, où des allèles dérivés sont entraînés car ils sont associés à l'allèle avantageux après recombinaison.	20
<u>Figure 6.</u> Exemples d'arbres généalogiques sous un modèle neutre standard avec une taille de population constante, après une expansion de population, dans une population structurée en deux sous populations, dans une population après la fixation d'une mutation avantageuse, et dans une population dans laquelle 2 allèles sont maintenus par sélection balancée.	25
<u>Figure 7.</u> Localisation des 24 populations échantillonnées sur l'ensemble de l'aire de distribution naturelle du pin maritime.	42
<u>Figure 8.</u> Gènes induits par le stress et leur fonction possible dans la réponse à la tolérance au stress osmotique.	46
<u>Figure 9.</u> Alignements de plusieurs membres de gènes de la famille des GRPs.	49
<u>Figure 10.</u> Voie de biosynthèse des lignines.	50
<u>Figure 11.</u> Voies de biosynthèse des carbohydrates, dont ceux de l'amidon, du sucrose, et de l'inositol phosphate.	53
<u>Figure 12.</u> Deux « zones contacts », Ouest et Sud-Est utilisées dans le logiciel Structure.	59
<u>Figure 13.</u> Assignement des individus en utilisant le logiciel Structure dans un modèle de « zones en contact » avec 2 zones.	59
<u>Figure 14.</u> Dendrogrammes construits avec les algorithmes « Ward » et « UPGMA » à partir des 8 et 6 SSRs.	62
<u>Figure 15.</u> Echantillon de 10 populations aux conditions climatiques contrastées et réparties sur l'aire naturelle de distribution du pin maritime.	64
<u>Figure 16.</u> Modèles de migration en îles et « stepping stone ».	75

<u>Figure 17.</u> Nombre de génération pour atteindre 95% de la valeur de F_{ST} à l'équilibre migration-dérive en fonction de la taille de la population et du taux de migration.	75
<u>Figure 18.</u> Modèle démographique sous-jacent aux simulations de coalescence sous « Simcoal2 ».	78
<u>Figure 19.</u> Valeurs du F_{ST} moyen pour 100 points simulés en fonction du nombre de générations depuis la scission de la population ancestrale.	80
<u>Figure 20.</u> Effet du modèle de mutation sur la forme de l'enveloppe neutre.	82
<u>Figure 21.</u> Effet de la taille d'échantillon sur la forme de l'enveloppe neutre simulée avec la méthode « Fdist2 »	82
<u>Figure 22.</u> Effet de la taille d'échantillon sur la forme de l'enveloppe neutre simulée avec la méthode « FstSNP ».	82
<u>Figure 23.</u> Modèle "separation-of-timescales" utilisé dans l'approche Bayésienne de Beaumont et Balding.	85
<u>Figure 25.</u> Dendrogrammes basés sur les matrices de pourcentage de différences entre paires d'haplotypes pour <i>PR-AGP4</i> , <i>CCoAOMT</i> , <i>Glucan</i> , <i>dhn1</i> , <i>GRP3</i> , <i>dhn2</i> .	90
<u>Figure 26.</u> Dendrogrammes basés sur les matrices de pourcentage de différences entre paires d'haplotypes pour <i>erd3</i> , <i>lp3-1</i> , <i>lp3-3</i> , <i>pp2C</i> , <i>Rd21A</i> .	91
<u>Figure 27.</u> Distribution des valeurs de F_{ST} pour chaque gène, chaque SNP au sein des gènes, et chaque SSR, en fonction de leur hétérozygotie moyenne (H_S), entre les 10 populations contrastées et enveloppes neutres simulées sous « Fdist2 » et « FstSNP ».	101
<u>Figure 28.</u> Déséquilibres de liaison entre SNPs polymorphes en excluant ceux n <i>DL</i> complet, et en considérant l'ensemble des gamètes sur les 10 populations pour les 11 gènes.	103
<u>Figure 29.</u> Structures haplotypiques et représentations schématiques de 5 gènes candidats.	108
<u>Figure 30.</u> Diagramme de dispersion du <i>DL</i> (estimé par r^2) entre sites polymorphes non rares en fonction de la distance en paires de bases, au sein de la métapopulation Atlantique sur 5 gènes.	110
<u>Figure 31.</u> Matrice des <i>p</i> -valeurs des tests exacts de Fisher pour les déséquilibres de liaison entre sites polymorphes non-rares estimés pour 4 gènes.	111
<u>Figure 32.</u> Statistiques <i>D</i> de Tajima et <i>H</i> de Fay et Wu le long des gènes pour 5 gènes.	114
<u>Figure 33.</u> Représentation graphique des diversités nucléotidiques silencieuses par paires de bases en moyenne chez les conifères, les arbres angiospermes, les plantes annuelles, les humains et <i>Arabidopsis thaliana</i> .	117

<u>Figure 34.</u> Régression de la diversité nucléotidique silencieuse en fonction du taux de divergence silencieux sur l'ensemble des 10 populations et au sein de la métapopulation atlantique.	124
<u>Figure 35.</u> Matrices de F_{ST} par paires de populations pour 3 gènes 2 SNPs « outliers » positifs.	125
<u>Figure 36.</u> Mécanismes d'évitement, d'échappement ou de tolérance sous-jacents à la résistance à la sécheresse à un stress hydrique.	136
<u>Figure 37.</u> Comment le rapport entre les isotopes ^{13}C et ^{12}C du carbone ($\delta^{13}C$) est-il relié à l'efficacité d'utilisation de l'eau ?	140
<u>Figure 38.</u> Plan de l'essai expérimental réparti en 2 parcelles.	146
<u>Figure 39.</u> Positions abaxiales et adaxiales des aiguilles le long d'une branche.	146
<u>Figure 40.</u> Schéma des prélèvements d'aiguilles pour la mesure du $\delta^{13}C$, et des mesures de diamètre et de hauteur.	149
<u>Figure 41.</u> Illustration des corrélations génétiques et des corrélations environnementales entre caractères.	155
<u>Figure 42.</u> Analyse en composantes principales sur les moyennes familiales, basée sur 7 caractères quantitatifs.	158
<u>Figure 43.</u> Evolution des valeurs d'héritabilité au sens strict de 3 caractères en fonction du nombre de répétitions.	159
<u>Figure 44.</u> Analyse en composantes principales à partir des moyennes des populations, sur 3 caractères.	161
<u>Figure 45.</u> Quelques graphiques de régression des moyennes familiales de $\delta^{13}C$ en fonction de <i>HT03</i> pour chaque lot d'analyse.	163
<u>Figure 46.</u> Analyse en composantes principales à partir des moyennes des populations et groupes de populations, sur les 3 caractères <i>HT03</i> , <i>TB</i> et $\delta^{13}C$.	167
<u>Figure 47.</u> Moyennes (et erreurs standard) des 18 populations analysées pour <i>HT03</i> , <i>TB</i> et $\delta^{13}C$.	168
<u>Figure 48.</u> Moyennes de <i>HT03</i> , <i>TB</i> et $\delta^{13}C$ en fonction des précipitations moyennes annuelles et des températures moyennes annuelles des lieux d'origine pour les populations espagnoles.	170
<u>Figure 49.</u> Héritabilités (et erreur standard) pour chaque population de <i>HT03</i> , <i>TB</i> et $\delta^{13}C$.	174
<u>Figure 50.</u> Héritabilités en fonction des moyennes par population pour les caractères <i>HT03</i> , <i>TB</i> et $\delta^{13}C$.	175

<u>Figure 51.</u> Moyennes familiales par population de <i>TB</i> en fonction de celles de <i>NB</i> , de <i>HT03</i> en fonction de <i>HT02</i> , de <i>TB</i> en fonction de <i>HT03</i> , de <i>DC</i> en fonction de $\delta^{13}C$, de <i>TB</i> en fonction de $\delta^{13}C$.	179
<u>Figure 52.</u> Représentation graphique des corrélations entre les fréquences alléliques de quelques marqueurs SNP et les moyennes des caractères <i>HT03</i> , <i>TB</i> et $\delta^{13}C$ sur l'ensemble des populations.	208
<u>Figure 53.</u> Stratégie utilisée au sein de cette étude faisant intervenir plusieurs disciplines.	216
<u>Figure 54.</u> Valeurs de Q_{ST} et F_{ST} chez 29 espèces.	226

Index des tableaux

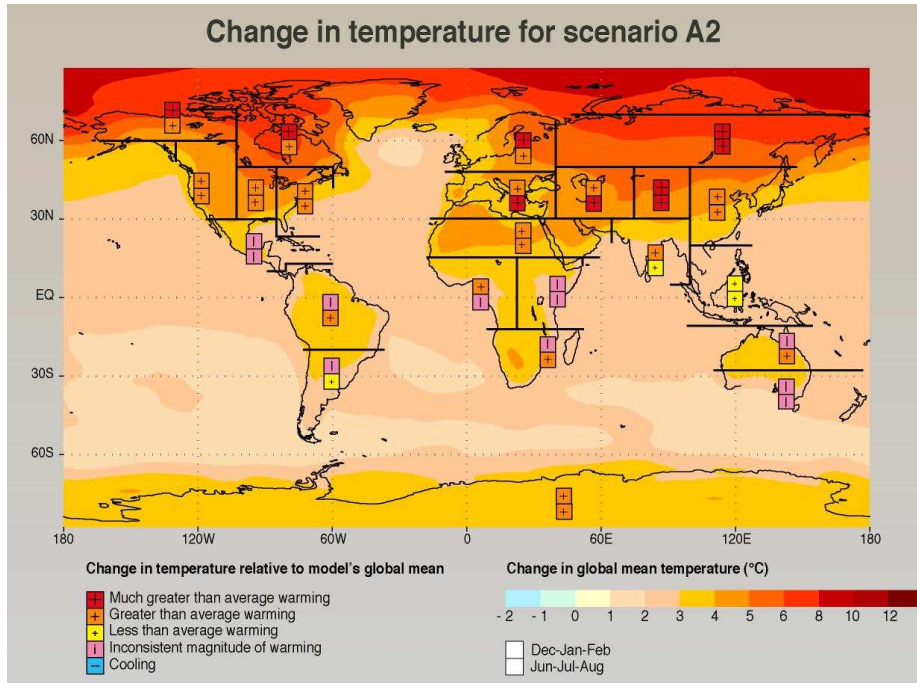
<u>Tableau 1.</u> Tableau synthétique des 5 grandes familles de tests de neutralité classés selon une échelle de temps de détection de la signature sélective.	24
<u>Tableau 2.</u> Quelques études mettant en évidence des écarts à la neutralité chez les espèces d'arbres forestiers et interprétés en terme de signatures moléculaires de sélection naturelle au sein d'approches gènes candidats.	36
<u>Tableau 3.</u> Etudes mettant en évidence des écarts à la neutralité chez les espèces d'arbres forestiers et interprétés en terme de signatures moléculaires de sélection naturelle au sein d'approches pan-génomiques.	37
<u>Tableau 4.</u> Coordonnées géographiques et données climatiques sur 15 populations.	43
<u>Tableau 5 :</u> Matrice des coefficients de détermination pour 6 variables climatiques.	43
<u>Tableau 6.</u> Description des 13 gènes candidats.	45
<u>Tableau 7.</u> Paires d'amorces utilisées pour l'amplification des fragments des 7 gènes candidats sélectionnés par le laboratoire INRA.	48
<u>Tableau 8.</u> Paires d'amorces, estimation de diversité et de différenciation pour les 8 SSRs.	57
<u>Tableau 9.</u> Matrice de F_{ST} par paires de populations entre les 24 populations basée sur 8 SSRs et 6 SSRs.	61
<u>Tableau 10.</u> Estimation des indices de différenciation pour 5 gènes candidats au sein de la métapopulation atlantique.	67
<u>Tableau 11.</u> "Outliers" détectés au niveau haplotypique et au niveau des SNPs par les méthodes "Fdist2", "FstSNP" ou "BayesFst" en considérant les 10 populations contrastées.	102
<u>Tableau 12.</u> Diversités nucléotidique et haplotypique sur les 5 gènes candidats au sein de la métapopulation.	106
<u>Tableau 13.</u> Estimateurs de la recombinaison par gène et tests de neutralité sélective sur les 5 gènes candidats.	107
<u>Tableau 14.</u> Revue de la littérature pour des estimations de diversité nucléotidique silencieuse, et de la diversité haplotypique chez plusieurs espèces de conifères, d'arbres angiospermes, de plantes annuelles, comparativement à quelques valeurs d'études chez les humains et <i>Arabidopsis thaliana</i> .	118

<u>Tableau 15.</u> Tableau récapitulatif sur les interprétations proposées des écarts à la neutralité en terme de signatures de sélection à l'échelle des 10 populations et au sein de la métapopulation atlantique sur 8 gènes candidats.	128
<u>Tableau 16.</u> Coordonnées géographiques et données climatiques disponibles sur 14 populations faisant parti du test de provenances.	144
<u>Tableau 17.</u> Caractères quantitatifs mesurés au sein du test de provenances / descendances.	151
<u>Tableau 18.</u> Tableau résumant l'augmentation ou la diminution des variances entre familles et des variances résiduelles pour le $\delta^{13}C$, lorsque l'on passe du lot d'analyses français au lot d'analyses USA.	163
<u>Tableau 19.</u> Moyenne, valeurs minimum et maximum, et coefficient de variation pour les 7 caractères quantitatifs.	165
<u>Tableau 20.</u> Coefficient de détermination de chaque effet déclaré dans le modèle d'ANOVA pour les 7 caractères quantitatifs.	165
<u>Tableau 21.</u> Estimation des variances entre populations, familles et résiduelles globalement sur l'ensemble des populations, et héritabilités au sens strict, erreur standard des héritabilités, et coefficients de différenciation quantitative entre populations (Q_{ST}) pour les 7 caractères quantitatifs.	172
<u>Tableau 22.</u> Corrélations génétiques au niveau des populations, familles et résiduelles.	178
<u>Tableau 23.</u> Corrélations génétiques additives et erreurs standards entre <i>HT03</i> , <i>TB</i> et $\delta^{13}C$ par population.	181
<u>Tableau 24.</u> Corrélations résiduelles et erreurs standard entre <i>HT03</i> , <i>TB</i> et $\delta^{13}C$ par population.	184
<u>Tableau 25.</u> Caractéristiques des 7 SNPs retenus parmi les 27 initialement choisis.	201
<u>Tableau 26.</u> Fréquences alléliques des 7 SNPs et moyennes phénotypiques de <i>HT03</i> , <i>TB</i> , $\delta^{13}C$ au sein de chaque population.	202
<u>Tableau 27.</u> Corrélations entre les fréquences alléliques des 7 marqueurs SNPs, les moyennes phénotypiques des caractères <i>HT03</i> , <i>TB</i> et $\delta^{13}C$, et les données climatiques sur les 18 populations.	207
<u>Tableau 28.</u> Résultats des tests des effets « classe génotypique » et « dose allélique » au niveau global et en hiérarchisant l'effet du SNP au sein de la population.	209

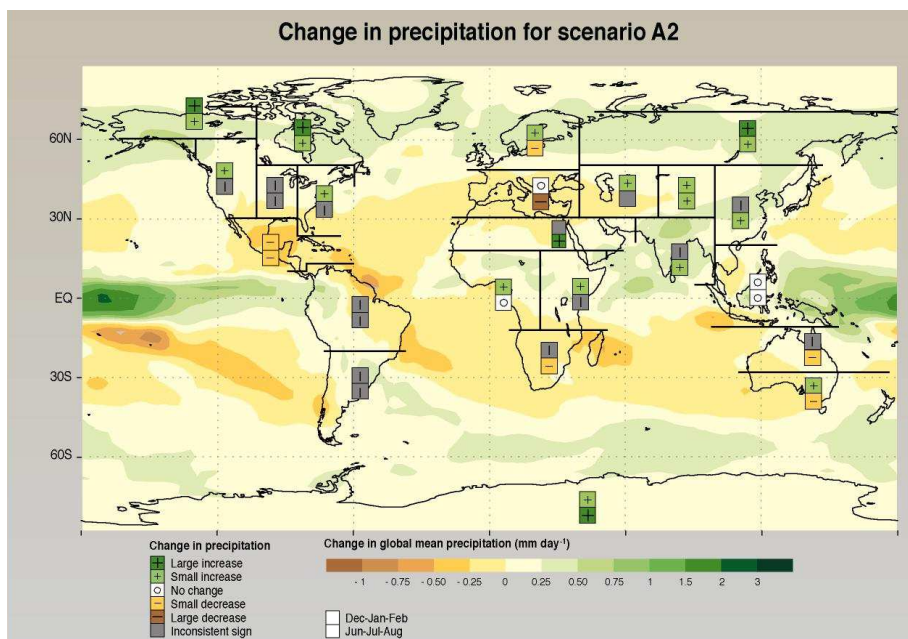
Introduction

Figure 1. Exemple de scénarios proposés par l'IPCC (rapport 2001) sur le changement des températures (A) et des précipitations (B) au niveau mondial d'ici 2100.

A)



B)



Une adaptation peut être définie à la fois comme un caractère phénotypique favorisé par la sélection naturelle, mais également comme le processus évolutif par lequel les caractères phénotypiques au sein d'une population seront favorisés et évolueront par sélection naturelle pour mieux répondre aux contraintes imposées par l'environnement local (Mitton 2001). Les variations phénotypiques observées entre populations d'une espèce dans différents environnements et entre les espèces reflètent souvent cette évolution adaptative (Kremer 1994, Rehfeldt *et al.* 2002, Howe *et al.* 2003). En effet, lorsque des facteurs environnementaux varient au sein de la distribution géographique d'une espèce, la sélection naturelle peut conduire à une différenciation adaptative entre populations, qui se traduira par des adaptations locales, c'est à dire que les populations auront la plus grande *fitness** dans l'environnement dans lequel elles ont évolué (Schemske 1984, Endler 1986, Linhart et Grant 1996, Kittelson et Maron 2001, Kawecki et Ebert 2004). En particulier, les facteurs climatiques comme les variations de température ou de précipitations sont considérées comme une source de pressions de sélection importantes pouvant conduire à des adaptations locales (Joshi *et al.* 2001, Thomas 2005), et pouvant aussi générer de la variation clinale pour des caractères adaptatifs (Huey *et al.* 2000, Jump *et al.* 2006, Garcia-Gil *et al.* 2003).

Dans le contexte des changements climatiques annoncés, les pressions de sélection d'origine environnementale risquent d'être encore plus fortes : augmentation des températures et précipitations moyennes au niveau mondial, étés plus humides au Nord de l'Europe et plus secs au Sud, augmentation dans la fréquence des événements de précipitations intenses et de sécheresse estivale (IPCC 2001, 2007, <http://ipcc-wg1.ucar.edu/wg1/wg1-report.html>, Parry 2000, Klein Tank *et al.* 2002, Figure 1). Pour faire face à ces changements, les plantes pourraient soit migrer et coloniser de nouveaux habitats aux conditions climatiques plus clémentes (Sykes *et al.* 1995, 1996), ou bien s'adapter aux nouvelles conditions climatiques de leur lieu d'origine. Les espèces d'arbres forestiers dont le temps de génération est plus long que les espèces annuelles ou bi-annuelles et qui ne pourraient pas facilement migrer vers de nouveaux habitats (par exemple dans le cas d'une distribution très fragmentée et/ou de distances de dispersion de graines limitées) pourraient être plus fortement affectées par les changements particulièrement rapides qui sont prédits (Hughes 2000, Davis et Shaw 2001, Jump et Penuelas 2005), bien que des données paléo-écologiques indiquent que les

* *fitness* : capacité des organismes à survivre et à se reproduire.

changements d'aires de distributions chez certaines espèces se soient produites sans perte apparente de diversité génétique (Hamrick 2004).

Chez ces espèces, une question fondamentale est donc de savoir si leur potentiel adaptatif (capacité d'une population à répondre aux pressions de sélection ou encore « évolvabilité » *sensu* Houle 1992) est suffisant pour leur permettre 1) tout d'abord de répondre et de survivre sur le temps de génération aux changements climatiques annoncés, et 2) de s'y adapter dans le futur. Cette capacité de réponse va dépendre de la variabilité existant pour les gènes contrôlant l'expression des caractères adaptatifs (Soulé 1987, Howe et Brunner 2005), mais aussi de tous ceux qui sont impliqués dans des caractères corrélés à la *fitness** qui peuvent contraindre ou limiter l'adaptation (notion d'« évolvabilité » au sens large, Houle 1992, Wagner et Altenberg 1996, Pigliucci 2008). Le potentiel adaptatif est directement lié au niveau de diversité génétique additive disponible pour des caractères adaptatifs, et a été traditionnellement étudié par l'héritabilité (rapport de la variance additive sur la variance phénotypique, Falconer et McKay 1996), ou par une mesure moins dépendante de l'environnement rendant mieux compte de l'évolvabilité, le coefficient de variation génétique additive (Houle 1992). Le potentiel adaptatif dépend également de la plasticité phénotypique, qui peut-être définie au niveau individuel comme la capacité d'un même génotype à produire différents phénotypes en réponse à différentes conditions environnementales (Schmalhausen 1949). Cette plasticité peut être considérée comme adaptative si elle a évolué en réponse à une sélection en milieux hétérogènes (Via et Lande 1985, Price *et al.* 2003, Pigliucci *et al.* 2006, Valladares *et al.* 2006). Néanmoins, l'évaluation du potentiel adaptatif, par l'étude de la variabilité génétique de caractères adaptatifs et de la plasticité phénotypique, nécessite la mise en place et le suivi de longs et coûteux essais expérimentaux.

Une seconde question fondamentale en génétique évolutive et écologique est de savoir quelle est l'importance relative des différentes forces évolutives (migration, mutation, *dérive génétique**, sélection naturelle) qui influence la variation adaptative (Mitchell-Olds *et al.* 2007), et plus particulièrement quel est le rôle de la sélection naturelle en interaction avec les autres facteurs mentionnés ci-dessus. Chez les arbres forestiers, et malgré des interactions avec les facteurs évolutifs liés à la démographie, le rôle déterminant de la sélection naturelle dans l'adaptation locale a été largement démontrée au sein d'essais expérimentaux (tests de

* *Dérive génétique* : changement aléatoire des fréquences alléliques au cours des générations dans des populations de taille finie.

provenances d'origines différentes): par exemple pour des caractères de résistance au froid et de débourrement chez *Pinus sylvestris* ou *Picea abies* (Hurme et al. 1997, Johnsen et Skroppa 2000, Savolainen *et al.* 2007), de croissance et de survie chez *Eucalyptus delegatensis* (Garnier-Géré et Ades 2001), de survie et de croissance chez le pin maritime (González-Martínez *et al.* 2002).

Avec le développement de la génomique, on a également cherché à étudier les bases moléculaires de l'adaptation. On s'est interrogé sur la nature, le nombre et la localisation de gènes ou mutations impliqués dans l'adaptation (Barton et Keightley 2000), ainsi que la distribution et l'amplitude des effets alléliques sur les phénotypes (Wright et Gaut 2005, Storz 2005). La recherche de gènes impliqués dans la variation des caractères adaptatifs a été traditionnellement abordée par la cartographie de loci à effet quantitatif (QTL), basée sur des associations statistiques significatives entre marqueurs moléculaires et variation de caractères quantitatifs au sein de pédigrés (Borevitz et Chory 2004) ou plus rarement au sein de populations naturelles (Slate 2005). Néanmoins, cette approche est limitée par les grands intervalles de confiance autour de la localisation du QTL, pouvant inclure des dizaines voire des centaines de gènes candidats différents (Flint et Mott 2001, Christians et Keightley 2002, Masle et al. 2005). Cela s'explique par les grands blocs de déséquilibre de liaison entre marqueurs, générés par le petit nombre d'évènements de recombinaison se produisant dans des pédigrés constitués généralement de 2 ou 3 générations.

Plus récemment, on a assisté au développement d'études d'association entre polymorphismes identifiés au sein de gènes candidats et variation phénotypique dans des populations naturelles. L'intérêt est qu'en travaillant sur du matériel issu de nombreux évènements de recombinaison (comme c'est souvent le cas en populations naturelles), les fenêtres de déséquilibre de liaison autour des hypothétiques sites importants pour l'adaptation sont plus étroites que celles obtenues lorsque l'on travaille sur les descendants de pédigrés correspondant au maximum à quelques générations de recombinaison (Cardon et Bell 2001). L'hypothèse de base est donc qu'une association statistique significative entre phénotype et génotype au locus est l'évidence d'une liaison physique entre le marqueur et un gène impliqué dans la variation du caractère. Les espèces d'arbres forestiers constituent *a priori* de bons modèles pour ces études (faible structuration en populations, large diversité nucléotidique, fenêtres étroites de déséquilibre de liaison, facilités de phénotypage) (Neale et Savolainen 2004, Gonzalez-Martinez et al. 2006b). Cependant, les applications restent encore très rares

(par exemple, Gonzalez-Martinez *et al.* 2007b chez le pin taeda et Thumma *et al.* 2005 chez l'eucalyptus) et l'on dispose de peu de recul concernant par exemple les effets des mutations sur les phénotypes, les tailles d'échantillons optimales ou la précision requise dans l'estimation des caractères phénotypiques et de leur impact sur les méthodes d'analyses utilisées.

On peut également inférer les processus évolutifs et identifier les loci ayant été soumis à la sélection directement à partir de méthodes analysant les profils de diversité nucléotidique et de *déséquilibre de liaison** (voir chapitre I et IV). Ces méthodes ont connu récemment un développement très important dans le cadre d'une nouvelle discipline, la génomique des populations que Luikart *et al.* (2003) définissent comme : « l'étude simultanée de nombreux loci ou régions du génome pour mieux comprendre les rôles des différentes forces évolutives qui agissent sur la variation au sein des génomes et des populations ». Cette discipline a largement profité des récents progrès de la génomique (séquençage massif d'*ESTs**, études d'expression, cartographie génétique et identification de QTL, clonage positionnel). Chez les espèces non-modèles, et notamment chez les espèces forestières (Gonzalez-Martinez *et al.* 2006b), ces progrès ont permis l'identification d'un nombre croissant de gènes candidats fonctionnels, expressionnels ou positionnels potentiellement impliqués dans la variation de caractères quantitatifs. De plus, des développements théoriques récents (théorie de la coalescence, méthodes bayésiennes) permettent de tester les écarts à la neutralité des patrons de diversité observés, en particulier pour détecter les effets possibles de la sélection naturelle au niveau moléculaire de façon beaucoup plus puissante qu'auparavant (Ford 2002, Luikart *et al.* 2003, Wright et Gaut 2005). Dans le cas où des gènes montreraient des patrons atypiques, il est important de caractériser les conséquences phénotypiques éventuelles des variants alléliques identifiés au sein de ces gènes sur la variation des caractères adaptatifs ciblés par des études d'association en populations naturelles (Gonzalez-Martinez *et al.* 2006b, 2007b, Mitchell-Olds *et al.* 2007).

Cette démarche a été suivie dans le projet européen (TREESNIPS : 2002-2007) dont cette thèse fait partie. L'objectif général de ce projet était d'identifier des polymorphismes de type SNPs (Polymorphismes de mutation ponctuelle) ou Indels (Polymorphismes d'insertion ou de délétion) au sein de gènes candidats à la variation de caractères d'adaptation aux

* *Déséquilibre de liaison* : association statistique entre allèles à différents loci.

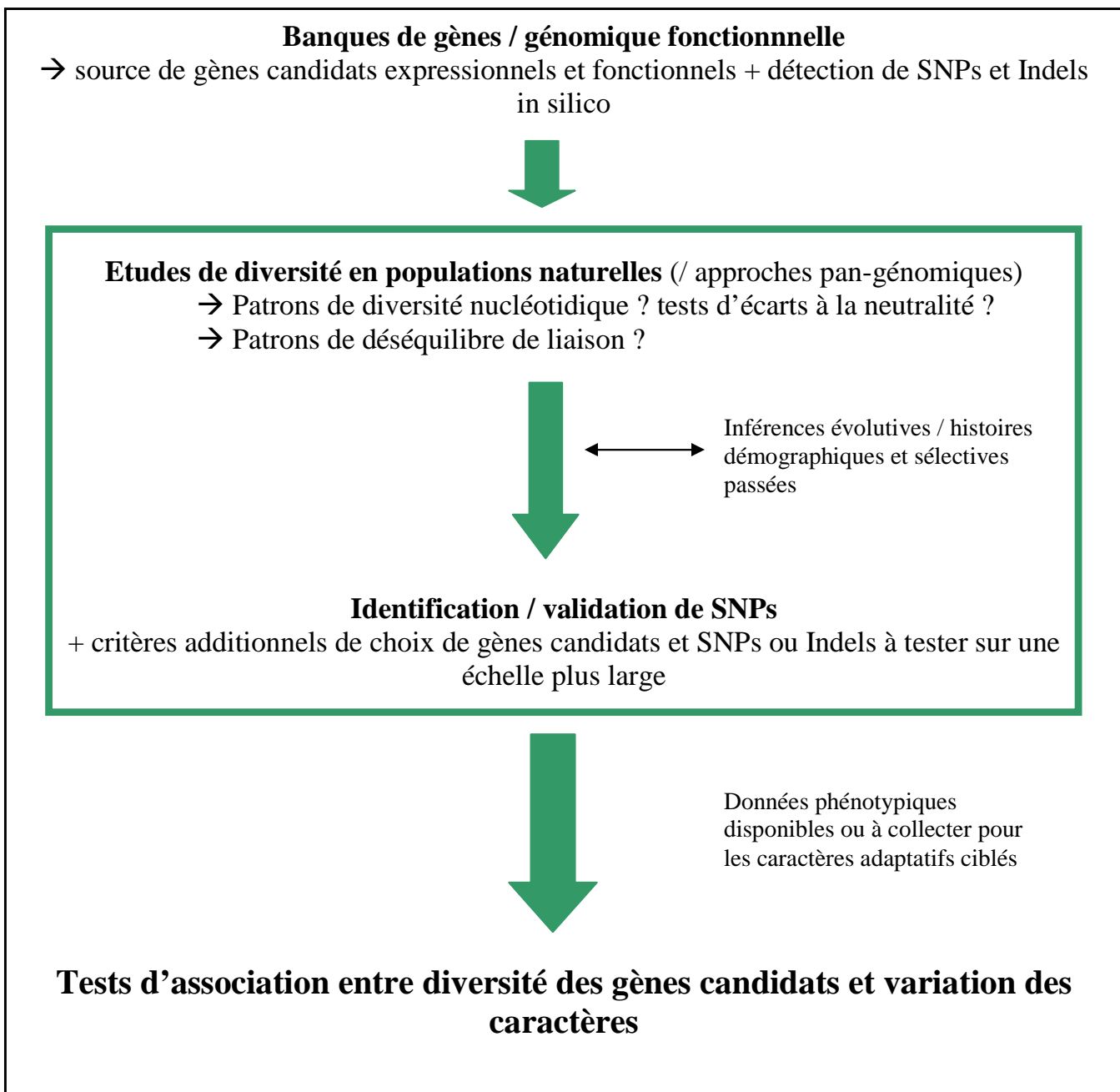
* *ESTs* : étiquettes de séquences exprimées.

changements climatiques (débourrement, stress hydrique ...) dans le but final de les valider en études d'associations (Figure 2). Pour cela, quatre espèces d'arbres forestiers ont été étudiées : *Pinus sylvestris*, *Pinus pinaster*, *Picea abies* and *Quercus petraea*, grâce aux efforts conjoints de 7 laboratoires européens. Cette thèse concerne plus précisément l'adaptation du pin maritime à la résistance au stress hydrique, en tant que réponse déterminante pour la santé et la productivité des forêts, dans le cas de déficits récurrents en eau prévus par les modèles de changements climatiques (Loustau *et al.* 2005). Une partie de ce travail a été réalisé grâce à la collaboration entre l'INRA et l'INIA (Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria). Il s'est appuyé sur l'échantillonnage de 24 populations différentes de pins maritimes dont les aires de collecte étaient soumises à des régimes de précipitations différents. Une des hypothèses de travail initiales était que ces populations auraient développé des stratégies d'adaptation locale pour la résistance au stress hydrique.

En suivant la stratégie illustrée sur la figure 2, les objectifs de ce travail étaient :

- d'identifier des gènes candidats expressionnels et fonctionnels polymorphes impliqués dans des mécanismes de résistance à un stress hydrique (voir chapitre II) ;
- d'étudier leurs patrons de diversité et de différenciation nucléotidiques pour tenter de faire des inférences sur les événements démographiques et sélectifs passés dans les populations échantillonnées. Nous avons aussi utilisé les données de 8 marqueurs microsatellites nucléaires (nuSSRs) qui ont permis d'estimer la différenciation *a priori* neutre entre populations (voir chapitre II). Des tests d'écarts à la neutralité ont été appliqués d'une part sur la base de la comparaison entre différenciation moléculaire des gènes candidats et différenciation neutre, et d'autre part, avec des méthodes basées sur des statistiques liées aux taux de mutation au sein d'une métapopulation atlantique *a priori* non-structurée (voir chapitres III et IV) ;
- d'évaluer la variabilité génétique (prérequis du potentiel adaptatif) au sein des populations, et sa répartition entre populations pour des caractères adaptatifs d'efficience d'utilisation en eau, de croissance et de biomasse, ainsi que la structure des corrélations entre ces caractères. Pour cela, un test de provenances-descendances a été mis en place sur un site des Landes humide à mésophile (voir chapitre V).

Figure 2. Approche gènes-candidats adoptée au sein du projet TREESNIPS.



- de tester des associations entre les variants alléliques de quelques marqueurs SNPs choisis au sein des gènes candidats et la variation des caractères quantitatifs étudiés, en prenant en compte la structure en populations naturelles (voir Chapitre VI).

Chapitre I

Détection de signatures de sélection au niveau moléculaire : méthodes et applications chez les arbres forestiers

Durant la dernière décennie, la génomique des populations a connu un essor important grâce à l'information croissante des données de séquences et à la densité de marquage du génome plus importante qui ont permis de développer des approches gènes candidats (Vigouroux *et al.* 2002, Akey *et al.* 2004) et pan-génomiques (Schmid *et al.* 2005). Il a également été possible d'analyser plus efficacement cette ressource exponentielle d'information et de faire des inférences sur l'histoire (notamment sélective) des populations grâce aux applications de la théorie de la coalescence (Luikart *et al.* 2003, Nordborg et Innan 2002).

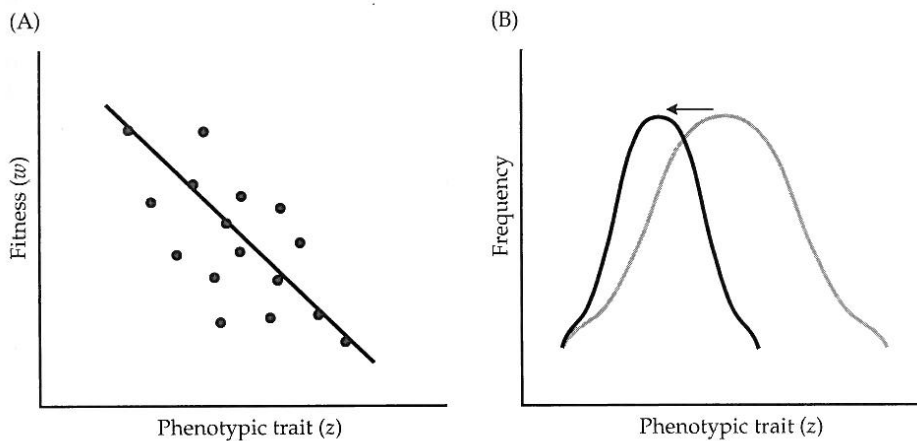
Dans ce chapitre bibliographique, nous présentons un panorama très synthétique des différentes méthodes permettant de tester les écarts à l'hypothèse d'un modèle d'évolution neutre. Nous développerons en particulier les méthodes utilisant des approches pan-génomiques, notamment celles basées sur la différenciation entre populations. Les applications récentes de ces tests chez les espèces d'arbres forestiers seront ensuite résumées et discutées (partie C). Les différentes formes de sélection naturelle et leurs conséquences aux niveaux phénotypique et moléculaire sont tout d'abord présentées.

A. Les différentes formes de sélection naturelle et leurs conséquences sur la diversité et la différenciation génétiques

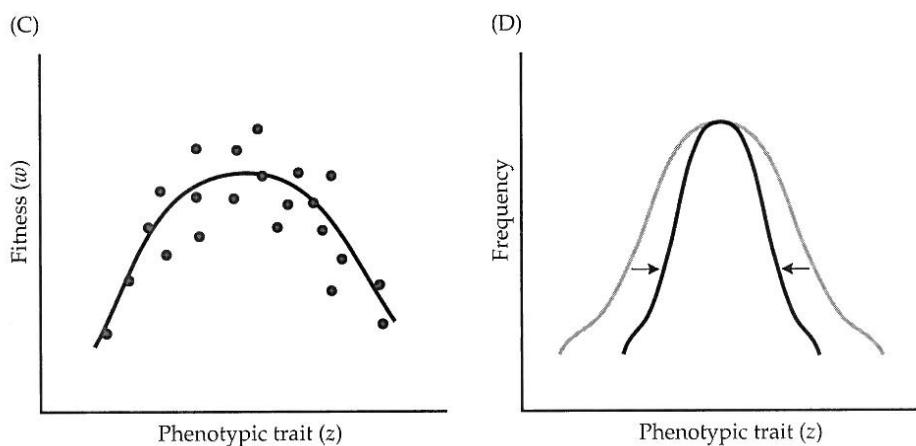
Endler (1986) propose de définir la sélection naturelle (terme introduit par Darwin en 1865) au sens large, comme un processus engendrant à la fois i) un changement de la distribution d'un caractère phénotypique au cours d'une génération, ce changement étant supérieur à ce que l'on pourrait attendre dans un cas d'ontogénie strict (développement progressif d'un organisme depuis sa conception jusqu'à sa forme mature puis sa mort), et aussi ii) un changement de distribution au fil des générations successives. Certains auteurs (par exemple, Lande et Arnold 1983, Futuyma 1998, Fox *et al.* 2001) distinguent le processus strict de sélection naturelle au sein d'une génération (i), de celui de l'évolution au fil des générations, qu'ils définissent plutôt comme une réponse à la sélection naturelle (ii). Pour qu'il y ait évolution par sélection naturelle au cours des générations, il doit y avoir de la variation phénotypique, les variants phénotypiques doivent avoir des valeurs de *fitness* différentes, et ces valeurs doivent être transmissibles (héritables) (Darwin 1865, Lewontin 1970).

Figure 3. Illustration des différentes formes de sélection au niveau phénotypique: relations entre fonctions de fitness et variation du caractère phénotypique (A, C et E) et conséquences sur la moyenne et la variance d'un caractère quantitatif (indiquées par des flèches) (B, D et F) pour les sélections directionnelle, stabilisatrice et divergente (d'après Conner et Hartl 2004).

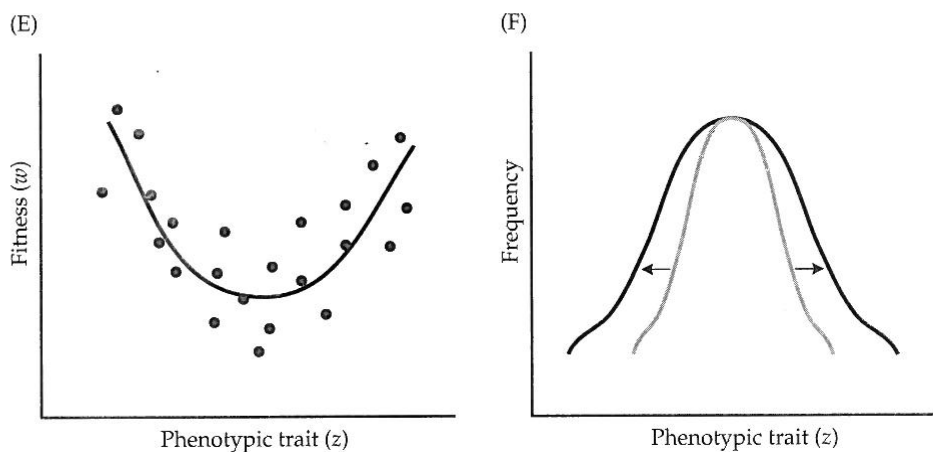
Sélection directionnelle



Sélection stabilisatrice



Sélection divergente



A.1. Différentes formes de sélection naturelle au niveau phénotypique

La sélection naturelle au niveau phénotypique peut prendre différentes formes (directionnelle, stabilisatrice et divergente (ou disruptive)) en fonction des différentes façons dont elle va affecter la moyenne et la variance des caractères phénotypiques (Figures 3B, 3D et 3F) (Conner et Hartl 2004). L'intensité et la forme de la sélection peuvent également être déterminées par la régression de la *fitness* sur la valeur du phénotype (appelée fonction de *fitness*), la pente de cette régression (dans le cas de sélection directionnelle) correspondant à l'intensité de sélection (Figures 3A, 3C, 3E) (Conner et Hartl 2004).

La sélection directionnelle change la moyenne et diminue la variance d'un caractère héritable (Figure 3B). Sa fonction de *fitness* peut être positive (si la *fitness* augmente avec la valeur phénotypique) ou négative (si la *fitness* diminue avec la valeur phénotypique, Figure 3A). Lors d'une sélection stabilisatrice, la *fitness* est plus élevée pour les phénotypes intermédiaires que pour les phénotypes extrêmes (Figure 3C). Dans ce cas, la moyenne du caractère phénotypique reste inchangée mais sa variance va diminuer en réponse à la sélection (Figure 3D). A l'inverse, lors d'une sélection divergente, la *fitness* est plus faible pour les phénotypes intermédiaires que pour les phénotypes extrêmes (Figure 3E). Cette forme de sélection ne change pas la moyenne de la population, tout comme la sélection stabilisatrice, mais peut augmenter sa variance (Figure 3F). Elle peut donc maintenir de la variation phénotypique au sein des populations, et peut conduire à une différenciation adaptative avec réduction de flux de gènes entre groupes d'individus proches géographiquement, qui peut conduire à de la spéciation sympatrique (évolution de populations non isolées géographiquement en espèces distinctes) (Via 2001).

A.2. Différentes formes de sélection naturelle au niveau moléculaire et conséquences sur les diversité et différenciation moléculaires

La sélection naturelle agit sur les phénotypes, mais comme les phénotypes sont en partie déterminés par des génotypes, la sélection naturelle peut donc indirectement affecter la variation au niveau moléculaire. Elle peut être de différentes formes : directionnelle, balancée et purifiante. En interaction avec d'autres forces évolutives (telles que la dérive génétique, la migration, la recombinaison, etc...), ces diverses formes de sélection peuvent affecter différemment la distribution des fréquences alléliques, le spectre des fréquences des sites polymorphes le long des séquences, ou bien le déséquilibre de liaison (*DL*) entre sites

proches, ainsi que la diversité génétique intra population et la différenciation génétique entre populations (voir Encadré 1 pour un rapide historique des différentes mesures et méthodes d'estimation de la différenciation génétique entre populations avec des marqueurs moléculaires).

A.2.1. Sélection directionnelle

On parle de sélection directionnelle (positive) lorsque la sélection entraîne l'augmentation en fréquence d'un allèle (soit une mutation au niveau d'un site nucléotidique ou bien une combinaison particulière de ces sites polymorphes le long d'une séquence d'ADN) car il confère un avantage sélectif, jusqu'à sa fixation (Figure 4A).

Une forte sélection directionnelle peut conduire à un balayage sélectif correspondant à une réduction de la diversité autour de l'allèle sélectionné par effet d'entraînement des variants neutres avoisinants qui sont en *DL* avec l'allèle sélectionné. En effet, durant le processus de substitution, les variants neutres sont soit perdus, soit fixés avec l'allèle avantageux, à moins que la recombinaison ne casse leur association. Le processus d'auto-stop génique a été tout d'abord proposé par Maynard Smith et Haigh en 1974, comme une cause potentiellement importante de réduction de l'hétérozygotie (diversité neutre) au sein de larges populations, puis a été réexaminé dans le cadre de la théorie de la coalescence avec la même conclusion (Kaplan *et al.* 1988, Kaplan *et al.* 1989) (Figure 4A).

D'autres caractéristiques du processus de sélection directionnelle avec effets d'entraînement ont par la suite été étudiées et mis en évidence. En particulier, le balayage sélectif entraînera une élévation du déséquilibre de liaison entre le site sélectionné et les sites avoisinants (Prezworski 2002, Kim et Nielsen 2004). Puis, au cours du temps, la diversité se régénère lentement : les premières mutations apparaissent et sont en faible fréquence, c'est pourquoi, on observera un excès d'allèles rares (Aguadé *et al.* 1989, Hudson 1990, Langley 1990), qui a été utilisé pour tester des patrons d'écarts à la neutralité (Tajima 1989, Braverman *et al.* 1995). La recombinaison intervient en permettant aux chromosomes d'échapper à l'effet d'entraînement, et le balayage sélectif est alors incomplet ou moins important et donc plus difficile à détecter. La signature au niveau moléculaire d'un balayage sélectif diminuera d'autant plus rapidement que le taux de recombinaison dans la région génomique où se situe le site sélectionné est élevé, et le coefficient de sélection est faible (Maynard-Smith et Haigh 1974, Kaplan *et al.* 1989). Par ailleurs, l'existence d'évènements de recombinaison lors d'une phase de sélection peut conduire à l'association de l'allèle avantageux avec des allèles dérivés (par rapport à l'allèle ancestral qui est par hypothèse celui

Encadré 1. Comment mesurer la différenciation génétique entre populations au niveau de marqueurs moléculaires? Panorama et historique.

La différenciation génétique entre populations, qui peut être due à une structure en populations (suite aux contraintes géographiques ou environnementales) est associée à des croisements non-aléatoires entre individus. Cet écart à la panmixie, connu sous le nom d'effet Wahlund (1928) se traduit par un déficit d'hétérozygotes observé dans la population dans son ensemble par rapport à la valeur théorique d'une population panmictique.

➤ Indices de fixation

Wright (1943, 1951, 1965) propose d'exprimer l'écart à la panmixie dans une population structurée en sous-populations au moyen de trois paramètres, également appelés indices de fixation ou statistiques F : (F_{ST} , F_{IT} et F_{IS}). Ces indices sont dérivés du coefficient de consanguinité F (probabilité que deux gènes identiques proviennent du même ancêtre). Il sont donc définis en termes de corrélation entre 2 gènes tirés dans une sous-population par rapport à deux gènes tirés aléatoirement dans la population totale (F_{ST} ou coefficient de différenciation génétique), corrélation entre gènes d'un même individu par rapport à deux gènes tirés aléatoirement dans la population totale (F_{IT}), et corrélation entre deux gènes d'un même individu par rapport à deux gènes tirés aléatoirement dans une sous-population (F_{IS}), et sont reliés par la formule :
$$1 - F_{IT} = (1 - F_{IS})(1 - F_{ST}).$$

➤ Identité par descendance

Par la suite, plusieurs estimations de ces paramètres seront proposées, notamment en définissant ces trois indices de fixation en fonction de probabilités d'identité par descendance (IBD) de deux gènes. Ainsi le F_{ST} est défini tel que $F_{ST} = (Q_S - Q_T)/(1 - Q_T)$, avec Q_S , la probabilité d'IBD de deux gènes entre individus d'une même sous-population, et Q_T , la probabilité d'IBD de deux gènes entre individus de différentes sous-populations (Rousset 2001).

➤ Hétérozygotie

En 1973, Nei propose de calculer les statistiques F en fonction des paramètres hétérozygoties (opposés des probabilités d'IBD) observés et attendus. Le coefficient de différenciation génétique est alors défini tel que : $F_{ST} = \frac{(H_T - H_S)}{H_T}$, avec H_T , l'hétérozygotie totale attendue et H_S , l'hétérozygotie intra population attendue.

Il généralise également le F_{ST} au cas de locus multi-alléliques (G_{ST}).

➤ Analyse de variance

Cockerham (1969, 1973) propose d'estimer ces statistiques F en utilisant une analyse de variance (ANOVA) hiérarchisée sur des variables indicatrices de la présence des allèles intra-individu, inter-individus à l'intérieur de sous-populations, et entre sous-populations à l'intérieur de la population totale. Une estimation du F_{ST} moyen pondéré sur l'ensemble des allèles et l'ensemble des loci est proposée (Weir et Cockerham 1984). Par la suite, plusieurs méthodes d'estimation de la différenciation génétique entre populations seront basées sur la même idée d'une analyse de variance des fréquences alléliques (Long 1986, Long *et al.* 1987).

➤ Temps de coalescence

Plus tard, Slatkin et Voelm (1991) montreront la relation entre les probabilités d'IBD et les temps de coalescence, en partant de l'hypothèse que la probabilité que deux gènes soient identiques par descendance correspond à la probabilité que deux gènes aient coalescé avant qu'il y ait eu une mutation. Le F_{ST} est alors reformulé tel que $F_{ST} = \frac{(\overline{t_N} - \overline{t_0})}{\overline{t_N}}$, avec $\overline{t_N}$, le temps de coalescence moyen de deux gènes pris dans deux sous-populations différentes, et $\overline{t_0}$, le temps de coalescence moyen de deux gènes pris dans une même sous-population. On comprend alors que plus la divergence entre deux gènes pris dans une même sous-population est récente, plus la différenciation génétique est élevée.

Figure 4. Illustration des différentes formes de sélection au niveau moléculaire au sein d'une population : sélection directionnelle et balayage sélectif (A), sélection négative ou purifiante (B), sélection balancée (C) (d'après McVean, <http://www.stats.ox.ac.uk/~mcvean/testing.pdf>).

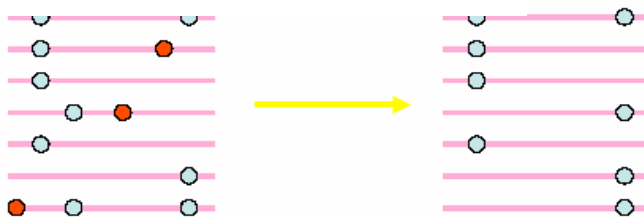
A) Sélection directionnelle suivie d'un balayage sélectif (sur les sites liés neutres) par effet d'entraînement



● Mutation avantageuse

○ Mutation neutre

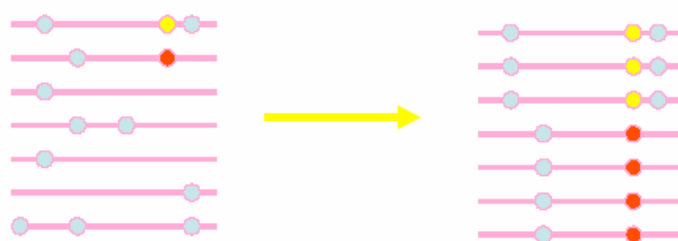
B) Sélection négative / purifiante



● Mutation délétère

○ Mutation neutre

C) Sélection balancée



●● Mutations soumises à de la sélection balancée

○ Mutations neutres

de l'espèce de référence) qui vont augmenter en fréquence plus rapidement qu'en absence de sélection (Fay et Wu 2000, Figure 5).

Néanmoins, ce modèle de balayage sélectif est relativement simple. Il suppose que l'allèle sélectionné vient d'apparaître (et donc sa fréquence est faible) lorsqu'il a été sélectionné. Quelques auteurs se sont également interrogés sur l'effet de la sélection directionnelle sur la variation déjà existante, et où l'allèle sélectionné n'était pas rare. Ils montrent ainsi que plus la fréquence de l'allèle avant l'évènement de sélection est élevée, plus la signature de balayage sélectif sera faible (Innan et Kim 2004, Przeworski *et al.* 2005).

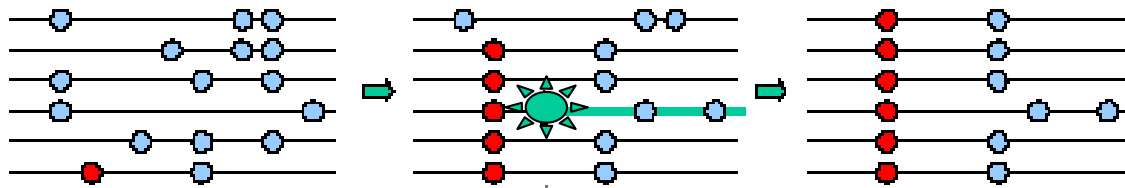
En présence d'une structure en populations, si la sélection naturelle favorise certains allèles dans certaines populations et d'autres allèles dans d'autres populations (dans le cas d'adaptations locales), et qu'elle s'accompagne d'une sélection récurrente contre les allèles introduits par migration qui sont localement mal-adaptés, elle conduira à une augmentation de la différenciation génétique entre populations (Charlesworth *et al.* 1997, Slatkin et Wiehe 1998, Petry 1983, Barton et Bengtsson 1986). Certains auteurs classent cette forme de sélection dite diversifiante entre populations, comme une forme de sélection balancée à l'échelle de l'espèce (Hedrick 2006b, Charlesworth 2006, voir partie A.2.3.).

La succession de balayages sélectifs au cours du temps a été proposée comme un des mécanismes permettant d'expliquer une corrélation positive entre diversité nucléotidique et taux local de recombinaison, qui a été observée pour la première fois dans le genre *Drosophila* (Begun et Aquadro 1992, Wiehe et Stephan 1993), et confirmée par la suite chez d'autres organismes dont les humains, le maïs, la tomate, ou la betterave (Stephan et Langley 1998, Nachman *et al.* 1998, Tenaillon *et al.* 2002, Baudry *et al.* 2001, Kraft *et al.* 1998). Cependant, d'autres auteurs suggèrent que cette corrélation positive pourrait aussi être expliquée dans le cadre d'un modèle neutre, et serait due en partie à l'effet mutagène de la recombinaison (Hellman *et al.* 2003)

A.2.2. Sélection purifiante

La sélection purifiante (ou négative) consiste en la contre-sélection d'allèles délétères entraînant la diminution de leur fréquence jusqu'à leur élimination complète (Figure 4B). Cette sélection purifiante peut s'accompagner d'une sélection « arrière-plan » : si certains sites sont soumis continuellement et fortement à de la sélection purifiante, la diversité aux sites avoisinants sera diminuée, créant une région génomique de faible diversité (Charlesworth *et al.* 1993, Hudson et Kaplan 1995). Certains auteurs parlent également de balayage sélectif au même titre que pour une sélection directionnelle (Stephan *et al.* 1992),

Figure 5. Cas d'un balayage sélectif incomplet, où des allèles dérivés sont entraînés car ils sont associés à l'allèle avantageux après recombinaison (d'après McVean, <http://www.stats.ox.ac.uk/~mcvean/testing.pdf>).



- Mutation avantageuse
- Mutations dérivées
- ★ Evènement de recombinaison

même si la sélection « arrière-plan » ne produit pas de déséquilibre de liaison. Charlesworth *et al.* (1993) ont suggéré que la sélection purifiante pouvait également être une des causes possibles d'une corrélation positive entre diversité et taux de recombinaison. Des études récentes ont tenté de voir : 1) comment on pouvait distinguer les effets des deux processus de sélection (directionnelle ou purifiante), sur la base des prédictions réalisées sur les spectres de distributions des sites polymorphes (Andolfo 2001, Fay et Wu 2000, 2001, Kim et Stephan 2002), et 2) quels étaient les impacts respectifs des taux de recombinaison et de mutation sur l'intensité de la corrélation attendue entre diversité génétique et recombinaison (Wiehe 1998, Tenailon *et al.* 2002).

De même, comme la sélection purifiante maintient les polymorphismes liés aux allèles délétères à des fréquences faibles, on observera tout comme pour la sélection directionnelle, un excès de polymorphismes en fréquences faibles (Bamshad et Wooding 2003, Nielsen 2005). En présence d'une structure en populations, la différenciation génétique entre populations aura tendance à augmenter, principalement à cause de la baisse de diversité au sein des populations soumises à la sélection purifiante (Charlesworth *et al.* 1997).

A.2.3. Sélection balancée

Dans le cas d'une sélection balancée (positive), deux ou plus de deux allèles à un locus sont maintenus au sein d'une population (Figure 4C). Le maintien de ce polymorphisme allélique est souvent le résultat de l'avantage sélectif d'un allèle rare, qui peut s'expliquer par de 1) la superdominance (ou avantage hétérozygote) ou de 2) la sélection fréquence-dépendante négative. Dans de rares cas de superdominance observés en populations naturelles, les allèles rares sont maintenus grâce aux hétérozygotes qui présentent un avantage sélectif supérieur à celui des homozygotes (exemple de l'anémie falciforme, Haldane 1949, Hill *et al.* 1991). Dans le cas de sélection fréquence-dépendante négative, l'allèle rare possède un avantage sélectif par rapport à un allèle plus commun (exemple du sex-ratio, Conover et Van Voorhees 1990, Conover *et al.* 1992, Charlesworth 2004, exemple du locus d'auto-incompatibilité, Charlesworth et Awadalla 1998, Nasrallah 2002). Il existe également des modèles plus complexes de sélection diversifiante maintenant du polymorphisme allélique, à cause d'une hétérogénéité spatiale ou temporelle (Hedrick *et al.* 1976, Hedrick 1986, 2006a et b).

Les conséquences de cette forme de sélection au niveau moléculaire sont l'excès de mutations en fréquences intermédiaires, ainsi que le maintien de niveaux de diversités élevés sur les sites environnants au locus sélectionné par effet d'entraînement (Kaplan *et al.* 1988,

Charlesworth *et al.* 1997, Charlesworth 2006). En présence d'une structure en populations, lorsqu'une sélection balancée maintient du polymorphisme au sein des populations, la majorité de la diversité se retrouve au sein des populations, et non entre populations, ce qui peut entraîner une diminution de la différenciation génétique entre populations (Charlesworth *et al.* 1997). Cependant, la structure en populations n'est parfois pas connue, et il peut exister des interactions entre l'effet de la sélection balancée au sein d'une population et l'effet de la structure « cachée » en sous-populations sur les profils de diversité génétique, cette interaction pouvant être non négligeable lorsque l'on se situe en dessous d'un seuil critique de migration entre sous-populations (Muirhead 2001).

La signature moléculaire de la sélection balancée peut être maintenue sur de longues périodes. Si les périodes de temps concernées sont supérieures à la divergence des populations en espèces, on peut même observer des polymorphismes trans-espèces (Klein 1987, Garrigan et Hedrick 2003, Bouillé et Bousquet 2005). La détection d'une signature de sélection balancée est donc beaucoup moins dépendante du temps depuis l'évènement de sélection, que celle de signatures dues à de la sélection positive. Elle dépendra davantage du taux de recombinaison entre le site sélectionné et les sites proches (Hudson et Kaplan 1988).

Néanmoins, il est aussi possible que le maintien de polymorphisme moléculaire ne soit pas uniquement dû à de la sélection balancée, mais plutôt à une alternance entre sélection directionnelle et migration, où un allèle est maintenu par sélection, et l'autre par migration (voir les simulations en populations humaines de Hiorns et Harrison, 1977). De plus quand la sélection directionnelle entraîne un balayage sélectif incomplet qui est interrompu quand l'allèle favorable est en fréquence intermédiaire, on observera également un maintien d'allèles en fréquence intermédiaire qui ressemblera à une signature de sélection balancée (Hedrick 2007).

B. Comment détecter de la sélection naturelle au niveau moléculaire ?

L'effet de la sélection naturelle peut être étudié par une approche directe consistant à suivre expérimentalement une population au cours du temps. Néanmoins, cette approche nécessite des données très difficiles à obtenir sur de longues périodes, et les coefficients de sélection sont sans doute trop faibles pour pouvoir être détectés sur de courtes périodes. Alternativement, l'approche indirecte consistant à repérer des signatures moléculaires de la sélection naturelle à partir des patrons de diversité observés en les testant par rapport à un modèle neutre, est dorénavant plus facilement réalisable avec l'obtention de données de

séquences, et les méthodes utilisées intègrent le processus généalogique sur des périodes beaucoup plus longues.

B.1. Panorama synthétique des tests de neutralité

Il existe de nombreuses méthodes disponibles permettant de détecter des signatures moléculaires de la sélection naturelle à partir des patrons de diversité nucléotidique observés. L'objectif de ce chapitre n'est pas de détailler toutes ces méthodes pour lesquelles de bons articles de synthèse existent (par exemple, Biswas et Akey 2006, Sabeti *et al.* 2006, Ford 2002, Garrigan et Hedrick 2003, Nielsen 2001, Nielsen 2005, Nordborg et Innan 2002, Wright et Gaut 2005). Nous résumons simplement ici sous forme de tableau, un classement de ces méthodes en 5 grandes familles de tests, selon le type de signatures de sélection qu'elles vont pouvoir détecter, qui sera différent en fonction de la période à laquelle les patrons seront observés depuis une phase de sélection (Sabeti *et al.* 2006, Biswas et Akey 2006, Garrigan et Hedrick 2003, Hedrick 2006a, Tableau 1). Nous avons vu en effet que l'efficacité de certains tests pouvait dépendre fortement du fait que les épisodes de sélection se soient produits récemment ou dans un passé beaucoup plus lointain (Przeworski 2002).

Ces différentes méthodes visent à tester généralement l'écart au modèle neutre standard (modèle de Wright-Fisher), qui repose sur une population d'individus diploïdes, de taille constante N ($2N$ allèles) à l'équilibre démographique, avec des générations non-chevauchantes, des individus se reproduisant en panmixie quelques soient la position géographique et le génotype, une reproduction sexuée avec possibilité d'autofécondation, aucune migration ni sélection, et des mutations neutres apparaissant à un taux μ constant par génération. L'écart à ce modèle peut donc être interprété non seulement en termes de signatures de sélection naturelle, mais également de non-équilibre démographique. Par exemple, une expansion de population peut générer des généalogies similaires à celles obtenues dans le cas d'un balayage sélectif (Nordborg et Innan 2002, Figure 6-b) et 6-d)). Par ailleurs, un goulot d'étranglement modéré ou une structure en populations pourront générer des généalogies comparables à celles issues d'une sélection balancée (Hein *et al.* 2005, Nordborg et Innan 2002, Figure 6-c)).

Les tests sensibles à l'histoire démographique des populations et à la structure en populations (voir Tableau 1), fournissent donc généralement des preuves limitées qu'un variant d'ADN ou un polymorphisme étroitement lié ait des conséquences directes sur le phénotype et la valeur adaptative (Vasemägi et Primmer 2005). Seuls les tests basés sur le

Tableau 1. Tableau synthétique de 5 grandes familles de tests de neutralité classés selon une échelle de temps depuis le moment où la sélection a opéré(d'après Sabeti *et al.* 2006, Biswas et Akey 2006, Garrigan et Hedrick 2003, Hedrick 2006a).

Temps présent	Signal de signature de sélection observé	Exemples de statistiques utilisées pour les tests de neutralité	Références	Type de données échantillonnées	Unilocus/multiloci	Contraintes/ inconvénients	Type de sélection détectée
Temps d'évolution depuis l'évènement de sélection	Augmentation du déséquilibre de liaison	test Z_{NS} , test LRH	Kelly 1997, Sabeti <i>et al.</i> 2002	Intra-population	unilocus	sensible à la structure en populations + sensible à l'hétérogénéité des taux de recombinaison	Sélections récentes de type directionnelle (ou purifiante) et sélection balancée
	Augmentation/Baisse de la différenciation génétique entre populations	Indices de différenciation (F_{ST} , $Snn...$)	Lewontin et Krakauer 1973, Beaumont et Nichols 1996, Vitalis <i>et al.</i> 2001, Beaumont et Balding 2004, Hudson 2000	entre populations ou espèces	multiloci	sensible à l'histoire démographique des populations, notamment goulot d'étranglement	Sélection récente directionnelle et balancée
	Augmentation de la fréquence des allèles dérivés	H de Fay et Wu	Fay et Wu 2000	Intra-population	unilocus	Sensible à la structure en populations	Sélection ancienne directionnelle ou balancée et sélection purifiante
	Baisse ou augmentation de la diversité génétique, notamment des allèles rares	D de Tajima ^(a) , F_S de Fu*, test HKA , F et D de Fu et Li	Tajima 1989, Fu 1997, Hudson <i>et al.</i> 1987, Fu et Li 1993	intra population	unilocus / multiloci - test d'hétérogénéité pour test HKA	sensible à la structure en populations, au changement dans la taille des populations (expansion de population...), au taux de mutation et recombinaison, à l'échantillonnage	Sélection ancienne directionnelle (ou purifiante) et balancée
	Augmentation des taux de mutations non-synonymes par rapport aux taux de mutations synonymes	Test d_N/d_S , K_A/K_S , test MK	McDonald et Kreitman 1991, Hughes et Nei 1988, Nielsen et Yang 1998, Yang <i>et al.</i> 2000, Yang et Nielsen 2002	entre espèces	multiloci - tests d'hétérogénéité	peu puissant	Sélection positive ancienne et récurrente

^(a) voir Encadré 2.

Encadré 2. Le test de Tajima (1989)

Test basé sur la comparaison entre deux estimateurs de la diversité nucléotidique θ :

- θ_S basée sur le nombre de sites polymorphes, S
- θ_π basée sur le nombre moyen de différences entre séquences, π

La différence entre ces deux estimateurs, normalisée par l'erreur standard, donne la statistique D :

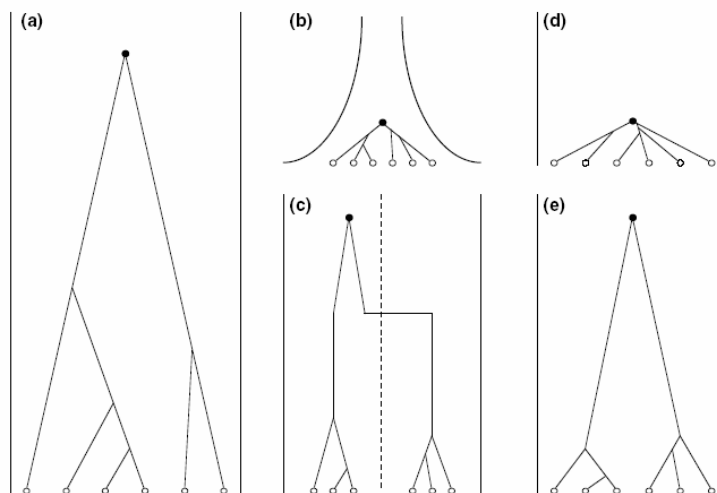
$$D = \frac{\hat{\theta}_\pi - \hat{\theta}_S}{\sqrt{\text{Var}(\hat{\theta}_\pi - \hat{\theta}_S)}}$$

Interprétation de la statistique D :

- Comme S est plus sensible que π aux allèles rares, $D < 0$ signifiera un excès de mutations rares pouvant être dû à une signature de balayage sélectif ou d'expansion de population.
- Réciproquement, π est peu affectée par les mutations rares mais plus par les allèles en fréquence intermédiaire. Donc $D > 0$ signifiera un excès d'allèles en fréquence intermédiaire pouvant être dû à une signature de sélection balancée ou un goulot d'étranglement.

Les régions critiques de la statistique sont obtenues par des simulations d'échantillons de coalescence

Figure 6. Exemples d'arbres généalogiques sous un modèle neutre standard avec une taille de population constante (a), après une expansion de population (b), dans une population structurée en deux sous-populations (c), dans une population après la fixation d'une mutation avantageuse (d), et dans une population dans laquelle 2 allèles sont maintenus par sélection balancée (e). L'ancêtre commun le plus récent est représenté par un point noir (d'après Nordborg et Innan 2002).



rapport entre taux de mutations synonymes et non-synonymes sont insensibles à l'histoire démographique et à la structure, mais ces derniers sont aussi très conservatifs (Bamshad et Wooding 2003, Biswas et Akey 2006, Nielsen 2001, Yang et Bielawski 2000, Nielsen 2005). Cependant, le développement de nouveaux tests de neutralité intra-population est un domaine de recherche en pleine expansion, car même s'ils sont sensibles à la démographie en général, ils peuvent être plus puissants s'ils sont appliqués dans de bonnes conditions. De nouveaux tests ont récemment été proposés, qui combinent des tests classiquement utilisés, en explorant les distributions conjointes des statistiques associées (Zeng *et al.* 2006, De Mita *et al.* 2007). De plus, ces tests seraient moins sensibles aux effets démographiques. En parallèle, on a aussi vu l'émergence rapide ces dernières années d'approches pan-génomiques dont l'objectif est d'essayer de distinguer les loci qui seraient soumis à des événements de sélection en comparaison d'autres loci représentatifs du génome et qui seraient affectés principalement par l'histoire démographique des populations (Luikart *et al.* 2003).

B.2. Les approches pan-génomiques

Ces approches consistent donc à cribler la diversité ou la différenciation moléculaire à de nombreux loci pour distinguer les effets de la démographie de ceux de la sélection. Elle repose sur le postulat que les processus démographiques affecteraient le génome dans son ensemble de façon relativement homogène, tandis que la sélection naturelle agirait sur certains loci en particulier (ceux impliqués dans la variation des caractères adaptatifs) (Cavalli-Sforza 1966).

B.2.1. L'approche pan-génomique basée sur la différenciation entre populations

- Développement de méthodes à partir de l'idée de Cavalli-Sforza

Cavalli-Sforza (1966) proposa de tester des signatures de sélection naturelle, en se basant sur l'hypothèse que des loci neutres devaient présenter des valeurs similaires de coefficient de consanguinité, F_{IS} , ainsi que les mêmes valeurs de F_{ST} (voir Encadré 1). A l'inverse, des loci soumis à de la sélection présenteraient des valeurs de F_{ST} s'écartant de celles observées sur l'ensemble du génome. Le premier test statistique exploitant l'idée de Cavalli-Sforza (1966) fut celui de Lewontin et Krakauer (1973), basé sur la variance de la statistique F_{ST} entre loci (voir Encadré 3). Néanmoins ce test fut sévèrement critiqué à cause des violations nombreuses des hypothèses sous-jacentes (absence de corrélation des

Encadré 3. Le test de Lewontin et Krakauer : principe et limites

➤ Principe

Suivant l'idée de Cavalli-Sforza (1966), Lewontin et Krakauer (1973) proposèrent un test de neutralité basé sur la variance de la statistique F_{ST} . La variance attendue des valeurs de F_{ST} entre les loci pour un allèle est estimée par $2 F_{ST}^2 / (n-1)$, où n est le nombre de populations. Cette variance suit une distribution de χ^2 moyen (χ^2 / ddl) où le nombre de degré de liberté (ddl) est égal au nombre de loci. La même approximation peut être étendue à plusieurs allèles, auquel cas, n est remplacé par $(d-1)(K-1)$, avec d , le nombre de populations et K , le nombre d'allèles.

Le test de neutralité de Lewontin et Krakauer propose de comparer le ratio de la variance observée sur la variance attendue aux valeurs critiques du χ^2 . Si le ratio est significativement plus grand que les valeurs critiques, l'hypothèse de l'évolution neutre des loci dans l'échantillon est rejetée, ce qui signifie qu'au moins un locus dans l'échantillon a été soumis à de la sélection naturelle.

➤ Limites

Mais dès 1975, ce test fut sévèrement critiqué. Robertson (1975 a, b) et Nei et Maruyama (1975) démontrèrent qu'une corrélation des fréquences alléliques entre populations pouvait augmenter la variance du F_{ST} par rapport à celle estimée par Lewontin et Krakauer (qui suppose une indépendance des fréquences alléliques entre populations). Deux processus peuvent introduire une telle corrélation : une migration plus élevée entre certaines populations (cas de la migration de type « stepping-stone ») qui peut produire une auto-corrélation spatiale, et l'histoire phylogénétique des populations, qui peut produire une auto-corrélation phylogénétique. La deuxième hypothèse du test de Lewontin et Krakauer qui fut critiquée par les mêmes auteurs est l'égalité des valeurs de F_{ST} pour tous les loci, en partant de l'hypothèse que les événements démographiques affectent tous les loci de la même façon. Cette hypothèse est facilement rejetée lorsqu'une mutation est plus récente qu'une autre, ce qui implique que sa fréquence allélique et donc sa valeur de F_{ST} soit différente. Enfin, les critiques portèrent aussi sur le type de test utilisé qui est une approximation du χ^2 , ce dernier étant peu fiable lorsque les tailles d'échantillons sont petites, le nombre d'allèles varie d'un locus à l'autre, et les F_{ST} sont inégaux entre populations (Nei et Maruyama 1975, Robertson 1975a, b, Nei et Chakravarti 1977, Nei et al. 1977).

fréquences alléliques entre populations, et égalité des valeurs de F_{ST} pour tous les loci neutres, voir Encadré 3), et fut donc rapidement abandonné.

Puis le développement de la théorie de la coalescence (approche rétrospective pour reconstruire la généalogie d'un échantillon de gènes d'une population, Kingman 1982) et son utilisation en tant qu'outil de simulations, permit de faire des prédictions théoriques jusqu'alors difficiles, et ainsi d'obtenir directement une distribution du F_{ST} et d'explorer différents scénarios démographiques (Beaumont 2005). Grâce à cet outil, plusieurs méthodes basées sur la même idée ont été par la suite développées, avec le souci constant de tester leur robustesse à plusieurs scénarios démographiques.

Bowcock *et al.* (1991) furent les premiers à reprendre et à affiner le test de Lewontin et Krakauer (1973), en proposant d'analyser la distribution simulée des valeurs de F_{ST} en fonction de la fréquence allélique dans un modèle d'admixture (où la population européenne provient d'un mélange de populations chinoises et africaines). La comparaison des valeurs de F_{ST} observées sur un ensemble de marqueurs RFLP (Restriction Fragment Length Polymorphisms) par rapport à la distribution simulée, permet d'identifier des marqueurs « outliers » s'écartant de cette distribution, qui sont interprétés comme une signature moléculaire de sélection naturelle.

Beaumont et Nichols (1996) proposèrent une approche légèrement différente, en simulant la distribution attendue du F_{ST} en fonction de l'hétérozygotie, plutôt que des fréquences alléliques, l'objectif étant d'utiliser des marqueurs multi-alléliques comme des SSRs (les hypothèses du modèle permettant de générer la distribution attendue sont développées dans le Chapitre III). Ils montrèrent que la distribution du F_{ST} conditionnelle à l'hétérozygotie était robuste à plusieurs modèles démographiques (modèles avec croissance exponentielle de la taille des populations, migration hétérogène entre populations) ainsi qu'à plusieurs tailles d'échantillonnage (10, 20, 50 et 100 individus haploïdes) et taux de mutation (0.01, 0.1 et 1).

Pour s'affranchir du problème de scénarios démographiques variables entre populations au sein d'une espèce (notamment lorsque les taux de migration entre populations sont hétérogènes), Vitalis *et al.* (2001) proposèrent de s'intéresser aux paires de populations plutôt qu'à un ensemble de populations qui auraient divergé d'une population ancestrale. L'idée de ce test est d'estimer un paramètre spécifique à chaque population. Ce paramètre est fonction de la probabilité d'identité par paires de gènes pris au sein ou entre populations, et peut être vu comme le ratio du temps de divergence de deux populations sur la taille efficace de ces populations. La distribution de ce paramètre est simulée sous divers scénarios neutres

de divergence, permettant de détecter les « outliers » qui s'écartent de cette distribution. Cette méthode permettrait de détecter des signatures d'adaptation locale qui seraient spécifique à une population en particulier en comparaison avec d'autres, mais serait globalement moins puissante car elle réduit l'examen de la divergence à des comparaisons de paires de populations. De plus, l'augmentation du nombre de tests (égal au nombre de comparaisons possibles) peut engendrer une augmentation du nombre de faux positifs (problème inhérent aux tests multiples, voir partie B.2.3.) (Beaumont 2005).

Porter (2003) a également proposé une approche utilisant non pas des simulations de coalescence mais du bootstrap paramétrique, sur la base de la distribution des fréquences alléliques sous un modèle en îles (modèle neutre) en fonction du taux de migration et de la taille efficace (Wright 1931) qui permet de déduire celle des F_{ST} . La comparaison de cette distribution attendue avec les F_{ST} observés permet d'identifier les loci « outliers » qui pourraient être soumis à de la sélection.

Plus récemment, Beaumont et Balding (2004) ont développé une méthode bayésienne, dont l'un des avantages est d'être beaucoup plus flexible concernant le nombre de paramètres à estimer, par exemple des taux de migration spécifiques à chaque population. Le principe et les hypothèses sous-jacentes à cette méthode sont détaillés dans le Chapitre III qui expose la méthodologie utilisée pour analyser nos données.

- De nombreuses applications en génétique humaine et animale

Les approches pan-génomiques ont été utilisées soit en comparant des distributions simulées par coalescence (utilisées par exemple dans les méthodes développées par Bowcock *et al.* 1991, Beaumont et Nichols 1996, Vitalis *et al.* 2001 et Beaumont et Balding 2004) à un petit jeu de données expérimentales, soit en utilisant de larges jeux de données permettant de construire les distributions empiriques de fréquences alléliques et autres paramètres statistiques décrivant les échantillons. Cette deuxième approche plus empirique a été avant tout développée chez l'homme, grâce à la disponibilité de jeux de données SNPs considérables, qui a permis d'estimer les diversités nucléotidique et haplotypique, et les taux de recombinaison le long du génome humain (projet HapMap :<http://www.hapmap.org/>). Ces données ont révélé notamment l'existence de larges fenêtres de DL chez cette espèce (Reich *et al.* 2001, Pritchard et Przeworsky 2001, Ardlie *et al.* 2002). Une application récente (Akey *et al.* 2002) a utilisé 26 530 marqueurs SNPs répartis sur l'ensemble du génome pour construire une distribution empirique des valeurs de F_{ST} représentative du génome, et a ainsi permis d'identifier 174 gènes candidats s'écartant de cette distribution. Ce type d'approche

empirique est cependant limité chez les espèces non-modèles car elle requiert le génotypage de nombreux loci le long du génome pour construire une distribution représentative (Luikart *et al.* 2003). Les études pan-génomiques basées sur des simulations de coalescence sont aussi en constant développement chez l'homme (Altshuler *et al.* 2005, Weir *et al.* 2005, Kayser *et al.* 2003, Storz *et al.* 2004), les animaux (Wilding *et al.* 2001, Campbell et Bernatchez 2004, Vasemägi *et al.* 2005, Nielsen *et al.* 2006, Beaumont et Nichols 1996, Storz et Nachman 2003, Storz et Dubach 2004, Storz *et al.* 2003, Pariset *et al.* 2006, Hoffman *et al.* 2006, Beaumont et Nichols 1996, Vitalis *et al.* 2001) et les végétaux (Vigouroux *et al.* 2002, Jump *et al.* 2006, Tsumura *et al.* 2007, Scotti-Saintagne *et al.* 2004, Acheré *et al.* 2005, Savolainen *et al.* 2006).

B.2.2. Autres approches pan-génomiques basées sur les patrons de diversité nucléotidique et l'étendue de déséquilibre de liaison intra-population

Il existe également de nombreux exemples d'approches pan-génomiques qui sont basées non plus sur la distribution de la variance du F_{ST} entre loci, mais sur celle d'autres statistiques liées aux niveaux de diversité ou au DL . Dans l'étude de Akey *et al.* (2004), plusieurs simulations de coalescence sont réalisées à partir des données observées, en testant différents paramètres démographiques qui permettent de générer des distributions sous différents scénarios démographiques (goulot d'étranglement, expansion, structure en îles, structure avec isolement). Les combinaisons de paramètres démographiques qui génèrent des données avec des statistiques de tests s'approchant le plus de celles sur les données observées (D de Tajima, D et F de Fu et Li, H de Fay et Wu) permettent d'identifier le modèle démographique qui est le plus probable. Puis cette combinaison de paramètres est utilisée pour évaluer le taux de significativité des différents tests, et ainsi identifier les loci dont les valeurs des statistiques s'écartent de la distribution neutre simulée.

Des approches pan-génomiques dans le cas d'échantillons relativement importants permettraient de comparer les profils de diversité et de DL entre populations, le principe étant que la détection d'une baisse de diversité ou d'une augmentation du DL d'une population par rapport à une autre peut par exemple indiquer un balayage sélectif local au sein d'une population (Storz 2005). Ce type d'approche est approprié pour les comparaisons par paire de populations dans lesquelles une seule des deux populations est sujette à une sélection positive sur une certaine région du génome. A l'inverse, dans le cas où deux populations isolées sont chacune soumises à des balayages sélectifs sur la même région impliquant la fixation d'allèles différents entre populations, le niveau de diversité sera réduit au sein de chaque

population, mais aucun écart entre les valeurs de diversité des deux populations ne sera observé (Storz 2005). C'est pourquoi, ces approches basées sur la diversité sont plus efficaces dans le cas d'études comparatives entre populations *a priori* contrastées sur le plan écologique mais soumises à des pressions différentes : ancestrales et populations dérivées ayant colonisé un nouvel environnement (par exemple chez *Drosophila melanogaster*, Harr *et al.* 2002, Kauer *et al.* 2003, Kauer *et al.* 2002, Schöfl et Schlötterer 2004), entre populations sensibles et populations résistantes (par exemple chez *Plasmodium falciparum*, Wootton *et al.* 2002, Nair *et al.* 2003), entre espèces domestiquées et leur ancêtre sauvage (par exemple chez *Zea mays*, Vigouroux *et al.* 2002, Wright *et al.* 2005), entre populations humaines (par exemple, Carlson *et al.* 2005, Schlötterer 2002, Wang *et al.* 2006, Voight *et al.* 2006), ou entre espèces proches (par exemple, entre humains et chimpanzés, Nielsen *et al.* 2005 et Bustamante *et al.* 2005).

Ces approches pan-génomiques sont également plus puissantes pour détecter la signature moléculaire d'un balayage sélectif récent que celle d'un balayage sélectif ancien (Storz 2005). En effet, lorsqu'un balayage sélectif s'est produit au sein d'une population et a entraîné une baisse de diversité, la différence des niveaux de diversité entre cette population et les autres sera plus prononcée immédiatement après le balayage sélectif. Puis l'introduction de nouvelles mutations dans la population où il y a eu balayage sélectif restaurera progressivement la diversité au sein de cette population, et la différence des niveaux de diversité entre populations diminuera (Wiehe 1998).

B.2.3. Limites des approches pan-génomiques

- Problème des tests multiples et faux positifs

L'approche pan-génomique présente néanmoins quelques limites, dont la plus importante est celle liée aux tests multiples. En effet, dans une approche pan-génomique, il existe autant de tests d'écart à la distribution simulée ou empirique que de loci étudiés. Dans le cas des méthodes développées par Bowcock *et al.* 1991, Beaumont et Nichols 1996, ou Vitalis *et al.* 2001, les tests réalisés sont de type «fréquentiste». En effet, si un locus ne se conforme pas à l'hypothèse nulle (en se situant en dehors de la distribution simulée ou empirique) à un seuil α (risque de première espèce) de 5% par exemple, cela signifie simplement que dans 5% des échantillons où l'hypothèse est vraie, elle sera rejetée à tort. On comprend alors que rejeter l'hypothèse nulle à un seuil α de 5% n'est pas la même chose lorsque l'on teste 100 ou 1000 loci: dans le premier cas, seuls 5 échantillons pourront être

théoriquement des faux positifs, tandis que dans le deuxième cas, 50 échantillons seront potentiellement des faux positifs.

Il est possible de corriger les probabilités par le nombre de tests grâce à des corrections de type Bonferroni (Sokal et Rohlf 1995). La *p-valeur* (valeur exacte du risque de première espèce dans une situation donnée) après correction de Bonferroni est égale à la *p-valeur* divisée par le nombre de tests. Or ces corrections rendent les tests particulièrement conservatifs (Perneger 1998, Verhoeven *et al.* 2005). Plus récemment, des méthodes plus puissantes ont été développées en utilisant le concept du FDR (« False Discovery Rate » correspondant à l'espérance du rapport entre le nombre de faux positifs sur le nombre d'hypothèses rejetées, Benjamini et Hochberg 1995), ce qui permet de contrôler le taux de faux positifs, et permet de calculer des analogues de *p-valeurs*, les *q-valeurs*, pour chaque test (Storey 2002, Storey et Tibshirani 2003).

Une autre alternative pour résoudre le problème lié aux tests multiples, est l'utilisation d'approches bayésiennes (cas des tests associés à la méthode de Beaumont et Balding 2004) dans lesquelles l'interprétation des probabilités est différente de celle associée aux méthodes fréquentistes. Dans le cas d'une approche Bayésienne, la probabilité est une mesure du degré de certitude ou de croyance d'un événement. Une analyse Bayésienne d'une hypothèse résulte donc précisément dans la probabilité que cette hypothèse soit vraie. Par exemple, pour une *p-valeur* de 5%, la probabilité que l'hypothèse nulle soit vraie sur un échantillon en particulier sera de 5% (O'Hagan et Luce 2003). Ce dernier point est cependant toujours débattu entre statisticiens de cultures fréquentiste ou bayésienne.

- Biais lié à l'échantillonnage des données de polymorphismes chez le modèle humain

L'application des approches pan-génomiques chez les humains a également permis de soulever le problème du biais lié à l'échantillonnage des données de polymorphismes (« ascertainment biais » ou biais de recrutement) (Clark *et al.* 2005). En effet, chez l'homme, les approches pan-génomiques sont largement basées sur les données de polymorphismes SNPs générées par le projet Hapmap. Or ces SNPs ont été initialement découverts sur un nombre limité d'individus, et par la suite génotypés dans un échantillon plus grand (270 individus répartis dans des populations du Nigeria, du Japon, de la Chine et des Etats-Unis, www.hapmap.org/abouthapmap.html.fr). Le biais engendré par l'échantillonnage initial, peut mener à des taux de faux positifs plus élevés. Kelley *et al.* (2007) démontrent par exemple grâce à des jeux de simulations, que le biais peut conduire à un excès d'allèles en fréquence

intermédiaire (les allèles rares étant sous représentés). Les moyennes des valeurs de D de Tajima sur l'ensemble des loci seront donc biaisées (la moyenne « réelle » étant probablement inférieure), et des « outliers » avec des valeurs de D de Tajima significativement négatives pourront en réalité être des faux positifs. Certes ce biais peut être pris en compte pour corriger les données (par exemple, Kimura *et al.* 2007, Nielsen *et al.* 2005), mais l'information sur la façon dont ont été obtenues les données dans les banques de données SNPs est généralement insuffisante pour l'estimer correctement (Nielsen 2005).

En conclusion, les approches pan-génomiques, malgré leurs limites, ont permis de montrer l'importance des signatures de sélection, notamment sur de nouveaux gènes, ou de nouvelles régions du génome considérées jusqu'à présent non-fonctionnelles, par exemple sur des pseudo-gènes (gène dont la séquence est encore voisine de celles de gènes de structure fonctionnels, mais qui ne s'exprime plus) ou sur de l'ADN non codant (Biswas et Akey 2006). Chez les espèces d'arbres forestiers, ce type d'approche est encore limité par manque de données.

C. Application des tests de neutralité chez les espèces d'arbres forestiers

C.1. L'approche gènes-candidats

L'approche gènes-candidats est largement répandue chez les espèces d'arbres forestiers, les conifères en particulier, pour tester d'éventuelles signatures moléculaires de sélection naturelle. Malgré les ressources génomiques croissantes chez de nombreuses espèces (banques d'ESTs, ...) cette approche est favorisée à cause d'une étendue du DL relativement plus faible que chez d'autres espèces, et une taille du génome chez les conifères particulièrement large. Ceci impliquerait le génotypage de nombreux polymorphismes pour avoir une idée du profil de diversité sur l'ensemble du génome, et pour pouvoir détecter des loci potentiellement soumis à la sélection par des méthodes pan-génomiques.

D.1.1 Faible étendue des DL

Les déséquilibres de liaison entre sites peuvent s'étendre jusqu'à 50 kb chez l'homme, (Reich 2001, Pritchard et Przeworsky 2001, Ardlie *et al.* 2002), et jusqu'à 150 kb et 100 kb chez des espèces autogames (*Arabidopsis thaliana* et *Oryza sativa* respectivement, Nordborg *et al.* 2002, Garris *et al.* 2003). A l'inverse, chez les espèces d'arbres forestiers, les estimations du DL uniquement disponibles au sein de quelques gènes, montrent qu'il s'étend

sur des régions beaucoup moins grandes. Par exemple, on observe des valeurs de r^2 (estimation du DL correspondant à un coefficient de corrélation entre sites nucléotidiques) inférieure à 0.20 sur moins de 100 pb chez *Picea abies* (estimé sur 22 loci, Heuertz *et al.* 2006), 400 pb chez *Pseudotsuga menziesii* (18 loci, Krutovsky et Neale 2005), 1500 pb chez *Pinus taeda* (19 loci, Brown *et al.* 2004), et inférieure à 0.05 sur moins de 500 pb chez *Populus tremula* (5 loci, Invargsson 2005). Comment expliquer ces faibles fenêtres de DL chez ces espèces ? La recombinaison est le principal processus qui érode le DL au cours du temps. Or le paramètre de recombinaison d'une population, ρ , est égal à $4N_e r$ (où N_e est la taille efficace de la population et r , le taux de recombinaison par paire de base par génération). Le DL est donc à la fois dépendant de N_e et de r . Neale et Savolainen (2004) suggèrent que le faible DL observé chez les arbres forestiers serait dû principalement aux grandes tailles efficaces des populations favorisées par l'allogamie de ces espèces. Brown *et al.* (2004) estime par exemple la taille efficace de population de *Pinus taeda* à 560 000 individus, alors que chez l'homme, Wall (2003) l'estime entre 10 000 et 15 000.

C.1.2. Large taille du génome des conifères

La seconde raison expliquant le choix d'approches gènes-candidats plutôt que pan-génomiques, est la large taille du génome des conifères. Par exemple, la taille moyenne du génome a été estimée à 25.7×10^9 pb par chromosome chez *Pinus pinaster* (soit 51.4×10^{18} pour cette espèce diploïde) (Chagné *et al.* 2002) et 21×10^9 pb par chromosome chez *Pinus taeda* (Arumuganathan et Earle 1991). Cette taille peut atteindre 63.5×10^{18} chez *Pinus lambertiana* (Murray 1998). Ces tailles sont donc largement supérieures à celles estimées chez d'autres espèces (par exemple, autour de 5×10^9 chez le maïs et le riz, 16×10^9 chez le blé, 18×10^7 chez la drosophile, 34×10^8 chez les hommes, voir base de données DOGS, <http://www.cbs.dtu.dk/databases/DOGS/GBgrowth.php>). Chez *Populus deltoides* et *Quercus robur*, ces tailles (55×10^7 et 90×10^7 respectivement) sont néanmoins plus faibles que chez les conifères et les autres plantes.

C.2.3. Ressources en gènes candidats

L'approche gènes-candidats a pu être utilisée chez ces espèces grâce aux ressources croissantes de données de séquences disponibles dans les banques d'ESTs telles que TIGR (<http://www.tigr.org/>) et NCBI (<http://www.ncbi.nlm.nih.gov/>). Ces nombreux ESTs proviennent principalement d'études d'expression différentielle (gènes candidats expressionnels), et sont annotées à partir d'informations disponibles chez les espèces

modèles. D'autres arguments peuvent confirmer l'intérêt potentiel de certains gènes candidats, comme l'existence de co-localisation avec des QTLs d'intérêt (gènes candidats positionnels), par exemple avec des QTLs de résistance au froid chez le *Pseudotsuga menziesii*, où des études de cartographie fine existent (Krutovsky et Neale 2005, Wheeler *et al.* 2005)

C.2. Tests d'écarts à la neutralité pour des approches gènes-candidats et pan-génomiques

L'utilisation d'approches gènes candidats, telle que celle présentée pour le projet TREESNIPs en introduction a été également suivie dans plusieurs projets récents concernant des espèces d'arbres forestiers (projets ADEPT, <http://dendrome.ucdavis.edu/adept>, DIGENFOR, <http://www.pierroton.inra.fr/biogeco/genetique/projets/europe/digenfor>, POPYOMICS (<http://www.soton.ac.uk/popyomic>, et Figure 2 de l'introduction). Néanmoins, les tests de neutralité utilisés dans ces projets sont pour la plupart sensibles à la structure en populations (voir Tableau 1) et ne peuvent donc être appliqués qu'intra-population, ou doivent prendre en compte cette structure dans la réalisation des tests (par exemple, Invargsson 2005 chez le peuplier où la structure en populations est intégrée dans les simulations de coalescence qui permettent d'estimer les valeurs critiques des statistiques des tests de neutralité sous un modèle neutre). De plus, distinguer les événements démographiques d'événements sélectifs reste délicat dans le cas où des données sur peu de gènes existent, et cela nécessiterait un travail plus important de simulations permettant d'intégrer l'histoire évolutive des populations. Les informations disponibles sur des loci *a priori* neutres permettent d'avoir un bon point de départ pour ces simulations mais sont encore peu réalisées (voir cependant Pyharjarvi *et al.* 2007). Finalement, peu d'études chez les espèces d'arbres forestiers proposent des interprétations en terme de signatures moléculaires de sélection naturelle. Les quelques-unes qui existent se basent sur les résultats combinés de plusieurs statistiques de tests de neutralité (Tableau 2), mais quelles que soient leurs conclusions, elles restent des hypothèses à valider.

Quant aux approches pan-génomiques pour la détection d'« outliers », elles sont encore peu répandues chez les arbres forestiers, pour les raisons évoquées précédemment (faible *DL*, large taille du génome chez les conifères). L'étude de Scotti-Saintagne *et al.* (2004), bien que sur un nombre encore limité de 389 marqueurs de types différents, fut une des premières approches à appliquer les méthodes associées aux études pan-génomiques pour

Tableau 2. Quelques études mettant en évidence des écarts à la neutralité chez les espèces d'arbres forestiers et interprétés en terme de signatures moléculaires de sélection naturelle (approches gènes candidats).

<i>Espèce</i>	<i>Gènes candidats (fonctions)</i>	<i>Profils de diversité et résultats de tests de neutralité</i>	<i>Hypothèses</i>	<i>Référence</i>
<i>Pinus sylvestris</i>	<i>PHYO</i> et <i>PHYP</i> (impliqués dans la photosynthèse)	→ Diversité nucléotidique non synonyme faible ($\theta_\pi^{(a)}$ non synonyme entre 0 et 8×10^{-4}) au sein des populations espagnoles et finlandaises	Sélection purifiante au sein de ces populations	Garcia-Gil <i>et al.</i> 2003
<i>Pinus radiata</i> et <i>Pinus pinaster</i>	<i>CesA3</i> et <i>Pp1</i> (impliqués dans la formation du bois)	→ Diversité nucléotidique plus basse chez <i>P. radiata</i> que chez <i>P. pinaster</i> (θ_π de 1.9×10^{-3} et 2.4×10^{-3} respectivement)	Taille efficace plus petite et/ou signature moléculaire d'une sélection artificielle sur la croissance chez <i>P. radiata</i>	Pot <i>et al.</i> 2005
<i>Pinus pinaster</i>	<i>Pp1</i> (impliqué dans la formation du bois)	→ Valeur de F_{ST} (0.23) plus élevée que pour marqueurs <i>a priori</i> neutres (0.098 pour des AFLPs ^(b) , et 0.184 pour des SSRs ^(c))	Sélection diversifiante entre populations de l'aire de distribution du pin maritime	Pot <i>et al.</i> 2005 Mariette <i>et al.</i> 2001
<i>Pinus pinaster</i>	<i>CesA3</i> (impliqué dans la formation du bois)	→ Valeur de F_{ST} (-0.05) non significative + Valeur D de Tajima positive significative au sein de la population Aquitaine et sur l'ensemble de l'aire de distribution + Forte structure haplotypique	Sélection balancée au sein des populations de l'aire de distribution du pin maritime	Pot <i>et al.</i> 2005
<i>Pinus densata</i>	<i>ARA</i> (Arabinoglactan-like protein)	→ Valeurs D de Tajima et F de Fu et Li positives significatives (contrairement à 6 autres loci) au sein d'une population à l'est du plateau tibétain.	Sélection balancée au sein de cette population	Ma <i>et al.</i> 2006
<i>Pinus densata</i>	<i>DEH</i> (dehydrine)	→ Valeurs D et F de Fu et Li négatives significatives et $d_N/d_S < 1$ au sein d'une population au centre du plateau tibétain.	Sélection purificatrice au sein de cette population	Ma <i>et al.</i> 2006
<i>Pinus densata</i>	<i>POD</i> (peroxydase)	→ Valeurs F de Fu et Li négative significative et $d_N/d_S > 1$ au sein d'une population à l'est du plateau tibétain.	Sélection positive au sein de cette population	Ma <i>et al.</i> 2006
<i>Populus tremula</i>	<i>TI1</i> et <i>TI2</i> (impliqués dans la défense contre les pathogènes)	→ $K_A/K_S > 1$ + Valeur de H de Fay et Wu négative significative (en prenant en compte la structure en populations) + Valeur de Z_{NS} positive significative (en prenant en compte la structure en populations)	Sélection positive avec effet d'entraînement au sein des populations de <i>Populus tremula</i>	Invargsson 2005
<i>Cryptomeria japonica</i>	<i>Acl5</i> (impliqué dans la croissance et le développement de la plante)	→ Test <i>HKA</i> significatif + Valeurs négatives de D de Tajima et D et F de Fu et Li au sein d'une population non structurée	Sélection positive avec balayage sélectif au sein cette population	Kado <i>et al.</i> 2003

^(a) estimateur de la diversité nucléotidique, ^(b) Amplified fragment length polymorphism, ^(c) Single sequence repeat.

Tableau 3. Etudes mettant en évidence des écarts à la neutralité chez les espèces d'arbres forestiers et interprétés en terme de signatures moléculaires de sélection naturelle (approches pan-génomiques).

<i>Espèce</i>	Marqueurs utilisés	Résultats des tests de neutralité	Hypothèses	Référence
<i>Quercus robur</i> et <i>Quercus petraea</i>	389 marqueurs (12 isozymes, 303 AFLPs ^(a) , 13 SCARs ^(b) , 38 SSRs ^(c) , 23 SNPs ^(d)) répartis sur différents groupes de liaison.	→ Identification de 12% de loci « outliers » positifs d'après la méthode de Beaumont et Nichols (1996) majoritairement localisés au sein de gènes ou régions non anonymes	Loci « outliers » impliqués dans la divergence entre les deux espèces	Scotti-Saintagne <i>et al.</i> 2004
<i>Picea abies</i>	150 marqueurs (125 AFLPs ^(a) , 25 SSRs ^(c) , 2 ESTs ^(e)) répartis sur différents groupes de liaison	→ Identification de 3% de loci « outliers » positifs d'après la méthode de Beaumont et Nichols (1996)	Loci « outliers » avec signatures de sélection diversifiante	Acheré <i>et al.</i> 2005
<i>Cryptomeria japonica</i>	148 marqueurs (CAPS ^(f))	→ Identification de 3% de loci « outliers » positifs et 1% de « outliers » négatifs d'après la méthode de Beaumont et Nichols (1996) → Identification de 8% de loci « outliers » par paires de populations d'après la méthode de Vitalis <i>et al.</i> (2001)	Loci « outliers » avec signatures de sélection diversifiante ou homogénéisante	Tsumura <i>et al.</i> 2007
<i>Howea forsteriana</i> et <i>Howea belmoreana</i>	274 marqueurs (AFLPs ^(a))	→ Identification de 1% de loci « outliers » positifs d'après la méthode de Beaumont et Nichols (1996)	Loci « outliers » impliqués dans la divergence entre les deux espèces	Savolainen <i>et al.</i> 2006

^(a) Amplified fragment length polymorphism, ^(b) Sequence characterized amplified region, ^(c) Single sequence repeat, ^(d) Single nucleotide polymorphism, ^(e) Expressed sequence tag, ^(f) EST-based cleaved amplified polymorphic sequence.

étudier la divergence entre deux espèces de chêne (*Quercus robur* et *Quercus petraea*)(Tableau 3). Elle a permis de détecter 12% de loci « outliers » interprétés comme des signatures de sélection divergente entre espèces. Ces « outliers » étaient majoritairement des marqueurs localisés au sein de gènes et impliquaient que la divergence entre ces deux espèces résiderait majoritairement dans les régions fonctionnelles du génome. Une étude comparable a été appliquée chez deux espèces de palmiers (Savolainen *et al.* 2006), et met en évidence 1% de loci « outliers » positifs s'écartant de la distribution du F_{ST} sur l'ensemble du génome (Tableau 3). La présence de ces loci « outliers » probablement liés à des gènes soumis à de la sélection divergente, permet aux auteurs d'avancer l'hypothèse que ces deux espèces résulteraient d'une spéciation sympatrique (sans isolement géographique) suite à une sélection divergente au niveau de quelques régions du génome, plutôt que d'une spéciation allopatrique (après isolement géographique) qui affecterait le génome dans son ensemble. L'étude d'Acheré *et al.* (2005) sur l'épicéa et celle de Tsumura *et al.* (2007) sur le cèdre du Japon ont permis de détecter seulement 3% de loci « outliers » sur la base du F_{ST} (Tableau 3). Ces auteurs soulignent le fait qu'il est nécessaire d'utiliser préférentiellement des marqueurs dans des régions codantes pour optimiser la probabilité de détecter des loci sous sélection.

En conclusion, il a été jusqu'à présent difficile d'obtenir des preuves solides de signatures moléculaires de sélection naturelle avec les approches gènes-candidats utilisées classiquement chez les espèces d'arbres forestiers. De plus, les approches pan-génomiques encore peu répandues chez ces espèces, ne permettent de détecter qu'une faible proportion de loci soumis à de la sélection naturelle. L'originalité de cette thèse est d'appliquer une méthodologie issue des approches pan-génomiques basées sur le F_{ST} , en utilisant non pas un lot de marqueurs pris sur l'ensemble du génome, mais un lot de gènes candidats *a priori* impliqués dans la résistance à la sécheresse, et un lot de marqueurs *a priori* neutres comme point de départ à la construction de la distribution attendue. Les tests de neutralité basés sur le F_{ST} ne permettant de détecter que certains types de signatures de sélection (voir paragraphe B.1., et Tableau 1), il est également judicieux de coupler ces tests avec ceux basés sur d'autres statistiques (liés à la distribution des fréquences alléliques, au DL , etc...), qui s'appliquent à des populations non-structurées.

Chapitre II

Matériel végétal, gènes candidats et marqueurs microsatellites

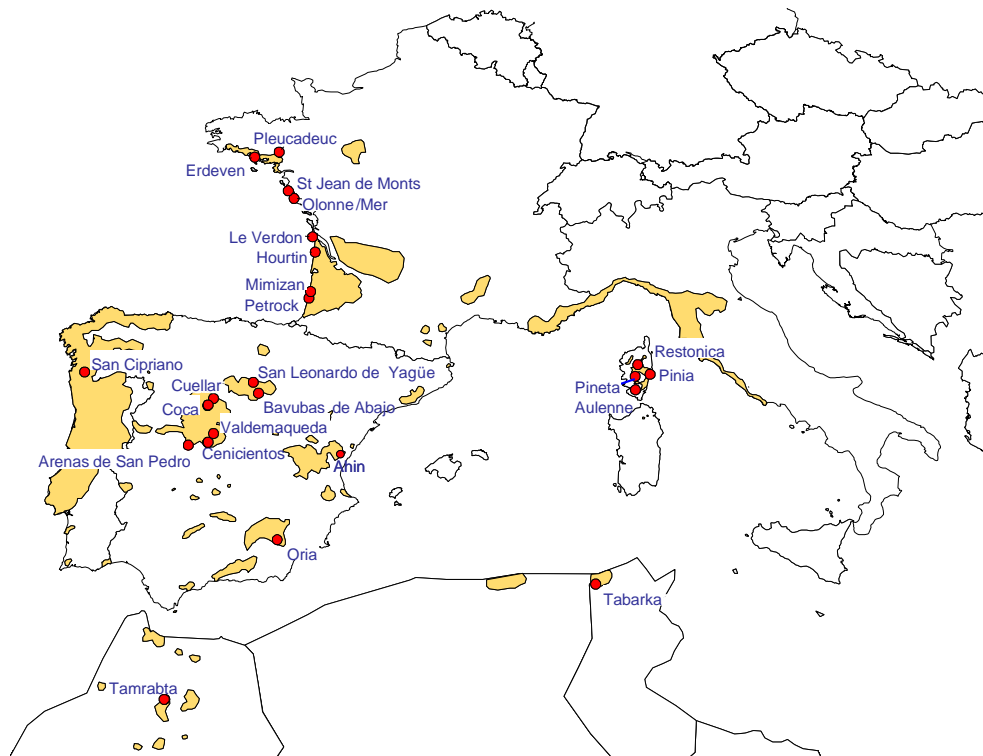
Afin d'étudier l'adaptation à la sécheresse chez le pin maritime, nous nous sommes intéressés à des populations naturelles de cette espèce provenant de milieux aux conditions climatiques contrastées. L'hypothèse de travail est que ces populations auraient évoluées différemment pour s'adapter aux conditions propres à leur habitat naturel. Dans ce chapitre, nous présentons dans un premier temps, l'échantillonnage des populations naturelles et quelques données climatiques relatives à leur lieu d'origine. Puis nous décrivons le choix des gènes candidats à la résistance à la sécheresse et l'obtention de séquences au sein de ces gènes. Enfin, nous présentons le choix des marqueurs microsatellites (SSRs) considérés comme neutres (n'ayant pas été soumis à la sélection naturelle) qui ont été utilisés pour estimer une différenciation moléculaire représentative de l'histoire démographique au sens large des populations de pin (dérive génétique colonisation, expansion démographique, etc...). Ils ont notamment été utilisés pour identifier des groupes homogènes de populations montrant peu ou pas de différenciation intra-groupe, dans le but de comparer la différenciation moléculaire entre groupes pour les SSRs et les gènes candidats, et ainsi de détecter d'éventuelles signatures moléculaires de sélection naturelle.

A. Populations échantillonnées et données climatiques

Des aiguilles et des graines ont été échantillonnées sur 24 populations couvrant l'ensemble de l'aire de distribution naturelle du pin maritime (Figure 7). Pour 15 de ces populations, des données climatiques relatives aux précipitations et températures ont pu être obtenues. A l'exception des populations tunisienne et marocaine, ces données couvrent des périodes de 37 années pour les précipitations et 23 années pour les températures, antérieures à l'année 2001, date d'échantillonnage au sein des populations (Tableau 4). Six variables expliquant au plus 70% de la variation de chacune des autres variables (Tableau 5) ont été par la suite choisies pour décrire les conditions climatiques du lieu d'origine des 15 populations.

Il existe une large gamme de variation pour ces conditions climatiques, qui ont pu entraîner des adaptations locales différentes. Les moyennes des précipitations totales annuelles (ANMR) varient de 348 mm pour Oria (Espagne) à 1343 mm pour Pineta (Corse) (Tableau 4). Les moyennes des précipitations totales du mois le plus sec (RDRYM) de 2 mm pour Oria et 3 mm pour Pinia (Corse) (comparé à une moyenne de 11.5 mm pour les autres populations) témoignent également de périodes estivales de sécheresse particulièrement sévères pour ces deux populations (Tableau 4). Parmi les variables liées à la température, les températures moyennes annuelles (ANMT) varient de 9.9° pour San Leonardo de Yagüe (Espagne) jusqu'à

Figure 7. Localisation des 24 populations échantillonnées sur l'ensemble de l'aire de distribution naturelle du pin maritime.



Région	Nom de la population	Code de la population
Corse	Restonica	2
Corse	Aullène	10
Corse	Pineta	11
Corse	Pinia	15
Espagne Centre	Cenicientos	20
Espagne Centre	Arenas de San Pedro	21
Espagne Centre	Coca	22
Espagne Centre	Cuellar	23
Espagne Centre	Valdemaqueda	24
Espagne Centre	San Leonardo de Yagüe	25
Espagne Centre	Bayubas de Abajo	26
Espagne Nord	San Cipriano	27
Espagne Est	Ahin	28
Espagne Sud	Oria	29
Maroc	Tamrabta	30
Côte atlantique française Ouest	Petrock	40
Côte atlantique française Ouest	Mimizan	41
Côte atlantique française Ouest	Hourtin	42
Côte atlantique française Ouest	Le Verdon	43
Côte atlantique française Ouest	Olonne/mer	44
Côte atlantique française Ouest	St Jean de Monts	45
Côte atlantique française Ouest	Pleucadec	46
Côte atlantique française Ouest	Erdeven	47
Tunisie	Tabarka	50

Tableau 4: Coordonnées géographiques et données climatiques sur 15 populations.

Nom de la population	Code de la population	Longitude ^(a)	Latitude (N) ^(b)	Altitude (m)	ANMR (mm) ^(c)	RDRYM (mm) ^(d)	ANMT (°C) ^(e)	ANMT variance (°C) ^(f)	MTCM (°C) ^(g)	MTWM (°C) ^(h)
Pineta	11	9°02' (E)	41°58'	750	1343	12	12.2	2.18	4.1	21.4
Pinia	15	9°29' (E)	42°01'	10	705	3	na (i)	na	na	na
Mimizan	41	1°18' (W)	44°08'	37	995	16	13.8	0.45	6.3	21.2
Pleucadeuc	46	2°20' (W)	47°47'	80	716	11	11.6	0.32	4.0	19.0
Arenas San Pedro	21	5°08' (W)	40°02'	1359	957	5	13.8	0.65	4.8	24.4
Coca	22	4°05' (W)	41°22'	788	428	7	12.2	0.26	3.2	22.5
San cipriano de Ribaterme	27	8°42' (W)	42°08'	310	661	18	15.1	0.26	6.9	24.3
Oria	29	2°37' (W)	37°52'	1232	348	2	12.5	0.23	3.9	22.9
Bayubas de Abajo	26	2°52' (W)	41°31'	955	405	6	12.0	0.31	3.2	22.2
Cenicientos	20	4°28' (W)	40°16'	na	579	6	12.4	0.26	3.9	22.7
Cuellar	23	4°19' (W)	41°24'	858	458	8	12.3	0.34	3.2	22.7
San Leonardo de Yague	25	3°4' (W)	41°49'	1050	671	14	9.9	0.29	1.9	19.5
Valdemaqueda	24	4°17' (W)	40°31'	926	594	8	10.7	0.40	2.6	21.0
Tamrabta	30	5°0' (W)	34°0'	1600	763	na	17.2	na	na	na
Tabarka	50	8°54' (E)	36°65'	66	999	na	17.9	na	na	na
Moyenne					708	9	13.12	0.50	4.0	22.0

^(a) Longitude Est (E) or Ouest (W).

^(b) Latitude Nord (N).

^(c) ANMR : Précipitations totales annuelles moyennées entre 1964 et 2001 pour les populations françaises, corses (MétéoFrance, <http://www.meteofrance.com/>) et espagnoles (A.F. Cancio, comm. personnelle) avec en moyenne 10% d'années manquantes. ANMR estimé de 1933 à 1963 pour la population Tamrabta (Wahid et al. 2004) et de 1901 à 1990 pour la population Tabarka (JBIC/SCET Tunisie, 2005)

^(d) RDRYM : Moyenne des précipitations totales du mois le plus sec.

^(e) ANMT : Températures annuelles moyennées entre 1978 et 2001 pour les populations françaises, corses et espagnoles (même source de données que pour ANMR) avec en moyenne moins de 1% de données manquantes. ANMT estimé sur 18 années dans la station de Meknes pour la population Tamrabta

(<http://www.weatherbase.com/>) et de 1901 à 1990 pour la population Tabarka (même source de données que pour ANMR).

^(f) ANMT Variance : Variance de ANMT entre années.

^(g) MTCM : Moyenne des températures du mois le plus froid.

^(h) MTWM : Moyenne des températures du mois le plus chaud.

⁽ⁱ⁾ Données non disponibles.

Tableau 5 : Matrice des coefficients de détermination r^2 (%) pour 6 variables climatiques.

	ANMR	RDRYM	ANMT	ANMT variance	MTCM	MTWM
ANMR		0.165	0.001	0.635	0.137	0.139
RDRYM			0.168	0.005	0.697	0.184
ANMT				0.075	0.626	0.326
ANMT variance					0.014	0.034
MTCM						0.002
MTWM						

15.1°C, 17.2° et 17.9° pour San Cipriano (Espagne), Tamrabta (Maroc) et Tabarka (Tunisie), respectivement (Tableau 4). Pour les populations dont les données sont disponibles, les températures moyennes des mois le plus froid (MTCM) et le plus chaud (MTWM) varient d'environ 5°C entre valeurs extrêmes (Tableau 4). Enfin, les variances des températures moyennes annuelles au cours des années (ANMT variance) restent inférieures à 0.7°C pour chacune des populations, excepté pour la population Pineta (ANMT variance de 2.18°C, Tableau 4) ce qui s'expliquerait par une augmentation d'environ 3°C des valeurs d'ANMT de cette population durant la période de 1997 à 2000.

B. Gènes candidats

B.1 Stratégie pour le choix des gènes candidats

Dans un premier temps, 13 gènes candidats ont été sélectionnés par le laboratoire d'accueil (INRA, Pierroton) et par le laboratoire espagnol (INIA, Madrid) partenaire du projet européen Treesnips (Tableau 6). Ces gènes candidats ont des fonctions déjà connues chez plusieurs espèces, et ont été choisis sur la base de leur rôle putatif dans la réponse au stress hydrique (Tableau 6), l'objectif étant d'avoir un ensemble de gènes intervenant dans différents mécanismes de réponse au stress hydrique, tels que le transport ionique, l'osmoprotection ou la protection des macromolécules (Figure 8 ; Yamaguchi-Shinozaki *et al.* 2002). Certains de ces gènes étaient également des candidats expressionnels, puisqu'ils ont été choisis sur la base de précédents résultats de profil d'expression sous différentes conditions de stress hydrique chez le pin maritime et chez d'autres espèces de conifères (Tableau 6 ; Dubos et Plomion 2003; Chaumeil 2006; Costa *et al.* 1998; Padmanabhan *et al.* 1997; Chang *et al.* 1996).

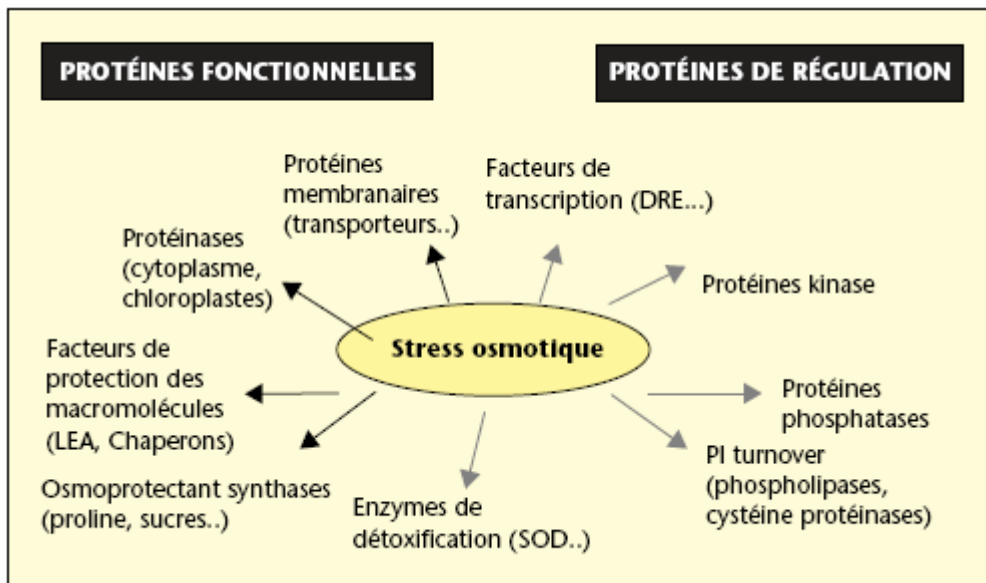
Des séquences homologues ont été ensuite recherchées par « Blastn » dans les banques d'ESTs, telles que NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>) et Pine Gene Index (<http://compbio.dfci.harvard.edu/tgi/cgibin/tgi/gimain.pl?gudb=pine>). Leur alignement grâce au logiciel Bioedit ver 7.0.5 (Hall 1999) a permis de rechercher le polymorphisme *in silico* de type SNP (mutation d'une seule paire de base) et Indel (insertion ou délétion de bases) et dessiner des amorces oligonucléotides dans des régions conservées. Le dessin des amorces a été réalisé à l'aide du logiciel Primer 3 (http://frodo.wi.mit.edu/cgibin/primer3/primer3_www.cgi) et les problèmes éventuels de formation de dimères d'amorces (hybridation de deux amorces due à leur similarité de séquences) et d'épingle à cheveu (« hairpin » en anglais, dû à l'hybridation d'une amorce sur elle-même) ont été vérifiés grâce au logiciel OligoCalc

Tableau 6. Description des 13 gènes candidats.

Gène	Code du gène	Accession / homologues TIGR ou GenBank ^(a)	Evidence d'expression différentielle en conditions de stress hydrique chez des espèces de conifères	Catégorie fonctionnelle	Rôle putatif dans la réponse à la sécheresse	N ^(b)	Longueur amplifiée (pb)				
							Total	Région codante	Région non codante		
									5' UTR	Intron	3' UTR
Put. Arabinogalactan/Proline Rich Protein	<i>PR-AGP4</i> ^(p)	AM501931/TC67837	Sous expression en condition de stress osmotique dans des racines de <i>P. pinaster</i> ^(c)	Formation de la paroi cellulaire	Inconnu	208	1630	360	88	708	474
Put. Arabinogalactan / Glycin-Rich Protein	<i>GRP3</i> ^(p)	AM502293/BX255640	Sous expression en condition de stress osmotique dans des racines de <i>P. pinaster</i> ^(c)	Formation de la paroi cellulaire	Inconnu	172	450	341			109
Caffeoyl-CoA-O-methyltransferase	<i>CCoAOMT</i> ^(p)	AM502291/TC73846	Sur expression en condition de stress hydrique dans des aiguilles de <i>P. pinaster</i> ^(d)	Biosynthèse des lignines	Epaissement de la paroi cellulaire et maintien de la pression osmotique intra cellulaire ^{(g) (h)}	204	1519	465	708	346	
Put. Caffeic-O-methyltransferase	<i>COMT</i> ^(p)	AM774402/TC66403		Biosynthèse des lignines	Epaissement de la paroi cellulaire et maintien de la pression osmotique intra cellulaire ^{(g) (h)}	20	1400	1019	35	346	
Put. Glucan-endo-1,3-beta-glucosidase precursor	<i>Glucan</i> ^(p)	AM502290/TC66277	Sur expression en condition de stress osmotique dans des racines de <i>P. pinaster</i> ^(c)	Métabolisme des carbohydrates (amidon et sucres)	Ajustement osmotique lié au métabolisme des sucres ⁽ⁱ⁾	158	1357	161		1004	192
Put. Inositol-3-phosphate synthase	<i>Ino3</i> ^(p)	AM774401/TC64990		Biosynthèse des RFO (Raffinose Family Oligosaccharides)	Ajustement osmotique lié au métabolisme des sucres ⁽ⁱ⁾	24	859	479		380	
Dehydrin 1	<i>dhn-1</i> ^(p)	AM502292/TC57901	Sur expression en conditions de froid, blessures et stress hydrique chez <i>Picea glauca</i> ^(m)	Protéines LEA (Late Embryogenesis Abundant)	Protection des structures cytoplasmiques ⁽ⁱ⁾	132	828	668	44	116	
Dehydrin 2	<i>dhn-2</i> ^(q)	EU020010/TC58333	Sur expression en condition de stress hydrique dans des racines de <i>P. taeda</i> ^(o)	Protéines LEA (Late Embryogenesis Abundant)	Protection des structures cytoplasmiques ⁽ⁱ⁾	56	513	402		111	
Early responsive to dehydration 3	<i>erd3</i> ^(q)	EU020011/TC65051		Putative methyltransferase possédant un domaine de liaison avec l'enzyme S-adenosyl methionine	Inconnu	58	796	554		201	41
Protein phosphatase 2C	<i>pp2c</i> ^(q)	EU020014/TC62065		Phosphatase similaire à <i>abil</i> chez <i>A. thaliana</i>	Voie de transduction du signal ^{(i) (k)}	53	591	414		177	
Cystein protease (pseudotzain)	<i>rd21A</i> ^(q)	EU020015/TC72918		Cystéine protease similaire à <i>rd21A</i> chez <i>A. thaliana</i>	Réparation des dégâts (dégradation des protéines) ⁽ⁿ⁾	54	942	561	186	195	
ABA-and-WDS-induced-gene-1	<i>lp3-1</i> ^(q)	EU020012/TC73436	Sur expression en condition de stress hydrique dans des racines de <i>P. taeda</i> ^{(e) (f)}	Protéine ASR (abscisic-, stress- and ripening-induced)	Inconnu	57	412	146			266
ABA-and-WDS-induced-gene-3	<i>lp3-3</i> ^(q)	EU020013/TC73474	Sur expression en condition de stress hydrique dans des racines de <i>P. taeda</i> ^(e)	Protéine ASR (abscisic-, stress- and ripening-induced)	Inconnu	54	407	240		167	
Total						1250	11704	5810	1061	3751	1082

^(a) Numéros d'accèsion des séquences soumises / Contig ou numéros d'accèsion des homologues dans les bases de données TIGR (<http://combio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=pine>) ou GenBank (<http://www.psc.edu/general/software/packages/genbank/genbank.html>); ^(b) Nombre de séquences; ^(c) Dubos *et al.* 2003; ^(d) Costa *et al.* 1998; ^(e) Padmanabhan *et al.* 1997; ^(f) Chang *et al.* 1996; ^(g) Cruz *et al.* 1992; ^(h) Joyce *et al.* 1983; ⁽ⁱ⁾ Ramanjulu *et al.* Bartels 2002; ^(j) Bray 1997; ^(k) Shinozaki *et al.* Yamaguchi-Shinozaki 2000; ^(l) Rodriguez 1998; ^(m) Richard *et al.* 2000; ⁽ⁿ⁾ Ingram *et al.* Bartels 1996; ^(o) Montré à travers les profils d'expression obtenus en utilisant l'outil "Magic Gene Discovery" (Laboratoire de Genomique et Bioinformatique, Université de Georgie; www.funngen.org); ^(p) Gènes candidats amplifiés et séquencés par la laboratoire d'accueil INRA, Pierroton; ^(q) Gènes candidats amplifiés et séquencés par le laboratoire partenaire du projet (INIA, Madrid).

Figure 8. Gènes induits par le stress et leur fonction possible dans la réponse à la tolérance au stress osmotique (d'après Yamaguchi-Shinozaki *et al.* 2002).



(<http://www.basic.northwestern.edu/biotools/OligoCalc.html>). Les amorces dessinées (Tableau 7) ont permis d'amplifier plusieurs fragments de gènes candidats entre 500 et 900pb, couvrant dans certains cas la longueur totale des gènes ou sinon, des régions localisés en 3'UTR (régions non transcrites) souvent plus spécifiques aux différents membres de familles multigéniques. Sur les 13 gènes, 7 ont été séquencés dans le cadre de cette thèse au laboratoire INRA de Pierroton, et 6 autres gènes l'ont été par le laboratoire partenaire (INIA, Madrid) du projet TREESNIPs (Tableau 6). J'ai analysé l'ensemble des données de séquences produites sur ces gènes dans un article sur la différenciation moléculaire entre populations (voir Chapitre IV et Annexe 1).

Ces 13 gènes candidats peuvent être répartis dans 8 catégories fonctionnelles. Même si ces catégories fonctionnelles, lorsqu'elles sont connues, ne sont pas directement liées à la résistance à la sécheresse, des preuves expérimentales (physiologiques, profils d'expression...) laissent supposer qu'elles peuvent intervenir de façon directe ou indirecte dans la réponse au stress hydrique.

B.2. Formation de la paroi cellulaire

Les gènes *PR-AGP4* et *GRP3* codent pour des protéines de la paroi cellulaire, appartenant à des familles de protéines parmi les plus abondantes, et souvent riches en un ou deux acides aminés, avec des domaines de séquence hautement répétés (Showalter 1993, Cassab 1998). La protéine codée par *PR-AGP4* fait partie à la fois des arabino-galactanes (« AGPs ») riches en différents acides aminés (Hydroxyproline, Alanine, Thréonine, Glycine et Sérine) et des « PRPs » riches en proline. Quant à *GRP3*, il correspond à un gène codant pour une protéine riche en glycine (classe des « GRPs »).

Les « PRPs », « GRPs » et « AGPs » seraient impliquées de manière générale dans le développement et la croissance de la plante. Les « GRPs » localisées dans les parois cellulaires sont des protéines de structure dont la fonction serait importante durant le développement vasculaire, en fournissant à la fois élasticité et robustesse aux tissus conducteurs des plantes (Showalter 1993). Certaines « PRPs » et « GRPs » joueraient un rôle lors de la réponse à un stress causé par des pathogènes et durant la cicatrisation des blessures (Showalter 1993, Cassab 1998). De même, certaines « AGPs » seraient très abondantes dans les gommés secrétées lors d'une blessure et constitueraient une barrière physique aux attaques des pathogènes (Cassab 1998).

Quelques études expérimentales mettent également en évidence une relation entre certaines de ces protéines et la réponse au stress hydrique, par exemple chez le maïs (Gomez *et al.* 1988), le riz (Mundy et Chua 1988), ou la tomate sauvage (sous-expression d'une « PTGRP », protéine riche en proline, thréonine et glycine, en conditions de stress, Yu *et al.* 1996). Dans ce dernier cas,

Tableau 7. Paires d'amorces utilisées pour l'amplification des fragments des 7 gènes candidats sélectionnés par le laboratoire d'accueil (INRA, Pierroton).

Code gène	Amorce "forward" (5'-3')	Amorce "Reverse" (5'-3')	Longueur (pb)
<i>PR-AGP4</i>	AAACTAGCCGAGCCATCAAG	AGGAGATCTGGACTCAACTTCC	570
	CTGTGTGGCGATAAGATCCA	CGGGTTCGATCTCGGTAAC	510
	AAACTAGCCGAGCCATCAAG	CGGGTTCGATCTCGGTAAC	890
	CCAATGGTTAAATGGGGAAG	CACTCTCTTCCACTGATGCAA	635
	GGTGAGGCTTCTGCAGAGGAAAT	CCCCTGAAATCCGTTTACTAC	600
	TGCCAGCAAAGCATTTACCG	CCCCTGAAATCCGTTTACTAC	890
<i>CCoAOMT</i>	TTGAGGCAACCTACATTCATTG	TGCTGAGAAGCGAGTAACCA	577
	GCATAAACCAAACTCAACTACGG	AAAAAGCTCAGAGATCAAGGGG	800
	TCCCTGGTAAGCCWCTCAAC	GGGCCCTCTCTGAAGTCAAT	380
	TTTTGCAGTTTTTGAGGAAATGGC	GGGCCCTCTCTGAAGTCAAT	814
<i>GRP3</i>	AGAGAGATTTTATTTTGCGACCC	TAAACCTAGACCGTTCGGAGC	460
	TCTGCTCGGAACGGTCTAGG	AAGCTTCCCCTAGGGAGAGG	90
<i>Glucan</i>	ATCTCAAGCCCGGTCTACT	TCCTACAAGGTCTCCCAATGA	679
	TGGGAAATCATCAAGGAAGG	CCTCCTGTGTAAGAAACCCTGT	801
	GGCCAGTGCTTTACTCCAAA	TAGTTTTGTGGGATGGGTGA	582
	CCCATCCCACAAAATATC	GATAACACCCACATCGCTAC	838
<i>Ino3</i>	ACAATCACCTCGGTAACAAC	GACCACCCTGTTTTTCTTCT	707
<i>Dhn1</i>	GTCGAACACGGGCATGGTC	CTCAGCTTCTTCAGGGGTGGT	116
	ACACGGGCATGGTCACGA	CAGAGCTGGAGCTGGAAGTGT	171
	CTTACCGCACGAACAGTT	GCCGGGGAGTTTTTCTTT	248
	GAAGAAAGGGTCGAAGGACAA	GTGCTTTCCATCACCAGG	622
<i>COMT</i>	GCATCTTCTGCGCCTCCAAA	TCAGACAGACTCTGGAAAAGCA	700
	TGAATTGTTTTTCAAATTAATCACATC	ACTATGCGCCACCATTACCA	500

Figure 9. Alignements de trois membres de la famille des « GRP ». La dernière séquence est une séquence obtenue dans la thèse, appartenant au membre 3. Les codons « start » et « stop » sont indiqués par des flèches. On peut voir clairement que la partie codante du gène est très conservée entre membres et entre espèces (même en 5'UTR pour les membres 2 et 3 de *P. pinaster*), mais avec des polymorphismes d'indels fixés entre membres dans la partie 3'UTR.

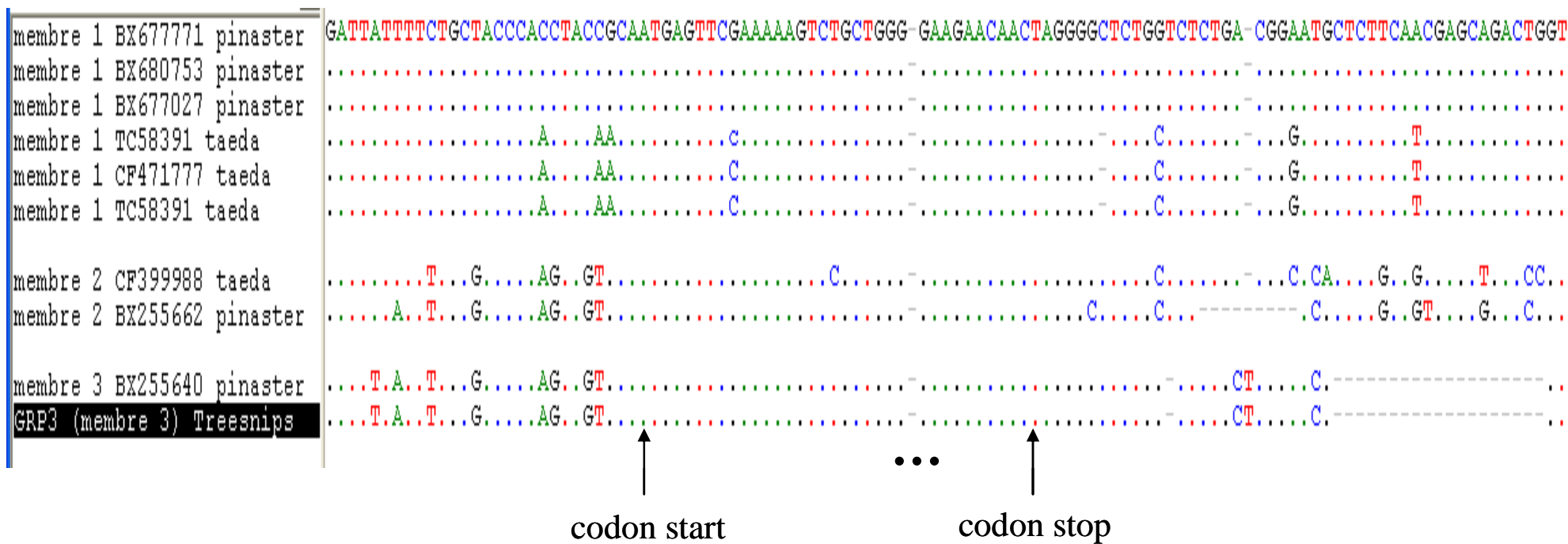
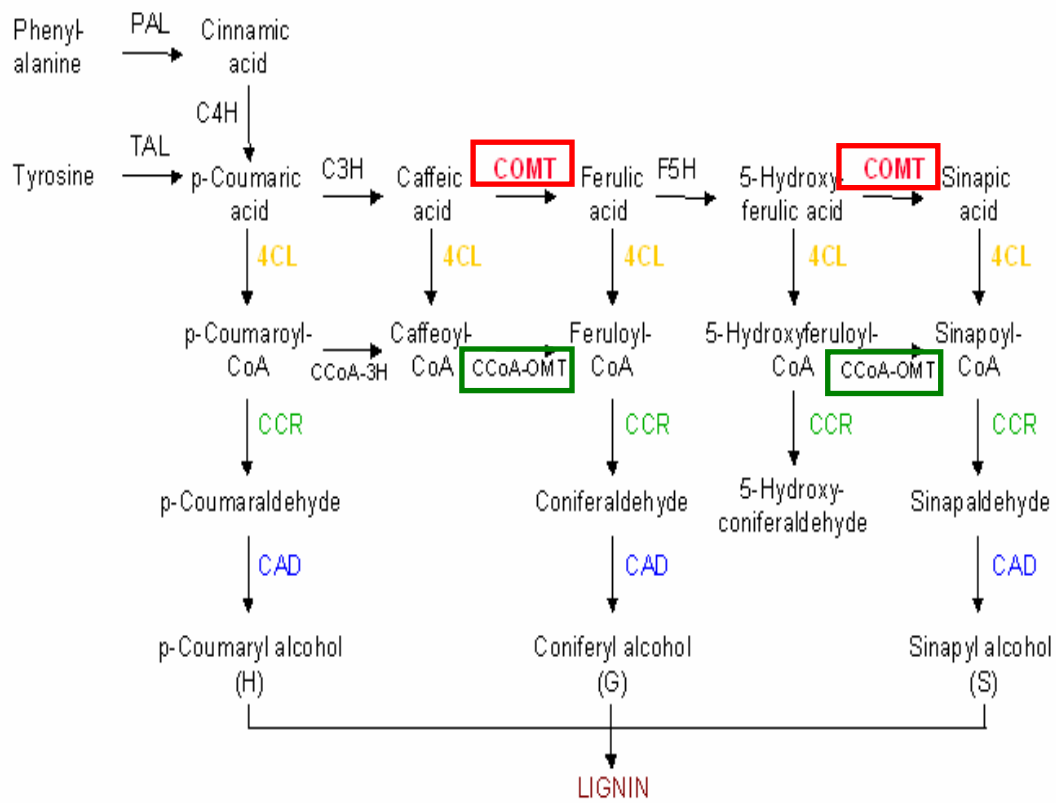


Figure 10. Voie de biosynthèse des lignines (source : base de données Pine Gene Index).



les résultats ont été interprétés comme une conséquence directe du remodelage des parois cellulaires lors d'un stress hydrique faisant intervenir la synthèse de protéines nouvelles et l'hydrolyse d'autres protéines telles que la PTGRP. Par ailleurs, une étude chez le pin taeda montre une diminution de l'expression de deux protéines « AGPs » dans le xylème en conditions de stress hydrique des plantules (No et Loopstra 2000). Le gène candidat choisi *PR-AGP4* est sous exprimé au niveau des racines et aiguilles de plantules de pin maritime stressés (Dubos *et al.* 2003). Yang *et al.* (2005) montre également que le gène homologue *ptaAGP4* chez le pin taeda est plus exprimé dans un écotype originaire d'une région à très faibles précipitations par rapport à un autre écotype du Sud de la Louisiane originaire d'une région plus humide. Le gène *GRP3* fait partie d'une petite famille multigénique qui a rendu difficile le dessin d'amorces et l'amplification du gène (Figure 9). Cependant, il a été choisi en tant que candidat expressionnel suite aux résultats de Dubos *et al.* (2003), qui ont montré que ce membre était fortement sous-exprimé dans les aiguilles et racines en conditions de stress hydrique.

B.3. Lignification

CCoAOMT et *COMT* sont deux gènes codant pour des enzymes de la voie de biosynthèse des lignines (Figure 10) qui sont les principaux composants de la paroi cellulaire avec la cellulose et l'hémicellulose. Des études physiologiques ont souligné l'implication du processus de lignification lors d'un stress hydrique : une étude sur les racines de sorgho a par exemple permis de relier la diminution de la conductance hydraulique lors d'un stress hydrique avec la lignification des parois cellulaires de différents tissus racinaires, contribuant ainsi à garder une pression osmotique intra cellulaire et limiter les pertes en eau (Cruz *et al.* 1992). Chez le radis, on a également mis en évidence une relation entre le processus de lignification et l'épaississement des parois cellulaires du parenchyme observé pendant l'application d'un stress hydrique (Joyce *et al.* 1983).

Chez les arbres forestiers, l'expression de transcrits de *CCoAOMT* et *COMT* peut varier de façon significative chez le pin maritime lors de la formation du bois de compression (dû à un accroissement accéléré en réponse à une sollicitation extérieure telle que le vent ou la recherche de lumière) (Plomion *et al.* 2000). Or ce bois présente une conductance hydraulique inférieure à celle du bois opposé (Mayr et Cochard 2003). Costa *et al.* (1998) met également en évidence une sur-expression de la protéine codée par *CCoAOMT* dans les aiguilles de pin maritime lors de l'application d'un stress hydrique.

B.4. Métabolisme des carbohydrates

Au sein du métabolisme des carbohydrates, et plus précisément celui de l'amidon et du sucre, l'enzyme Glucan-endo-1,3-beta-glucosidase (produite par le gène noté *Glucan*) participe à la synthèse du dextrose (autrement nommé D-glucose) (Figure 11) comptant parmi les nombreux sucres solubles. L'inositol-3-phosphate synthase (codé par le gène *Ino3*) est également une enzyme de la voie de biosynthèse des carbohydrates, qui intervient au sein du métabolisme de l'inositol-phosphate (Figure 11). Cette voie métabolique est celle des RFOs (« Raffinose Family Oligosaccharides »), principaux sucres solubles au sein des graines et autres tissus végétatifs des plantes (Mundree *et al.* 2002).

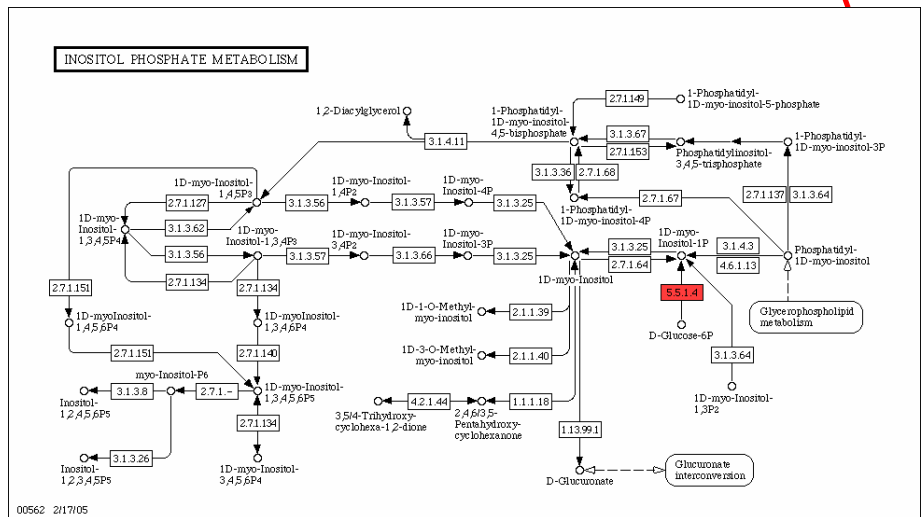
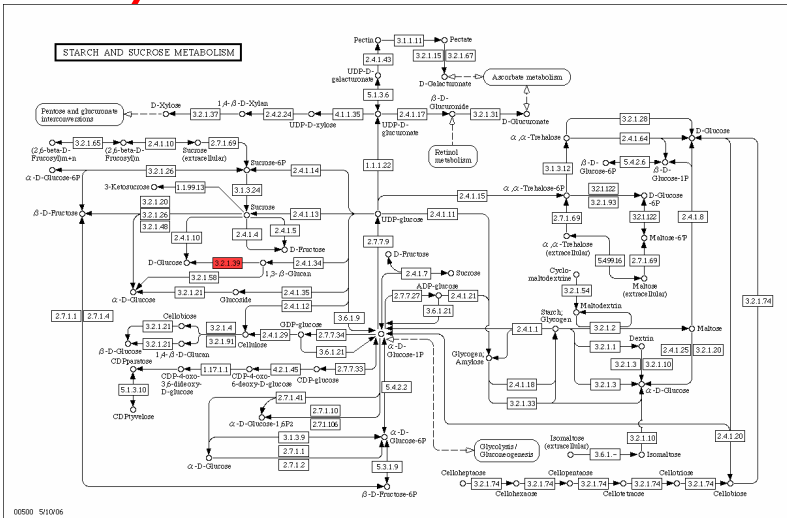
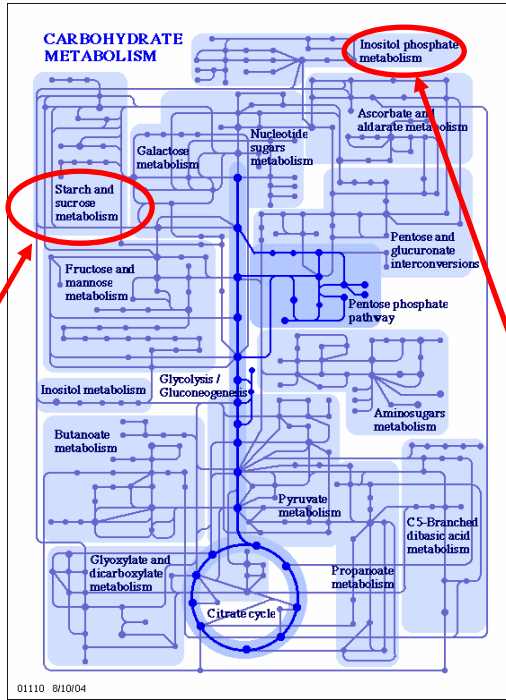
Or, l'accumulation des sucres solubles dans le cytoplasme cellulaire des plantes est très souvent observée lors d'un déficit en eau. Cette accumulation de solutés au sein des cellules entraînerait un différentiel de concentration entre l'intérieur de la cellule et l'extérieur, et favoriserait la rentrée d'eau à l'intérieur de la cellule et le maintien de la turgescence tissulaire (Ramanjalu et Bartels 2002, Bray 1993, Bray 1997, Bohnert et Sheveleva 1998). Ces sucres pourraient également intervenir dans la protection des membranes, agissant comme des agents protecteurs qui stabiliseraient les protéines lors d'un stress hydrique (Ramanjalu et Bartels 2002). Au niveau moléculaire, Dubos *et al.* (2003) met en évidence une sur-expression du gène *Glucan* lors d'un déficit hydrique dans les racines de pin maritime.

B.5. Protéines LEA

Dhn1 et *dhn2* sont deux gènes codant pour des déhydrines, qui appartiennent au groupe des protéines LEA (« Late Embryogenesis Abundant ») initialement identifiées au cours des phases de maturation et de dessiccation pendant le développement des graines de coton (Baker *et al.* 1988). Parmi ces protéines LEA, les déhydrines sont les plus largement étudiées et plusieurs fonctions putatives en réponse à un stress hydrique ont été proposées, telles que la stabilisation des structures membranaires de la cellule (Dure 1993, Close 1996, 1997), ou la régulation du potentiel osmotique cellulaire (Nylander *et al.* 2001, Dure 1993). On sait également que la transcription de ces protéines pourrait être régulée par une perte hydrique cellulaire, mais également par l'ABA (acide abscissique), une hormone omniprésente chez les plantes qui interviendrait en particulier lors d'une réponse au stress hydrique (voir Paragraphe B.6 ci-dessous) (Ingram et Bartels 1996, Allagulova *et al.* 2003).

Chez les arbres forestiers, plusieurs études montrent une accumulation des transcrits de gènes de déhydrines en période de stress hydrique. Chez le peuplier, Caruso *et al.* (2002) mettent en évidence la sur-expression d'un homologue de *dhn1*, durant des stress hydriques ou relatifs au

Figure 11. Voies de biosynthèse des carbohydrates, dont celui de l'amidon et du sucrose, et celui de l'inositol phosphate (source : base de données Pine Gene Index).



froid. De même, chez l'épicéa blanc, Richard *et al.* (2000) observent l'accumulation des transcrits de l'homologue de *dhn1* dans les aiguilles de *Pinus pinaster* après 48h de stress hydrique. Chez le pin taeda, l'étude de Watkinson *et al.* (2003) note également la sur-expression d'un homologue de *dhn2* dans les aiguilles en conditions de stress hydrique léger, qui a ensuite été validée par PCR en temps réel (RT-PCR).

B.6. Un facteur de régulation de l'ABA

L'hormone ABA peut intervenir lors de la transduction d'un signal (c'est-à-dire la perception du signal par une cellule jusqu'à la régulation de l'expression de gènes) suite à plusieurs stress abiotiques (sécheresse, salinité, gel, refroidissement, blessure, hypoxie, lumière, pathogènes) (Bray 1997). Lors d'un stress hydrique, elle peut agir comme signal pour déclencher des changements dans la conductance stomatale (Bray 1997), mais également déclencher ou réguler l'expression de certains gènes, comme les gènes *rd22* et *rd29B* identifiés en réponse au stress hydrique chez *Arabidopsis thaliana* (Shinozaki et Yamaguchi-Shinozaki 2000). Des expériences sur des mutants d'*A. thaliana* ont montré que des protéines contenant un domaine phosphatase 2C, de deux gènes *ABI1* et *ABI2*, pouvaient jouer un rôle de régulation de la voie de transduction du signal par l'ABA (Leung *et al.* 1994, 1997, Meyer *et al.* 1994, Rodriguez 1998, Merlot *et al.* 2001). Le gène qui nous intéresse, *pp2c*, codant pour une phosphatase 2C, est un homologue du gène *ABI1*, qui ferait partie d'une voie de phosphatase/kinase modulant la sensibilité des canaux ioniques des cellules « sentinelles » face aux cascades de signaux induites par l'ABA (Leung *et al.* 1998, Hugouvieux *et al.* 2001).

B.7. Des gènes aux fonctions inconnues, dont l'expression est induite par un stress hydrique

Les fonctions de gènes *erd3*, *rd21A*, *lp3-1* et *lp3-2* restent encore méconnues, mais ces gènes ont été choisis en raison des résultats d'expression différentielle lors d'un stress abiotique. *Rd21A* code pour une cysteine-proteinase, et a une expression qui est régulée par l'ABA, contrairement au gène *erd3* (Shinozaki et Yamaguchi-Shinozaki 2000). Ces deux gènes ont été obtenus par homologie avec deux ESTs de *Pinus taeda*, provenant d'un unigène basé sur 6 bibliothèques d'ESTs du xylème (Gonzalez-Martinez *et al.* 2006a).

Lp3-1 et *lp3-3* font partie d'une autre famille de gènes connue sous le nom d'ASR (« ABA/water stress/ripening »), qui peuvent être induits par un stress hydrique ou par la maturation du fruit. Ces gènes furent d'abord identifiés chez la tomate (Thompson et Corlett 1995), puis il a été montré que des gènes similaires aux gènes *ASR* s'exprimaient dans le pollen de lys en dessiccation (Wang *et al.* 1998) et le maïs sous stress hydrique (Riccardi *et al.* 1998). Les

gènes *lp3-3* et *lp3-1* amplifiés chez le pin maritime sont tous deux homologues d'ESTs de *Pinus taeda*, le premier présentant une sur-expression dans les aiguilles en conditions de stress hydrique extrêmes (Padmanabhan *et al.* 1997), et le deuxième étant issu d'une banque racinaire d'ESTs obtenue après application d'un stress hydrique (Gonzalez-Martinez *et al.* 2006a).

B.8. Obtention des séquences de gènes candidats

B.8.1. Extraction d'ADN à partir de tissu haploïde

L'ADN a été extrait à partir des méga-gamétophytes, qui correspondent à l'endosperme haploïde d'origine maternelle qui entoure l'embryon diploïde. L'intérêt est d'avoir directement accès à la phase haploïde, sans passer par une étape de clonage. Les méga-gamétophytes ont été prélevés après germination des graines en chambre de culture dans de la vermiculite, afin de minimiser la quantité des protéines de réserve contenus dans le méga-gamétophyte. L'extraction d'ADN a été adaptée à partir du protocole de Plomion *et al.* (1995), conçu pour l'extraction d'ADN dans une matrice chargée en protéines et polysaccharides. Le dosage et la qualité des échantillons ont été par la suite estimés par spectrophotométrie en mesurant les densités optiques à 260 nm et 280 nm, correspondant à l'absorbance des ADNs (ou ARNs) et à celle des protéines, respectivement. En moyenne, entre 20ng/μl et 1500ng/μl ont été obtenus, et seuls les échantillons avec un ratio DO260/DO280 supérieur à 1.8 ont été exploités.

B.8.2. Amplification, séquençage des fragments de gènes candidats et détection des polymorphismes

L'amplification des fragments de gènes candidats à partir des amorces dessinées (Tableau 7) a été réalisée par une approche PCR (polymerase chain reaction) par paliers de températures (touchdown) pour optimiser le test de la gamme des températures possibles. Après purification des produits issus de la PCR, le séquençage a été réalisé en utilisant le kit DYEnamicTM ET Terminator (Amersham Biosciences Inc., Uppsala, Suède). Les produits séquencés ont été chargés sur différents séquenceurs capillaires (soit MegaBACE 1000, Amersham Biosciences Inc, ou ABI 3130 XL, Applied Biosystems). L'identification des bases s'est faite à partir des chromatogrammes bruts en utilisant le logiciel Sequence Analyser (Amersham Biosciences Inc.). Le logiciel CodonCode Aligner ver. 1.5.1 (Codon Code Corporation, <http://www.codoncode.com/>) a permis d'aligner les séquences et d'attribuer un indice de qualité à chaque base, le Phred score. Chaque polymorphisme SNP ou Indel a été vérifié visuellement sous CodonCode et par la suite validé lorsque les valeurs de Phred score dépassaient 20 (correspondant à une probabilité d'erreur

de 1%). D'autre part, l'observation d'un singleton issu d'une erreur d'amplification de la Taq Polymérase (taux d'erreur estimé à environ 1/1000pb) et dont le score Phred serait supérieur à 20 est quasiment impossible car il ne serait possible que si l'erreur était intervenue dans les premiers cycles de PCR et au même endroit pour toutes les copies génomiques amplifiées (dont le nombre est estimé à environ 2^n où n est le nombre de cycles de dénaturation, hybridation des amorces et d'élongation depuis le début de la PCR).

C. Choix de marqueurs microsatellites et définition d'une référence « neutre »

C.1. Choix des marqueurs microsatellites et génotypage

Parmi les marqueurs microsatellites disponibles chez le pin maritime, 8 SSRs nucléaires, *NZPR413*, *NZPR1078*, *ctg64* (Chagné *et al.* 2004), *ctg275*, *FRPP91*, *FRPP94*, *ITPH4516* (Mariette *et al.* 2001) et *A6F03* (Guevara *et al.* 2005), ont pu être amplifiés (Tableau 8) et génotypés à partir de l'ADN d'aiguilles (tissu diploïde) échantillonnées sur les arbres mères des 24 populations de pin maritime. Ce travail a été réalisé par Valérie Léger (du pôle de génotypage/séquençage du laboratoire d'accueil) en partenariat avec le laboratoire partenaire du projet (INIA, Madrid). Ces SSRs ont été sélectionnés pour leur localisation dans différents groupes de liaison à partir de 2 pédigrés de pin maritime, leur richesse allélique (avec un minimum de 4 allèles) et la possibilité d'amplification multiplex (amplification de plusieurs SSRs à la fois).

C.2. Estimation de la diversité moléculaire des SSRs et identification de groupes homogènes de populations

La diversité moléculaire a été estimée pour chaque SSR au moyen du logiciel Genepop accessible sur le web (http://wbiomed.curtin.edu.au/genepop/genepop_op6.html) par différents paramètres : le nombre d'allèles observés (A), la diversité allélique intra population (H_S), et le coefficient de consanguinité (F_{IS} , Weir et Cockerham 1984) moyenné sur l'ensemble des populations (voir Chapitre I Encadré 1). Des tests exacts pour les F_{IS} ont été réalisés au sein de chaque population, afin de tester d'éventuels écarts à l'équilibre d'Hardy-Weinberg.

Nous avons exploré la possibilité de regrouper les populations en pools relativement homogènes, dans le but de disposer de tailles d'échantillons plus importantes pour l'analyse de la différenciation nucléotidique estimée sur les gènes candidats. Pour cela, nous avons combiné deux méthodes. La première méthode repose sur une classification hiérarchique effectuée à partir d'une matrice d'indices de différenciation moléculaire (F_{ST} , Weir et Cockerham 1984, Chapitre 1 Encadré 1) estimés par paire de populations avec Genepop. Les indices de différenciation pouvant

Tableau 8. Paires d'amorces, estimation de diversité et de différenciation pour les 8 SSRs.

locus SSR	Amorce « forward »	Amorce « Reverse »	24 populations			
			A ^(d)	H _S	F _{IS}	F _{ST}
NZPR413 ^(a)	TGAACCTCGATGGAATAGCC	CCCGCCTTGCATCAATTA	7	0.453	-0.033	0.189
NZPR1078 ^(a)	TGGTGATCAAGCCTTTTTCC	GTTGATGAGTGATGGCATGG	6	0.617	0.055	0.042
FRPP91 ^(b)	GTACTCCCACATAAAATGAGACTT	CCGAAATACATTGCAGGTTA	29	0.773	0.215*** ^(e)	0.130
FRPP94 ^(b)	GGCAAACCTCTTTTAGAGTGC	TTTGTGCGATTTTCTTGAAATCTAA	16	0.692	0.032	0.084
ITPH4516 ^(b)	TGATGCAAACAAGTTCCATG	AGCACTCGCTAAACTATGAAGG	16	0.786	0.104*** ^(f)	0.104
ctg64 ^(a)	GGAAGCTGTTACAAGTGCGG	ATCGAGAAGAGAGGAAGGGC	6	0.538	0.041	0.098
A6F03 ^(c)	CCTGAAAATCGACGGATCG	ATGGTATTTTGCGGGTTGC	12	0.616	0.038	0.122
ctg275 ^(a)	ACGGAGATATATTGCTGGCG	AAAGAATAACGTGAAACAAACCC	23	0.727	0.027	0.089
Moyenne			14.375	0.520	0.069	0.104

^(a) Chagné et al. 2004.

^(b) Mariette et al. 2001.

^(c) Guevara et al. 2005.

^(d) Nombre d'allèles observés

^(e) Test exact de Fisher significatif à $P < 0.001$ au sein des populations Mimizan, Pleucadeuc, Arenas et Oria.

^(f) Test exact de Fisher significatif à $P < 0.001$ au sein des populations Mimizan et Tamrabta.

être considérés comme des distances génétiques, ces matrices ont servi à la construction de dendrogrammes à l'aide de la fonction «hclust» du logiciel R ver 2.0.1 (Ihaka et Gentleman 1996) en utilisant deux algorithmes : 1) l'algorithme de Ward (Ward 1963) qui permet un regroupement des populations par étape, en minimisant à chaque étape la variation au sein des nouveaux groupes formés (en somme des écarts au carré relativement à la moyenne du groupe) par rapport à celle présente entre les groupes, et 2) l'algorithme «UPGMA» (Sneath et Sokal 1973) qui produit un arbre ultramétrique.

La seconde méthode est basée sur une approche bayésienne implémentée dans le logiciel Structure ver 2.0 (Pritchard *et al.* 2000). Nous avons utilisé un modèle de “zone en contact” tel que le suggère Evanno *et al.* (2005) lorsque les taux de migration entre populations ne sont *a priori* pas homogènes. Les deux zones analysées séparément sont une zone Ouest incluant toutes les populations espagnoles et françaises à l'exception de la Corse, et une zone Sud-Est constituée des populations Marocaine, Tunisienne et Corses (Figure 12). Ces 2 zones correspondent aux principaux groupes de populations détectés lors de précédentes études phylogéographiques en utilisant des marqueurs mitochondriaux et chloroplastiques (Burban et Petit 2003, Bucci *et al.* 2007). La statistique Δk (Evanno *et al.* 2005) a été utilisée pour estimer le nombre optimal de groupes de populations dans chaque zone. Cette statistique est basée sur l'importance du changement entre vraisemblances correspondant au nombre optimal de groupes en passant successivement d'un nombre de groupes de populations k à un autre. Cette statistique prend également en compte la variance des valeurs de différentes simulations et le mode de la distribution de cette statistique correspond au nombre de groupes optimal (Evanno *et al.* 2005). Les simulations au sein du programme ont été paramétrées en utilisant des modèles d'« admixture » (mélange de différentes populations) supposant une corrélation des fréquences alléliques (ou DL) possible entre loci, ainsi qu'une période de « burn-in » de 10^6 itérations (période des états initiaux de la chaîne du processus Markov Chain Monte Carlo (MCMC) pendant laquelle on suppose que les données convergent vers une distribution stationnaire).

C.3. Résultats préliminaires sur la diversité moléculaire des SSRs et définition d'une référence neutre

Les estimations de diversité obtenues sont similaires aux précédentes estimations sur des pédigrés ou différents échantillons de populations naturelles (Mariette *et al.* 2001, Derory *et al.* 2002, Chagné *et al.* 2004, Guevara *et al.* 2005). Les coefficients F_{IS} sont proches de zéro (moyenne=0.07 sur l'ensemble des 24 populations), à l'exception de *FRPP91* et *ITPH4516*, qui présentent des déficits d'hétérozygotes significatifs au sein de plusieurs populations (Tableau 8).

Figure 12. Deux « zones contacts », Ouest (Zone O) et Sud-Est (Zone S-E) utilisées dans le logiciel Structure.

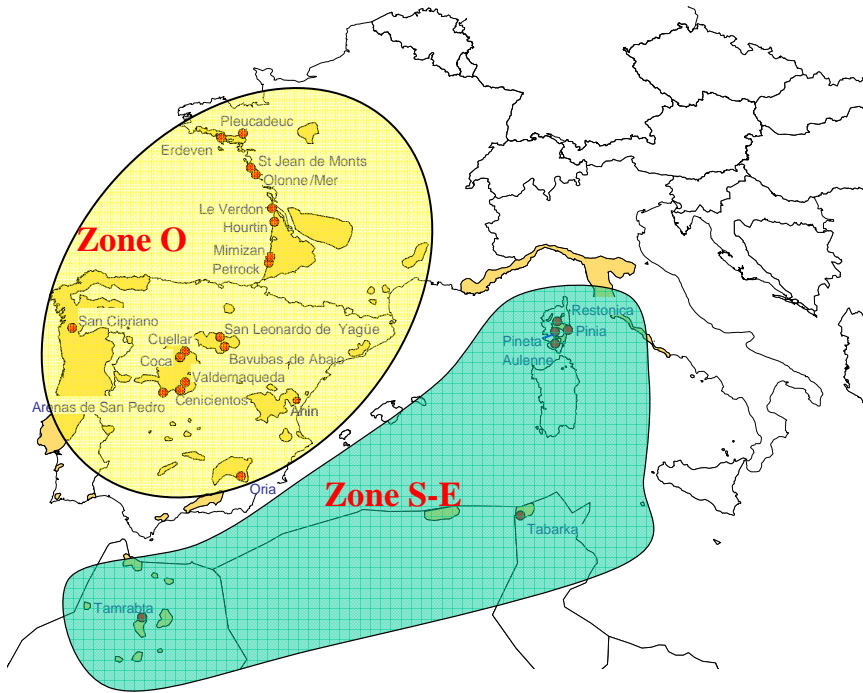
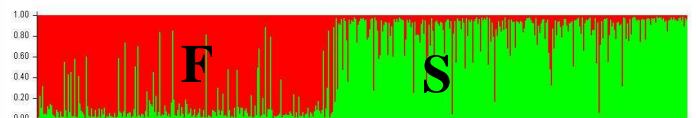


Figure 13. Assignement des individus en utilisant le logiciel Structure dans un modèle de « zones en contact » avec 2 zones : une zone Ouest en utilisant $\Delta k=2$ (A) et une zone Sud-Est en utilisant $\Delta k=2$ (B) et $\Delta k=3$ (C). C: populations Corses, F: populations de la France continentale, S: populations Espagnoles du Nord-Ouest et du centre, O: population Oria, M: population Marocaine, T: population Tunisienne.

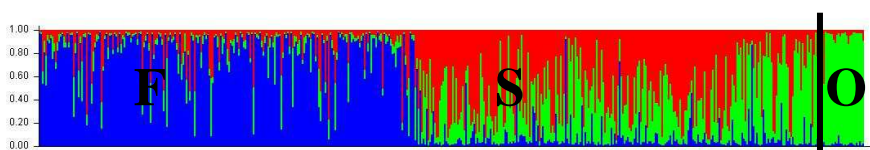
A) Zone Sud-Est
 $\Delta k = 2$



B) Zone Ouest
 $\Delta k = 2$



C) Zone Ouest
 $\Delta k = 3$



L'approche bayésienne a permis d'identifier uniquement deux groupes de populations au sein de la zone Sud-Est : d'un côté les populations Marocaine et Tunisienne et d'un autre côté, les populations Corses (Figure 13A). Cependant, ce résultat n'est pas cohérent avec la matrice de différenciation par paire de populations, le F_{ST} étant de 0.29 entre populations marocaine et tunisienne (Tableau 9). Sur le dendrogramme, on les distingue nettement l'une de l'autre avec les populations corses (Figure 14). L'attribution de ces 2 populations dans des groupes différents est indiscutable sur la base de ces résultats, malgré le résultat obtenu avec le logiciel Structure dans la zone Sud-Est.

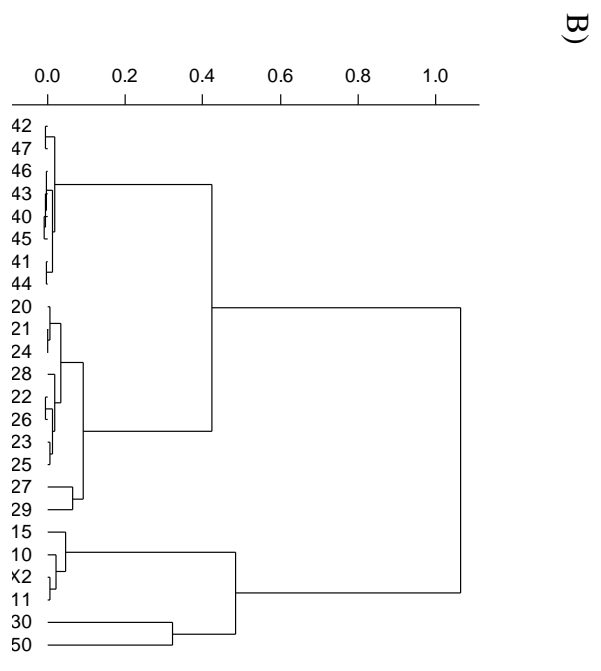
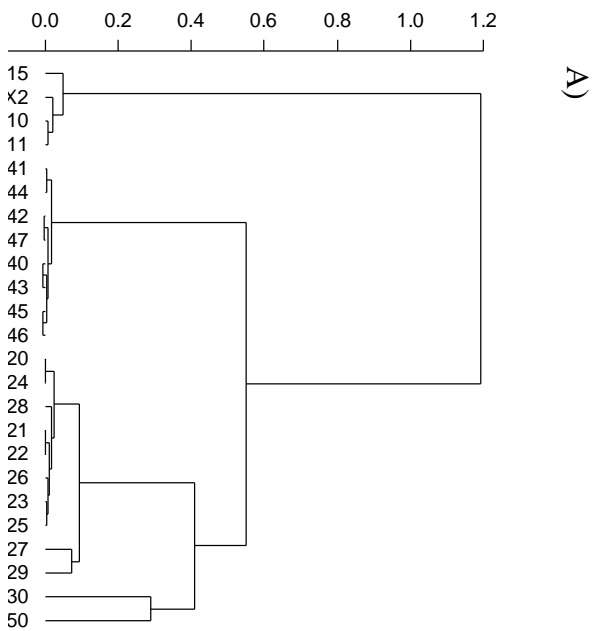
Au sein de la zone Ouest, l'approche bayésienne distingue deux groupes de populations : d'un côté les populations françaises continentales, et d'un autre côté l'ensemble des populations espagnoles (Figure 13B). Cependant pour une faible variation de la valeur de Δk à une valeur supérieure, les individus de la population Oria sont assignés préférentiellement à un troisième groupe. Ce regroupement est plus cohérent avec ce que l'on peut observer à partir des dendrogrammes (Figure 14), et également avec les résultats obtenus par González-Martínez *et al.* (2007a) sur des marqueurs allozymes, où les flux de gènes estimés entre la population Oria et d'autres populations voisines sont très faibles, ce qui pourrait s'expliquer par la présence de montagnes qui l'isole des autres.

En conclusion, 6 groupes homogènes de populations ont été retenus dont les valeurs des F_{ST} par paires de populations intra groupe était au plus de 0.06 (Tableau 9): La Corse (groupe C), la France continentale (groupe F), l'Espagne du nord-ouest et du centre (groupe S), et trois groupes chacun constitué d'une seule population: Oria (Sud de l'Espagne, groupe O), Tamrabta (Maroc, groupe M) et Tabarka (Tunisie, groupe T).

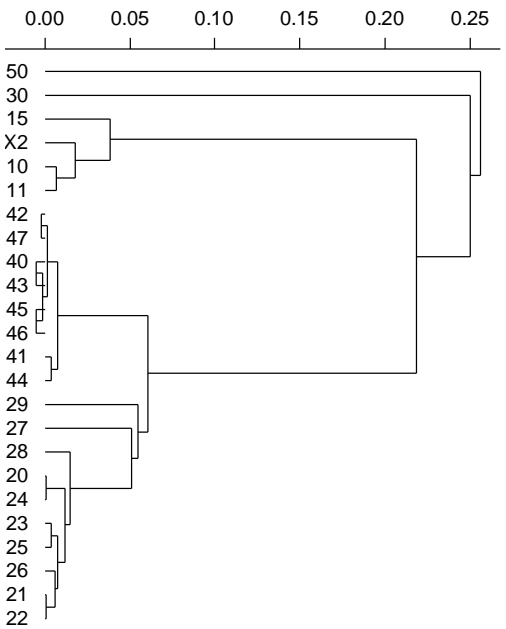
Tableau 9. Matrice de F_{ST} par paires de populations entre les 24 populations basée sur les 8 SSRs (valeurs en dessous de la diagonale) et les 6 SSRs sans *FRPP91* et *ITPH4516* (valeurs au dessus de la diagonale).

Code Population	2	10	11	15	40	41	42	43	44	45	46	47	20	21	22	23	24	25	26	27	28	29	30	50
2	*	0.030	0.008	0.034	0.278	0.294	0.256	0.265	0.261	0.272	0.257	0.249	0.110	0.154	0.176	0.159	0.162	0.146	0.181	0.224	0.189	0.151	0.305	0.295
10	0.023	*	0.009	0.051	0.244	0.266	0.215	0.231	0.237	0.244	0.218	0.212	0.099	0.114	0.144	0.120	0.140	0.122	0.150	0.167	0.157	0.120	0.284	0.195
11	0.013	0.007	*	0.024	0.241	0.256	0.209	0.230	0.223	0.236	0.218	0.204	0.098	0.121	0.138	0.131	0.129	0.119	0.140	0.191	0.162	0.121	0.304	0.240
15	0.027	0.051	0.038	*	0.234	0.245	0.202	0.219	0.209	0.227	0.204	0.198	0.109	0.121	0.128	0.117	0.137	0.107	0.135	0.167	0.158	0.102	0.274	0.283
40	0.280	0.275	0.288	0.235	*	-0.003	0.002	-0.007	0.002	-0.008	-0.003	0.001	0.078	0.048	0.044	0.061	0.060	0.057	0.044	0.080	0.053	0.109	0.307	0.321
41	0.303	0.300	0.308	0.253	0.003	*	0.012	0.004	-0.002	-0.007	0.009	0.014	0.095	0.073	0.061	0.087	0.082	0.069	0.057	0.104	0.073	0.120	0.317	0.343
42	0.277	0.266	0.278	0.223	0.000	0.010	*	0.003	0.013	-0.001	0.001	-0.004	0.071	0.029	0.021	0.043	0.041	0.035	0.026	0.059	0.038	0.089	0.302	0.304
43	0.270	0.263	0.277	0.223	-0.005	0.006	0.000	*	0.011	-0.005	-0.005	0.006	0.062	0.036	0.027	0.046	0.046	0.039	0.030	0.071	0.038	0.101	0.282	0.312
44	0.272	0.272	0.278	0.221	0.006	0.004	0.013	0.012	*	0.001	0.005	0.005	0.081	0.064	0.059	0.074	0.071	0.064	0.053	0.092	0.078	0.113	0.303	0.337
45	0.281	0.279	0.288	0.233	-0.002	0.000	0.001	-0.003	0.001	*	-0.004	0.005	0.076	0.049	0.044	0.069	0.059	0.051	0.050	0.088	0.050	0.108	0.305	0.327
46	0.271	0.260	0.275	0.220	0.004	0.015	0.002	-0.003	0.007	-0.005	*	0.006	0.070	0.032	0.030	0.047	0.054	0.036	0.039	0.063	0.046	0.098	0.290	0.320
47	0.261	0.253	0.264	0.209	0.001	0.011	-0.002	0.002	0.006	0.002	0.006	*	0.071	0.041	0.033	0.048	0.054	0.039	0.031	0.062	0.048	0.086	0.279	0.303
20	0.158	0.169	0.189	0.138	0.068	0.091	0.074	0.059	0.070	0.069	0.070	0.058	*	0.009	0.024	0.024	0.001	0.019	0.026	0.066	0.024	0.061	0.210	0.215
21	0.189	0.181	0.202	0.148	0.050	0.072	0.044	0.042	0.059	0.052	0.046	0.039	0.005	*	-0.003	0.007	0.001	0.013	0.009	0.027	0.013	0.062	0.253	0.236
22	0.195	0.191	0.201	0.145	0.045	0.060	0.033	0.031	0.053	0.044	0.039	0.031	0.021	0.001	*	0.009	0.003	0.003	-0.005	0.043	0.009	0.058	0.253	0.257
23	0.180	0.171	0.194	0.136	0.051	0.075	0.044	0.040	0.062	0.057	0.044	0.036	0.018	0.008	0.005	*	0.013	0.006	0.007	0.030	0.016	0.044	0.216	0.221
24	0.195	0.201	0.212	0.156	0.059	0.084	0.054	0.048	0.066	0.057	0.058	0.048	0.001	0.002	0.004	0.011	*	0.018	0.007	0.061	0.012	0.078	0.272	0.247
25	0.171	0.172	0.186	0.131	0.056	0.072	0.048	0.043	0.060	0.051	0.043	0.038	0.016	0.014	0.005	0.004	0.018	*	0.010	0.043	0.013	0.040	0.199	0.240
26	0.203	0.200	0.210	0.154	0.052	0.070	0.048	0.039	0.056	0.054	0.049	0.033	0.018	0.011	0.002	0.006	0.007	0.008	*	0.045	0.020	0.065	0.245	0.244
27	0.237	0.222	0.250	0.178	0.076	0.112	0.069	0.072	0.096	0.089	0.077	0.066	0.068	0.045	0.045	0.037	0.060	0.052	0.052	*	0.038	0.066	0.257	0.272
28	0.203	0.198	0.217	0.167	0.047	0.071	0.046	0.040	0.065	0.048	0.051	0.042	0.016	0.014	0.011	0.014	0.016	0.014	0.020	0.047	*	0.043	0.225	0.223
29	0.165	0.158	0.173	0.119	0.102	0.113	0.096	0.095	0.104	0.101	0.099	0.082	0.056	0.060	0.049	0.042	0.069	0.039	0.062	0.072	0.042	*	0.175	0.192
30	0.295	0.290	0.319	0.265	0.282	0.296	0.285	0.260	0.275	0.281	0.271	0.259	0.192	0.233	0.224	0.202	0.244	0.191	0.227	0.247	0.203	0.156	*	0.323
50	0.327	0.284	0.333	0.303	0.287	0.316	0.285	0.281	0.296	0.293	0.293	0.269	0.175	0.200	0.222	0.196	0.209	0.205	0.203	0.247	0.191	0.183	0.289	*

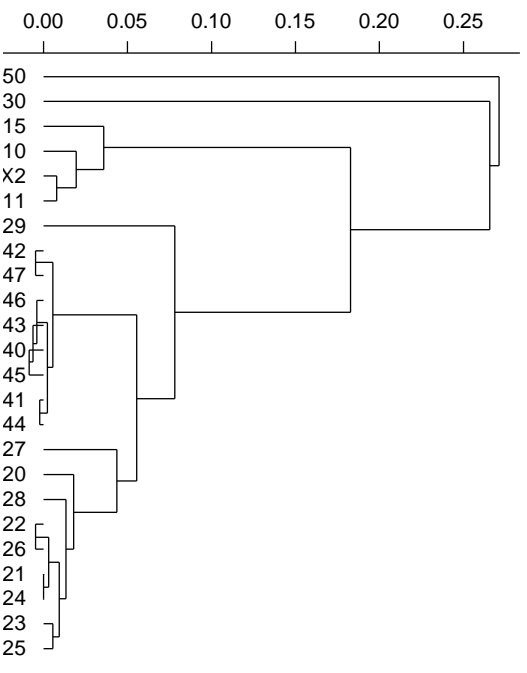
Figure 14. Dendrogrammes construits avec l'algorithme « Ward » à partir des 8 SSRs (A) et des 6 SSRs sans *FRPP91* et *ITPH4516* (B) et avec l'algorithme « UPGMA » à partir des 8 SSRs (C) et des 6 SSRs sans *FRPP91* et *ITPH4516* (D).



C)



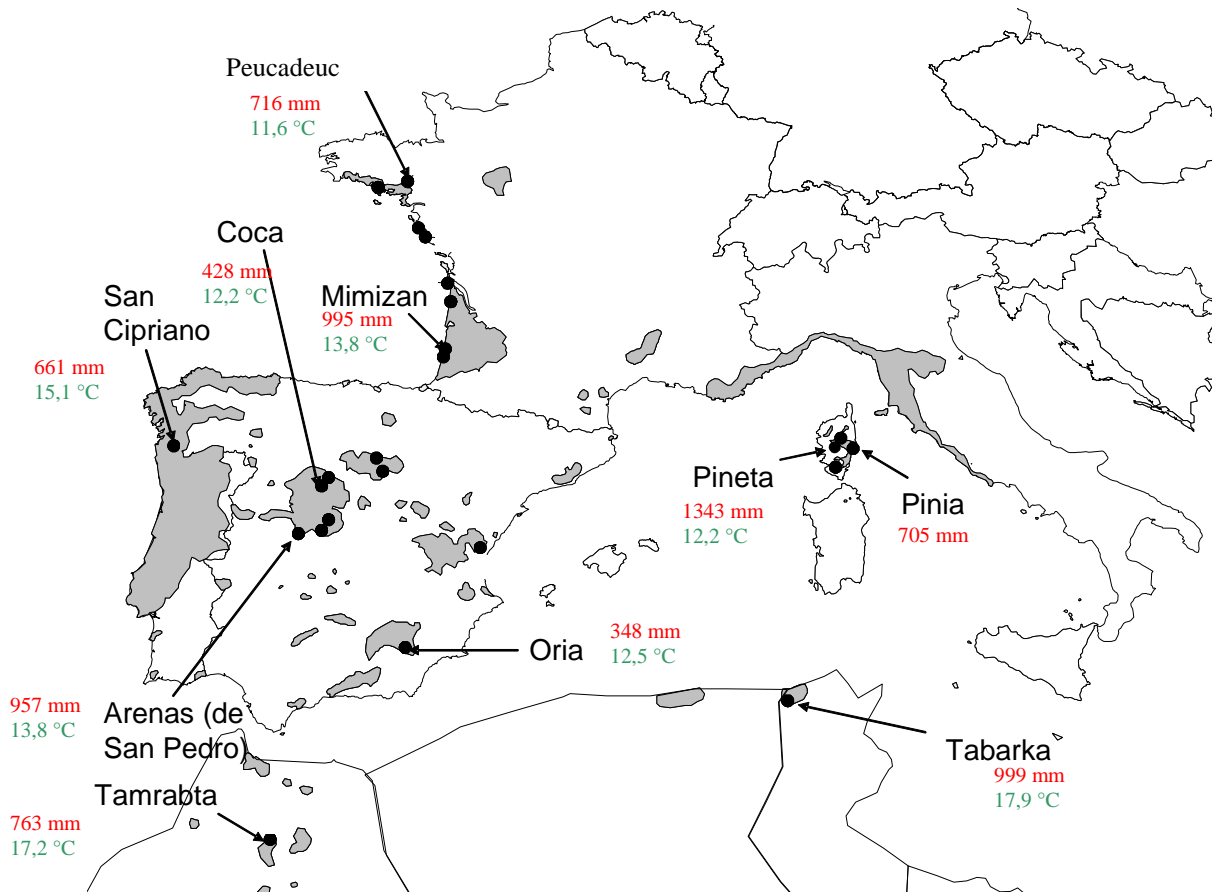
D)



Chapitre III

Méthodologie d'étude des patrons de diversité et différenciation moléculaires

Figure 15. Echantillon de dix populations aux conditions climatiques contrastées et réparties sur l'aire naturelle de distribution du pin maritime. Les précipitations totales annuelles et les températures moyennes annuelles pour chaque population sont indiquées en rouge et vert respectivement. Pour la population Pinia seulement, la température moyenne annuelle n'était pas disponible.



Ce chapitre présente les méthodes utilisées pour étudier les patrons de diversité et de différenciation moléculaires des gènes candidats. Nous détaillons tout d'abord l'estimation des différents paramètres génétiques (taux de mutations, taux de recombinaison, déséquilibres de liaison, différenciation) sur 10 populations représentatives de la collection décrite au chapitre II, ainsi qu'au sein d'une métapopulation échantillonnée plus largement. Puis nous présentons les trois méthodes de détection d' « outliers » que nous avons utilisées pour tester l'écart des patrons de différenciation génétique aux distributions attendues sous un modèle neutre. Quelques résultats préliminaires sont présentés et permettent de discuter de la robustesse des méthodes à différents paramètres. Enfin, nous décrivons les tests de neutralité que nous avons utilisés pour explorer plus finement les écarts éventuels au modèle neutre au sein de la métapopulation atlantique.

A. Echantillonnage

Les analyses moléculaires ont été réalisées sur 10 populations parmi les 24 populations décrites précédemment (Chapitre II) qui étaient représentatives de l'aire de distribution naturelle du pin maritime et de ses conditions climatiques contrastées (Figure 15) : deux populations échantillonnées sur la côte atlantique française (Mimizan et Pleucadeuc), deux populations corses (Pinia et Pineta), quatre populations espagnoles provenant du nord-ouest (San Cipriano), du centre (Arenas et Coca), et du sud (Oria), une population marocaine (Tamrabta) et une population tunisienne (Tabarka). Une vingtaine de séquences par gène a tout d'abord permis de sélectionner des fragments de 11 gènes candidats polymorphes parmi les 13 initialement choisis (représentant environ 11 kb): *PR-AGP4*, *CCoAOMT*, *GRP3*, *dhn-1*, *Glucan* (dont j'ai obtenu les séquences ensuite pour une dizaine de gamètes par population) et *erd3*, *rd21A*, *dhn-2*, *pp2c*, *lp3-1*, *lp3-3* (dont les séquences ont été obtenues par l'INIA pour 5 à 7 séquences par population). Les deux gènes restants (*COMT* et *Ino3*) étaient monomorphe (sur 1400 pb pour 20 gamètes pour *COMT*) ou quasi monomorphe (1 seul SNP sur 850 pb pour 24 gamètes pour *Ino3*) respectivement.

Un échantillonnage plus important a été réalisé au sein d'une métapopulation afin d'explorer des tests de neutralité basés sur la théorie de la coalescence et utilisant l'information de distribution des fréquences alléliques et du *DL* entre sites identifiés au sein des gènes candidats (voir Tableau 1 Chapitre I). Ces tests peuvent être plus puissants pour détecter d'autres signatures de sélection à différentes échelles de temps (voir Chapitre I). Cependant, nous avons vu au chapitre I qu'une littérature abondante avait été aussi consacrée

à l'impact de l'histoire démographique sur un grand nombre de ces tests, et montré qu'ils étaient sensibles non seulement à la structuration, mais aussi à des événements tels que goulot d'étranglement ou expansion de population.

Pour se mettre dans les meilleures conditions d'application de ces tests, une trentaine de séquences par gène (pour *PR-AGP4*, *dhn1*, *GRP3*, *CCoAOMT* et *Glucan*) ont été obtenues pour une métapopulation « atlantique » regroupant les provenances suivantes : Le Verdon, Hourtin, Mimizan et Petrock en Aquitaine, et Olonne/Mer, St Jean de Monts et Pleucadeuc situées au nord de l'Aquitaine (Figure 15). Cette métapopulation a été choisie car :

1) Les populations la constituant ont une origine commune (à partir de la voie de colonisation d'une zone refuge espagnole après la période glaciaire, Salvador *et al.* 2000),

2) Aucune barrière géographique n'existe entre provenances le long de la côte atlantique, permettant des flux de gènes *a priori* importants chez cette espèce (dispersion de pollen sur des dizaines de kilomètres),

3) Les conditions climatiques et environnementales sont relativement homogènes entre provenances (ANMR de 776 et 995 mm et ANMT de 11.6 à 13.8°C pour Pleucadeuc et Mimizan respectivement, et une altitude ne dépassant pas 80m, voir Tableau 4 Chapitre II), ce qui supposerait l'absence d'adaptation locale.

4) Des études précédentes montrent une absence de structuration phénotypique entre provenances du massif Aquitain (Danjon 1994)

D'autre part, nous avons vérifié qu'il n'existait aucune différenciation génétique significative entre les différentes provenances de cette métapopulation sur la base des données SSRs (valeurs de F_{ST} par paires de populations inférieures à 0.01). De même pour les 5 gènes candidats étudiés, aucune valeur de N_{ST} ou de F_{ST} n'était significative (Tableau 10), confirmant l'absence de différenciation moléculaire au sein de cette métapopulation.

B. Estimation des diversités nucléotidique et haplotypique

La diversité nucléotidique a été estimée pour chaque gène par deux estimateurs : 1) θ_π , basé sur le nombre moyen de nucléotides différents entre paires de séquences (Nei 1987) et 2) θ_S , basé sur le nombre de sites en ségrégation (Watterson 1975). Ces estimations ont pris en compte SNPs et Indels (considérés comme des sites non-synonymes lorsqu'ils étaient localisés dans les régions codantes) et ont porté soit sur l'ensemble des sites, soit sur les sites non synonymes et silencieux (incluant à la fois les sites synonymes de la région codante et tous les sites de la région non-codante). L'ensemble de ces estimateurs, ainsi que le nombre d'haplotypes (combinaison de SNPs ou Indels) (N_h) et la diversité haplotypique (H_d , Nei

Tableau 10. Estimation des indices de différenciation pour 5 gènes candidats au sein de la métapopulation atlantique.

Code du gène	N ^(a)	N_{ST}	Gamme de valeurs du F_{ST}
<i>PR-AGP4</i>	12	-0.073 ns ^(b)	-0.111 - 0.052 ns
<i>CCoAOMT</i>	11	-0.124 ns	-0.041 - 0.097 ns
<i>Glucan</i>	11	-0.138 ns	-0.123 - 0.094 ns
<i>dhn1</i>	14	-0.020 ns	-0.118 - 0.041 ns
<i>GRP3</i>	9	-0.069 ns	-0.179 - 0.056 ns

^(a) Nombre moyen de séquences par population

^(b) Non significatif

1987) ont été calculés avec les logiciels Arlequin ver 3.01 (Excoffier *et al.* 2005) et DNAsp v4.10 (Rozas *et al.* 2003). Des intervalles de confiance à 95% pour la diversité ont été calculés selon Tajima (1983):

$$SD = \text{racine}((1/(L*3*(N-1)))*(N+1)*\theta + (2*(N^2+N+3)*\theta^2)/(9*N*(N-1)))$$

avec L , la longueur totale de la séquence, N , le nombre total de séquences, et θ , l'estimation de la diversité nucléotidique par site.

Pour tester si ces statistiques étaient significativement différentes de ce que l'on attendrait sous un modèle neutre, nous avons réalisé des simulations de coalescence avec les logiciels DNAsp v4.10 et Proseq v2.91 (Filatov 2002). Ces simulations ont été faites sur la base du nombre de sites en ségrégation pour θ_π et H_d , ou sur la base du θ_π observé pour le θ_S . Certaines ont été réalisées en faisant l'hypothèse d'une structuration en 10 dèmes de taille constante et en l'absence de recombinaison intra génique pour les 10 populations, et d'autres en intégrant le taux de recombinaison ρ intra-gène intra-population pour les tests concernant la métapopulation atlantique (voir partie suivante).

C. Estimation de l'importance des évènements de recombinaison

Les mesures directes des taux de recombinaison sont classiquement effectuées par cartographie fine mais cette approche reste longue et coûteuse si l'on veut une résolution suffisante nécessaire à une estimation précise (Stumpf et McVean 2003). Pour inférer les taux de recombinaison, nous avons utilisé les méthodes statistiques indirectes développées et appliquées récemment chez les humains, dans le cadre du projet HapMap (McVean *et al.* 2002, McVean *et al.* 2004) à partir du polymorphisme de séquences d'ADN.

C.1. Nombre minimum d'évènements de recombinaison

Le nombre minimum d'évènements de recombinaison par gène dans l'histoire de l'échantillon, R_M , a été estimée à l'aide du logiciel DNAsp, par un test « 4-gamètes », basé sur le comptage du nombre de gamètes entre paires de loci bi-alléliques, et l'observation que les 4 haplotypes possibles soient présents ou non, sous l'hypothèse d'un modèle de mutations à sites infinis (Hudson et Kaplan 1985). Cette approche est néanmoins limitée car un évènement de recombinaison risque de ne pas être détecté pour des petites tailles d'échantillon qui ne contiennent pas l'allèle le plus rare (Stumpf et McVean 2003). Des simulations ont d'ailleurs montré que ce type de méthodes sous-estimaient généralement le niveau réel de recombinaison (Posada et Crandall 2001, Wiuf *et al.* 2001).

C.2. Estimation du taux de recombinaison d'une population

Nous avons également estimé le taux de recombinaison de la population, ρ , défini comme : $\rho = 4N_e r$ ($2N_e r$ pour des organismes haploïdes), avec N_e , la taille efficace de la population, et r , le taux de recombinaison par génération, correspondant à la probabilité qu'un évènement de recombinaison intervienne durant la méiose. Cet estimateur qui dépend de N_e et donc de l'histoire démographique des populations, est estimé exclusivement intra-population. Le paramètre ρ a été estimé avec une extension de la méthode de Hudson (2001) par McVean *et al.* (2002) pour un modèle à sites finis approprié pour des marqueurs de type SNPs, grâce au logiciel LDhat ver 2.0 (McVean *et al.* 2002). Cette méthode consiste à regarder les distributions alléliques de deux loci à la fois, et à multiplier toutes les densités de probabilité pour obtenir une probabilité dite composite, permettant un gain de temps considérable par rapport aux approches « full-likelihood » qui exploitent toute l'information de séquence. Par défaut, le paramètre ρ est estimé sur la base du θ_S observé (Watterson 1975) pour reconstruire la généalogie du gène. Nous avons donc vérifié que l'estimateur θ_S ne s'écartait pas de la neutralité. Dans le cas contraire, nous l'avons ré-estimé sur la base de simulations sous modèle neutre pour chaque gène, étant donné le nombre de sites polymorphes observés.

D. Estimation des déséquilibres de liaison entre sites

Au sein de chaque gène, le DL a été estimé entre paires de sites polymorphes (avec une fréquence allélique de l'allèle le plus rare $> 5\%$), par l'indice r^2 qui est la corrélation au carré des fréquences alléliques (Hill et Robertson 1968), avec le logiciel DNAsp. Cet estimateur a l'avantage d'être moins sensible que d'autres estimateurs du DL (par exemple D') à la taille de l'échantillon. L'existence d'une association significative a été déterminée par un test exact de Fisher, puis par une correction séquentielle de Bonferroni pour les tests multiples.

Nous avons également exploré la diminution du DL (r^2) en fonction de la distance physique au sein de chaque gène, par une régression non-linéaire du r^2 sur la distance (en pb) entre sites adjacents (Remington *et al.* 2001). Pour ajuster les données à une fonction non-linéaire, nous avons utilisé l'espérance de r^2 selon Hill et Weir (1988) sous les hypothèses d'un équilibre dérive-recombinaison, d'un faible niveau de mutation, et d'un ajustement à la taille des échantillons n :

$$E(r^2) = \left[\frac{10 + C}{(2 + C)(11 + C)} \right] \left[1 + \frac{(3 + C)(12 + 12C + C^2)}{n(2 + C)(11 + C)} \right], \quad (1)$$

avec C , le taux de recombinaison de la population (c'est-à-dire $C = \rho = 4N_e r$). L'équation 1 a donc été utilisée pour estimer le paramètre C sur nos données, en utilisant l'algorithme de Gauss-Newton implémenté dans la procédure PROC NLIN de SAS.

E. Estimation de la différenciation génétique entre populations

Nous avons vu dans le chapitre I que la différenciation génétique pouvait être estimée par de nombreux indices (voir Encadré 1 Chapitre I). Nous avons utilisé principalement des estimateurs basés sur des approches d'ANOVA adaptés aux différents types de données moléculaires obtenues, dont nous détaillons ci-dessous le principe et les hypothèses sous-jacentes.

E.1. Approche basée sur les fréquences alléliques

L'approche de Weir et Cockerham (1984) a été utilisée pour estimer la différenciation génétique par l'indice F_{ST} entre les 10 populations et entre les 6 groupes homogènes de populations définis au Chapitre II. Cet indice est également à la base du test de neutralité de Beaumont et Nichols (1996) utilisé par la suite (voir partie F.1. de ce chapitre).

L'estimation du F_{ST} selon Weir et Cockerham (1984) est basée sur une ANOVA hiérarchisée en 1) populations, 2) individus au sein des populations, et 3) allèles au sein des individus. La valeur du F_{ST} est donc exprimée en fonction des trois composantes de variance des fréquences alléliques pour un locus à deux allèles

$$F_{ST} = \frac{\sigma_a^2}{(\sigma_w^2 + \sigma_b^2 + \sigma_a^2)}$$

avec σ_a^2 la composante due aux différences entre populations, σ_w^2 celle due aux différences entre gènes à l'intérieur des individus, et σ_b^2 celle due aux différences entre individus à l'intérieur des populations. Il est également possible de généraliser cet estimateur à plusieurs loci multialléliques (Weir et Cockerham 1984).

L'estimation du F_{ST} repose sur l'hypothèse que les populations échantillonnées sont de même taille et descendent d'une seule population ancestrale dont les fréquences alléliques sont à l'équilibre d'Hardy-Weinberg et en équilibre de liaison (Weir et Cockerham 1984). Cet estimateur est néanmoins corrigé par rapport au biais dû aux différences de distributions des fréquences alléliques entre les échantillons des populations (Weir et Cockerham 1984, Cockerham et Weir 1993, Weir et Hill 2002). Ce paramètre est également indépendant du

nombre de populations et d'individus échantillonnés, et donc plus adapté que le paramètre de Nei (1973) lorsque les tailles d'échantillon sont faibles (Weir et Cockerham 1984).

Par contre, cette estimation fait partie d'une catégorie de statistiques uniquement basée sur les fréquences alléliques et qui considère tous les allèles comme équidistants les uns des autres. Ces statistiques seraient relativement peu puissantes notamment pour détecter de la différenciation génétique pour des marqueurs dont le taux de mutation est élevé, comme dans le cas des microsatellites (Hudson *et al.* 1992).

E.2. Approche basée sur les distances entre allèles

C'est pourquoi, nous avons également estimé la différenciation génétique entre populations par des méthodes basées sur la distance entre haplotypes, qui sont adaptées aux données de séquences nucléotidiques des gènes candidats et à celles des SSRs dont les tailles d'allèles varient en fonction du nombre de répétitions de motifs. Grâce à ces méthodes, la structure phylogéographique éventuelle des haplotypes, ainsi que le taux de mutation peuvent être pris en compte (Lynch et Crease 1990, Pons et Petit 1995, Slatkin 1995)

E.2.1. L'approche d'Excoffier, Smouse et Quattro (1992)

Excoffier *et al.* (1992) ont développé un estimateur adapté aux haplotypes issus de séquences et relativement flexible aux différentes hypothèses sur l'évolution de ces séquences. Partant du fait que les sommes de carrés d'écarts à une moyenne (utilisées dans une ANOVA) peuvent s'écrire comme des sommes de différences au carré entre paires d'haplotypes (Li 1976), ils construisent directement une analyse de variance moléculaire (AMOVA) hiérarchisée (en populations, individus au sein des populations, etc...) à partir de la matrice de distances au carré entre paires d'haplotypes. Ces p -distances sont définies comme le nombre de sites nucléotidiques différents entre haplotypes (Nei et Kumar 2000). Ils définissent la différenciation génétique entre populations (noté Φ_{ST}) comme suit :

$$\Phi_{ST} = \frac{\sigma_a^2}{\sigma_T^2}$$

avec σ_a^2 , la composante de la covariance due aux différences entre populations, et σ_T^2 , la variance totale.

L'avantage de cette méthode est qu'aucune hypothèse sur l'histoire phylogéographique des polymorphismes n'est nécessaire puisque cette information est incluse dans la matrice de distance (Excoffier *et al.* 1992). Sur la matrice des p -distances, cet

estimateur est similaire au γ_{ST} de Nei (1982), au N_{ST} de Lynch et Crease (1990) ou de Pons et Petit (1996), et au $\langle F_{ST} \rangle$ de Hudson *et al.* (1992).

E.2.2. L'approche de Slatkin

Slatkin (1995) propose d'estimer un indice R_{ST} qui est un analogue du G_{ST} pour les données microsatellites et qui prend en compte les différences de tailles alléliques. Il est défini comme :

$$R_{ST} = \frac{\bar{S} - S_w}{\bar{S}}$$

avec S_w et \bar{S} , les différences moyennes au carré de la taille des allèles entre paires de gènes pris au sein des populations, et entre paires de gènes pris sur l'ensemble des populations.

Cet indice R_{ST} est l'analogue de l'estimateur Φ_{ST} , et Michalakis et Excoffier (1996) montrent qu'au niveau d'un locus, Φ_{ST} peut être relié à R_{ST} par :

$$\Phi_{ST} = \frac{(1-c)R_{ST}}{1-cR_{ST}}$$

avec c , un coefficient fonction du nombre d'individus et de populations.

E.3. Estimation des paramètres

Les estimations du F_{ST} selon Weir et Cockerham (1984) pour les gènes candidats ont été réalisées à l'aide du logiciel Arlequin 3.01 (Excoffier *et al.* 2005), en recodant tous les haplotypes de façon à ce qu'ils soient équidistants. Les tests de signification sont réalisés dans le logiciel par des permutations non-paramétriques. Les estimations du F_{ST} pour les SSRs ont été réalisées à l'aide du logiciel Genepop sur le Web (<http://genepop.curtin.edu.au/>). Enfin, les estimations du Φ_{ST} pour les gènes candidats en prenant en compte leurs distances en nombre de différences (comparable au N_{ST}), et pour les SSRs (analogue au R_{ST}) ont été réalisées à l'aide du logiciel Arlequin et Genepop respectivement.

F. Détection d' « outliers » basée sur la différenciation génétique

Trois méthodes ont été utilisées et comparées pour la détection d' « outliers » en utilisant l'information sur la différenciation entre populations pour les polymorphismes SNPs ou pour les haplotypes identifiés de chaque gène candidat : 1) la méthode développée par Beaumont et Nichols (1996), 2) une extension de cette approche aux marqueurs bialléliques

SNPs, à l'aide de simulations réalisées avec le logiciel « Simcoal2 » (Laval et Excoffier 2004, et 3) une méthode bayésienne proposée par Beaumont et Balding.

F.1. Approche de Beaumont et Nichols implémentée dans « Fdist2 »

F.1.1 Principe

Cette approche consiste à construire une enveloppe neutre basée sur la distribution des valeurs de F_{ST} (estimées selon Weir et Cockerham 1984) conditionnellement à l'hétérozygotie dont il dépend (voir Encadré 1 Chapitre 1). La distribution du F_{ST} en fonction de H_S (moyenne des hétérozygoties intra-population sur l'ensemble des populations) permet donc de comparer des loci à niveaux d'hétérozygotie variés. Les valeurs de F_{ST} en dehors de l'enveloppe neutre peuvent être interprétées comme des signatures de sélection diversifiante (pour les F_{ST} plus élevés) ou homogénéisante (pour les F_{ST} moins élevés) entre populations.

Un F_{ST} moyen expérimental a d'abord été estimé à partir des 8 marqueurs SSRs *a priori* neutres et non liés. Cette valeur ainsi qu'une taille moyenne d'échantillon ont ensuite été utilisées dans des simulations effectuées par « Fdist2 » pour générer des données haploïdes dont la moyenne des F_{ST} était proche du F_{ST} neutre expérimental. Le logiciel calcule ensuite plusieurs quantiles de cette distribution neutre du F_{ST} construite en fonction de H_S par classes d'hétérozygotie chevauchantes de 400 données. Ces quantiles permettent de tester si les valeurs de F_{ST} aux SSRs, gènes candidats et SNPs s'écartent de l'enveloppe simulée à un risque α donné. Par exemple, pour tester des « outliers » avec un risque α de 5%, les quantiles de 2.5% et 97.5% ont été calculés, le test étant bilatéral car aucune hypothèse *a priori* n'a été faite sur le type de sélection (directionnelle ou homogénéisante) affectant un loci (bien que cela eut été possible sur certains gènes, d'après les données de la littérature).

En plaçant les valeurs des 8 SSRs sur le graphe des F_{ST} en fonction des H_S , le locus *NZPR413* montrait une valeur de F_{ST} qui dépassait la borne supérieure de l'enveloppe neutre. L'objectif étant d'obtenir une enveloppe la plus représentative d'une distribution attendue sous la neutralité, nous avons donc simulé de nouveau la distribution neutre sans ce SSR (voir Eveno *et al.* 2008 en Annexe 1). Par ailleurs, l'une des hypothèses de cette approche étant l'indépendance des loci entre eux, nous avons exclu les SNPs en complet *DL* de l'analyse.

F.1.2. Robustesse de l'approche à plusieurs modèles démographiques

Cette approche repose sur un modèle qui approxime le modèle en îles infinies de Wright (1931) où le nombre de populations est infini, les populations sont de taille efficace

identique et constante, la migration entre populations est symétrique (Figure 16A et B), la sélection est absente, et les populations ont atteint un équilibre entre la migration et la dérive. Dans leur approche, Beaumont et Nichols utilisent un modèle avec 100 îles (où populations) à partir desquelles un nombre donné de populations peut être échantillonné (par exemple 10 populations ou 6 groupes homogènes de populations dans notre cas). Ces hypothèses sont certainement éloignées de la réalité biologique, notamment dans le cas de distribution fragmentée ou pour de nombreuses espèces d'arbres forestiers où les modèles de migration dépendent des distances entre individus ou populations (modèles « stepping stone », Figure 16C), ou des « voisinages » (Austerlitz *et al.* 1999). De plus, une diminution des taux de migration peut conduire à une augmentation de la variance des valeurs de F_{ST} entre loci (Lewontin et Krakauer 1973). L'hypothèse d'un modèle de migration symétrique, alors qu'il existe des populations plus isolées pourrait conduire à l'identification de faux « outliers ».

L'hypothèse que les populations de pin maritime aient atteint un équilibre migration-dérive est également peu probable car il n'y aurait eu que quelques centaines de générations (avec un temps de génération d'environ 10-15 ans) depuis la dernière glaciation (il y a 10 000 ans). De plus, Whitlock et McCauley (1999) ont montré que le nombre de générations nécessaire pour atteindre la valeur de F_{ST} à l'équilibre migration-dérive pouvait être particulièrement importante si la taille de population était large (cas de ces espèces majoritairement allogames) et/ou les taux de migration étaient faibles (probable entre groupes de populations de différentes régions de l'aire naturelle du pin maritime) (Figure 17).

Beaumont et Nichols (1996) ont cependant montré la robustesse de leur méthode à plusieurs modèles démographiques (modèles avec croissance exponentielle de la taille des populations, migration hétérogène entre populations) ainsi qu'à plusieurs tailles d'échantillonnage (10, 20, 50 et 100 individus haploïdes) et plusieurs valeurs de taux de mutation (0.01, 0.1 et 1).

Figure 16. Modèles de migration en îles (A, B) et modèle de migration « stepping stone » (C).

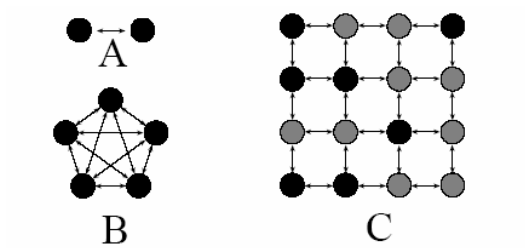
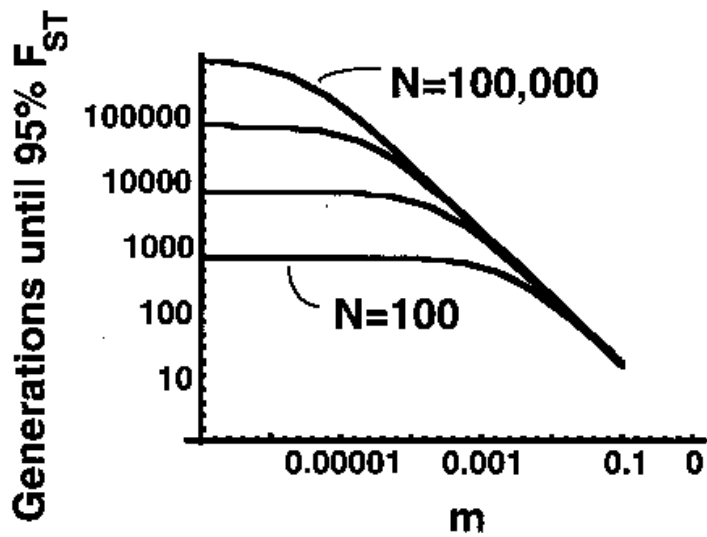


Figure 17. Nombre de génération pour atteindre 95% de la valeur de F_{ST} à l'équilibre migration-dérive en fonction de la taille de la population, N , et du taux de migration, m (d'après Whitlock et McCauley 1999).



Une des principales limites de leur approche serait les modèles de mutation proposés qui ne sont pas adaptés aux marqueurs bialléliques SNPs dans notre cas. Ces modèles sont 1) un modèle de mutation pas à pas (Kimura et Otha 1978, modèle SMM, Encadré 4), qui serait relativement adapté au processus mutationnel pour les 8 marqueurs microsatellites, et 2) un modèle de mutation en allèles infinis (Kimura 1968, modèle IAM, Encadré 4) généralement utilisé pour modéliser les polymorphismes de protéines, et donc relativement bien adapté aux haplotypes des gènes candidats uniquement lorsqu'ils sont considérés comme équidistants. A l'inverse, que ce soit le modèle SMM ou IAM, ils ne sont pas adaptés au taux de mutation plus faible conduisant à des marqueurs bialléliques comme les SNPs.

F.2. Extension de l'approche à des marqueurs bi-alléliques (méthode « FstSNP »)

C'est pourquoi, nous avons adapté cette méthode à des marqueurs bialléliques en construisant une enveloppe neutre à partir de données bialléliques générées par simulations de coalescence à l'aide du logiciel « Simcoal2 » (Excoffier *et al.* 2000). Le logiciel est relativement flexible car il est possible de tester différents scénarios démographiques (fission de population, changement de taille de population, etc...) et d'intégrer des matrices de migration non-symétriques. Il permet aussi de simuler des données bialléliques, et nous avons choisi de considérer comme dans « Fdist2 » le modèle en îles avec des taux de migration symétrique et des tailles de populations identiques et constantes. Cependant, les simulations réalisées considéraient un modèle dans lequel, à partir d'une population ancestrale unique, plusieurs sous populations se sont formées, et toutes ont été échantillonnées en prenant en compte l'hétérogénéité des tailles d'échantillons que nous avons dans chaque population (Figure 18). Le paramètre déterminant pour construire la distribution neutre étant le produit $N_e m$ (avec N_e , la taille efficace des populations, et m , le taux de migration moyen par génération), nous l'avons d'abord déduit du F_{ST} moyen estimé à partir des SSRs, selon la formule de Wright (1931) sous l'hypothèse d'un modèle en îles infinis (en considérant un taux de mutation négligeable par rapport aux taux de migration) :

$$F_{ST} \sim 1/(1+4N_e m)$$

Encadré 4. Les différents modèles de mutation (d'après Excoffier 2007, http://cmpg.unibe.ch/teaching/PG_05_Selection%20Mutations%2007.pdf).

➤ Modèle de mutation en allèles infinis (IAM)

Dans ce modèle, chaque mutation crée de nouveaux allèles qui n'existaient pas auparavant. Ce modèle de mutation a été développé à l'origine pour expliquer l'évolution des protéines.

➤ Modèle de mutation à sites infinis (ISM)

Ce modèle implique que chaque mutation apparaît à une position différente le long du chromosome. Un nucléotide donné ne peut donc pas être sujet à plus d'une mutation. Ce modèle a été créé à l'origine pour étudier des séquences d'ADN de grande longueur.

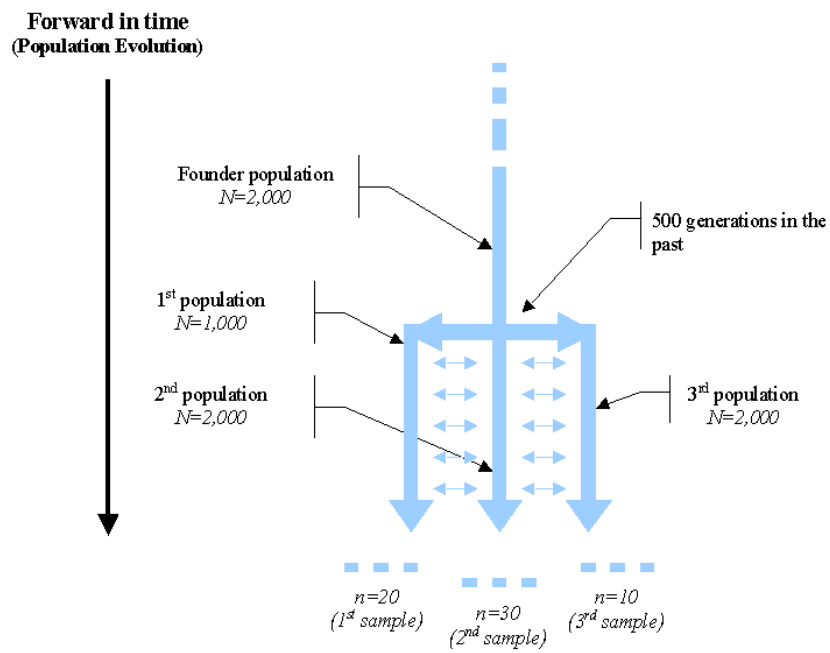
➤ Modèle de mutation « pas à pas » (SMM)

Le modèle de mutation « pas à pas » tente d'approximer le processus mutationnel de marqueurs microsatellites, où la variation microsatellite résulterait d'une erreur lors de la réplication, qui vraisemblablement entraînerait l'ajout ou la délétion d'une seule unité de répétition à chaque fois (Valdes *et al.* 1993, Weber et Wong 1993. Dans sa forme la plus simple (modèle symétrique à une étape), seuls les changements (apparition ou disparition) d'une seule unité de répétition peuvent survenir à un taux de mutation μ par microsatellite et par génération. Ce modèle suppose l'existence d'une relation entre les différences de taille d'allèle et le temps depuis leur divergence. Néanmoins il ne permet pas de distinguer l'identité par descendance de l'identité par état (homoplasie) dû par exemple à des phénomènes de réversion.

➤ Modèle de mutation à sites finis (FSMM)

Parce que les séquences d'ADN ont une taille finie, et parce qu'il existe des régions où le taux de mutation est particulièrement élevé et où les mutations peuvent donc apparaître plusieurs fois sur le même site (homoplasie), le modèle FSMM est plus vraisemblable. Il est à noter que si plusieurs mutations apparaissent aux mêmes sites, on peut alors sous-estimer le nombre de mutations ayant apparu durant l'évolution des séquences.

Figure 18. Modèle démographique sous-jacent aux simulations de coalescence sous « Simcoal2 ». Exemple d'une population ancestrale en 3 sous-populations (extrait de l'aide en ligne du logiciel Simcoal2, <http://cmpg.unibe.ch/software/simcoal2/>).



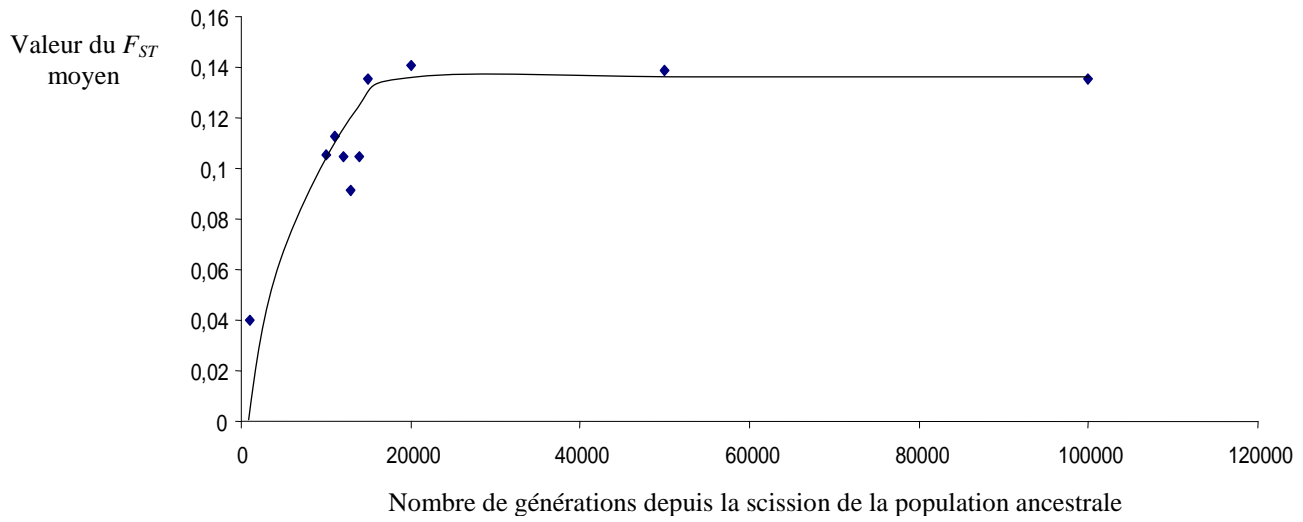
Ainsi avec un F_{ST} moyen neutre de 0.138 entre les 10 populations, $N_e m$ valait 0.17, ce qui a permis de choisir une taille efficace (fixée à 10 000) et un taux m supposé constant au fil des générations, (par déduction de 0.17×10^{-4}), à partir desquels les données SNPs ont été simulées sous l'hypothèse d'un modèle en îles équilibré avec une matrice de migration symétrique. Nous avons fait varier le nombre de simulations de coalescence ainsi que le nombre de générations, G , depuis le clivage des populations de façon à obtenir un équilibre et de retrouver une valeur de F_{ST} neutre proche de 0.138. En effet, les premières simulations en fixant arbitrairement G à 1 000 générations, ont conduit à une moyenne de F_{ST} très inférieure au F_{ST} observé, ce qui était sans doute dû au fait que l'équilibre migration-dérive n'avait pas encore été atteint. En testant plusieurs valeurs de G pour 10 jeux de simulations sur 100 points simulés, on observe que les valeurs moyennes de F_{ST} simulées atteignent un plateau entre 14 000 et 15 000 générations correspondant approximativement à la valeur moyenne du F_{ST} « neutre » (Figure 19). Par la suite, toutes les simulations ont été réalisées en fixant 15 000 générations depuis le clivage de la population ancestrale, après avoir vérifié que le F_{ST} moyen simulé était proche du F_{ST} moyen neutre. Un jeu de données d'une centaine de SNPs indépendants a été simulé et analysé avec le logiciel Arlequin pour l'estimation des valeurs de F_{ST} (Weir et Cockerham 1984) et des hétérozygoties moyennes intra population, dans le but de construire les distributions attendues.

Enfin, nous avons calculé les quantiles de l'enveloppe neutre par classes discrètes d'hétérozygotie de 0.05 avec le logiciel R ver 2.0.1. (<http://www.r-project.org/>).

F.3. Quelques résultats de simulations avec « Fidst2 » et « FstSNP »

Nous présentons ici quelques simulations préliminaires que nous avons réalisées pour explorer l'impact de différentes valeurs de paramètres sur la forme de l'enveloppe neutre simulée. Une discussion sur l'efficacité de ces deux méthodes à détecter des « outliers » est présentée dans l'article publié dans *Molecular Biology and Evolution* (Eveno *et al.* 2008, Annexe 1)

Figure 19. Valeurs du F_{ST} moyen pour 100 points simulés en fonction du nombre de générations depuis la scission de la population ancestrale (F_{ST} neutre de 0.13, N fixé à 10 000 et m de $0.17 (10^{-4})$).



F.3.1. Effet du modèle de mutation

A taille identique d'échantillon, les enveloppes neutres simulées sous « Fdist2 » sont relativement similaires sous un modèle de mutation SMM ou IAM. Par exemple, dans le cas d'un échantillon de 6 gamètes et d'un F_{ST} neutre de 0.138, la différence entre les quantiles supérieurs n'excède pas 0.05 (Figure 20). En revanche, le quantile supérieur simulé pour les données SNPs (méthode « FstSNP ») est beaucoup plus élevé pour des valeurs de H_S inférieures à 0.37 (différence parfois supérieure à 0.1, Figure 20). Cette large différence dans la forme des enveloppes illustre le fait que la variance entre valeurs de F_{ST} est en moyenne plus élevée pour des marqueurs bi-alléliques par rapport à des marqueurs multialléliques, notamment pour de faibles hétérozygoties.

F.3.2. Effet de la taille d'échantillon

La forme des enveloppes neutres dépend également de la taille des échantillons (n). A l'inverse des simulations réalisées par Beaumont et Nichols (1996) (n entre 10 et 100), la gamme de variation de n dans notre étude se situe entre 5 et 32 gamètes pour une valeur de F_{ST} similaire à notre valeur expérimentale. Pour tenir compte de la différence principale des tailles d'échantillons entre 2 groupes de gènes, nous avons réalisé des simulations adaptées à nos 2 jeux de données: un premier groupe de gènes (*PR-AGP4*, *CCoAOMT*, *GRP3*, *Glucan* et *dhn1*) avec en moyenne 9 gamètes par population (données haplotypiques sans données manquantes) ou 14 gamètes (dans le cas des SNPs où plus de données étaient en moyenne disponibles pour chaque SNP), et un deuxième groupe de gènes (*erd3*, *dhn2*, *lp3-1*, *lp3-3*, *pp2c* et *rd21A*) avec en moyenne 6 gamètes par population. Les simulations sous « Fdist2 » montrent que l'enveloppe neutre diminue sensiblement lorsque l'on considère successivement 6, 9, 14 et 50 gamètes par population au-delà de laquelle, la forme de l'enveloppe reste très similaire (Figure 21). On constate également une forte diminution de l'enveloppe simulée sous « FstSNP » lorsque l'on augmente la taille de l'échantillon (Figure 22), notamment au niveau de la borne inférieure de l'enveloppe pour l'ensemble des valeurs d'hétérozygotie, et au niveau de la borne supérieure pour des valeurs d'hétérozygotie inférieures à 0.35. Ces observations montrent l'impact non négligeable de la taille des échantillons en deçà de 20, malgré la robustesse de l'approche de Beaumont et Nichols à des tailles d'échantillons supérieures.

Figure 20. Effet du modèle de mutation sur la forme de l'enveloppe neutre (10 populations considérées, F_{ST} neutre de 0.138, enveloppes à 10%, IAM : infinite allele mutation model, SMM : stepwise mutation model, SNP : modèle de mutation adapté à la production de données SNPs).

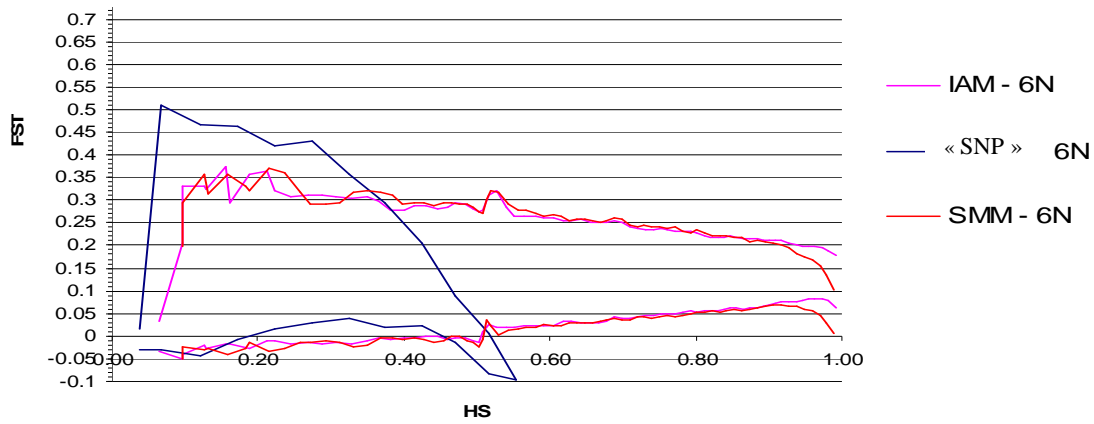


Figure 21. Effet de la taille d'échantillon sur la forme de l'enveloppe neutre simulée avec la méthode « Fdist2 » (10 populations considérées, F_{ST} neutre de 0.138, enveloppes à 10%).

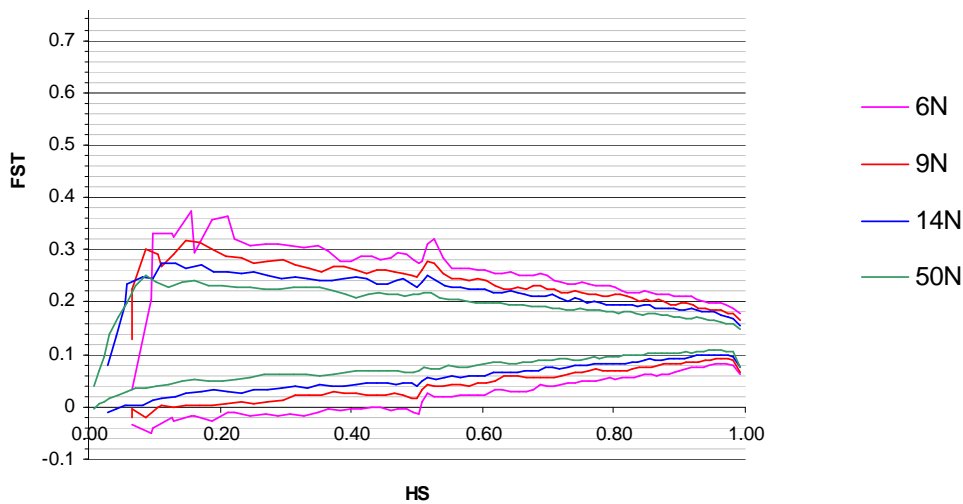
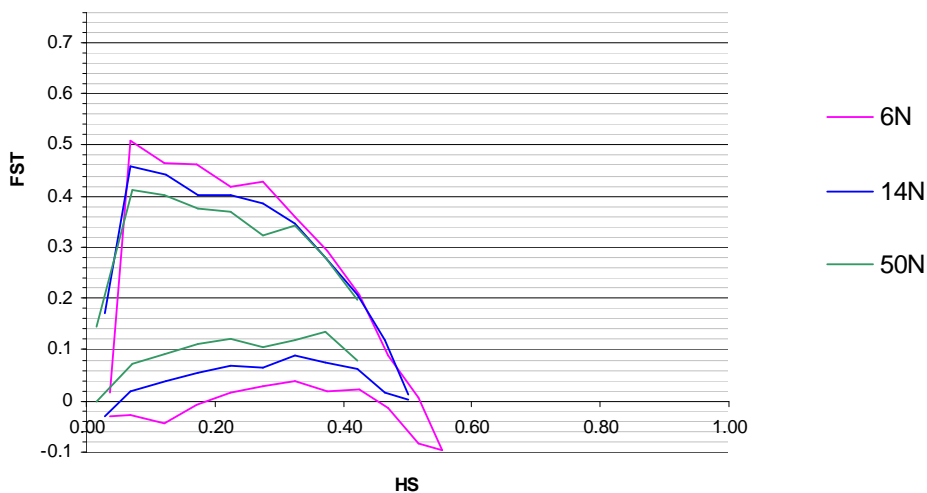


Figure 22. Effet de la taille d'échantillon sur la forme de l'enveloppe neutre simulée avec la méthode « FstSNP » (10 populations considérées, F_{ST} neutre de 0.138, enveloppes à 10%).



F.4. Détection d' « outliers » par l'approche Bayésienne de Beaumont et Balding (2004)

Les deux méthodes précédentes reposent sur des hypothèses démographiques fortes, notamment celles concernant un modèle de migration symétrique entre populations, même si elles sont *a priori* robustes à des écarts à ces hypothèses. D'autres méthodes ont été proposées qui intègrent ou permettent de s'affranchir du problème d'hétérogénéité des taux de migration (Vitalis *et al.* 2001, voir Chapitre I, Beaumont et Balding 2004). Malgré des essais préliminaires, nous n'avons pas réussi à appliquer la méthode de Vitalis dans le cas des marqueurs SNPs, et le nombre de comparaisons par paires de populations aurait été également très important. Nous avons donc choisi d'appliquer la méthode de Beaumont et Balding (2004) qui permettait une analyse conjointe des populations et des données moléculaires (haplotypes et SNPs), et qui en plus d'être flexible à l'hétérogénéité des taux de migration entre populations, comporte également d'autres avantages qui sont développés ci-dessous.

F.4.1. Particularité de l'approche bayésienne

Cette approche consiste à combiner deux sources d'information : l'information apportée par les données *via* l'expression d'une vraisemblance mais également une information *a priori* sur les paramètres, afin d'obtenir la distribution *a posteriori* des paramètres à estimer. L'inférence Bayésienne repose sur le théorème de Bayes (1763) où la probabilité des paramètres θ conditionnellement aux données D , et notée $P(\theta | D)$, est définie par :

$$P(\theta | D) = [P(\theta)P(D | \theta)] / P(D)$$

avec $P(\theta)$, la probabilité *a priori* des paramètres, $P(D | \theta)$, la fonction de vraisemblance des données conditionnellement aux paramètres, et $P(D)$, la probabilité totale des données obtenue en sommant le produit de la probabilité *a priori* et la fonction de vraisemblance sur toutes les valeurs de θ .

Un des avantages de cette approche est que l'information *a priori* peut permettre d'améliorer l'estimation de ces paramètres notamment en présence de nombreuses données manquantes, ces données étant traitées comme des paramètres inconnus (Shoemaker *et al.* 1999).

F.4.2. Principe de l'approche de Beaumont et Balding (2004)

F.4.2.1. Modèle généalogique sous-jacent

Le modèle généalogique sous-jacent à la méthode est le modèle de « separation-of-timescales » adapté de Wakeley (1999) (Figure 23). La généalogie se divise en 2 phases : i) la généalogie récente de chaque dème appelée « phase de dispersion » (« scattering phase ») par Wakeley (1999), et ii) la généalogie ancestrale appelée « phase de collecte » (« collecting phase ») se comportant comme une seule population. Dans cette généalogie, on suppose que les mutations se produisent durant la phase de collecte et non dans la phase de dispersion. De plus, la moyenne et la variance des fréquences des gènes au sein de chaque dème, dépend uniquement des paramètres de migration λ_{ij} spécifiques à chaque population i et chaque locus j (qui sont fonction des tailles efficaces et des taux de migration de chaque population), et des fréquences aux locus dans le pool immigrant. Ces fréquences peuvent être estimées à partir d'échantillons de dèmes indépendants, sans avoir à modéliser l'histoire démographique du pool immigrant de la phase de collecte. L'avantage de ce type de modèle généalogique, est qu'aucune hypothèse restrictive n'est faite sur les taux de migration ou la taille des populations, les valeurs de λ_{ij} pouvant varier d'un dème à l'autre.

F.4.2.2. Comptage des allèles et expression du paramètre F_{ij}

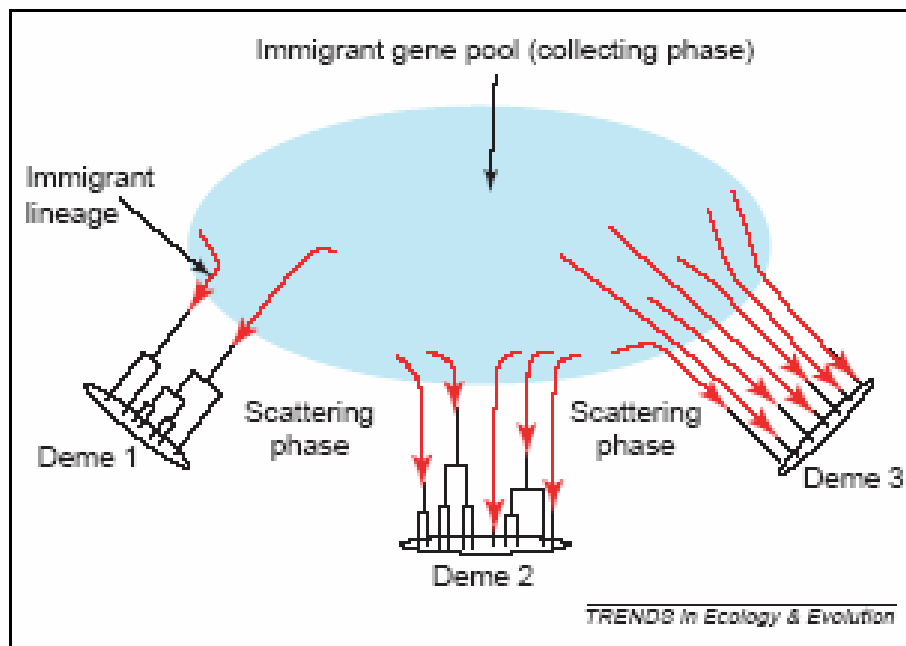
Selon le modèle généalogique que nous venons de décrire, le comptage des allèles à un locus donné au sein d'une population est modélisé par une vraisemblance qui suit une loi multinomiale Dirichlet :

$$L_{ij} \equiv P(a_{ij1}, \dots, a_{ijK_i} | \lambda_{ij}, x_{i1}, \dots, x_{iK_i}) \quad (1)$$

$$= \frac{\Gamma(\lambda_{ij})}{\Gamma(n_{ij} + \lambda_{ij})} \prod_{k=1}^{K_i} \frac{\Gamma(a_{ijk} + \lambda_{ij} x_{ik})}{\Gamma(\lambda_{ij} x_{ik})}$$

avec n_{ij} , la taille de l'échantillon au locus i dans la population j , a_{ijk} , le nombre d'allèles k au locus i dans la population j (ce paramètre devient a_{ij} dans le cas bi-allélique où $k=2$), x_{ik} , un paramètre de nuisance représentant la fréquence de l'allèle k au locus i dans le pool de migrants (voir Figure 23), et λ_{ij} , le paramètre d'intérêt correspondant au nombre de migrants dans la population j . Dans le cas d'un modèle neutre, le paramètre λ_{ij} est constant sur l'ensemble des loci i . Les valeurs aberrantes de λ_{ij} dans une ou plusieurs populations indiqueraient que le locus i serait potentiellement soumis à la sélection.

Figure 23. Modèle “separation-of-timescales” adapté de Wakeley (1999) utilisé dans l’approche bayésienne de Beaumont et Balding (d’après Beaumont 2005).



Les généalogies des échantillons à un locus sont représentées pour trois dèmes, et correspondent à la phase de dispersion. L’ancêtre de chaque lignée est un immigrant venant du pool de gènes. Les fréquences des gènes dans le pool de gènes immigrant sont déterminées par la généalogie de la phase de collecte et le modèle de mutation. Si le F_{ST} est élevé, l’échantillon au sein du dème aura peu d’ancêtres immigrants (cas du dème 1), et les fréquences au sein de l’échantillon tendront à être différentes de celles au sein du pool de gènes immigrant. A l’inverse, si le F_{ST} est faible (cas du dème 3), les fréquences au sein de l’échantillon se rapprocheront de celles au sein du pool de gènes immigrant.

Balding (2003) montre que ce paramètre peut s'exprimer en fonction de F_{ij} :

$$\lambda_{ij} = 1/F_{ij} - 1 \quad (2)$$

avec F_{ij} représentant ici la probabilité que deux gènes choisis au hasard dans une population possèdent un ancêtre commun au sein de cette population sans qu'il y ait eu migration ou mutation, et correspondant à la définition du F_{ST} en tant que coefficient de consanguinité.

F.4.2.3. Régression du paramètre F_{ij}

Balding (2003) propose également de décomposer le paramètre F_{ij} en des effets spécifiques au locus (qui seront affectés par la mutation et la sélection), et en des effets spécifiques aux populations (qui pourront dépendre de leurs tailles efficaces, des taux de migration, et du système de reproduction), de la façon suivante :

$$\text{Logit}(F_{ij}) = \log(F_{ij}/(1-F_{ij})) = \alpha_i + \beta_j + \gamma_{ij}$$

ou

$$F_{ij} = \exp(\alpha_i + \beta_j + \gamma_{ij}) / (1 + \exp(\alpha_i + \beta_j + \gamma_{ij})) \quad (3)$$

avec α_i , l'effet du locus i , β_j , l'effet de la population j , et γ_{ij} , l'erreur correspondant à l'effet d'interaction entre le locus i et la population j . On parle alors de modèle bayésien hiérarchique. Ce type de modèle est beaucoup plus flexible qu'un modèle bayésien standard, car il tient compte de la variabilité des F_{ST} entre loci (Beaumont et Rannala 2004).

F.4.2.4. Choix de la distribution *a priori* des paramètres α_i , β_j et γ_{ij} .

Afin de construire la distribution *a priori* de chaque paramètre de régression (puis de chaque F_{ij}), Beaumont et Balding (2004) proposent des valeurs par défaut pour les moyennes et variances *a priori* de ces paramètres. Pour α_i , β_j , et γ_{ij} , les valeurs des moyennes sont fixées à 0, -2 et 0 respectivement, et les valeurs des écart-types à 1, 1.8 et 0.5 respectivement. Les moyennes de 0 pour α_i , et γ_{ij} correspondraient à un modèle neutre en absence de sélection, où il n'y a aucun effet locus ou effet d'interaction entre locus et population. La moyenne est fixée à -2 pour le paramètre β_j afin que la distribution de F_{ij} soit une fonction croissante de β_j .

F.4.2.5. Implémentation de la chaîne MCMC

Pour obtenir les distributions *a posteriori* de ces paramètres de régression à partir de la fonction de vraisemblance des données et de la distribution *a priori*, les équations (1), (2) et (3), ainsi que les valeurs des moyennes et écart-types des distributions *a priori* des paramètres

de régression, ont été intégrées dans une chaîne de Markov Monte Carlo (« Markov Chain Monte Carlo », MCMC). Cette méthode d'échantillonnage statistique permet d'explorer la distribution *a posteriori* des paramètres en générant des échantillons sur la base du modèle de coalescence sous-jacent (Beerli 2005). Un algorithme Metropolis-Hastings a été utilisé pour explorer l'espace des paramètres de façon efficace et pour accepter ou rejeter les valeurs qui seront retenues pour construire leurs distributions postérieures. Cet algorithme présente l'avantage de maximiser l'efficacité des simulations notamment lorsque l'espace des états (c'est-à-dire tous les états possibles que peut prendre un ensemble de variables) est grand (Beaumont et Rannala 2004).

Les échantillons générés qui se suivent dans la chaîne étant auto-corrélés (car correspondant à de toutes petites modifications des valeurs des paramètres), il faut faire tourner la chaîne de Markov pendant une période suffisante pour sélectionner des échantillons qui soient indépendants, et pour qu'une convergence de la chaîne soit atteinte. C'est cette période de convergence que l'on cherche généralement à atteindre, après avoir éliminé les états initiaux de la chaîne pendant une période dite de « burn-in » (Beerli 2005).

F.4.2.6. Interprétation des distributions a posteriori des paramètres

Le paramètre qui nous intéresse en premier lieu est le paramètre spécifique aux effets des loci α_i : dans le cas de neutralité sélective, l'effet du locus i sera nul (et donc $\alpha_i = 0$), alors qu'une valeur positive de α_i peut être interprétée comme de la sélection diversifiante sur le locus i , et une valeur négative peut être due à de la sélection balancée. Beaumont et Balding (2004) proposent de considérer l'effet α_i comme significativement différent de zéro à un seuil P , lorsque l'intervalle $100(1-P)\%$ de la distribution *a posteriori* de α_i exclut zéro.

Les deux autres paramètres de régression β_j et γ_{ij} peuvent également être interprétés en termes de sélection. Les effets β_j ou γ_{ij} seront considérés comme significatifs lorsque l'intervalle $100(1-P)\%$ de leur distributions *a posteriori* exclut -2 et 0 respectivement (qui sont les moyennes de leur distribution *a priori*). Une valeur de β_j positive et significative indiquerait un flux de gènes moins important au sein de cette population, ou une taille efficace plus petite. A l'inverse, une valeur négative s'interpréterait comme un flux de gène plus important ou une taille efficace plus grande. Quant au paramètre γ_{ij} , une valeur significative et positive pourrait refléter un effet d'interaction entre une population et un locus dû à la sélection qui aurait par exemple conduit à la fixation d'un allèle localement, tandis que les effets de la sélection sur ce locus seraient quasi inexistant dans les autres populations (Beaumont et Balding 2004).

F.4.3. Applications préliminaires de la méthode « BayesFst »

Les simulations ont été réalisées à l'aide du programme « BayesFst » (Balding, comm. Pers.). Nous avons appliqué la méthode en combinant les jeux de données haplotypiques aux gènes candidats et 293 SNPs ou Indels identifiés au sein de ces gènes. Les données des 8 SSRs n'ont pas été intégrées car la distribution de Dirichlet utilisée pour estimer la vraisemblance des données n'est pas appropriée au modèle de mutation et à la diversité élevée des SSRs (Graham *et al.* 2000).

F.4.3.1. Déclaration du facteur d'interaction

La méthode implémentée dans « BayesFst » est basée sur l'estimation de trois facteurs: l'effet locus (α_i), l'effet population (β_j), et l'effet interaction « locus-populations » (γ_{ij}). En intégrant ce dernier facteur, on augmente considérablement le nombre de paramètres à estimer. Nous avons vérifié qu'aucun effet γ_{ij} n'était significatif sur l'ensemble des 304 loci analysés pour les 10 populations. D'autre part, nous avons comparé les modes (maximum de la densité) et les écart-types des distributions *a posteriori* de α_i et β_j en déclarant ou non le facteur d'interaction, et vérifié que la déclaration d'un facteur d'interaction n'avait pas d'impact sur les distributions *a posteriori* des autres facteurs du modèle de régression.

De plus, outre le fait que le nombre de paramètres à estimer conditionne la puissance informatique nécessaire, il est également important de s'interroger sur l'information contenue dans les données pour la construction de la distribution *a posteriori* des paramètres. Le risque est que cette distribution *a posteriori* soit basée en grande partie sur la distribution *a priori* et non sur la vraisemblance des données. Lorsque l'on déclare l'effet d'interaction, un calcul simple indique que l'on a environ 45 600 données (~15 individus par population * 10 populations * ~ 304 loci) et environ 3 354 paramètres à estimer (10 populations + ~ 304 loci + ~ 10*304 interactions). Dans ce cas, la proportion du nombre de paramètres à estimer par rapport au nombre de données s'élève à 7%. Par contre, lorsque l'on ne déclare pas l'interaction, cette proportion diminue considérablement (~ 45 600 données pour 314 paramètres à estimer). Cette proportion de 0.7% se rapproche davantage de la proportion de 1% utilisée dans les simulations de Beaumont et Balding (2004) qui exploitent au minimum 300 000 données pour estimer 3 500 paramètres.

F.4.3.2. Impact du facteur de corrélation

Bien qu'ayant exclu les SNPs en *DL* complet de l'analyse, il restait toutefois les SNPs en *DL* partiel. Par exemple pour *PR-AGP4*, 65% des SNPs présentent un *DL* minimum de 0.8 avec d'autres SNPs. Si dans l'approche précédente de Beaumont et Nichols, il n'existe aucun moyen d'intégrer cette corrélation dans les simulations de coalescence, il est possible dans l'approche de Beaumont et Balding de fixer une corrélation moyenne entre marqueurs adjacents. Cette corrélation, telle qu'elle est définie dans le programme « BayesFst » peut être prise en compte directement dans la distribution *a priori* du paramètre α_i , en élargissant ou en diminuant sa variance. En faisant varier cette valeur de corrélation (valeurs testées : 0, 0.25 et 0.5), nous n'avons pas observé de changement dans le nombre d'« outliers » détectés. Cela semble confirmer que la distribution *a priori* du paramètre α_i contribue peu à la construction de la distribution *a posteriori* de ce paramètre. Nous avons utilisé pour la suite des analyses une corrélation de 0 par défaut.

F.4.3.3. Regroupement des haplotypes

Malgré sa flexibilité, la méthode bayésienne reste néanmoins basée sur l'estimation d'un indice ne tenant pas compte des distances entre haplotypes. Nous avons donc effectué des simulations après avoir regroupé les haplotypes proches (distants par un petit nombre de mutations et sans aucun évènement de recombinaison) en « haplo-groupes ». Le fait d'intégrer une partie de l'information sur les distances entre haplotypes permet de penser que l'on se rapproche d'un estimateur similaire au N_{ST} (voir partie E.2.1.). Le nombre d'haplotypes étant moins important, cela permettait aussi d'augmenter la précision de l'estimation des fréquences alléliques de chacun. Pour cela, la matrice de *p*-distance basée sur le nombre de mutations différentes entre haplotypes, a été obtenue avec Arlequin. Cette matrice a été transformée en une matrice basée sur le pourcentage de différences entre haplotypes par rapport au nombre total de mutations. Des dendrogrammes ont été construits en utilisant l'algorithme « UPGMA » sur ces matrices de distances. Un premier seuil arbitraire de 10% de différences a été utilisé pour regrouper les haplotypes (Figures 25 et 26).

Figure 25. Dendrogrammes basés sur les matrices de pourcentage de différences entre paires d'haplotypes pour *PR-AGP4* (A), *CCoAOMT* (B), *Glucan* (C), *dhm1* (D), *GRP3* (E), *dhm2* (F). Le seuil pour regrouper les haplotypes en « haplogroupes » fixé à 10% est indiqué par une ligne.

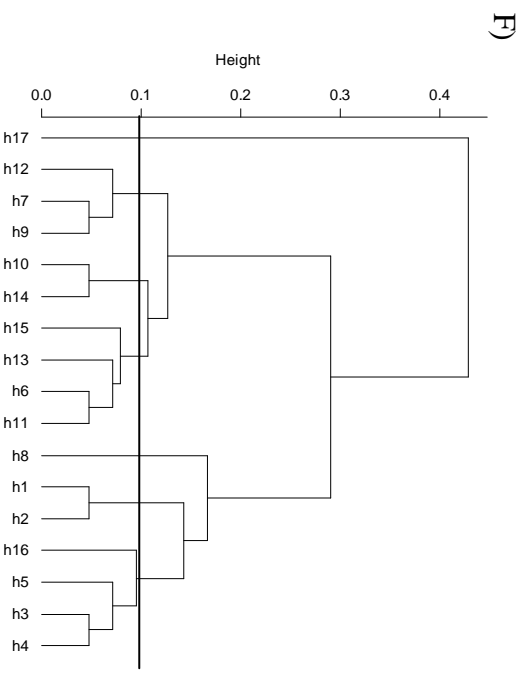
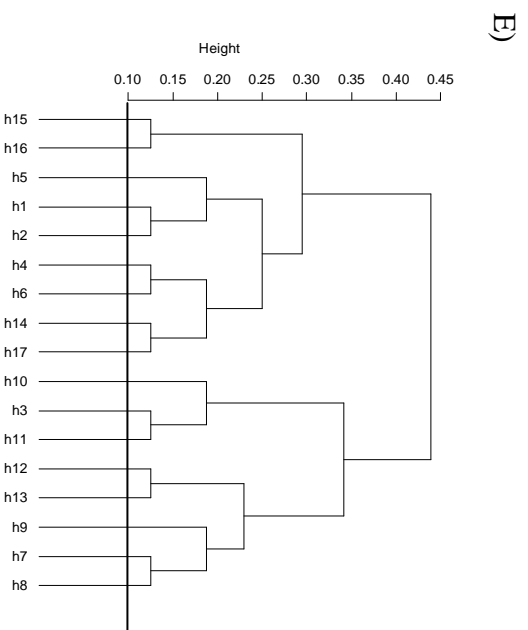
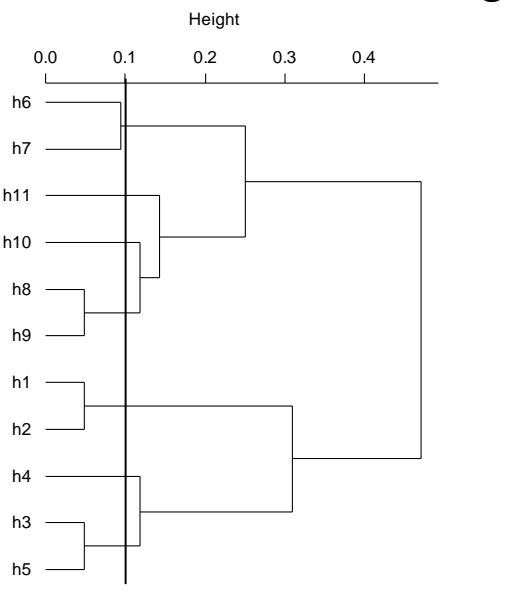
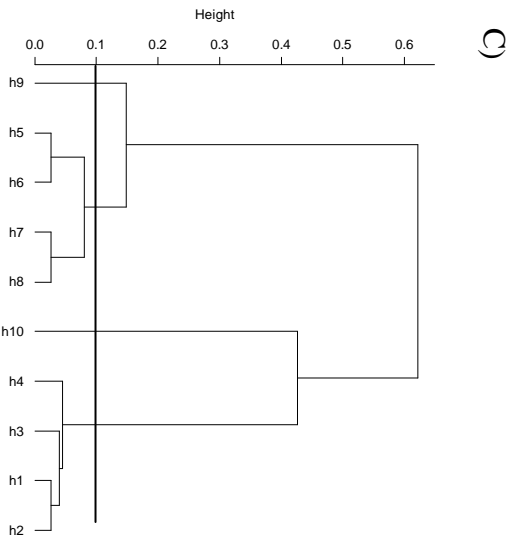
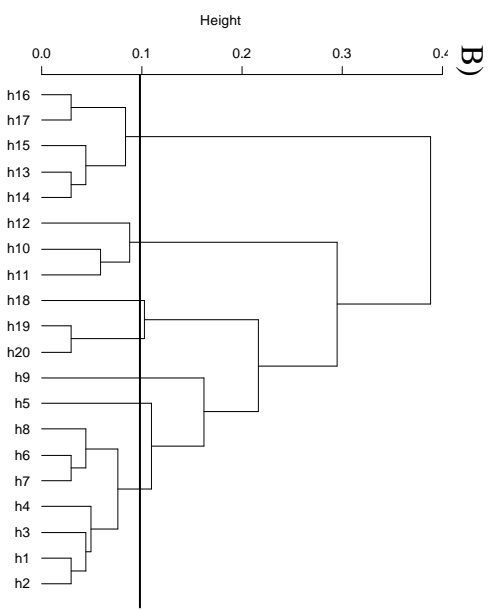
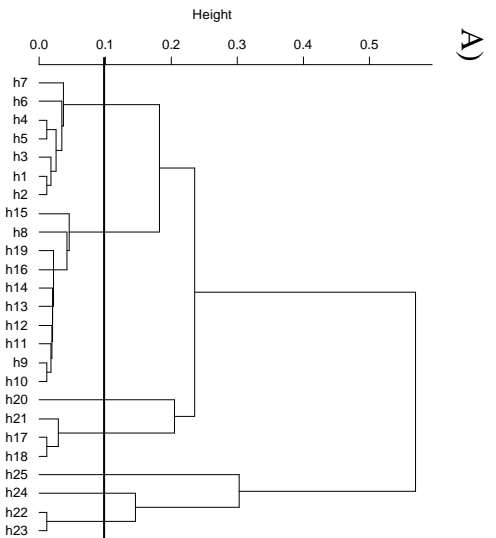
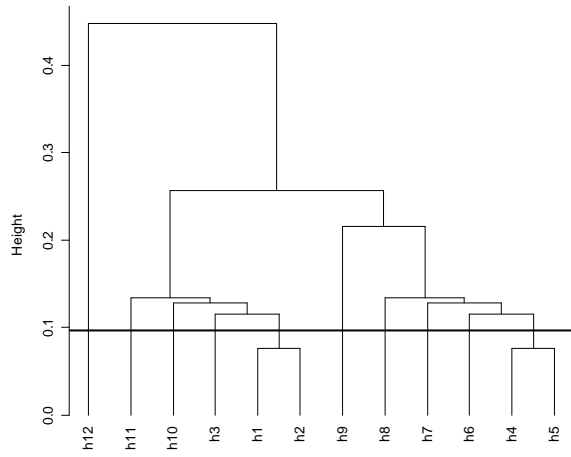
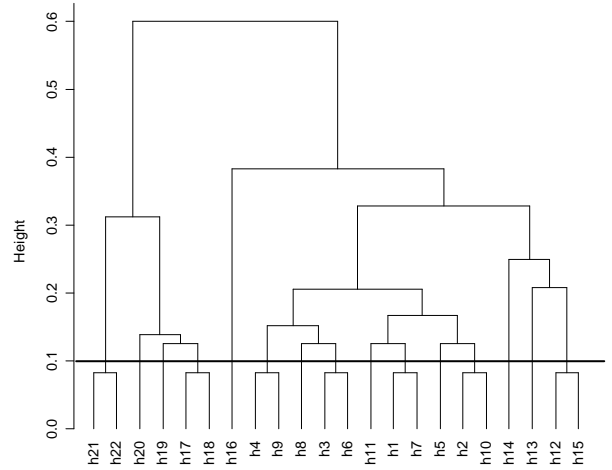


Figure 26. Dendrogrammes basés sur les matrices de pourcentage de différences entre paires d'haplotypes pour *erd3* (G), *lp3-1* (H), *lp3-3* (I), *pp2C* (J), *Rd21A* (K). Le seuil pour regrouper les haplotypes en « haplogroupes » fixé à 10% est indiqué par une ligne.

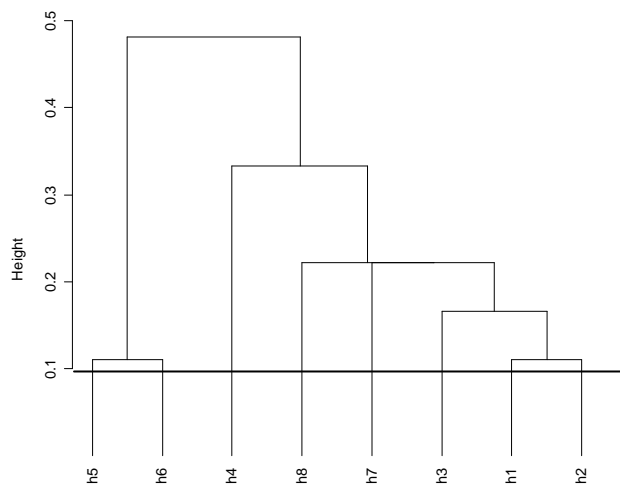
G)



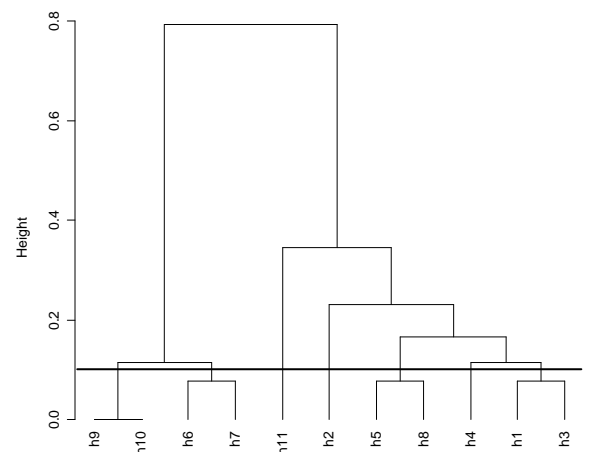
H)



I)



J)



G. Tests de neutralité au sein de la métapopulation atlantique

Au sein de la métapopulation atlantique, quatre tests de neutralité intra-population ont été appliqués. Les statistiques de ces tests ont été estimées à l'aide de DNAsp :

1) La statistique D de Tajima (1989) compare les estimateurs θ_π et θ_S selon la relation $D = (\theta_\pi - \theta_S) / (\text{Var}(\theta_\pi - \theta_S))^{1/2}$. Des valeurs significativement négatives reflètent un excès d'allèles rares, qui pourrait résulter d'un événement de sélection directionnelle et/ou une expansion démographique. A l'inverse, des valeurs significativement positives indiqueraient un excès d'allèles en fréquence intermédiaire, et seraient consistants avec une signature de sélection balancée et/ou de goulot d'étranglement.

2) La statistique F_S de Fu (1997) est basée sur la probabilité que le nombre attendu d'haplotypes sous un modèle neutre, étant donnée la diversité observée, soit plus élevé ou égal au nombre d'haplotypes observés. L'interprétation de ce test est similaire à celui du test de Tajima.

3) La statistique H de Fay et Wu (2000) compare les estimateurs θ_π and θ_H du taux de mutation, ce dernier étant basé sur la fréquence de mutations dérivées. Ces mutations dérivées ont été déduites de séquences de *Pinus taeda*, une espèce séparée de *Pinus pinaster* par un taux de divergence d'environ 0.028 (Brown *et al.* 2004). Les séquences de *Pinus taeda* ont été obtenues à partir de la banque de données Genbank (numéros d'accèsion 4F101790, AF036095, AAFP75823 pour les gènes *PR-AGP4*, *CCoAOMT* et *GRP3* respectivement) ou directement séquencées à partir d'ADN de mégagamétophyte de *Pinus taeda* en utilisant les mêmes couples d'amorces que pour *Pinus pinaster* (pour les gènes *dhn-1* et *Glucan*). Une valeur significativement négative de la statistique H pourrait refléter un excès de mutations dérivées, résultat d'une sélection directionnelle sur une mutation avantageuse avec effet d'entraînement des mutations dérivées associées à elle après recombinaison.

4) la statistique Z_{ns} (Kelly 1997) est une moyenne des estimations de la corrélation au carré entre toutes les paires de sites possibles. Plus la valeur de cette statistique est grande, plus il y a de sites polymorphes en déséquilibre. Un maintien de déséquilibre élevé peut être interprété comme de la sélection directionnelle avec effets d'entraînements ou comme de la sélection balancée.

Les intervalles de confiance et seuils de signification de l'ensemble de ces statistiques ont été obtenus par des simulations de coalescence à l'aide de DNAsp. Ces simulations sont basées sur le nombre de sites en ségrégation, et intègrent le taux de recombinaison ρ . Nous avons également utilisé des fenêtres coulissantes pour les statistiques D et H (par pas de 100 pb toutes les 25 pb) afin de localiser plus précisément les régions des gènes où les statistiques

s'écartaient éventuellement de la neutralité. Le niveau de significativité de la statistique D à chaque fenêtre est donné par des tests bilatéraux en supposant que la statistique suit une distribution β (Tajima 1989).

Chapitre IV

Patrons de diversité et différenciation moléculaires : résultats et discussion

Ce chapitre synthétise dans une première partie les principaux résultats obtenus avec les méthodes de détection d' « outliers » basées sur la différenciation génétique entre populations pour 11 gènes candidats polymorphes. Ces résultats ont fait l'objet d'une publication dans *Molecular Biology and Evolution* (Eveno *et al.* 2008, Annexe 1). Dans cette étude, que l'on considère les 10 populations représentatives de la distribution du pin maritime séparément ou bien regroupées en 6 pools homogènes d'après la différenciation des microsatellites, les résultats sont très similaires en termes de proportions d' « outliers » détectés, au niveau haplotypique (celui de la séquence entière pour chaque gène) ou bien au niveau des SNPs. Certains « outliers » identifiés dans le cas des 10 populations étaient même « masqués » en considérant les 6 groupes. Par conséquent, nous synthétisons ici les résultats sur les 10 populations (voir l'Annexe 2 pour les résultats des 6 groupes, informations supplémentaires de Eveno *et al.* 2008). Les questions de robustesse des méthodes à la taille des échantillons et à la variation des taux de migration, ou celles liées aux tests multiples sont développées dans l'article (Annexe 1).

Nous présentons ensuite les résultats sur la diversité moléculaire et les tests de neutralité au sein de la métapopulation atlantique pour les gènes *PR-AGP4*, *dhn1*, *CCoAOMT* et *Glucan*, pour lesquels un minimum de 30 séquences haploïdes étaient disponibles (sauf pour *GRP3* pour lequel 18 séquences étaient disponibles).

L'ensemble des résultats est ensuite discuté, en essayant notamment de comparer les niveaux de diversité observés 1) aux prédictions théoriques que l'on peut faire sur les taux d'évolution, 2) aux données expérimentales chez d'autres espèces proches, et 3) aux informations historiques disponibles chez l'espèce étudiée. Nous tentons aussi de proposer des scénarios évolutifs sur les différents gènes, leur rôle potentiel dans l'adaptation du pin maritime étant discuté dans le chapitre VII (Discussion générale).

A. Synthèse des résultats sur l'approche de détection d' « outliers »

Rappelons que cette approche était basée sur la différenciation génétique entre populations à deux niveaux : au niveau de chaque SNP au sein des gènes candidats et au niveau de chaque gène grâce à l'information haplotypique. Les haplotypes ont été considérés soit comme équidistants les uns des autres, soit plus ou moins proches selon le nombre de polymorphismes qui les séparait. Trois méthodes ont été utilisées pour détecter des SNPs ou gènes dont les niveaux de différenciation s'écartaient d'une distribution neutre : « Fdist2 », « FstSNP » et « BayesFst » (voir Chapitre III).

Au niveau haplotypique, 5 gènes (*PR-AGP4*, *erd3*, *dhn-2*, *lp3-1* et *dhn1*) parmi les 11 ont été identifiés comme étant des « outliers » à un seuil de 5% avec l'une des deux méthodes utilisées : « Fdist2 » ou « BayesFst » (Figure 27 et Tableau 11). Parmi ces gènes, *dhn-1* et *dhn-2* ont été identifiés par les deux méthodes. Même après avoir corrigé les *p-valeurs* dans la méthode fréquentiste « Fdist2 », en adoptant une approche minimisant le taux de faux positifs (FDR) (voir *q-valeurs* dans le Tableau 11 et le Chapitre I Paragraphe B.2.3. sur les problèmes liés aux tests multiples), *dhn-2* était encore « outlier » à 9%.

Parmi les 5 gènes « outliers », 2 gènes (*PR-AGP4*, *erd3*), présentaient une différenciation plus élevée que celle attendue sous un modèle neutre que ce soit en considérant chaque haplotype séparément ou en regroupant ceux distants d'un nombre minimum de mutations. A l'inverse, trois gènes (*dhn-1*, *dhn-2* et *lp3-1*) présentaient une différenciation plus faible que celle attendue sous l'hypothèse de neutralité.

Parmi les 94 SNPs non redondants et ayant une fréquence allélique minimale de 5%, 15 ont été détectés comme « outliers » à 10% avec « FstSNP » ou « BayesFst ». Parmi eux, seuls deux « outliers » positifs ont été détectés avec les deux méthodes. Aucun des 8 « outliers » détectés avec « FstSNP » ne restaient significatifs après avoir corrigé les *p-valeurs* par une approche basée sur le FDR (voir *q-valeurs* dans Tableau 11). Cependant, la correction effectuée dans le cas de tests multiples fait l'hypothèse que tous les tests sont indépendants, ce qui correspondrait à une situation où les loci sont tirés aléatoirement sur l'ensemble du génome. Or, dans notre cas, il s'agit de SNPs au sein de gènes candidats choisis pour leur éventuelle implication lors d'une réponse au stress hydrique, une telle correction ne serait donc pas complètement adaptée. D'autre part, les « outliers » détectés au niveau des SNPs, sans faire la correction des tests multiples, sont cohérents avec les résultats au niveau haplotypique. Pour *PR-AGP4*, plusieurs SNPs en déséquilibre ($r^2 > 0.8$, Figure 28) sont détectés comme « outliers » positifs, et pour *erd3* un SNP non synonyme très différencié expliquerait le profil « outlier » au niveau haplotypique. A l'inverse, plusieurs SNPs sont détectés comme « outliers » négatifs pour *dhn-1* et *dhn-2* (Tableau 11).

Figure 27. Distribution des valeurs de F_{ST} pour chaque gène, chaque SNP au sein des gènes, et chaque SSR, en fonction de leur hétérozygotie moyenne (H_S), entre les 10 populations contrastées (F_{ST} moyen utilisé pour les simulations de 0.138). Les quantiles à 5% et 95% des enveloppes neutres simulées sont représentés par des lignes continues et pointillées (correspondant à un test bilatéral à 10%) respectivement pour « Fdist2 » et « FstSNP ». Les lignes en noir représentent les limites des enveloppes pour le premier lot de gènes (avec une moyenne de 14 gamètes par population pour les SNPs et de 9 gamètes pour les gènes sans données manquantes), et les lignes en gris, celles pour le deuxième lot de gènes (avec une moyenne d'environ 6 gamètes par population pour les SNPs et les gènes). Les gènes détectés comme « outliers » avec « Fdist2 » sont en gras et les SNPs détectés comme « outliers » avec « FstSNP » sont indiqués par des flèches (d'après la figure 2 tirée de Eveno *et al.* 2008, Annexe 1).

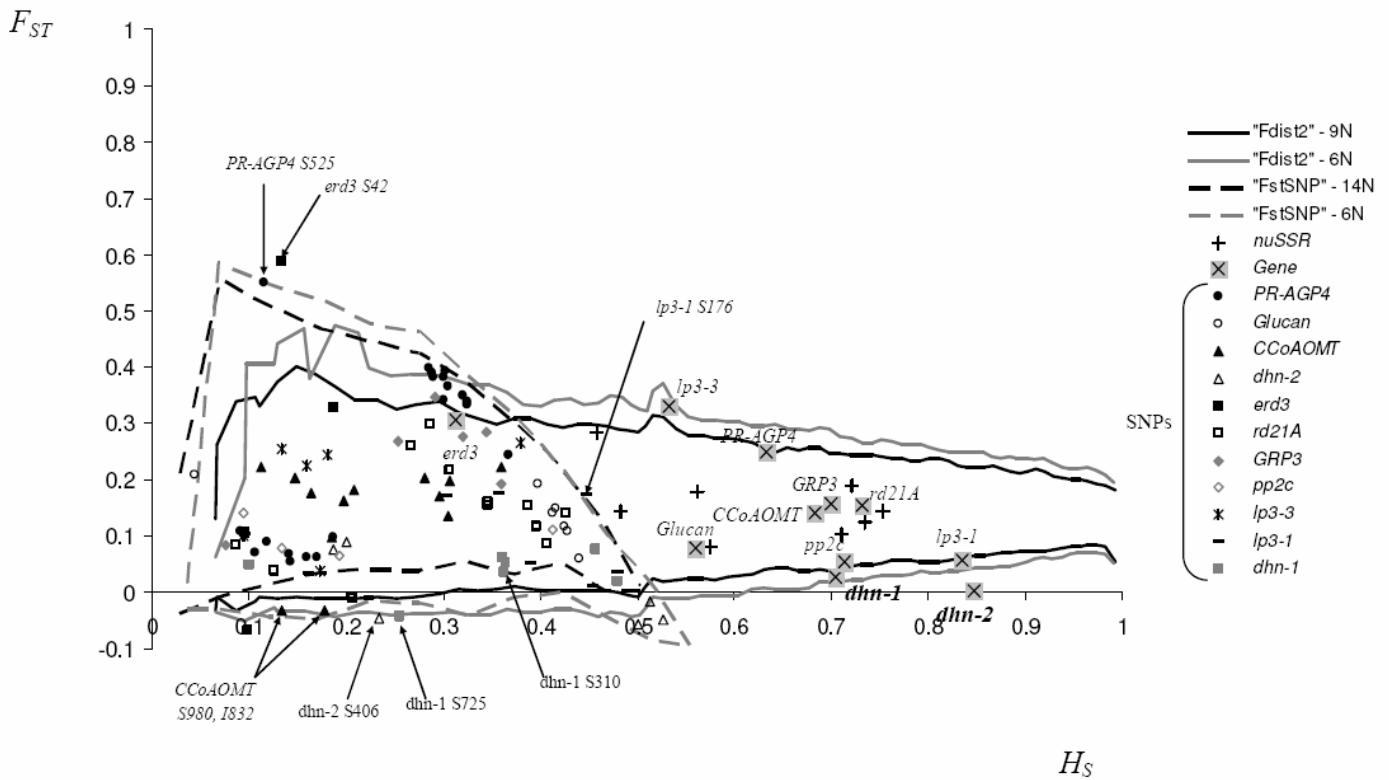


Tableau 11. “Outliers” détectés (A) au niveau haplotypique , et (B) au niveau des SNPs ayant une hétérozygotie supérieure à 0.1, en considérant les 10 populations contrastées pour les méthodes “Fdist2”, “FstSNP” ou “BayesFst” (d’après le tableau 1 de Eveno *et al.* 2008, Annexe 1).

A)

Gène	F_{ST}	"Fdist2"		"BayesFst"					
		P -valeur (c)	q -valeur (g)	"HG" (a)			"H" (b)		
				P -valeur (c)	α_i (d)	SD (e)	P -valeur (c)	α_i (d)	SD (e)
<i>PR-AGP4</i>	0.249			0.04	0.67	0.47			
<i>erd3</i>	0.306			0.10	1.29	0.67	0.04	1.22	0.58
<i>GRP3</i>	0.158			0.09	0.63	0.40			
<i>dhn-2</i>	0.002	0.01	0.09	0.03	-1.14	0.68	0.05	-0.39	0.63
<i>lp3-1</i>	0.058			0.11	-0.58	0.57	0.05	-0.75	0.60
<i>dhn-1</i>	0.027	0.06	0.27	0.02	-1.12	0.60	0.01	-1.28	0.58

B)

SNP	Région du gène / type de mutation (f)		"Fdist2"		"FstSNP"		"BayesFst"		
	F_{ST}	P -valeur (c)	P -valeur (c)	q -valeur (g)	p -valeur (c)	α_i (d)	SD (e)		
<i>PR-AGP4</i>	S182	intron / -	0.383	0.08			0.05	0.95	0.50
	S331	intron / -	0.398	0.06			0.05	0.98	0.49
	S472	intron / -	0.382	0.08			0.08	0.62	0.50
	S525	intron / -	0.550	0.02	0.07	0.51	0.08	0.74	0.62
	S666	intron / -	0.341				0.09	0.95	0.55
	S762	intron / -	0.339	0.09			0.08	0.73	0.52
	S1177	exon / ns	0.333	0.09					
	S1181	exon / s	0.350	0.08					
	S1409	3'UTR / -	0.365	0.08			0.04	0.97	0.52
	S1480	3'UTR / -	0.394	0.06					
<i>GRP3</i>	S1488	3'UTR / -	0.391	0.08					
	S110	exon / ns	0.285	0.08					
	S164	exon / s	0.347	0.06			0.04	1.21	0.6
<i>erd3</i>	S633	3'UTR / -	0.277	0.10			0.03	1.55	0.71
	S42	exon / ns	0.587	0.04	0.06	0.49			
<i>CCoAOMT</i>	I832	exon / ns	-0.034	0.01	0.01	0.33			
	S980	exon / ns	-0.031	0.02	0.02	0.33			
<i>dhn-2</i>	S202	exon / ns	-0.047	0.04					
	S332	intron / -	-0.056	0.07					
	S406	exon / s	-0.045	0.08	0.05	0.49			
<i>lp3-1</i>	S176	3'UTR / -	0.174		0.04	0.49			
<i>dhn-1</i>	S310	exon / s	0.063		0.06	0.49			
	S725	exon / ns	-0.043	0.02	0.01	0.33			

(a) Groupes d'haplotypes incluant les haplotypes avec moins de 10% de polymorphismes de différence relativement au nombre total de polymorphismes le long de chaque gène.

(b) Tous les haplotypes sont considérés séparément comme des allèles indépendants.

(c) Niveau de significativité (P -valeur) pour les “outliers” détectés (en gras pour $P < 0.05$).

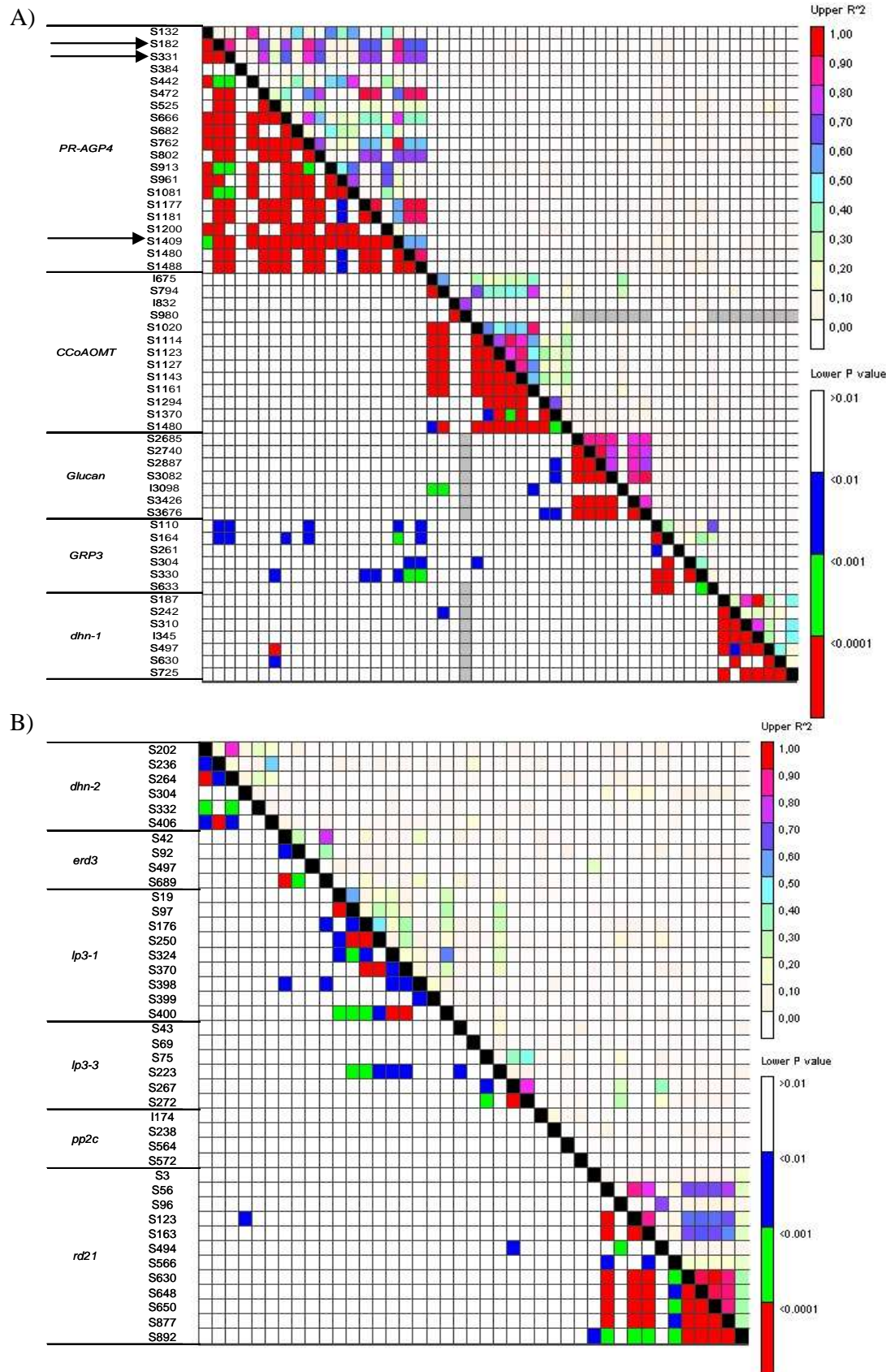
(d) Mode de la distribution a posteriori du paramètre α_i .

(e) Ecart-type de la distribution a posteriori du paramètre α_i .

(f) s: polymorphisme synonyme, ns: polymorphisme non synonyme.

(g) les q -valeurs sont les taux de faux positifs analogues aux p -valeurs. Elles ont été obtenues pour chaque gène “outlier” en considérant 11 hypothèses indépendantes correspondant aux 11 gènes candidats, et pour chaque SNP, en considérant 94 hypothèses indépendantes correspondant aux 94 polymorphismes testés.

Figure 28. Déséquilibres de liaison entre SNPs polymorphes (fréquence allélique minimum de 5%) en excluant ceux qui sont en *DL* complet, et en considérant l'ensemble des gamètes sur les 10 populations pour les gènes *PR-AGP4*, *CCoAOMT*, *Glucan*, *GRP3* et *dhn1* (A), et les gènes *dhn2*, *erd3*, *lp3-1*, *lp3-3*, *pp2c* et *rd21* (B). Les estimations du *DL* par le paramètre r^2 (corrélation au carré entre sites) sont indiquées au dessus de la diagonale, et les probabilités du test exact de Fisher sont indiquées en dessous de la diagonale. Les 3 SNPs détectés comme « outliers » positifs sur *PR-AGP4* sont indiqués par des flèches.



D'autre part, pour deux gènes non « outliers » au niveau haplotypique (*CCoAOMT* et *GRP3*), quelques SNPs ou Indels ont été détectés comme « outliers » négatifs (Indel I832 et SNP S980 pour *CCoAOMT*, Tableau 11 et Figure 27) et comme « outliers » positifs (SNP S164 pour *GRP3*, Tableau 11). Pour ces gènes, on pourrait supposer qu'il existe un effet direct dû à la mutation, d'autant plus que certaines de ces mutations peuvent être non synonymes.

B. Résultats sur les patrons de diversité nucléotidique et tests de neutralité au sein de la métapopulation atlantique

B.1. Polymorphismes détectés

Au total, plus de 4,5 kb d'ADN ont été séquencés pour une trentaine de gamètes sur les 5 gènes (*PR-AGP4*, *CCoAOMT*, *GRP3*, *dhn1* et *Glucan*). Sur ces séquences, 96 sites polymorphes ont été identifiés, dont 88 SNPs et 8 Indels, représentant en moyenne un site polymorphe tous les 47 pb. Parmi ces sites polymorphes, 18 possédaient un allèle en une seule copie (singletons), la plupart appartenant au gène *PR-AGP4*. Sur les 88 SNPs, 26% sont des mutations non-synonymes, les Indels correspondant tous à des mutations silencieuses (Tableau 12).

B.2. Variation des patrons de diversités nucléotidique et haplotypique entre gènes

La diversité nucléotidique moyenne par site, θ_π , est de $5.48 (x 10^{-3})$ sur l'ensemble des sites, et varie de $2.92 (x 10^{-3})$ pour *GRP3* à $10.03 (x 10^{-3})$ pour *dhn1* (Tableau 12). La diversité nucléotidique aux sites silencieux est supérieure à celle estimée aux sites non-synonymes pour les 5 gènes, avec deux fragments de gènes (*CCoAOMT* et *Glucan*) ne présentant aucun site non-synonyme sur les régions codantes séquencées (335 pb et seulement 72 pb pour *CCoAOMT* et *Glucan* respectivement). La diversité nucléotidique moyenne basée sur le nombre de sites polymorphes, θ_S , est similaire à θ_π sur l'ensemble des sites. Entre les 5 gènes, les valeurs de chaque estimateur ne sont pas significativement différentes à cause de larges intervalles de confiance inhérents au processus généalogique sous-jacent (Tajima 1983). Des variations importantes sont cependant observées pour *PR-AGP4* entre θ_S et θ_π (voir partie B.4 plus bas). Pour le gène *dhn-1*, la valeur de θ_π est significativement supérieure à celle attendue sous un modèle neutre (d'après les simulations de coalescence intégrant le taux de recombinaison ρ , Tableau 13), et significativement inférieure pour le gène *PR-AGP4* (Tableau 12).

La diversité haplotypique est en moyenne de 0.546 et varie du simple au double entre gènes (de 0.323 pour *Glucan* à 0.774 pour *dhn1*, Tableau 12). Trois gènes (*PR-AGP4*, *CCoAOMT* et *Glucan*) présentent des valeurs H_d significativement inférieures à celles attendues sous un modèle neutre, indiquant de fortes structures haplotypiques (illustrées par les Figures 29 A, B et C). Parmi ces trois gènes, *PR-AGP4* possède six haplotypes dont un en fréquence majoritaire (76%) (Figure 29A). En regroupant les haplotypes proches (quelques mutations différentes seulement) il reste 4 groupes d'haplotypes très distants (au moins 10 mutations de différence) dont un qui est majoritaire (84%), un en fréquence intermédiaire (11%), et deux autres en fréquence beaucoup plus faible (<4%). Le second gène qui possède une très forte structure haplotypique (*Glucan*), est représenté par seulement deux haplotypes distants de 11 mutations, dont l'un a une fréquence de 80% (Figure 29B). Enfin, le gène *CCoAOMT* possède 6 haplotypes répartis en deux groupes d'haplotypes proches ayant des fréquences de 79% et 20% (Figure 29C). Ces deux groupes d'haplotypes sont différents d'au moins 11 mutations. Pour *dhn1* et *GRP3*, le nombre d'haplotypes (Figures 29D et E) et les valeurs de H_d observés sont plus importants (0.774 et 0.627 respectivement, Tableau 12).

B.3. Variation du déséquilibre de liaison entre sites polymorphes intra-gène

Parmi les 1426 comparaisons possibles entre paires de sites intra-gène (après avoir exclu les polymorphismes dont la fréquence de l'allèle le plus rare était inférieure à 5%), 54% (770 comparaisons) présentaient des valeurs de r^2 significatives à 5% avec un test exact de Fisher. Parmi elles, presque toutes sont également significatives à 1%, et 17% restent significatives à 5% après une correction de Bonferroni. Ce niveau élevé de DL se traduit également par des statistiques Z_{ns} significatives (simulations de coalescence réalisées en intégrant ρ) pour tous les gènes sauf pour *GRP3* (Tableau 13).

Néanmoins, on observe une grande hétérogénéité des patrons de DL entre gènes, avec des valeurs de Z_{ns} allant de 0.177 pour *GRP3* à 1 pour *Glucan*. La baisse globale du DL observée sur l'ensemble des 5 gènes n'est également pas représentative de celle observée sur chacun des gènes (Figure 30) : pour *GRP3*, des sites proches de 50 pb peuvent être statistiquement indépendants, alors que le DL ne baisse pas ou très peu le long des fragments de *CCoAOMT* et *Glucan*.

Le gène *Glucan* avec 2 haplotypes dont tous les sites sont en DL complet dans la métapopulation atlantique, a un taux de recombinaison ρ , estimée à 0 (Tableau 13). Quelques rares mutations ont cependant été observées sur l'ensemble des 10 populations.

Tableau 12. Diversités nucléotidique et haplotypique sur les 5 gènes candidats au sein de la métapopulation.

Code du gène	N ^(a)	Ensemble des sites								Sites non synonymes					Sites silencieux ^(b)					N _h ^(f) (singl.)	H _d ^(e)
		L ^(c)	SNPs (singl.)	Indels (singl.)	θ_π ^(e)	95% CI	θ_s ^(e)	95% CI	L ^(d)	SNPs (singl.)	Indels (singl.)	θ_π ^(e)	θ_s ^(e)	L ^(d)	SNPs (singl.)	Indels (singl.)	θ_π ^(e)	θ_s ^(e)			
<i>PR-AGP4</i>	62	1383	42 (15)		3.66* ⁽⁻⁾	-0.22 - 7.60	6.58	2.62 - 10.54	397	12 (5)		2.11	6.44	986	30 (10)		4.26	6.48	9 (6)	0.417*** ⁽⁻⁾	
<i>CCoAOMT</i>	39	973	14	1	4.71	-0.41 - 9.83	3.68	0.98 - 6.38	335			0.00	0.00	637	14	1	7.16	5.57	6 (2)	0.590* ⁽⁻⁾	
<i>Glucan</i>	31	1127	15	6	6.06	-0.29 - 12.41	4.72	1.37 - 8.07	72			Na ^(g)	Na ^(g)	1053	15	6	6.43	4.99	2	0.323*** ⁽⁻⁾	
<i>GRP3</i>	18	436	5 (2)		2.92	0.05 - 5.79	3.35	0.06 - 6.64	251	3 (2)		1.73	3.49	185	2		2.57	1.59	6 (3)	0.627	
<i>dhn-1</i>	47	604	12 (1)	1	10.03*** ⁽⁺⁾	-0.42 - 20.48	4.93	1.26 - 8.60	385	8 (1)		8.90	4.71	215	4	1	11.87	5.26	6 (1)	0.774	
Total	197	4523	88 (18)	8					1440	23 (8)				3076	65 (10)	8			29 (12)		
Moyenne					5.48		4.65					3.18	3.66				6.46	4.78		0.546	

^(a) Nombre de séquences analysées;

^(b) Sites silencieux incluant les polymorphismes synonymes dans les régions codantes et les polymorphismes dans les régions non codantes;

^(c) Longueur totale analysée en paires de bases (pb);

^(d) Longueur estimée en pb;

^(e) Diversité nucléotidique θ_π (Nei 1987) et θ_s (Watterson 1975) par site ($\times 10^{-3}$) avec leur intervalle de confiance à 95%, et diversité haplotypique H_d (Nei 1987). La significativité des estimations de diversités nucléotidique et haplotypique a été testée par simulations de coalescence en présence de recombinaison ρ et en fonction du nombre de sites polymorphes pour θ_π et H_d et en fonction de la moyenne θ_π simulée sous un modèle neutre pour θ_s . Niveaux de significativité: ***: $P < 0.001$, **: $P < 0.01$, *: $P < 0.05$. ⁽⁻⁾ and ⁽⁺⁾: diversités nucléotidique ou haplotypique observées inférieures et supérieures à l'attendu neutre respectivement ;

^(f) Nombre d'haplotypes;

^(g) Diversité nucléotidique non estimée due à une longueur de séquence analysée inférieure à 150 pb.

Tableau 13. Estimateurs de la recombinaison par gène et tests de neutralité sélective sur les 5 gènes candidats.

Code du gène	N ^(a)	L ^(b)	S ^(c)	R_M ^(d)	ρ ^(e)	Z_{NS} de Kelly	D de Tajima	F_S de Fu	H de Fay et Wu
<i>PR-AGP4</i>	62	1383	42	1	2.142	0.274* (+)	-1.451* (-)	3.319* (+)	-15.298** (-)
<i>CCoAOMT</i>	39	973	15	1	0	0.676** (+)	0.912	4.431* (+)	-2.216
<i>Glucan</i>	31	1127	21	0	0	1***(+)	1.005	15.781*** (+)	-2.452
<i>GRP3</i>	18	436	5	2	15.265	0.177	-0.404	-1.737	-4.523** (-)
<i>dhn-1</i>	47	604	13	1	3.411	0.422* (+)	3.126*** (+)	7.155	0.959
Total	197	4523	96						

^(a) Nombre de séquences analysées;

^(b) Longueur totale analysée en paires de bases (pb);

^(c) Nombre total de polymorphismes incluant SNPs et Indels;

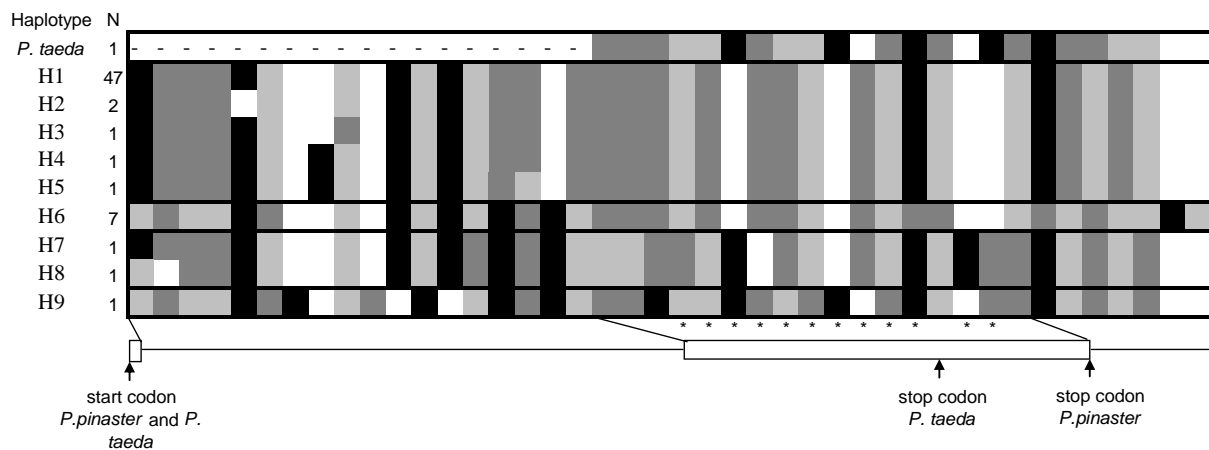
^(d) Estimation du nombre minimum par gène d'évènements de recombinaison dans l'histoire de l'échantillon par un test « 4-gamètes » (Hudson et Kaplan 1985);

^(e) Estimation du taux de recombinaison de la population par gène en utilisant LDhat ver 2.0 (McVean et al. 2002);

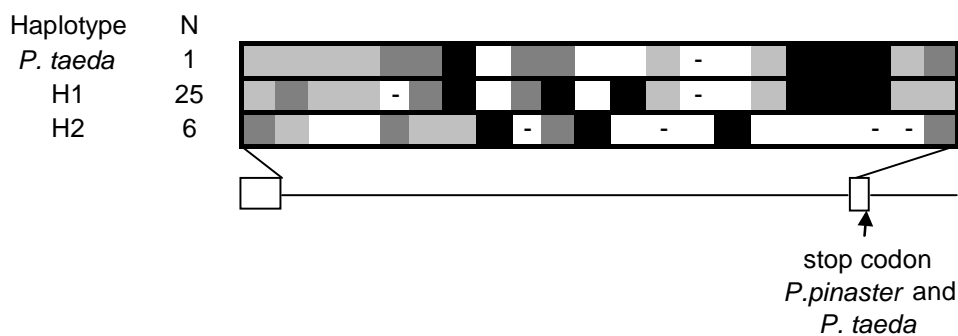
^(f) La significativité des statistiques des tests de neutralité (Z_{ns} , D , F_S et H) a été testée par simulations de coalescence en présence de recombinaison ρ et en fonction du nombre de sites polymorphes. Niveaux de significativité: ***: $P < 0.001$, **: $P < 0.01$, *: $P < 0.05$. ⁽⁻⁾ and ⁽⁺⁾: diversités nucléotidique ou haplotypique observées inférieures et supérieures à l'attendu neutre respectivement ;

Figure 29. Structures haplotypiques et représentations schématiques de *PR-AGP4* (A), *Glucan* (B), *CCoAOMT* (C), *dhn-1* (D), et *GRP3* (E). Le premier haplotype représente l'haplotype de la séquence de *Pinus taeda* qui est utilisée pour inférer les mutations dérivées et les taux de substitutions. Les haplotypes proches de *Pinus pinaster* (voir Figure 30) sont groupés et encadrés en noir. Le nombre de chaque haplotype est également indiqué. Les étoiles indiquent les mutations non-synonymes. Les rectangles blancs ainsi que les lignes noires sous la structure haplotypique représentent les régions codantes et non codantes respectivement. Les positions des codons « start » et « stop » pour les séquences de *Pinus pinaster* et *Pinus taeda* sont indiquées par des flèches (lorsqu'ils sont inclus dans le fragment séquencé).

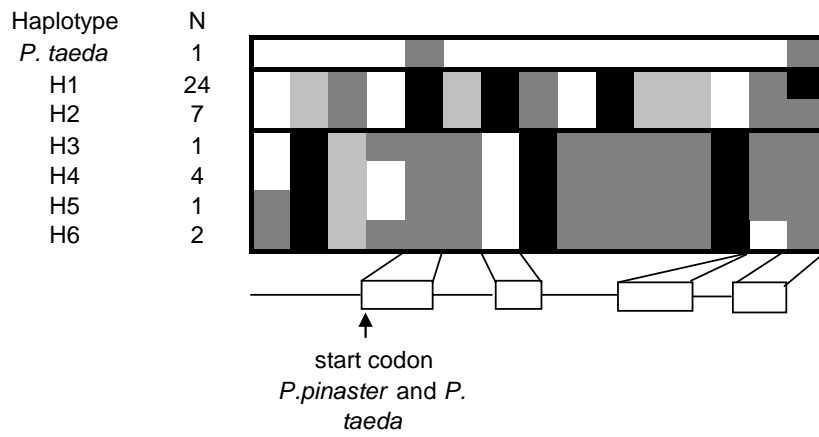
A) *PR-AGP4*



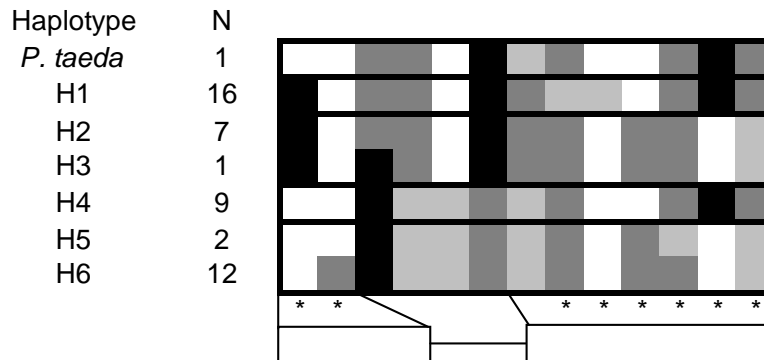
B) *Glucan*



C) *CCoAOMT*



D) *dhn1*



E) *GRP3*

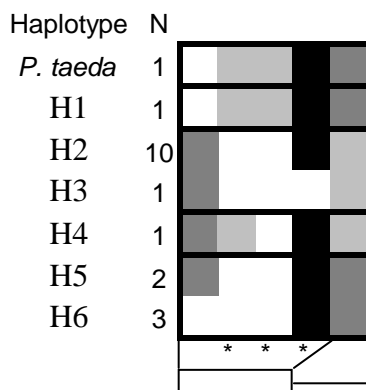


Figure 30. Diagramme de dispersion du *DL* (estimé par r^2) entre sites polymorphes non rares (fréquence allélique minimum > 5%) en fonction de la distance en paires de bases. Ces diagrammes sont réalisés au sein de la métapopulation atlantique sur 5 gènes avec une taille d'échantillon supérieure à 18, et en moyenne de 37 (*PR-AGP4*, *CCoAOMT*, *Glucan*, *dhn1* et *GRP3*). Les courbes de régression non-linéaire du *DL* en fonction de la distance sont représentées pour un ajustement sur l'ensemble des 5 gènes, ainsi que pour chaque gène.

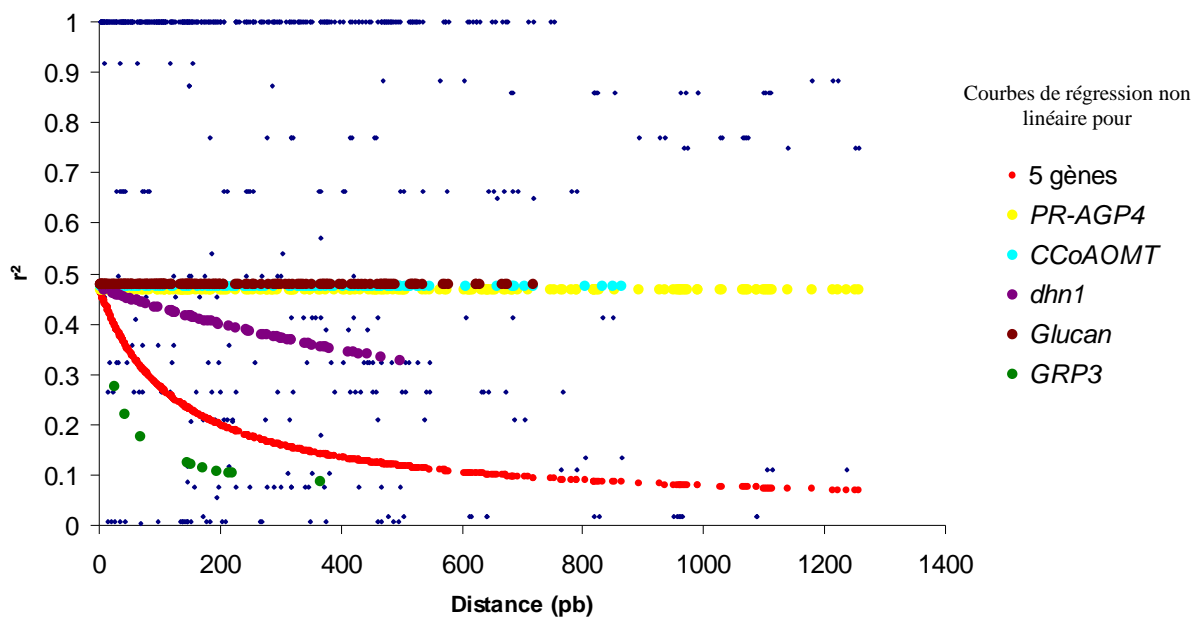
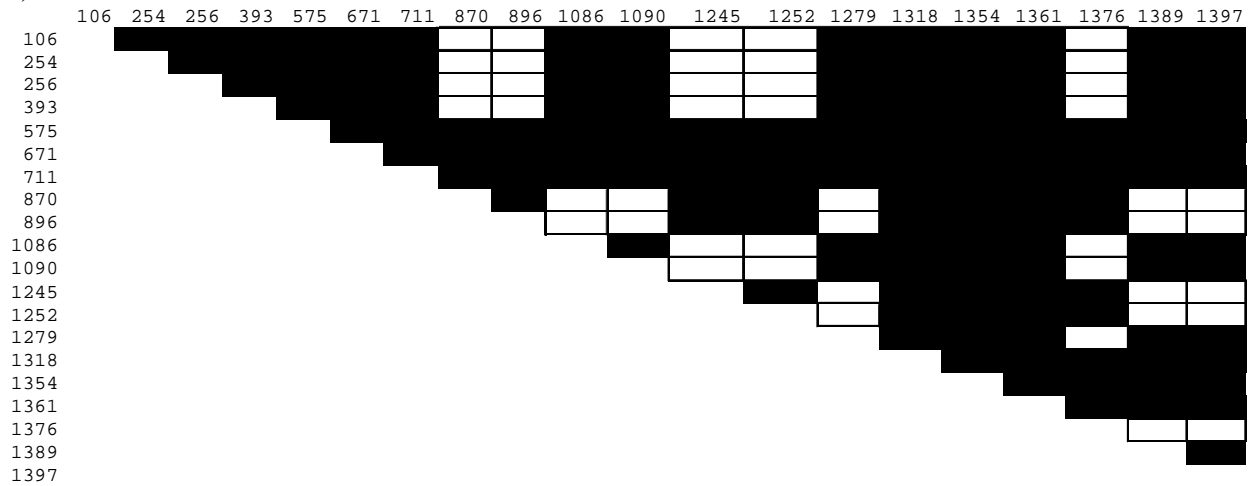
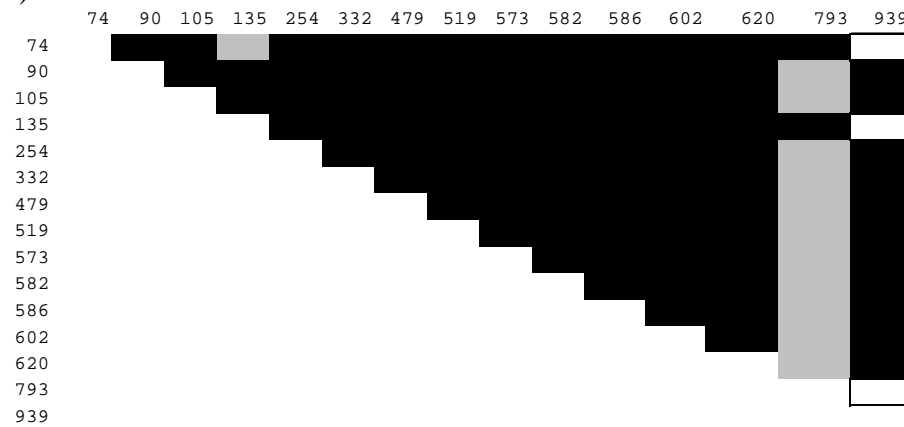


Figure 31. Matrice des p -valeurs des tests exacts de Fisher pour les déséquilibres de liaison entre sites polymorphes non-rares estimés pour *PR-AGP4* (A), *CCoAOMT* (B), *dhn-1* (C) et *GRP3* (D). Pour le gène *Glucan*, la matrice n'est pas représentée car tous les sites sont en déséquilibre de liaison complet. Pour les autres gènes, les cases noires correspondent à des p -valeurs inférieures à 1%, et les cases grises à des p -valeurs inférieures à 5%. Les nombres indiquent les positions des polymorphismes le long des gènes.

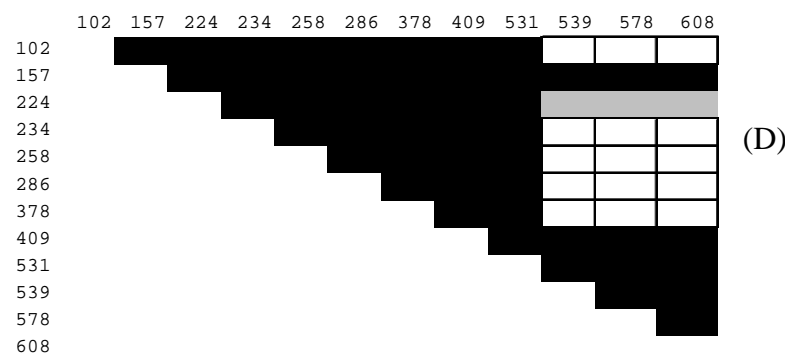
A) *PR-AGP4*



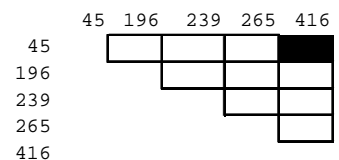
B) *CCoAOMT*



C) *dhn1*



D) *GRP3*



Le gène *CCoAOMT* présente également un excès de *DL* (Z_{ns} de 0.676) avec plus de 50% des comparaisons par paires de sites présentant un r^2 de 1, et 97% avec des valeurs de r^2 significatives à 5% (Figure 31B). Au sein de ce gène, on observe un bloc en *DL* complet de 366 pb (avec 9 SNPs des positions 254 à 620) (Figure 31B). On note aussi que malgré l'observation d'au moins un évènement de recombinaison dans l'histoire de l'échantillon ($R_M=1$) du à la présence de deux sites très peu polymorphes qui ne sont pas en *DL* complet, l'estimation du taux de recombinaison ρ est de zéro car la quasi-totalité des loci polymorphes sont en *DL* complet.

Les gènes *PR-AGP4* et *dhn1* présentent moins de sites en *DL* complet, comme le montrent des valeurs de Z_{ns} (0.274 et 0.422 respectivement) qui sont inférieures à celles des gènes *Glucan* et *CCoAOMT*, bien que significatives (Tableau 13). De plus, on observe un nombre d'haplotypes plus élevé au sein de ces gènes (Figure 29A et D). Les *DL* entre sites sont également très variables (r^2 de 0.01 à 1), et un tiers des comparaisons au sein de ces deux gènes montrent des *DL* significatifs à 5% (Figures 31 A et C). La présence de plusieurs petits blocs de sites en *DL* complet (5 blocs pour *PR-AGP4* et 3 pour *dhn1*) indique l'existence de plusieurs évènements de recombinaison passés. Les taux de recombinaison par gène estimé par la méthode de « vraisemblance composite » développée dans LDhat2 sont de 2.1 et 3.4 pour *PR-AGP4* et *dhn1* respectivement, bien qu'ils sous-estiment probablement le taux de recombinaison réel. Le gène *GRP3* est le seul pour lequel le test Z_{ns} n'est pas significatif (Tableau 13), et sur les 5 sites les plus polymorphes, aucun n'est en *DL*, exceptée une paire de sites à 5% (r^2 de 0.57).

B.4. Des écarts à la neutralité révélant des profils hétérogènes selon les gènes

Toutes les simulations de coalescence pour déterminer les valeurs critiques des statistiques des tests sous un modèle neutre ont été réalisées en intégrant l'estimateur du taux de recombinaison de la population par gène, ρ obtenu avec LDhat2. Parmi les cinq gènes, *PR-AGP4* est le seul à présenter un D de Tajima (noté D_T) négatif et significatif ($D_T=-1.451$) (Tableau 13), qui s'explique par une valeur de θ_S presque deux fois plus élevée que la valeur du θ_π (Tableau 12). Cela signifie qu'il existe un excès de mutations rares dans l'échantillon. En effet, 15 singletons sont observés dans ce gène dont 13 appartiennent à un haplotype unique distant des autres (H9, Figure 29A). Cet haplotype rare ou un haplotype proche est également présent sur l'ensemble des 10 populations en fréquence faible (<5%). Quelques mutations non-singletons (7) sont en fréquence faible (<4%) et réparties de façon uniforme sur le reste des haplotypes, principalement en amont du gène (Figure 29A).

L'analyse en fenêtre coulissante indique les régions des gènes où la statistique D_T s'écarte significativement de la neutralité. Pour *PR-AGP4*, cette statistique est négative et significative à 5% dans deux régions (avec des valeurs minimales de D_T de -2.012 et -1.969) (Figure 32A) : une première de 124 pb dans l'intron et une seconde de 140 pb dans l'exon de la 2ème partie du gène, qui comprennent toutes deux 5 SNPs singletons localisés sur l'haplotype unique H9, dont ceux de l'exon sont tous non-synonymes. Il est à noter que sur cette même région exonique, il existe une mutation polymorphe (>11%) qui est non synonyme (Figure 29A).

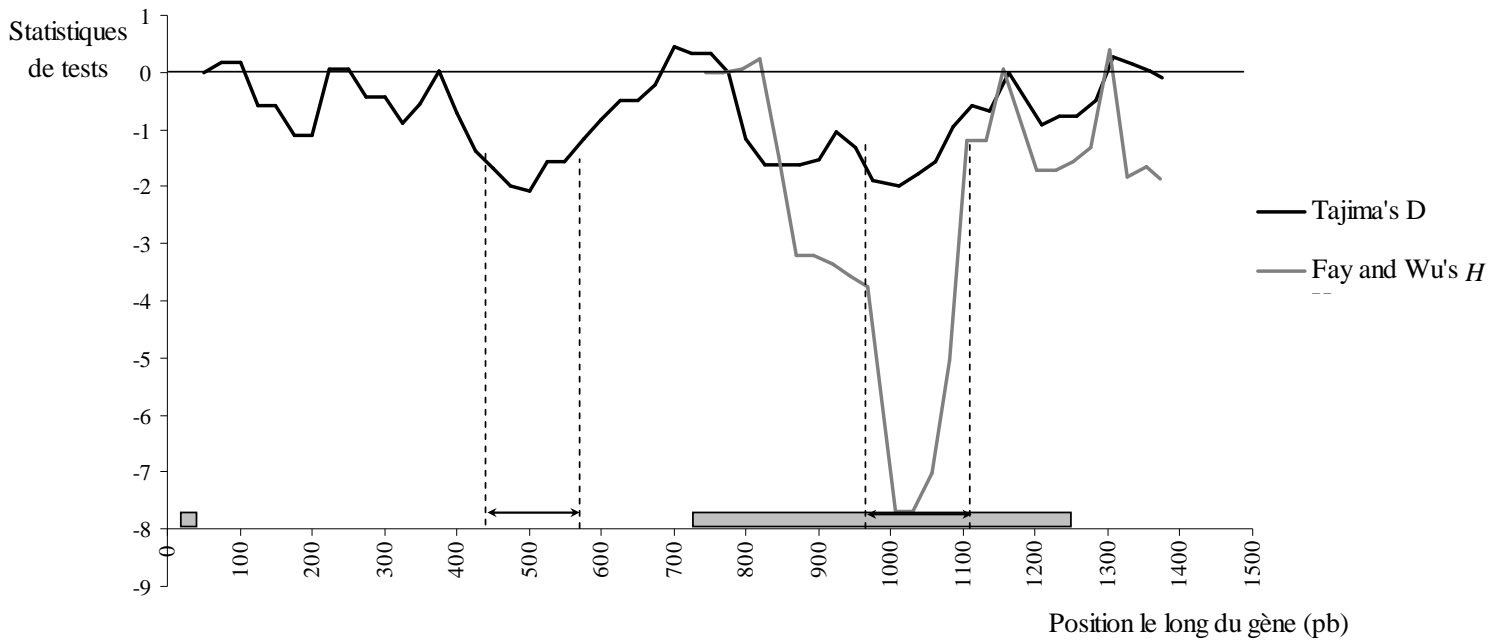
A l'inverse, le gène *dhn1* montre une valeur D_T significative à 5% et positive ($D_T=3.126$, Tableau 13), qui s'explique par une valeur de θ_π plus de deux fois supérieure à celle du θ_S (Tableau 12), et reflète un excès de polymorphismes en fréquence intermédiaire. Pour ce gène, D_T est significatif à 5% quasiment tout le long du gène. Une première fenêtre significative pour cette statistique couvre 229 pb de régions exonique et intronique (avec une valeur D_T maximum de 2.887), la région exonique incluant notamment un SNP non-synonyme polymorphe (>19%). Les autres SNPs ou Indels sur cette première fenêtre sont synonymes ou dans des régions non codantes, tous en *DL* complet et à une fréquence de 54%. Sur deux autres régions plus petites contiguës à la première, les statistiques D_T sont positives mais significatives à 10% seulement. Ces régions couvrent chacune 154 pb et 106 pb majoritairement de parties codantes, et comprennent 7 SNPs non synonymes et 1 SNP dans la région intronique, quasiment tous en *DL* complet.

Des valeurs significatives et positives du F_S de Fu pour trois gènes (*PR-AGP4*, *CCoAOMT* et *Glucan* avec des valeurs respectives de 3.319, 4.431 et 15.781, Tableau 13) indiquent que le nombre d'haplotypes observé est inférieur à celui attendu sous un modèle neutre. Ces résultats sont cohérents avec les fortes structures haplotypiques observées au sein de ces gènes (Figure 31).

Pour *PR-AGP4* et *GRP3*, il existe également un excès de mutations dérivées, comme l'attestent les valeurs négatives et significatives à 5% du H de Fay et Wu, noté H_{FW} (-15.298 et -4.523 pour *PR-AGP4* et *GRP3* respectivement, Tableau 13). Le long du gène *PR-AGP4*, la statistique H_{FW} diminue jusqu'à des valeurs de -7.710 dans la même région exonique où la statistique de D_T est négative (Figure 32A). Si l'on considère uniquement la région sur laquelle ces valeurs sont les plus faibles, seuls deux haplotypes sont présents (du 26^{ème} au 30^{ème} SNP), dont l'un (H9) contient uniquement des mutations dérivées en faible fréquence (variants identiques aux variants ancestraux de l'haplotype de *Pinus taeda*) (Figure 29A). Ces mutations rares dérivées correspondent pour la plupart aux mutations rares qui ont engendré un D_T négatif. Aux alentours de la région où la statistique H_{FW} est faible, le

Figure 32. Statistiques D de Tajima et H de Fay et Wu le long des gènes pour *PR-AGP4* (A), *Glucan* (B), *CCoAOMT* (C), *dhn-1* (D) et *GRP3* (E) estimées avec une “fenêtre” de 100 pb, centrée toutes les 25 pb. Les régions des gènes délimitées par des lignes discontinues représentent les régions où la statistique D de Tajima est significative à 5%. Les rectangles grisés indiquent les régions codantes.

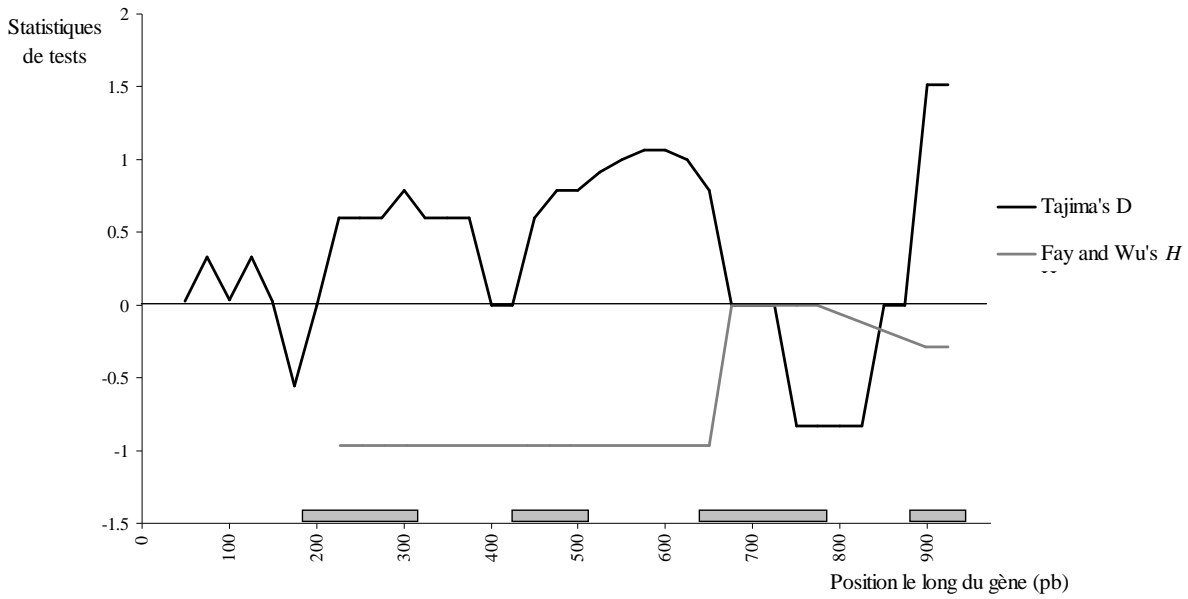
A) *PR-AGP4*



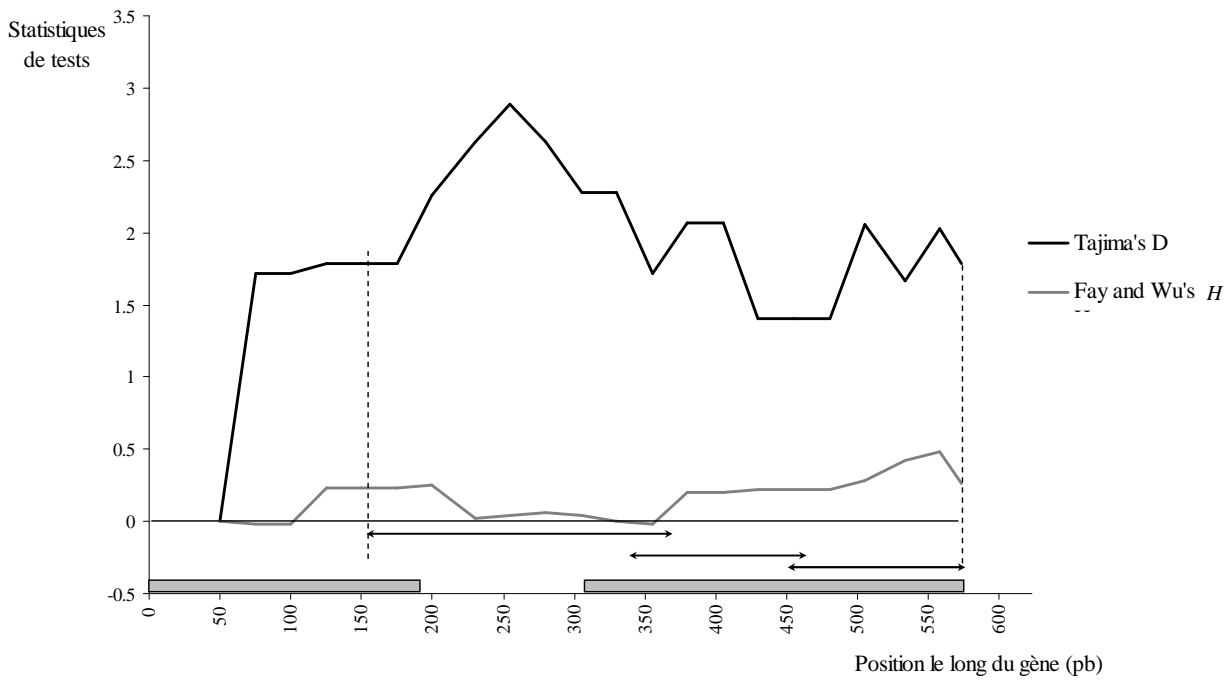
B) *Glucan*



C) *CCoAOMT*

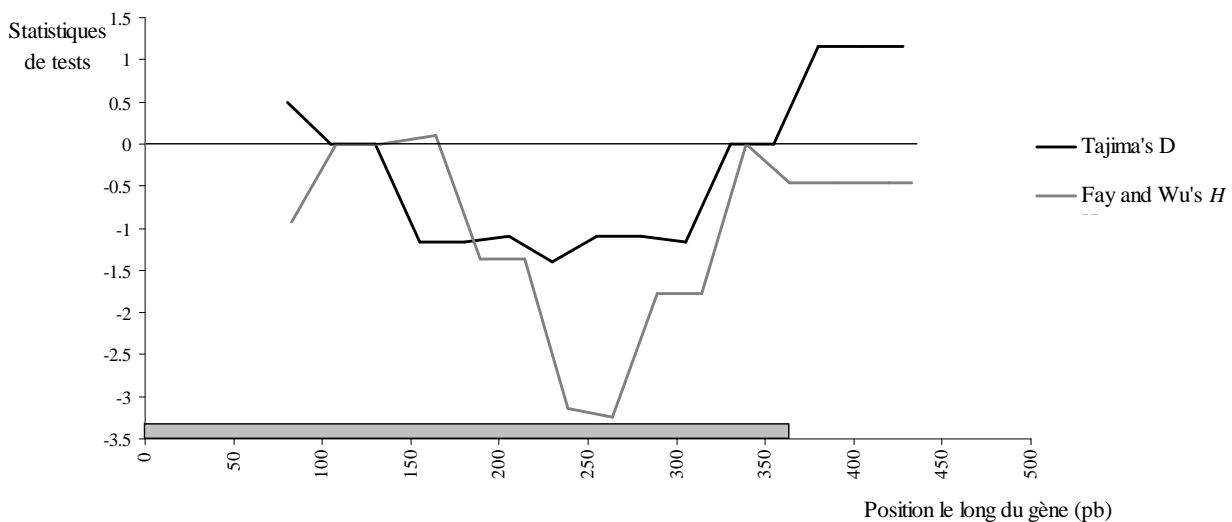


D) *dhn1*



E) *GRP3*

GRP3



nombre d'haplotypes augmente jusqu'à 9. Le long du gène *GRP3*, la statistique H_{FW} atteint des valeurs de -3.242 également dans une région exonique où la valeur D_T est la plus négative (Figure 32-E).

Cependant, la situation est légèrement différente de celle pour *PR-AGP4*, puisque l'excès de mutations dérivées caractéristique d'une valeur négative de H_{FW} , est principalement due à un seul variant dérivé non-synonyme (4^{ème} SNP, Figure 29E) qui est en fréquence très élevée (94%)

C. Discussion sur les niveaux de diversité et différenciation nucléotidiques chez le pin maritime

C.1. Observation d'une faible diversité nucléotidique silencieuse chez le pin maritime ?

La moyenne de diversité nucléotidique silencieuse ($\theta_{\pi S}$) (mutations synonymes et régions non codantes) observée au sein de cette étude est presque aussi élevée au sein de la métapopulation Atlantique (6.46×10^{-3}) que sur les 10 populations (7.85×10^{-3}) en utilisant les mêmes gènes (Eveno *et al.* 2008). Ces moyennes sont également du même ordre de grandeur que celles précédemment estimées chez cette espèce à partir de gènes candidats impliqués dans la formation du bois (Pot *et al.* 2005, Garnier-Géré *et al.* 2005, Tableau 14). Ce niveau de diversité est généralement comparable à celui estimé chez d'autres espèces de conifères (allant en moyenne de 0.72×10^{-3} pour *Pinus radiata* jusqu'à 12.2×10^{-3} pour *Pinus densata*) (Figure 33). D'autre part, les diversités chez le pin maritime sont environ deux fois inférieures à celles chez le Chêne sessile (comm. pers. J. Derory, projet Européen Digenfor sur une vingtaine de gènes candidats au débourrement). En général, il a été constaté que la diversité chez les conifères semble plus faible que chez les espèces d'arbres forestiers angiospermes et encore plus faible que chez les plantes annuelles (Figure 33, et voir références du Tableau 14).

C.2. Pourtant, une diversité plus élevée a été observée pour d'autres marqueurs...

Le consensus établi chez la majorité des espèces forestières avant l'ère de la génomique et des résultats sur des séquences d'ADN était qu'il existait une grande diversité génétique intra-population pour plusieurs types de marqueurs (allozymes, mais aussi RFLP, RAPDs, AFLPs) associée à une faible différenciation génétique entre populations (Hamrick *et al.* 1989, 1991, 1992). Ces résultats étaient interprétés comme la conséquence d'un mode de reproduction majoritairement allogame avec une capacité de dispersion du pollen pouvant

Figure 33. Représentation graphique des diversités nucléotidiques silencieuses $\theta_{\pi S}$ par paires de bases en moyenne chez les conifères (gris), les arbres angiospermes (blanc), les plantes annuelles (noir), les humains (rouge) et *Arabidopsis thaliana* (vert). Le nombre de loci à partir desquels les moyennes d'estimations ont été réalisées est indiqué entre parenthèses.

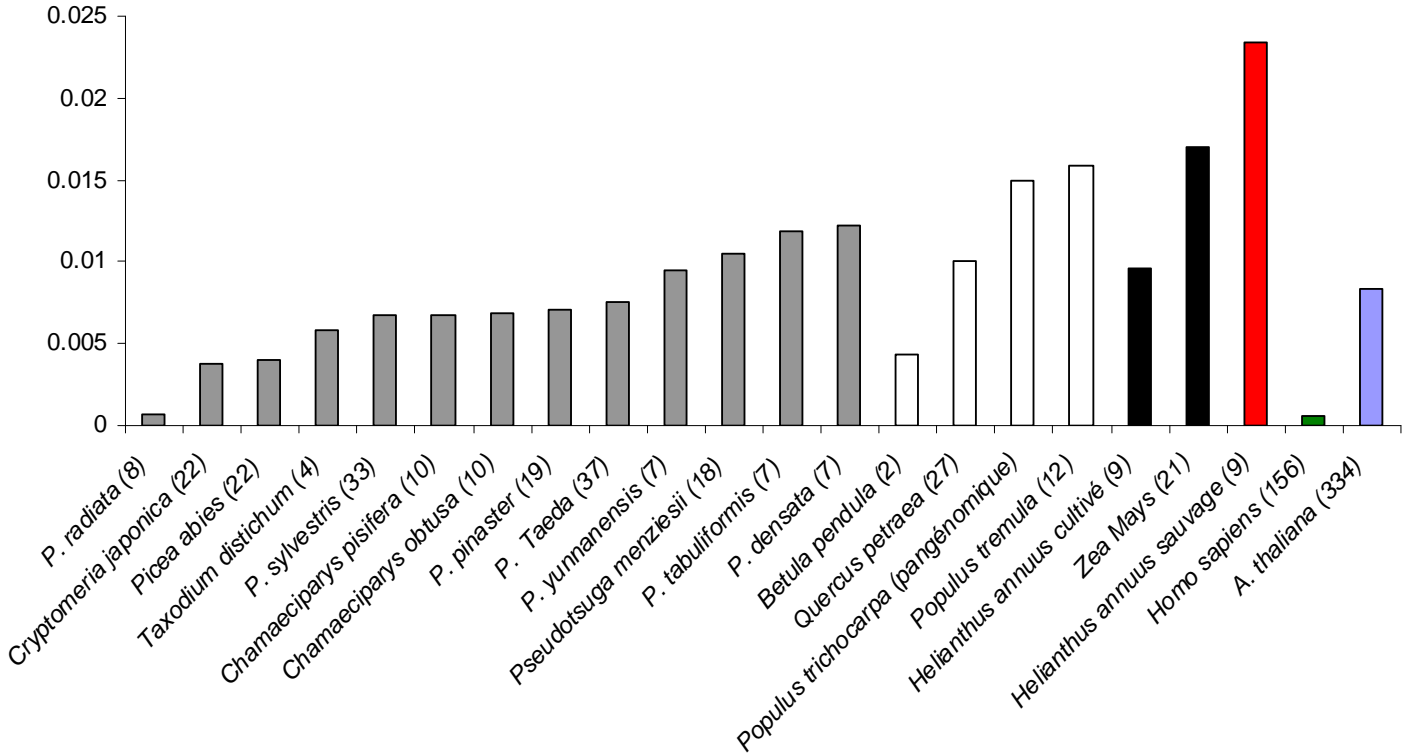


Tableau 14. Revue de la littérature pour des estimations de diversité nucléotidique silencieuse, $\theta_{\pi S}$, et de la diversité haplotypique, H , chez plusieurs espèces de conifères, d'arbres angiospermes, de plantes annuelles, comparativement à quelques valeurs d'études chez les humains et *Arabidopsis thaliana*. Ces estimations ont été réalisées pour la plupart à partir de séquences de gènes candidats dont les catégories fonctionnelles sont indiquées.

	<i>Espèce</i>	<i>Catégorie fonctionnelle des loci</i>	<i>Nombre de loci</i>	<i>Moyenne du $\theta_{\pi S}$</i>	<i>H</i>	<i>Références</i>
Conifères	<i>P. pinaster</i>	Formation du bois	8	0.0062	0.850	Garnier-Géré <i>et al.</i> 2005 (en préparation)
		Formation du bois	8	0.00136	0.425	Pot <i>et al.</i> 2005
		Résistance à la sécheresse	11	0.00852	0.747	Eveno <i>et al.</i> 2008
	<i>P. radiata</i>	Formation du bois	8	~0.0035	0.376	Pot <i>et al.</i> 2005
		Formation du bois	10	NA	0.603	comm. pers. Tibbits
	<i>P. taeda</i>	Formation du bois	19	0.00658	NA	Brown <i>et al.</i> 2004
		Résistance à la sécheresse	18	0.00853	0.680	González-Martínez <i>et al.</i> 2006a
	<i>P. sylvestris</i>	Débourrement	2	0.002400	NA	García-Gil <i>et al.</i> 2003
		Formation du bois	1	0.004900	0.950	Dvornyk <i>et al.</i> 2002
		Dehydrines et gènes candidats liés au froid	14	0.007700	0.690	Wachowiak <i>et al.</i> 2008 (soumis)
	<i>P. densata</i>	Nombreuses fonctions	16	0.006485	NA	Pyhajarvi <i>et al.</i> 2007
		Nombreuses fonctions	7	0.0122	NA	Ma <i>et al.</i> 2006
	<i>P. tabuliformis</i>	Nombreuses fonctions	7	0.0119	NA	Ma <i>et al.</i> 2007
	<i>P. yunnanensis</i>	Nombreuses fonctions	7	0.0095	NA	Ma <i>et al.</i> 2008
	<i>Pseudotsuga menziesii</i>	Résistance au froid et formation du bois	18	0.01055	0.827	Krutovsky et Neale 2005
	<i>Cryptomeria japonica</i>	Nombreuses fonctions	7	0.00383	NA	Kado <i>et al.</i> 2003
		Nombreuses fonctions	5	0.00350	NA	Kado <i>et al.</i> 2006
		Nombreuses fonctions	10	0.00383	0.486	Kado <i>et al.</i> 2008
	<i>Chamaeciparys obtusa</i>	Nombreuses fonctions	10	0.00689	0.453	Kado <i>et al.</i> 2008
	<i>Chamaeciparys pisifera</i>	Nombreuses fonctions	10	0.00675	0.792	Kado <i>et al.</i> 2008
	<i>Taxodium distichum</i>	Nombreuses fonctions	4	0.0058	NA	Kado <i>et al.</i> 2006
	<i>Picea abies</i>	Photopériode ou vernalization + loci de contrôle	22	0.00399	0.545	Heuertz <i>et al.</i> 2006
		Nombreuses fonctions	3	NA	NA	Bouillé et Bousquet 2005
<i>Picea glauca</i>	Nombreuses fonctions	3	NA	NA	Bouillé et Bousquet 2005	
<i>Picea mariana</i>	Nombreuses fonctions	3	NA	NA	Bouillé et Bousquet 2005	
<i>Abies kawakamii</i>	cytosolic glyceraldehyde 3-phosphate dehydrogenase	1	NA	0.774	Shih <i>et al.</i> 2007	
Angiospermes	<i>Quercus peduncula</i>	Phénologie du bourgeon et floraison	27	NA	NA	comm. pers. Garnier-Géré du projet Digenfor
	<i>Quercus petraea</i>	Phénologie du bourgeon et floraison	27	0.01	NA	comm. pers. Garnier-Géré du projet Digenfor
	<i>Populus tremula</i>	Stress abiotique et synthèse d'hormones	5	0.01600	NA	Ingvarsson 2005a
		Proteases inhibiteurs induits lors de blessures	6	0.01705	NA	Ingvarsson 2005b
	<i>Populus trichocarpa</i>	Arrêt de croissance et débourrement	1	0.00854	NA	Ingvarsson <i>et al.</i> 2006
		Pan-génomique		0.015	NA	Tuskan <i>et al.</i> 2006
<i>Betula pendula</i>	Morphogénèse	2	0.0043	NA	Jarvinen <i>et al.</i> 2003	
Plantes annuelles	<i>Zea Mays</i>	Nombreuses fonctions	21	0.017	NA	Tenaillon <i>et al.</i> 2001
	<i>Helianthus annuus</i>	Nombreuses fonctions	9	0.0234	NA	Liu et Burke 2006
	<i>Helianthus annuus</i>	Nombreuses fonctions	9	0.0096	NA	Liu et Burke 2007
Humains	<i>Homo sapiens</i>	Fragments de séquence non codante	50	0.00088	NA	Yu <i>et al.</i> 2002
	<i>Homo sapiens</i>	Maladies cardio-vasculaires, endocrinologie, neuropsychiatrie	106	0.00041	NA	Cargill <i>et al.</i> 1999
A. thaliana	<i>A. thaliana</i>	Loci aléatoires	334	0.0083	NA	Schmid <i>et al.</i> 2005

aller au-delà de plusieurs dizaines de kilomètres (Bacles *et al.* 2005, Robledo-Arnuncio et Gil 2004, Kaufman *et al.* 1998), et induisant des flux de gènes importants entre populations. Ces observations seraient cohérentes aussi avec des estimations de grande taille efficace chez ces espèces (par exemple, taille efficace de *Pinus taeda* estimée à 560 000 individus, Brown *et al.* 2004, revue à la baisse à moins de 100 000 par Willyard *et al.* 2007, mais toujours importante). Ce système de reproduction allogame serait associé chez ces espèces à des mécanismes tels que l'auto-incompatibilité et la dépression de consanguinité (Petit et Hampe 2006). Ce dernier processus peut être particulièrement fort chez les arbres forestiers (par exemple, Koelewijn *et al.* 1999 chez *Pinus sylvestris*) et a été interprété comme la conséquence d'une proportion non négligeable d'allèles délétères récessifs qui seraient dus à de forts taux de mutation (Lande *et al.* 1994).

Dans leurs différentes revues, Hamrick *et al.* (1989, 1991, 1992) montrent ainsi que la diversité génétique intra-population estimée à partir d'alozymes est en moyenne plus élevée chez les plantes allogames *versus* autogames (H_S de 0.25 *versus* 0.18) et chez les plantes pérennes *versus* annuelles (H_S de 0.22 *versus* 0.25), tandis que la différenciation génétique entre populations est en moyenne plus faible chez les plantes pérennes *versus* annuelles (G_{ST} de 0.08 *versus* 0.25), et chez les plantes allogames *versus* autogames (G_{ST} de 0.10 *versus* 0.51). Chez *Pinus pinaster* dont la distribution est plus fragmentée que celle d'autres conifères (Petit *et al.* 1995, Bucci *et al.* 2007), les valeurs de différenciation (G_{ST}) ont été estimées entre 0.12 et 0.17 pour des isozymes (Petit *et al.* 1995, Barba *et al.* 2001), autour de 0.10 pour des marqueurs d'ADN chloroplastique, mitochondriale et nucléaire (Petit *et al.* 2005), et autour de 0.12 pour des SSRs (Eveno *et al.* 2008, après transformation du F_{ST} en G_{ST} selon Cockerham et Weir 1987). De plus, la diversité génétique intra-population au niveau de marqueurs SSRs a été estimée autour de 0.6 (Chagné *et al.* 2004, Mariette *et al.* 2001, Derory *et al.* 2002, informations supplémentaires de Eveno *et al.* 2008).

On peut donc se poser la question de savoir pourquoi, malgré les valeurs de diversité élevées pour ces marqueurs *a priori* neutres ou anonymes, on observe une diversité nucléotidique apparemment plus faible? Une façon de réconcilier ces résultats serait de comparer plutôt la diversité haplotypique des gènes candidats avec le paramètre correspondant de diversité allélique estimée sur les alozymes ou les RFLPs. On observe une diversité haplotypique en moyenne très élevée pour les arbres forestiers (H moyen de 0.657, Tableau 14), tout à fait comparable aux niveaux de diversité élevés publiés par Hamrick pour les marqueurs alozymes chez les arbres et autres espèces (Hamrick *et al.* 1989, 1992) ou les

marqueurs RFLPs. Ce nombre d'haplotypes élevé, malgré un niveau de diversité relativement faible, pourrait être dû en partie au faible DL observé en moyenne intra-locus chez ces espèces (malgré une grande variation entre gènes), puisqu'un nombre limité de sites polymorphes statistiquement indépendants peut conduire à de nombreuses combinaisons alléliques (Dvornyk *et al.* 2002, Heuertz *et al.* 2006).

C.3. Facteurs susceptibles d'affecter les niveaux de diversité nucléotidique

La variation des niveaux de diversité nucléotidique θ égale à $4N_e\mu$ entre espèces peut donc s'expliquer en théorie par deux facteurs principaux, la taille efficace de la population, N_e (qui est affectée par l'histoire démographique des populations), et le taux de mutation, μ . Il est important aussi de considérer quel pourrait être l'effet de différents types de sélection sur l'estimation d'un niveau moyen de diversité nucléotidique, et si l'échantillon de gènes étudiés peut être considéré comme représentatif de ce que l'on observerait sur le génome entier.

C.3.1. Le taux de mutation

La première hypothèse pouvant expliquer un niveau de diversité nucléotidique relativement faible, serait un taux de mutation plus faible. La théorie de l'évolution neutre (Kimura 1983) prédit que le taux de mutation (probabilité d'apparition d'une mutation par génération) est égal au taux de substitution (temps qu'il faut pour qu'un allèle soit substitué par un autre) dans des populations évoluant sous un modèle neutre de Wright-Fisher. Ce taux de substitution nucléotidique peut être estimé en théorie par le niveau de divergence entre espèces (K_S) divisé par 2 fois le temps depuis leur divergence (T) (Li et Graur 1991). Si le taux de divergence peut être relativement facile à estimer sur la base de données de séquences entre plusieurs espèces, les estimations de temps de divergence basées sur des données fossiles plus rares peuvent être beaucoup plus approximatives. Cependant, les quelques estimations disponibles de taux de substitution montreraient des niveaux plus faibles chez les arbres forestiers que chez les espèces annuelles, qui pourraient être expliquées notamment par un temps de génération plus long et un taux métabolique plus lent (Petit et Hampe 2006). Chez les conifères, ces taux de substitution par site et par année ont été par exemple estimés à 1.17×10^{-10} chez *Pinus taeda* (Brown *et al.* 2004), entre 0.05 et 0.15×10^{-9} chez *Pinus sylvestris* (Dvornyk *et al.* 2002), entre 2.23 et 3.32×10^{-10} chez *Picea glauca*, *abies* et *mariana* (Bouillé et Bousquet 2005). Dans notre cas, avec un taux de substitution K_S en moyenne de 0.042 sur 11 gènes, et un temps de divergence estimé à 45 millions d'années entre les sous-genres *Pinus* et *Strobus* (Willyard *et al.* 2007), le taux de substitution par site et

par année est estimé à 4.72×10^{-10} chez *Pinus pinaster*, donc du même ordre de grandeur que chez les autres conifères. Ces estimations sont effectivement plus faibles que celles chez les plantes annuelles, tels que *Arabidopsis* (1.5×10^{-8} Koch *et al.* 2000). Cependant, Heuertz *et al.* (2006) suggèrent que ces valeurs sont généralement sous estimées à cause d'une sur-estimation du temps de divergence entre espèces. D'autre part, quand on ramène ce taux de substitution par génération pour les espèces d'arbres forestiers (Petit et Hampe 2006, Savolainen et Pyhäjärvi 2007), il est du même ordre de grandeur que celui de plantes annuelles. Par conséquent, de faibles taux de mutation ne seraient pas la raison de niveaux de diversité nucléotidique faibles chez les espèces d'arbres forestiers, et notamment le pin maritime.

C.3.2. Démographie des populations

L'autre facteur ayant pu affecter la diversité nucléotidique serait donc la taille efficace reliée directement à l'histoire démographique des populations. Certaines populations peuvent avoir subi au cours de leur histoire démographique, des goulots d'étranglement (réduction de la taille efficace) entraînant une baisse de diversité. Ces populations n'auraient pas encore retrouvé leur état d'équilibre entre mutation et dérive. Cet argument fut avancé par Heuertz *et al.* (2006) et Pyhäjärvi *et al.* (2007) pour expliquer les faibles niveaux de diversité nucléotidique observés chez *Picea abies* et *Pinus sylvestris*. Chez *Pinus pinaster*, des études phylogénétiques semblent montrer l'existence d'au moins trois populations refuges durant la dernière glaciation (Burban et Petit 2003). Parmi ces populations, celles du Sud-Est et du centre de l'Espagne constitueraient les plus larges centres de ré-expansion des espèces après la période glaciaire (Carrion *et al.* 2000, Carrion 2002, Carrion *et al.* 2003). Ces populations refuges seraient le point de départ de la recolonisation à la fois vers l'Ouest et vers le Nord (Salvador *et al.* 2000) où les populations de la région Atlantique se seraient formées. Selon González-Martínez *et al.* (2002), la dispersion de cette espèce le long de cette voie de migration fut rapide, puisque une faible différenciation adaptative est observée le long de cette voie, malgré des conditions écologiques contrastées. Les niveaux de diversité aussi élevés au sein de la métapopulation atlantique que sur l'ensemble des populations (Tableau 12) semblent montrer que cette population se serait établie sans fort effet de fondation, et avec une immigration récurrente du Sud et de l'Est de l'Europe (comme suggéré par les données d'ADN et de SSRs chloroplastiques, Vendramin *et al.* 1998, Bucci *et al.* 2007). De plus, le pin maritime serait apparu depuis 8000 ans sur la côte atlantique française (Paquerau, 1964) et aurait eu le temps d'évoluer pendant 500 à 700 générations (si on considère la maturité de

reproduction entre 11 et 16 ans). Donc l'hypothèse d'une faible diversité nucléotidique expliquée par un goulot d'étranglement reste peu probable pour cette espèce, du moins au sein des populations de la côte atlantique française.

C.3.3. Effets possibles de la sélection

Parmi les processus de sélection qui pourraient affecter le génome des plantes, l'importance de la sélection purifiante maintenant une faible diversité nucléotidique non-synonyme par rapport à la diversité synonyme était une hypothèse de la théorie neutraliste, et a été observée chez de nombreuses espèces (Ford 2002). On observe également un nombre croissant d'études ayant mis en évidence de la sélection positive sur des gènes candidats (Wright et Gaut 2005, Ford 2002). Ce type de sélection pourrait être d'autant plus importante sur des gènes impliqués dans des fonctions biologiques directement affectées par des pressions environnementales (sécheresse, froid, etc..Howe et Brunner 2005). De plus, par effets d'auto-stop génique, la diversité nucléotidique silencieuse pourrait être soit diminuée dans le cas d'une sélection directionnelle sur des loci proches (Maynard-Smith et Haigh 1974), soit augmentée dans le cas de sélection balancée (Delden *et al.* 1978). Cependant, quelques gènes candidats ne sont sans doute pas représentatifs de l'ensemble du génome, et les conséquences réelles des effets d'auto-stop génique sur la diversité au niveau de l'ensemble du génome au sein de populations naturelles de plantes est encore mal évaluée, malgré quelques prédictions théoriques (Maynard-Smith et Haigh 1974, Gillespie 2000, Barton 2000). Chez *Picea abies*, l'hypothèse d'effets importants au sein des gènes candidats étudiées est peu probable (Heuertz *et al.* 2006), car l'étendu du *DL* est limité (r^2 diminuant jusqu'à des valeurs inférieures à 0.20 sur environ 100pb), et il n'y a pas d'excès de mutations dérivés conduisant à des valeurs négatives de *H* de Fay et Wu, qui serait le signal d'effets d'auto-stop génique (Fay et Wu 2000). Chez le pin maritime, même si les *DL* entre paires de sites intra-gène sont faibles dans l'ensemble, ils varient beaucoup le long des gènes et entre gènes. Par exemple tous les sites chez *Glucan* et une grande majorité pour *CCoAOMT* sont en *DL* complet au sein de la métapopulation Atlantique sur plus de 1kb. D'autre part, des écarts à la neutralité sont significatifs et importants pour plusieurs statistiques de tests intra-population chez 4 des 5 gènes étudiés (voir Paragraphe C.4.).

La théorie neutraliste prédit aussi que le niveau de polymorphisme d'un gène au sein d'une espèce est proportionnel au niveau de divergence de ce gène entre espèces (Hudson *et al.* 1987). Si l'on fait une régression de la diversité nucléotidique silencieuse ($\theta_{\pi S}$) sur la divergence silencieuse (K_s), on constate que certains gènes (*CCoAOMT* et *lp3-3*) s'écartent

davantage d'une relation linéaire que les autres gènes, leurs diversités nucléotidiques étant faibles par rapport à leurs taux de divergence (Figure 34). Ces observations, combinées avec celles d'écarts à la neutralité significatifs observés pour ceux-ci et d'autres gènes sur l'ensemble des analyses inter- et intra-population seraient consistantes avec des effets de la sélection sur les patrons de diversité moléculaire observés.

C.4. Quels scénarios évolutifs au sein des populations naturelles de pin maritime ?

Nous avons vu que l'on pouvait faire l'hypothèse d'un faible impact de la démographie (d'une absence possible de goulot d'étranglement) au moins dans certaines populations, telles que la métapopulation atlantique et que la sélection pourrait avoir des effets non négligeables sur certains gènes. Si l'on accepte l'hypothèse de sélection, quelles interprétations peut-on alors faire concernant les scénarios évolutifs des différents gènes, sur la base des résultats obtenus à la fois à l'échelle des 10 populations contrastées et au sein de la métapopulation atlantique ?

C.4.1. Signatures de sélection diversifiante entre populations, et directionnelle au sein de la métapopulation atlantique

Les tests d'écart à la neutralité basés sur la différenciation génétique révèlent que le gène *PR-AGP4* ainsi que quelques SNPs au sein de ce gène présentent une différenciation supérieure à l'attendue neutre. Sachant que ces tests sont robustes aux hypothèses démographiques (et notamment la structure), ce profil correspondrait à de la sélection diversifiante entre populations, c'est-à-dire que dans les populations les plus divergentes, on aurait une sélection directionnelle favorisant différents haplotypes ou sites. Les populations qui contribuent le plus à ce niveau élevé de différenciation, sont les populations Tamrabta (Maroc), Tabarka (Tunisie) et Oria (Sud de l'Espagne) (voir Figure 35 et Annexe 12). Elles correspondent à des populations aux conditions climatiques les plus arides (précipitations les plus faibles pour Oria, et températures les plus élevées pour Tamrabta et Tabarka, voir Tableau 4 Chapitre II) et relativement isolées (voir Figure 7 Chapitre II). On peut donc supposer que pour *PR-AGP4*, se soient exercées au sein de ces populations des pressions de sélection suffisamment fortes et que les flux de gènes entre ces populations et les autres aient été suffisamment limités pour que l'adaptation locale ait pu être efficace (voir les résultats de la méthode de Beaumont et Balding sur les effets spécifiques aux populations, Eveno *et al.*

Figure 34. Régression de la diversité nucléotidique silencieuse ($\theta_{\pi S}$) en fonction du taux de divergence silencieux (K_S) (estimé à partir d'une séquence de *Pinus taeda*), sur l'ensemble des 10 populations (cercles noirs) et au sein de la métapopulation atlantique (cercles blancs). Seul le gène *GRP3* n'est pas inclus dans l'analyse car aucune séquence orthologue de *Pinus taeda* n'était disponible.

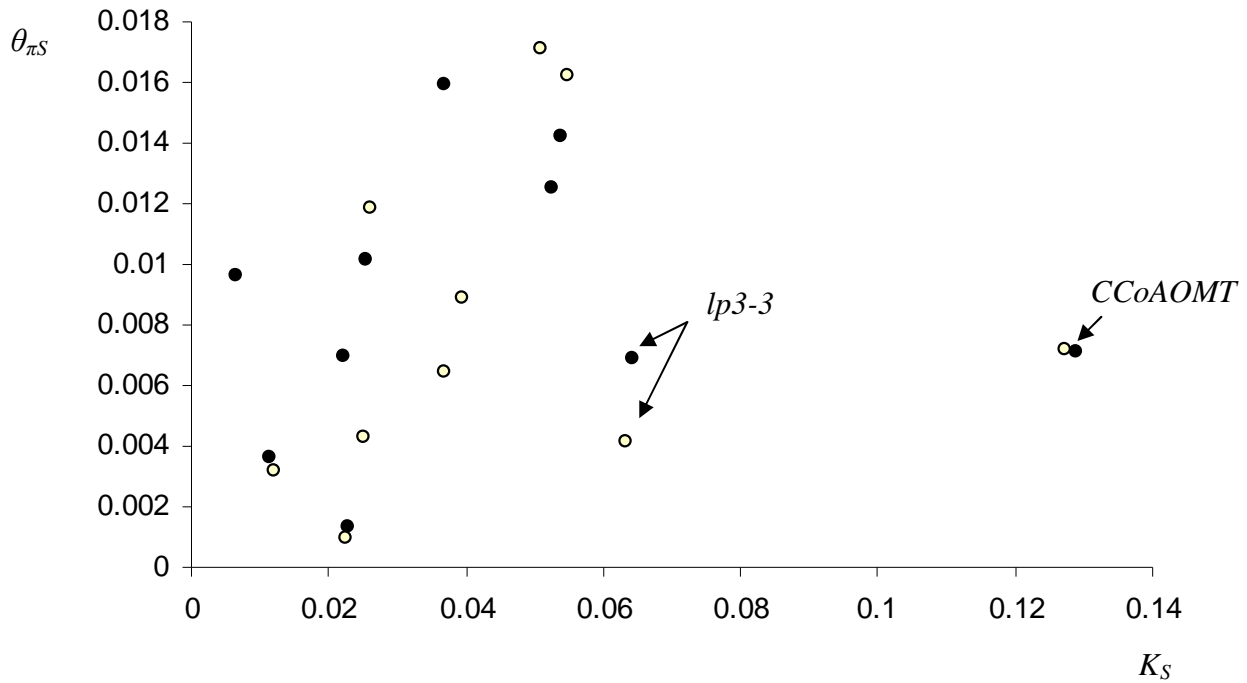


Figure 35. Matrices de F_{ST} par paires de populations pour les gènes (sans données manquantes) *PR-AGP4* (A), *erd3* (B), *GRP3* (C), et les SNPs (avec données manquantes) *GRP3 S164* (D) et *erd3 S42* (E). Les valeurs supérieures à 0.2 sont surlignées en bleu, et les valeurs supérieures à 0.5 sont surlignées en jaune (voir aussi Annexe 12).

A)

<i>PR-AGP4</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadeuc	Tabarka
Pineta										
Pinia	-0.144									
Arenas	0.301	0.197								
Coca	0.003	-0.017	0.228							
San Cipriano	0.377	0.231	-0.049	0.264						
Oria	-0.012	0.040	0.412	0.108	0.478					
Tamrabta	0.040	0.039	0.450	0.039	0.544	0.086				
Mimizan	0.218	0.125	-0.024	0.163	-0.019	0.365	0.359			
Pleucadeuc	0.172	0.077	-0.052	0.119	-0.024	0.324	0.323	-0.080		
Tabarka	0.309	0.195	-0.074	0.236	-0.059	0.409	0.464	-0.022	-0.056	
Mean	0.140	0.082	0.154	0.127	0.194	0.246	0.260	0.121	0.089	0.156

B)

<i>erd3</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadeuc	Tabarka
Pineta										
Pinia	0.000									
Arenas	0.000	0.000								
Coca	NA	NA	NA							
San Cipriano	0.000	0.000	0.000	NA						
Oria	0.262	0.228	0.336	NA	0.262					
Tamrabta	0.567	0.534	0.640	NA	0.567	0.038				
Mimizan	0.040	0.000	0.126	NA	0.040	0.071	0.352			
Pleucadeuc	0.040	0.000	0.126	NA	0.040	0.071	0.352	-0.111		
Tabarka	0.000	0.000	0.000	NA	0.000	0.228	0.534	0.000	0.000	
Mean	0.239	0.220	0.279	NA	0.239	0.182	0.413	0.113	0.124	0.196

C)

<i>GRP3</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadeuc	Tabarka
Pineta										
Pinia	0.12									
Arenas	0.103	0.032								
Coca	0.366	0.2538	0.0779							
San Cipriano	0.099	-0.056	-0.016	0.1621						
Oria	0.07	0.0334	0.0875	0.2513	0.0629					
Tamrabta	0.029	0.0651	0.1428	0.3392	0.1107	-0.11				
Mimizan	0.138	0.115	-0.037	0.0485	0.0126	0.0954	0.1711			
Pleucadeuc	0.49	0.3091	0.1366	-0.063	0.2351	0.3113	0.4154	0.091		
Tabarka	0.062	0.1159	0.2222	0.4135	0.1968	-0.041	-0.085	0.2642	0.515	
Mean	0.164	0.110	0.083	0.205	0.090	0.084	0.120	0.100	0.271	0.185

D)

<i>GRP3 S164</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadeuc	Tabarka
Pineta										
Pinia	0.066									
Arenas	-0.092	0.152								
Coca	0.302	0.633	0.175							
San Cipriano	0.120	0.458	0.080	0.000						
Oria	0.066	-0.125	0.152	0.633	0.458					
Tamrabta	0.203	-0.096	0.285	0.737	0.588	-0.096				
Mimizan	0.312	0.642	0.180	0.000	0.000	0.642	0.745			
Pleucadeuc	0.198	0.535	0.122	0.000	0.000	0.535	0.655	0.000		
Tabarka	0.171	-0.097	0.252	0.683	0.536	-0.097	-0.110	0.691	0.600	
Mean	0.149	0.241	0.145	0.395	0.280	0.241	0.323	0.357	0.331	0.292

E)

<i>erd3 S42</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadeuc	Tabarka
Pineta										
Pinia	0.000									
Arenas	0.000	0.000								
Coca	NA	NA	NA							
San Cipriano	0.000	0.000	0.000	NA						
Oria	0.233	0.233	0.308	NA	0.233					
Tamrabta	1.000	1.000	1.000	NA	1.000	0.525				
Mimizan	0.000	0.000	0.072	NA	0.000	-0.052	0.800			
Pleucadeuc	0.040	0.040	0.126	NA	0.040	-0.111	0.774	-0.220		
Tabarka	0.000	0.000	0.000	NA	0.000	0.200	1.000	-0.035	0.000	
Mean	0.253	0.253	0.278	NA	0.253	0.190	0.789	0.130	0.132	0.241

2008). Par contre, lorsque l'on considère les populations Mimizan et Pleucadeuc qui ont des conditions de milieux intermédiaires, la différenciation entre ces populations et les autres est moins importante pour ce gène (Figure 35 et Annexe 12).

Au sein de la métapopulation atlantique, les tests d'écart à la neutralité appliqués sont tous significatifs. Le H_{FW} significatif et négatif indique par exemple un excès d'allèles dérivés (qui sont répartis sur plusieurs haplotypes de la population), et peut s'interpréter comme une conséquence d'effets d'entraînement des mutations dérivées sur l'haplotype correspondant, suite à des événements de recombinaison pendant la phase de sélection directionnelle. Cette interprétation est également cohérente avec un test D_T significatif et négatif, notamment sur deux régions (en amont et en aval), dû à un excès d'allèles rares. Cependant, cet excès de mutations rares n'entraîne pas d'excès d'haplotypes, car les mutations sont répartis sur quelques haplotypes en amont, mais proviennent majoritairement d'un haplotype en aval (Figure 29A). Donc la structure haplotypique est forte surtout en aval du gène ce qui induit une diversité haplotypique H_d significativement négative et un F_S de Fu significativement positif sur l'ensemble du gène. Donc l'apparent paradoxe entre un D_T négatif et un F_S de Fu positif traduirait un patron complexe, avec peut-être une succession d'évènements de sélection avec effets d'entraînement, ou bien la conséquence, en interaction avec la recombinaison, d'une signature plus forte en aval du gène, avec un haplotype portant des mutations non synonymes pouvant être la cible de la sélection.

Dans les autres populations, on observe des signatures comparables pour la population tunisienne Tabarka (D_T , F_S de Fu et H_{FW} hautement significatifs, résultats non présentés), ainsi qu'une structure haplotypique forte pour Oria et Tamrabta (F_S de Fu significatifs mais signatures moins fortes que pour Tabarka). Cependant, les 3 haplotypes rares sont tous différents entre eux et de celui de la population atlantique (ce qui explique la divergence entre toutes ces populations). Donc pour ce gène, on peut observer des signatures de sélection significatives à la fois au niveau inter-population et au sein de plusieurs populations (Tableau 15). Ceci a été vérifié avec un échantillonnage plus important dans d'autres populations espagnoles dans le cadre du projet Digenfor (<http://www.pierroton.inra.fr/biogeco/genetique/projets/europe/digenfor/index.html>).

Le gène *GRP3* présente 3 SNPs avec une différenciation supérieure à 0.27, dont 2 non-synonymes et un synonyme « outlier » positif. Au niveau haplotypique, le gène est outlier à 10% avec la méthode de Beaumont et Balding, et nous avons vu que la valeur du N_{ST} de ce

Tableau 15. Tableau récapitulatif des interprétations proposées pour les écarts à la neutralité en termes de signatures de sélection à l'échelle des 10 populations et au sein de la métapopulation atlantique.

Code du gène	Scénarios proposés pour chaque gène	
	Parmi les 10 populations contrastées	Au sein de la métapopulation Aquitaine
<i>PR-AGP4</i>	Sélection diversifiante – adaptation locale malgré le flux de gène notamment dans la population Tamrabta (Maroc) et Tabarka (Tunisie) Au niveau haplotypique + au niveau de 4 SNPs	Succession de plusieurs évènements de sélection directionnelle le long du gène avec entraînement de mutations dérivées suite aux recombinaisons
<i>CCoAOMT</i>	2 SNPs non synonymes avec une forte signature de sélection homogénéisante .	Sélection balancée relativement récente avec très peu de recombinaisons
<i>Glucan</i>		Sélection balancée forte avec maintien d'un allèle en fréquence faible avec très peu de recombinaison: sélection fréquence-dépendante
<i>GRP3</i>	3 SNPs avec un fort F_{ST} (dont un « outlier » positif), et une forte différenciation pour le gène (F_{ST} , N_{ST}) indiquant une sélection diversifiante entre populations	Sélection directionnelle ayant entraîné notamment une mutation dérivée en forte fréquence après recombinaison
<i>dhn-1</i>	sélection homogénéisante au niveau haplotypique + au niveau d'1 SNP non synonyme	Sélection balancée , accompagnée de nombreux évènements de recombinaison
<i>dhn-2</i>	sélection homogénéisante au niveau haplotypique + au niveau d'1 SNP synonyme	Sélection balancée au sein de la métapopulation Aquitaine ?
<i>erd3</i>	Sélection diversifiante due à une adaptation locale au sein de Tamrabta au niveau haplotypique + au niveau d'1 SNP non synonyme	Sélection directionnelle au sein de la métapopulation Aquitaine ?
<i>lp3-1</i>	sélection homogénéisante au niveau haplotypique + au niveau d'1 SNP en 3'UTR	Sélection balancée au sein de la métapopulation Aquitaine ?

gène était quasiment 2 fois supérieure au F_{ST} , ce qui suggérerait une relative inefficacité des méthodes basées uniquement sur le F_{ST} pour rendre compte de cette observation (voir aussi la discussion de Eveno *et al.* 2008). Ces résultats suggèrent une sélection diversifiante entre les 10 populations contrastées. Cette différenciation très élevée existe entre plusieurs paires de populations appartenant à des régions géographiques différentes (côte atlantique, Corse, Tunisie, Maroc et Espagne, Figure 35 et Annexe 12). Même s'il ne sont pas tous détectés comme « outliers », il est intéressant de noter sur ce gène la présence de plusieurs sites non-synonymes (3 sur 576 pb) qui ne sont pas statistiquement liés et forment au total 5 haplotypes différents dans la métapopulation atlantique (Figure 30D). Dans la mesure où ces sites ne sont pas rares sur l'ensemble des populations, on peut supposer qu'il y ait eu une sélection positive sur certains de ces SNPs au moins dans quelques populations. En effet, dans le cas contraire, ces mutations qui changent la fonction des protéines et qui peuvent donc être délétères, auraient été éliminées par de la sélection purifiante. Dans la métapopulation atlantique, cette interprétation serait cohérente avec le H_{FW} négatif et significatif. Cependant, comme le nombre de sites polymorphes au sein de ce gène est faible, il faut rester prudent quant à l'interprétation des tests intra-population, le signal de différenciation entre les populations restant le plus informatif. Pour confirmer une hypothèse de sélection positive intra-population, il pourrait être judicieux de séquencer une plus grande longueur de fragment pour identifier plus de mutations, ce qui augmenterait la puissance de tests comme le D_T , sensible au nombre de mutations qui ségrégent (Pluzhnikov et Donnelly 1996).

Un autre gène (*erd3*) relativement polymorphe (θ_π de 0.00151, Eveno *et al.* 2008) présente néanmoins une signature de sélection diversifiante au niveau haplotypique et au niveau d'un SNP non-synonyme (F_{ST} de 0.59). Cette signature est principalement due à la différenciation très élevée de la population Tamrabta par rapport aux autres populations, Tamrabta étant fixée pour un des allèles de ce SNP, alors que les 2 populations Corse (Pinia, Pineta), les 2 populations espagnoles (Arenas, San-Cipriano) et la population Tabarka sont fixées pour l'autre allèle (Annexe 11). Par contre, ces populations étant toutes monomorphes, on ne pourrait pas détecter de signatures intra-population. De même, on ne détecte pas de signature de sélection au sein de la métapopulation atlantique, même si elle est polymorphe pour ce SNP, aucun des tests de distribution de fréquences allélique n'étant significatif.

C.4.2. Signatures de sélection homogénéisante sur les 10 populations, et sélections balancées au sein de la métapopulation atlantique?

Quatre autres gènes (les 2 déhydrines *dhn1* et *dhn2*, et *lp3-1* aux niveaux haplotypique et SNPs, et *CCoAOMT* pour quelques SNPs seulement) montrent une différenciation entre les 10 populations significativement inférieure à celle des attendus neutres selon différentes méthodes. Cette faible différenciation est due au partage d'allèles sur l'ensemble de l'aire de distribution du pin maritime pour ces différents gènes. Ces signatures ont été interprétées comme la conséquence possible d'effets de sélection homogénéisante entre les différentes populations et/ou de sélection balancée intra-population (Eveno *et al.* 2008). Parmi les processus pouvant être invoqués, la sélection balancée au niveau moléculaire pourrait résulter d'une forte hétérogénéité environnementale au cours des années, avec des périodes de sécheresse récurrentes. Une telle hétérogénéité temporelle pourrait favoriser alternativement un haplotype puis l'autre entre différentes générations. Ce type de variabilité climatique a été par exemple proposé par Jump *et al.* (2006) pour expliquer les patrons de polymorphismes balancés observés à un locus potentiellement lié à l'effet de la température chez *Fagus Sylvatica*. On peut aussi imaginer que cette hétérogénéité temporelle pourrait être combinée à une hétérogénéité spatiale, qui pourrait augmenter la diversité allélique à certains gènes impliqués dans la variation phénotypique au sein d'une population (Rueffler *et al.* 2006). Il est par ailleurs peu probable que l'on soit dans une situation où un allèle serait maintenu par sélection, tandis que d'autres seraient apportés par migration (voir les simulations de Hiorns et Harrison 1977 en populations humaines), car dans ce cas, on s'attendrait à une variation beaucoup plus importante des allèles non-sélectionnés entre populations proches *versus* distantes.

Nous développons ci-dessous l'interprétation des patrons observés dans la métapopulation plus largement échantillonnée pour *CCoAOMT* et *dhn1*. Pour les gènes *dhn2* et *lp3-1* moins bien échantillonnés dans cette métapopulation, les statistiques D_T étaient positifs mais non significatives (environ 1,5 pour les 2 gènes), allant dans le sens d'un excès de mutations en fréquences intermédiaires et d'une signature de sélection balancée.

CCoAOMT

Pour le gène *CCoAOMT*, les 2 sites non différenciés sont un SNP et un Indel non-synonymes, qui sont en fort *DL* ($r^2 > 0.8$), et se retrouvent en fréquence intermédiaire (entre

11% et 22%) dans les populations corses (Pinia et Pineta) et les populations espagnoles (Coca, San Cipriano et Oria). Pour les autres sites de ce gène, la différenciation est plus élevée et se situe dans la distribution de l'attendue neutre (voir Figure 2 de l'article en Annexe 1). Ceci explique que le gène ne soit pas un « outlier » négatif au niveau haplotypique.

Au sein de la métapopulation atlantique, on observe également pour ce gène une forte structure haplotypique (H_d négatif et significatif et F_S de Fu positif et significatif) avec deux groupes d'haplotypes majoritaires très différents. Le D_T est positif bien que significatif à 10% seulement, ce qui va aussi dans le sens d'un excès d'allèles en fréquence intermédiaire. Sur la région du gène qui a été séquencée, le déséquilibre de liaison est relativement élevé (Z_{ns} significatif et positif). Cependant, des résultats supplémentaires sur une région de ce gène plus en aval indiquent que le déséquilibre de liaison entre sites adjacents diminue, à cause d'une recombinaison plus élevée (Audigeos 2005). Ce profil serait compatible avec une signature forte de sélection balancée, maintenant ces deux groupes d'haplotypes en fréquence intermédiaire, avec un signal qui diminuerait en aval du gène à cause de la recombinaison.

La proposition d'une signature de sélection balancée est la même que celle proposée par Gonzalez-Martinez *et al.* (2006a) sur le gène orthologue chez *Pinus taeda* (échantillon provenant du Sud-Est de la distribution naturelle non fragmentée de cette espèce), qui met en évidence deux haplotypes très divergents en fréquence similaire, induisant un DL plus élevé au sein de ce gène que sur d'autres gènes candidats à la résistance au stress hydrique. Par contre, le profil de diversité au sein des populations espagnoles du projet Digenfor ne ressemble pas à celui observé au sein de notre métapopulation, avec plutôt un excès d'allèles rares dans les populations Coca et Bayubas de Abajo (D_T négatif et significatif), indiquant plutôt une signature de sélection directionnelle.

Dhn1

Le premier résultat marquant pour *dhn1* est l'existence de nombreux sites non synonymes (12 sur 23), avec un ratio entre taux de mutations non-synonymes et taux de mutations synonymes largement supérieur à 1 (1,87), ce qui indiquerait une signature forte de sélection positive. Au sein de la métapopulation atlantique, l'excès d'allèles en fréquence intermédiaire (D_T significatif et positif) peut être interprétée comme une signature de sélection balancée, de façon cohérente avec ce qui est observé au niveau des 10 populations. Indépendamment des arguments historiques cités plus haut, l'hypothèse d'un goulot d'étranglement est difficilement réconciliable avec l'observation des diversités nucléotidique et haplotypique au sein de cette métapopulation qui sont les plus élevées parmi les 5 gènes.

On note également que de nombreux sites non-synonymes ne sont pas en *DL* complet (r^2 en moyenne inférieure à 0.5). Il est difficile de prédire si tous les haplotypes sont maintenus par la sélection avec la création par recombinaison de nouveaux haplotypes équivalents, ou si certains sont plus importants que d'autres sur le plan fonctionnel, avec une variation possible entre populations. Il serait alors intéressant de tester les effets des différents haplotypes sur la variation phénotypique en étude d'associations. Enfin, ces profils sont également comparables avec ceux observés au sein d'autres populations espagnoles au sein du projet Digenfor, où l'on est également en présence d'un excès d'allèles et d'haplotypes en fréquence intermédiaire (D_T et F_S de F_u positifs et significatifs) qui serait cohérent avec de la sélection balancée.

Glucan

Au sein de la métapopulation atlantique, *Glucan* est le gène présentant la plus forte structure haplotypique en comparaison des autres gènes, avec tous ses sites en *DL* complet, formant deux haplotypes dont un majoritaire. Il a également été vérifié sur un plus petit nombre de séquences que la structure haplotypique était conservée en amont du gène (et donc au total sur plus de 3 kb), consistant avec un effet d'entraînement possible sur une longue distance autour du site hypothétique soumis à la sélection. Dans la métapopulation, ce profil pourrait donc s'interpréter comme la conséquence d'une sélection balancée soit récente, soit relativement plus forte que pour les autres gènes, pour expliquer l'absence totale de mutations rares ou d'haplotypes recombinants. N'ayant pas identifié de sites non-synonymes le long du gène, on peut se poser la question de la localisation du ou des sites sélectionnés. De tels sites pourraient peut-être se situer dans la partie 5'UTR au niveau de motifs de reconnaissance de facteurs de transcription. Sur l'ensemble des 10 populations, aucun outlier n'a été détecté pour ce gène, l'haplotype majoritaire de la métapopulation atlantique étant en fréquence plus faible dans d'autres populations (autour de 20% dans les populations San Cipriano et Tamrabta), ce qui conduit à une légère différenciation sur l'ensemble des populations, malgré les forts signaux de sélection balancée dans toutes les populations (F_S de F_u positifs et significatifs dans les 10 populations, résultats non montrés). Cette homogénéité des signaux de sélection intra-population serait difficilement explicable uniquement par leur histoire démographique, puisqu'on a vu qu'elle était très variable entre certaines populations d'après leurs effets spécifiques forts dans l'analyse de Beaumont et Balding (2004) (Eveno *et al.* 2008). Ces mêmes profils se retrouvent dans d'autres populations espagnoles au sein du projet

Digenfor, avec un excès d'allèles et d'haplotypes en fréquence intermédiaire (D_T et F_S de Fu positifs et significatifs).

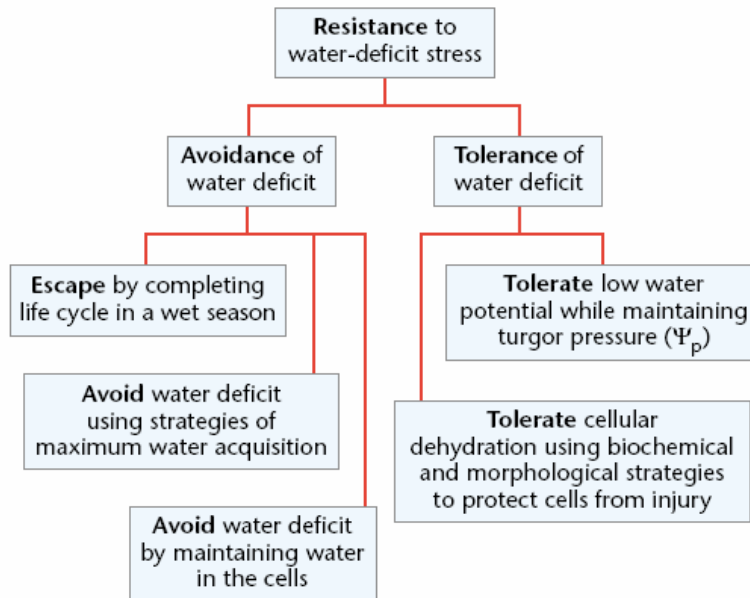
C.4.3. Conclusion

Pour les résultats basés sur les méthodes utilisant la différenciation entre populations, l'observation conjointe d' « outliers » à la fois positifs et négatifs semblerait plus cohérente avec une interprétation en termes d'effets de sélection spécifiques à chaque locus plutôt qu'à l'impact d'évènements démographiques. Au sein de la métapopulation atlantique, on observe aussi des signatures contrastées entre gènes (par exemple, *PR-AGP4* et *Glucan*). De plus, en combinant ces résultats inter- et intra-population, des interprétations cohérentes et complémentaires de signatures sélectives sont possibles, avec d'une part de la sélection diversifiante ou homogénéisante entre populations et d'autre part de la sélection directionnelle ou balancée intra-population (Tableau 15). Au final, le fait que nous n'observions aucun signal homogène sur l'ensemble de ces gènes candidats en intégrant tous les résultats (Tableau 15) serait aussi cohérent avec l'hypothèse que des évènements autres que démographiques aient affecté la diversité et la différenciation nucléotidiques de ces gènes.

Chapitre V

Variabilité génétique de la composition isotopique en carbone et relations avec hauteur et biomasse dans un test de provenances de pin maritime

Figure 36. Mécanismes d'évitement (« avoidance »), d'échappement (« escape ») ou de tolérance (« tolerance ») sous-jacents à la résistance à un stress hydrique (d'après Bray 2001).



A. Introduction

Nous avons vu que le potentiel adaptatif des populations naturelles, qui est leur capacité de réponse face aux pressions de sélection (Houle 1992), dépendait de la plasticité phénotypique des individus qui les composaient, de la variabilité génétique additive pour des caractères adaptatifs, et de la variabilité existant pour les gènes contrôlant l'expression de ces caractères. Nous présentons ici l'étude de la variation de caractères quantitatifs potentiellement liés à l'adaptation, au sein d'un large échantillon de populations naturelles de pin maritime provenant de milieux aux conditions climatiques contrastées. Plusieurs caractères de croissance et de biomasse ont été étudiés, ainsi qu'un caractère écophysologique permettant d'évaluer de façon indirecte l'efficacité d'utilisation en eau de la plante. Ce chapitre a été rédigé sous forme d'article, la partie de résultats préliminaires et problèmes méthodologiques sera intégrée en informations supplémentaires de cet article. Sur les deux tests de provenance/descendances initialement mis en place, l'un à Pierroton (Cestas, France), l'autre en Espagne sur un site aux conditions plus sèches, celui en Espagne n'a pas pu être exploité en raison de la sécheresse de 2003 causant une forte mortalité sur les arbres. N'ayant à notre disposition des données que dans un seul lieu de test, nous n'estimerons pas de plasticité phénotypique au niveau des individus ou des populations. Cependant, la comparaison des performances des populations dans le lieu de test, les relations entre ces performances et les caractéristiques de leur milieu d'origine, ainsi que l'interprétation des corrélations environnementales entre caractères pourra donner des éléments de réponse quant au comportement des génotypes quand le milieu varie et pourra permettre de faire des prédictions sur leurs performances dans d'autres milieux.

A.1. Résistance à la sécheresse et relation entre production de biomasse et consommation en eau

Les plantes disposent de différentes stratégies de résistance face à un stress environnemental comme un stress hydrique : l'échappement (par exemple, en complétant leur cycle biologique pendant une période sans stress hydrique), l'évitement (par exemple, en augmentant l'extraction d'eau par un système racinaire plus développé ou en retenant l'eau à l'intérieur des cellules) et la tolérance (par exemple, en développant des morphologies foliaires particulières ou des processus physiologiques qui leur permettent de supporter une sécheresse) (Bray 2001, Figure 36). Bien que ces stratégies ne soient pas exclusives, les études sur le pin maritime font plutôt référence à des comportements d'évitement (Nguyen-Queyrens *et al.* 1998, El Hadj Moussa 1989) avec une modulation de la transpiration selon les

disponibilités en eau chez la provenance des Landes, ou un système racinaire plus développé pour une provenance marocaine.

Parmi ces mécanismes d'évitement, celui de la modulation de la transpiration grâce à l'ouverture et la fermeture des stomates, présente néanmoins l'inconvénient de générer une baisse de production de biomasse. En effet, l'assimilation de CO₂ pour la photosynthèse et la perte d'eau par évapotranspiration sont deux processus qui interviennent au niveau des stomates. La fermeture des stomates en période de stress hydrique peut donc entraîner un coût physiologique en diminuant la fixation du carbone, limitant ainsi l'activité photosynthétique et la productivité (Bray 2001, Rhodes et Nadolska-Orczyk 2002, Araya 2007). La plante doit donc chercher un compromis entre l'acquisition de CO₂ pour la production de composés organiques et sa croissance, et la limitation de la perte d'eau, notamment en conditions de déficits en eau. Le résultat de ce compromis peut se définir par l'efficacité d'utilisation en eau (*EUE*) qui correspond au niveau de la plante entière au rapport entre la biomasse produite par la plante et la quantité d'eau qu'elle consomme. Ce caractère physiologique est donc potentiellement très important pour les stratégies d'évitement développées par les plantes en réponse à un déficit en eau.

A.2. Calcul de l'*EUE* par l'approche isotopique

Néanmoins, la mesure de l'*EUE* au niveau de la plante est particulièrement longue et difficile notamment pour les plantes pérennes, puisqu'elle impliquerait une estimation précise de l'eau consommée par la plante et une destruction des plants (arrachage et pesée de la biomasse totale). C'est pourquoi, une approche basée sur la discrimination isotopique du carbone par la plante a été développée par Farquhar *et al.* (1989) dans l'objectif de proposer une estimation de l'*EUE* plus facile à mesurer.

On définit tout d'abord l'*EUE* instantanée (*EUE_i*) comme le rapport entre l'assimilation nette du CO₂ (*A*) et la transpiration foliaire (*E*), cette dernière dépendant à la fois de la différence de pression de vapeur d'eau air-feuille (*VPD*) et de la conductance stomatique à diffusion de vapeur d'eau (*g_s*). Sachant que la *VPD* est très sensible à l'environnement et difficile à estimer, c'est plutôt l'*EUE* intrinsèque foliaire (*EUE_{int}*) que l'on utilise car il ne retient que le ratio *A / g_s*.

Ce ratio *A / g_s* est relié selon la loi de diffusion de Fick (Guehl *et al.* 1996) aux concentrations respectives en CO₂ de l'intérieur de la feuille (*C_i*) et de l'air atmosphérique (*C_a*) par :

$$(e1) \ EUE_{int} = A/g_s = C_a (1 - C_i/C_a)/1,6$$

où le coefficient 1,6 est représentatif des échanges gazeux d'eau et de CO₂ *via* les stomates. Ces échanges sont donc dépendants de la fermeture des stomates et de la disponibilité en eau.

D'autre part, durant l'assimilation de CO₂ pour la photosynthèse, les plantes fixent préférentiellement l'isotope ¹²C à l'isotope ¹³C, le premier étant plus léger et donc plus facilement mobilisé (Craig 1954). En 1982, Farquhar *et al.* proposent de mesurer la discrimination isotopique du carbone par la plante (Δ) à partir des ratios d'abondance isotopique ($R = ^{13}\text{C}/^{12}\text{C}$) dans l'air (R_a) et dans la plante (R_p), tel que:

$$(e2) \Delta = R_a/R_p - 1.$$

R_a étant constante à l'échelle d'une expérience, la discrimination Δ sera donc proportionnelle à la quantité de ¹²C et inversement proportionnelle à la quantité de ¹³C absorbées dans la plante (Figure 37).

Or, ces auteurs ont également montré que chez les plantes en C3 (dont font parti les arbres) cette discrimination Δ pouvait être reliée aux compositions de carbones C_i et C_a de la façon suivante :

$$(e3) \Delta = a + (b-a)C_i/C_a$$

où a est un coefficient de fractionnement des isotopes du carbone durant la diffusion du CO₂ atmosphérique au travers des stomates, et b , un coefficient de fractionnement durant la diffusion au travers des cellules de la feuille jusqu'au stroma des chloroplastes.

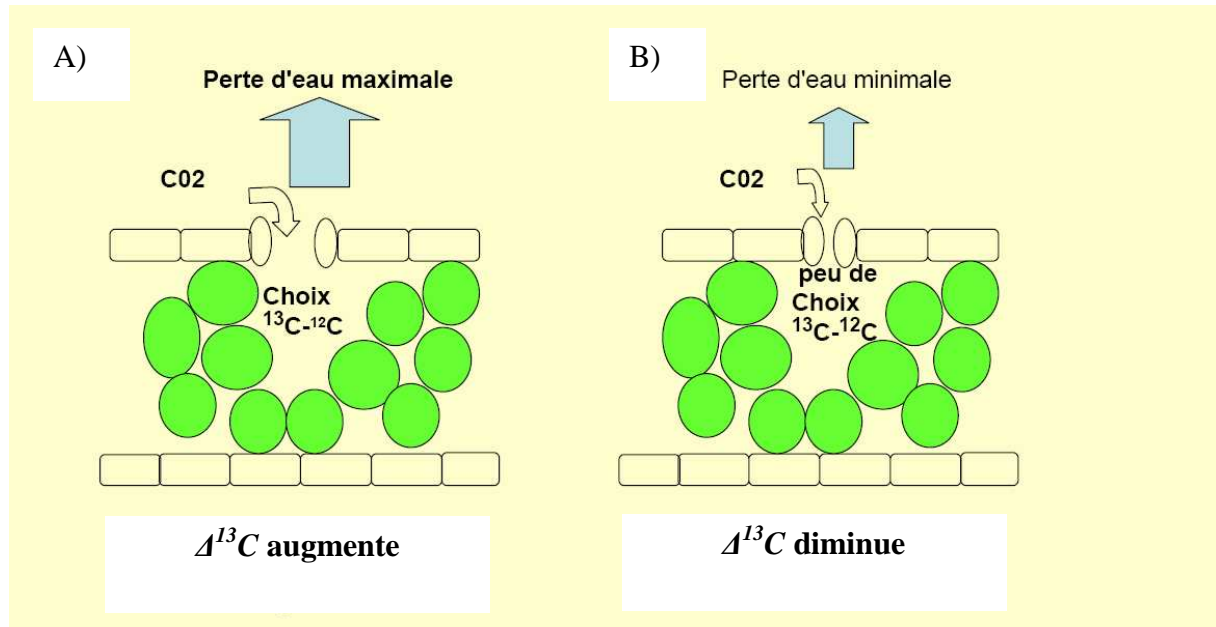
Sur la base des relations (e1) et (e3) qui sont des fonctions de C_i et C_a , on montre que la discrimination isotopique du carbone durant l'assimilation du CO₂ est reliée négativement à l' EUE_{int} foliaire : plus la plante discrimine contre le ¹³C, moins l'efficience est importante, ce qui correspond à une situation où les stomates *a priori* plus ouverts ne limitent pas la fixation du ¹²C (Figure 37). A l'inverse, durant un stress hydrique, la fermeture des stomates se traduira par une diminution du rapport C_i/C_a et du Δ (cf équation (e2)) et donc par une augmentation en théorie de l' EUE_{int} . Cette relation négative entre Δ et EUE_{int} a été montrée initialement par Farquhar et Richards (1984) chez le blé.

Néanmoins, dans la formule du Δ (e2), la valeur absolue R_p reste difficile à mesurer directement, donc c'est l'écart de la composition isotopique de la plante par rapport à un standard international, la Pee Dee Belemnite (PDB), que l'on mesure. Cette mesure notée $\delta^{13}\text{C}$, est définie comme la composition isotopique de la façon suivante:

$$(e4) \delta^{13}\text{C} = (R_p - R_s) / R_s$$

où R_s est le ratio d'abondance isotopique du carbone pour le standard. La valeur de discrimination isotopique du carbone de la plante, Δ , correspond alors à l'écart entre

Figure 37. Comment le rapport entre les isotopes ^{13}C et ^{12}C du carbone ($\delta^{13}\text{C}$) est-il relié à l'efficacité d'utilisation de l'eau ? Lorsque l'ouverture des stomates est maximum, la plante peut choisir de fixer préférentiellement l'isotope ^{12}C qui se traduira par une augmentation de la discrimination $\Delta^{13}\text{C}$ (A). Lorsque l'ouverture des stomates diminue, la plante aura moins de choix et fixera les deux isotopes, qui se traduira par une baisse de la discrimination $\Delta^{13}\text{C}$ (B) (d'après Jean-Louis Prioul, Réunion Scientifique du 24 Jun 2005, Société Française des Isotope Stables).



compositions isotopiques de l'air ($\delta^{13}C_a$) et de la plante ($\delta^{13}C_p$). En remplaçant R_p et R_a par leurs valeurs d'après (e4), on obtient:

$$(e5) \Delta = (\delta^{13}C_a - \delta^{13}C_p) / (1 + \delta^{13}C_p)$$

avec $\delta^{13}C_a$ environ égale à -8‰ bien que cette valeur puisse varier par exemple avec la distance par rapport au sol ou le changement de saison (Farquhar *et al.* 1989). Cette équation montre que Δ est reliée négativement à $\delta^{13}C_p$, donc on s'attend à une relation positive entre $\delta^{13}C_p$ et EUE_{int} . Cette relation positive a été mise en évidence expérimentalement chez plusieurs espèces forestières (par exemple, chez le chêne, Ponton *et al.* 2001, le mélèze occidental, Zhang et Marshall 1994, Zhang *et al.* 1994, le gommier bleu, Osório *et al.* 1998, et le pin ponderosa, Cregg *et al.* 2000). Dans la pratique, comme dans notre expérience, c'est souvent le paramètre de composition isotopique $\delta^{13}C_p$ qui sera mesuré.

A.3. Utilisation du caractère EUE_{int} dans des programmes de sélection

En 1977, Passioura propose d'intégrer la composante physiologique que constitue l' EUE dans un modèle décomposant le rendement comme suit : $Y = E * EUE * HI$. (Y , rendement, E , quantité d'eau évapotranspirée par la culture, HI , indice de récolte). A cause des relations potentielles entre EUE_{int} et la croissance ou bien la survie en milieu limité en eau (compétition avec d'autres arbres dans un même peuplement), ce caractère pourrait être un critère de sélection potentiellement intéressant pour des programmes visant à améliorer la tolérance des plantes à la sécheresse (Condon *et al.* 2002, Condon *et al.* 2004, Dreyer *et al.* 2004, Pita *et al.* 2001).

A.4. Questions de recherche

Une large variabilité dans les moyennes de $\delta^{13}C_p$ observée entre provenances d'espèces d'arbres forestiers a pu être reliée chez certaines espèces aux conditions climatiques dans lesquelles avaient évolué ces provenances (par exemple, Lauteri *et al.* 1997 chez le Châtaigner, Li *et al.* 2000 chez l'Eucalyptus). Pour le pin maritime, il n'existe que quelques études basées sur des échantillons limités de populations naturelles (au maximum 5) provenant de Corse, de la côte atlantique française, du Portugal, du Maroc et d'Italie (Guehl *et al.* 1996, Nguyens-Queyrens *et al.* 1998, Brendel *et al.* 2002, Tognetti *et al.* 2000). Dans ces études, une variabilité des moyennes de $\delta^{13}C_p$ importante a été observée entre provenances. Certaines de ces provenances présentant différentes caractéristiques morphologiques et physiologiques pouvant affecter leur sensibilité à la sécheresse (par exemple, en termes de croissance racinaire et de régulation osmotique notamment entre la population marocaine

Tamjoute et une population Landaise, Guyon et Kremer 1982, Sarrauste 1982, Nguyen et Lamant 1989), cette variabilité a notamment été interprétée comme la conséquence de stratégies différentes d'adaptations à des conditions environnementales contrastées. Dans cette étude, nous avons voulu savoir pour un échantillon plus large de populations naturelles représentatif de l'aire de distribution naturelle de cette espèce, si l'on observait des différences entre valeurs de $\delta^{13}C_p$, ainsi que pour d'autres caractères adaptatifs tels que la croissance en hauteur ou la biomasse. Nous nous sommes notamment demandé quelle était la part de la variation qui pouvait être expliquée par la structure en populations. Nous avons aussi essayé de voir si les conditions des milieux d'origine de ces populations permettaient de mieux comprendre les variations observées, et si l'on pouvait parler d'adaptation locale pour l' EUE_{int} chez certaines populations de pin maritime.

D'autre part, quelques études mettent en évidence une variabilité génétique additive significative sur la base d'un indice d'héritabilité au sens strict (rapport de la variance génétique additive sur la variance phénotypique) pour le $\delta^{13}C_p$ chez des espèces d'arbres forestiers aux distributions peu fragmentées (par exemple, héritabilité de 0.31 +/- 0.07 dans des populations naturelles de *Castanea sativa*, Lauteri *et al.* 2004 ; héritabilité de 0.54 +/- 0.26 dans un diallèle de *Picea mariana*, Johnsen *et al.* 1999). Cette large variabilité génétique additive est un pré-requis du potentiel adaptatif de l'espèce, mais pour une espèce où il existe une forte structure en populations comme le pin maritime, est-ce que cette variance génétique additive est maintenue au sein de chaque population pour ce caractère (héritabilité du $\delta^{13}C_p$ estimée à 0.17 +/- 0.06 dans un semi-diallèle composé de seulement 12 familles d'une population des Landes, Brendel *et al.* 2002) et si ce n'est pas le cas, comment varie-t-elle entre les différentes populations naturelles?

Par ailleurs, le potentiel adaptatif et la variabilité génétique peuvent aussi dépendre de l'importance et de la structure des corrélations génétiques entre caractères liés à l'adaptation, ces corrélations pouvant être causées par plusieurs facteurs (pléiotropie, déséquilibre de liaison, sélection...). Comment le caractère $\delta^{13}C_p$ est-il alors relié à la croissance et à la biomasse ? Cette relation est-elle la même au sein de toutes les populations naturelles ? Il a été montré que la croissance des plants pouvait être positivement ou négativement corrélée aux valeurs de $\delta^{13}C_p$ foliaire selon que la variation dans la discrimination isotopique du carbone était plutôt associée à des changements dans l'assimilation nette du CO₂ (A) (traduisant la capacité photosynthétique) ou la conductance stomatique (g_s) (Farquhar *et al.*

1989). Les relations entre valeurs de $\delta^{13}C_p$ et croissance peuvent également dépendre des conditions environnementales. Par exemple, Condon et Richards (1993) ont montré chez des génotypes de blé que ces relations pouvaient changer quand les plants étaient cultivés dans des sites recevant plus ou moins de pluie. Chez le pin maritime, des corrélations phénotypiques positives (r^2 entre 0.4 et 0.6) entre l' EUE_{int} (mesurée par des échanges gazeux), le $\delta^{13}C_p$ et des caractères de croissance (biomasse totale, hauteur totale ou circonférence moyenne de cerne) ont été mises en évidence au sein d'une population Landaise par Guehl *et al.* (1996) (dans un test implanté au Nord Est de la France) et Brendel *et al.* (2002) (dans un test implanté dans les Landes). De même, Tognetti *et al.* (2000) montrent qu'il existe une forte corrélation phénotypique positive (r^2 de 0.7) entre la hauteur moyenne des arbres et la valeur moyenne de $\delta^{13}C_p$ pour 5 populations provenant du Portugal, de la Corse et de l'Italie, et implantées dans quatre stations différentes d'Italie. Toutefois au niveau des corrélations génétiques, Brendel *et al.* (2002) montrent que celle entre biomasse et EUE_{int} n'est pas significative au sein de la population Landaise dans le test implanté dans les Landes. Ils suggèrent qu'au sein de cette population au moins, le contrôle génétique de ces deux caractères serait indépendant. Les relations entre $\delta^{13}C_p$ et croissance semblent donc complexes et il est important de les étudier dans un échantillon plus large de populations implantées dans un environnement commun.

A.5. Objectifs

Dans le but de mieux comprendre l'adaptation à la sécheresse chez le pin maritime, notre premier objectif était donc de quantifier la variabilité génétique du caractère d'efficacité d'utilisation en eau par une mesure de $\delta^{13}C_p$, dans un large échantillon de populations naturelles de pin maritime (18 populations avec en moyenne une 20aine de familles par population) provenant de régions aux conditions climatiques contrastées (Tableau 16). Le second objectif était d'estimer les corrélations génétiques entre le $\delta^{13}C_p$ et les caractères liés à la croissance en hauteur et/ou biomasse au sein de chaque population. Pour cela, nous avons analysé les données du test de provenance/descendances mis en place à Pierroton (Cestas, France).

B. Matériels et méthodes

B.1. Description du dispositif expérimental

Le test de provenances / descendances a été mis en place à partir de graines semées en mottes tourbe en Juin 2003. Puis les plantules ont été installées en plein champ en Décembre

Tableau 16. Coordonnées géographiques et données climatiques disponibles sur 14 populations faisant parti du test de provenances.

Nom de la population	Code de la population	Longitude ^(a)	Latitude (N) ^(b)	Altitude (m)	ANMR (mm) ^(c)	RDRYM (mm) ^(d)	ANMT (°C) ^(e)	ANMT variance (°C) ^(f)	MTCM (°C) ^(g)	MTWM (°C) ^(h)
Pleucadeuc	46	2°20' (W)	47°47'	80	716	11	11.6	0.32	4.0	19.0
Mimizan	41	1°18' (W)	44°08'	37	995	16	13.8	0.45	6.3	21.2
San cipriano de Ribaterme	27	8°42' (W)	42°08'	310	661	18	15.1	0.26	6.9	24.3
Pinia	15	9°29' (E)	42°01'	10	705	3	na (i)	na	na	na
San Leonardo de Yague	25	3°4' (W)	41°49'	1050	671	14	9.9	0.29	1.9	19.5
Bayubas de Abajo	26	2°52' (W)	41°31'	955	405	6	12.0	0.31	3.2	22.2
Cuellar	23	4°19' (W)	41°24'	858	458	8	12.3	0.34	3.2	22.7
Coca	22	4°05' (W)	41°22'	788	428	7	12.2	0.26	3.2	22.5
Valdemaqueda	24	4°17' (W)	40°31'	926	594	8	10.7	0.40	2.6	21.0
Cenicientos	20	4°28' (W)	40°16'	na	579	6	12.4	0.26	3.9	22.7
Arenas San Pedro	21	5°08' (W)	40°02'	1359	957	5	13.8	0.65	4.8	24.4
Oria	29	2°37' (W)	37°52'	1232	348	2	12.5	0.23	3.9	22.9
Tabarka	50	8°54' (E)	36°65'	66	999	na	17.9	na	na	na
Tamrabta	30	5°0' (W)	34°0'	1600	763	na	17.2	na	na	na
Moyenne					708	9	13.12	0.50	4.0	22.0

^(a) Longitude Est (E) or Ouest (W).

^(b) Latitude Nord (N).

^(c) ANMR : Précipitations totales annuelles moyennées entre 1964 et 2001 pour les populations françaises, corses (MétéoFrance, <http://www.meteofrance.com/>) et espagnoles (A.F. Cancio, comm. personnelle) avec en moyenne 10% d'années manquantes. ANMR estimé de 1933 à 1963 pour la population Tamrabta (Wahid et al. 2004) et de 1901 à 1990 pour la population Tabarka (JBIC/SCET Tunisie, 2005)

^(d) RDRYM : Moyenne des précipitations totales du mois le plus chaud.

^(e) ANMT : Températures annuelles moyennées entre 1978 et 2001 pour les populations françaises, corses et espagnoles (même source de données que pour ANMR) avec en moyenne moins de 1% de données manquantes. ANMT estimé sur 18 années dans la station de Meknes pour la population Tamrabta (<http://www.weatherbase.com/>) et de 1901 à 1990 pour la population Tabarka (même source de données que pour ANMR).

^(f) ANMT variance : Variance de ANMT entre années.

^(g) MTCM : Moyenne des températures du mois le plus froid.

^(h) MTWM : Moyenne des températures du mois le plus chaud.

⁽ⁱ⁾ Données non disponibles.

2003, et suivies durant 3 saisons de végétation (jusqu'en décembre 2006). Ce test était localisé dans la pépinière de l'Unité Expérimentale de l'INRA Pierroton (Gironde, France, 44°44'N, 0°44'W), sur un podzol de lande humide à mésophile (semi-humide). Les plants n'ont pas été irrigués, car en se basant sur les conditions de températures et de précipitations de cette station, on pouvait supposer qu'elles permettraient de discriminer suffisamment les provenances, tout en minimisant le taux de mortalité des plants. Les plants ont été installés selon un dispositif en parcelles mono-arbre réparties de façon aléatoire, avec 15 blocs dont 5 complets établis sur deux grandes parcelles très proches (Figure 38). Chaque bloc comprenait un descendant de chacun des 20 à 30 arbres mères constituant les 24 populations collectées, les descendants d'un arbre mère (correspondant à une famille de demi-frère) comprenant au maximum 15 individus dans l'essai. La densité de plantation était d'environ 11000 plants par hectare, avec environ 600 plants par bloc, soit un total d'environ 9000 plants sur l'ensemble du dispositif.

B.2. Modalités de mesure et stratégie d'échantillonnage de la composition isotopique du carbone

B.2.1. Modalités de mesure

L'efficacité d'utilisation en eau intrinsèque foliaire des plantes (EUE_{int}) a été évaluée par une approche isotopique, en mesurant la composition isotopique du carbone de chaque plante (voir introduction) que nous avons notée ici $\delta^{13}C$. Les mesures de composition isotopique du carbone peuvent être réalisées sur aiguilles, bois, ou cellulose, mais les mesures sur bois et cellulose sont plus longues et plus coûteuses (étapes de carottage et extraction de cellulose). Notre choix s'est donc porté sur une mesure à partir d'aiguilles. Ce type de mesure est cependant assez sensible à la variation environnementale donc nous en avons tenu compte dans la stratégie d'échantillonnage développée ci-dessous.

Les échantillons d'aiguilles ont été mis à l'étuve à 60°C pendant 24 heures, puis broyés à l'aide d'un broyeur Retsch MM2000 (Hann, Allemagne). Puis un premier lot d'analyses de la composition isotopique du carbone a été réalisé à l'INRA de Nancy (France), et un deuxième lot géré par nos collègues de l'INIA dans le cadre du projet TREESNIPs, a été mesuré dans un laboratoire de l'Université du Colorado (USA). Pour le premier lot d'analyses, nous avons pesé une quantité de 1 mg +/- 0.1 de poudre de chaque échantillon dans une micro-capsule en aluminium de 4x6 mm (Elemental Microanalysis Ltd.) selon le protocole de Brendel *et al.* (2003). Nous avons réparti les capsules dans des plaques ELISA

Figure 38. Plan de l'essai expérimental réparti en 2 parcelles (distance en mètres).

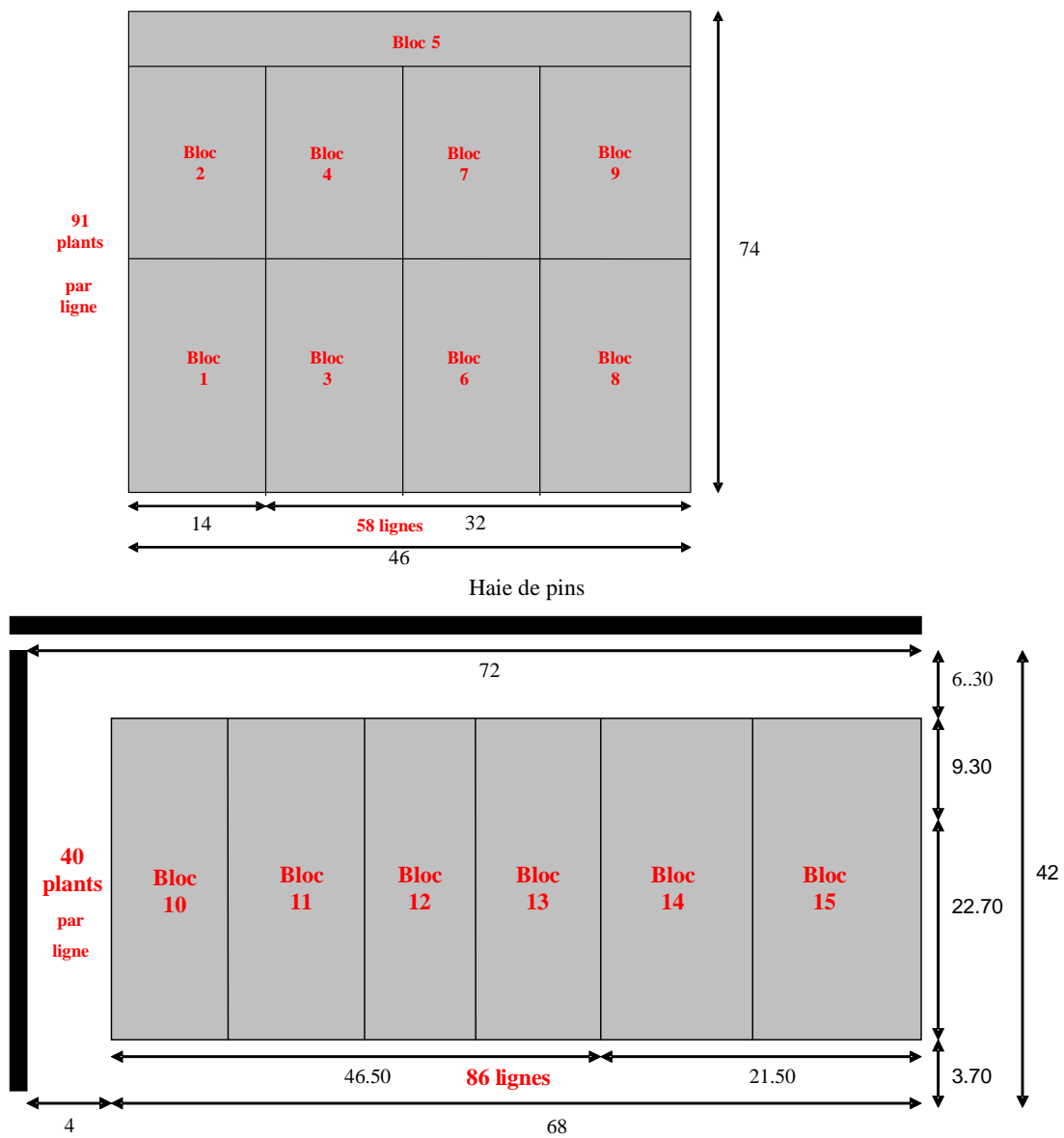
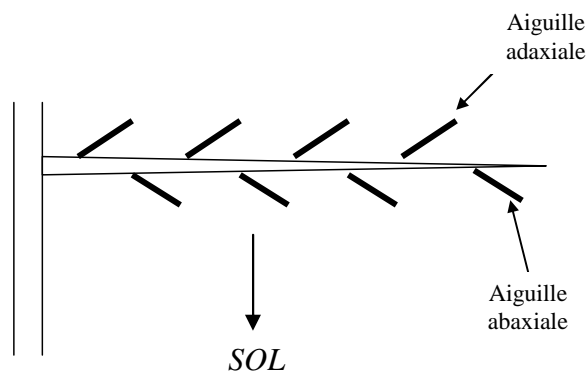


Figure 39. Positions abaxiales et adaxiales des aiguilles le long d'une branche.



de 96 puits. Au sein de chaque plaque, nous avons inclus 11 échantillons de référence d'une poudre d'aiguilles ayant une composition isotopique connue. Ces échantillons ont été par la suite utilisés comme standards internes au sein de chaque série de mesures (correspondant aux 96 échantillons d'une plaque). Pour le deuxième lot d'analyse, les modalités de mesure étaient les mêmes, mais la quantité de poudre d'aiguilles utilisée pour chaque mesure était deux fois moins importante (0.5 mg au lieu de 1 mg). Les mesures de composition isotopique du carbone ont été réalisées pour les deux lots d'analyse par chromatographie en phase gazeuse couplée à de la spectrométrie de masse (Delta S ; Finnigan MAT, Bremen Allemagne).

Il peut se produire une dérive des valeurs de $\delta^{13}C$ obtenues sur une série de mesures des échantillons constituant une plaque, due à l'appareil (par exemple à cause de la fin de vie de la bouteille contenant le gaz porteur). C'est pourquoi, au sein de chaque série de mesures, nous avons réalisé une correction de la valeur de chaque échantillon, consistant à faire une régression linéaire des 12 valeurs de standard au sein de la série de mesures, puis à utiliser le coefficient de pente de la droite pour obtenir une valeur corrigée de chaque échantillon le long de la série de mesures.

B.2.2. Stratégie d'échantillonnage

Bien que plusieurs études menées chez des espèces d'arbres forestières et citées en introduction, mettent en évidence une variation génétique de la composition isotopique du carbone, $\delta^{13}C$, plusieurs sources de variation d'origine non-génétique peuvent affecter cette mesure :

- 1) une variation spatiale au sein de l'essai,
- 2) une variation positionnelle en fonction de la position à laquelle est prélevée l'aiguille sur la plante,
- 3) et une variation temporelle (au cours d'une journée notamment) ou entre différentes dates de prélèvement qui pourraient être affectées par des événements climatiques différents.

Brendel *et al.* (2003) montrent en effet chez *Pinus sylvestris* qu'il existe une variation positionnelle et temporelle non négligeable dans les valeurs de $\delta^{13}C$ mesurées à partir d'aiguilles. Par exemple, les aiguilles abaxiales (orientées vers le sol, Figure 39) présenteraient des valeurs de $\delta^{13}C$ plus élevées que les aiguilles adaxiales (orientées vers la lumière, Figure 39) d'une même branche. Cette observation pourrait s'expliquer par la présence de bois de compression dans la partie abaxiale de la branche qui conduirait à une diminution de la conductance stomatique jusqu'aux aiguilles. Cette étude met également en

évidence une variation temporelle dans les valeurs de $\delta^{13}C$ entre aiguilles prélevées au fil des mois: les valeurs de $\delta^{13}C$ augmentent d'environ 0.6 (correspondant à 60% de la variation totale dans leur essai au cours du temps) durant le printemps (coïncidant avec un pic de rayonnement solaire et une augmentation du potentiel d'évapotranspiration de la plante) alors que les valeurs de $\delta^{13}C$ varient au plus de 0.2 (20% de la variation totale) au cours des autres mois. Il faut donc tenir compte de ces variations spatiales et temporelles afin d'optimiser la stratégie de prélèvement des aiguilles, sachant que la variation spatiale est *a priori* partiellement prise en compte par un plan en blocs complets ou incomplets.

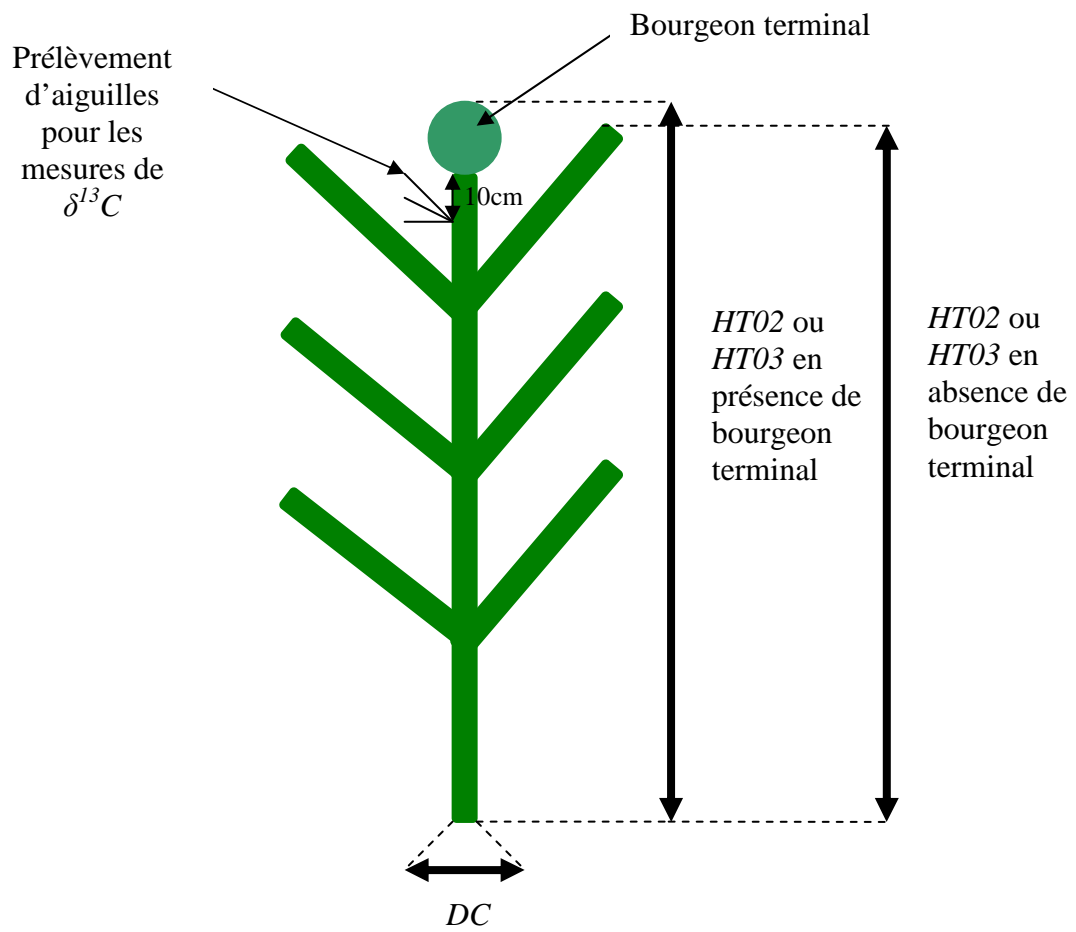
Les aiguilles ont donc été collectées dans la partie apicale de l'arbre (à 10 cm en dessous du bourgeon terminal) et avant le lever du jour, afin de minimiser les variations des valeurs de $\delta^{13}C$ dues à la position le long de l'arbre, et celles dues à la variation des conditions d'éclairage au cours de la journée (Figure 40).

D'autre part, nous avons comme contrainte économique de pouvoir effectuer un maximum de 2300 mesures. L'année précédant la grosse campagne de mesures, nous avons donc essayé d'obtenir une estimation préliminaire des différentes sources de variations contrôlées du dispositif pour le $\delta^{13}C$, celles des effets d'origine génétique (effets populations et familles), et environnementale (intra-arbre, entre blocs), afin d'optimiser leur estimation l'année suivante sur l'ensemble des plants. Le test préliminaire a été réalisé en Juillet 2004 sur une centaine d'échantillons.

Pour tester s'il existait une variation positionnelle intra-arbre significative, 5 aiguilles ont été prélevées à 5 orientations différentes selon les points cardinaux (toujours à 10 cm en dessous du bourgeon terminal) sur 4 individus dans 4 familles d'une population (Mimizan), tandis qu'un mélange d'aiguilles provenant de 3 orientations différentes (Sud, Nord-Ouest, Nord-Est) a été collecté sur 4 individus pour une 5^{ème} famille. La variation du $\delta^{13}C$ a été analysée à l'aide d'un modèle d'analyse de variance (ANOVA) (procédure PROC GLM de SAS (1999)) en déclarant les effets famille, bloc, position, et interaction entre famille et bloc. L'effet position intra-arbre était non significatif, et la variation résiduelle (reste de la variation intra-arbre) était faible (12%) relativement à la variation entre arbres (72%). Il a donc été décidé de faire par la suite un mélange d'aiguilles échantillonnées à deux orientations aléatoires sur chaque plant.

Dans un deuxième temps, nous avons analysé l'importance relative des effets population, bloc et famille (hiérarchisé à l'effet population), dans la variation totale du caractère. Pour cela, 5 populations d'origines différentes et présentant des conditions climatiques très contrastées (Tableau 16) ont été choisies: Arenas de San Pedro (Espagne), Coca (Espagne),

Figure 40. Schéma des prélèvements d'aiguilles pour la mesure du $\delta^{13}C$, et des mesures de diamètre (DC), et de hauteur ($HT02$, $HT03$) en présence ou en absence de bourgeon terminal.



Mimizan (Landes), Pinia (Corse), Tabarka (Tunisie). Un mélange d'aiguilles a été collecté sur 5 familles de chacune de ces populations, et la variation du $\delta^{13}C$ a été analysée par ANOVA selon un modèle comprenant les effets population, famille, bloc et interaction entre famille et bloc. L'effet population était significatif et expliquait 10% de la variation totale en somme des carrés des écarts (SCEt). D'autre part, l'effet bloc également significatif expliquait 13% de la variation totale, tandis que l'effet d'interaction entre bloc et population était non significatif, indiquant qu'une partie de la variation environnementale était contrôlée de façon efficace par le dispositif en blocs. Cependant, l'effet entre famille hiérarchisée à l'effet population était non significatif même s'il expliquait 15% de la variation totale, à cause probablement d'un manque de puissance dû à l'échantillonnage limité. Par conséquent, la grosse campagne de mesures a été planifiée pour Août 2005 sur 18 populations couvrant la distribution de l'aire naturelle du pin maritime, avec 1) un nombre maximum de familles possibles (20 à 30) au sein de chaque population pour mieux estimer l'effet famille et la variance entre familles et 2) un nombre de répétitions par famille (ou nombre de blocs) suffisant (7) pour estimer plus précisément la moyenne de chaque famille. Pour atteindre cet objectif, en plus des 2300 mesures initialement prévues, 900 mesures supplémentaires ont été réalisées par l'INIA dans un laboratoire des USA (voir plus haut).

B.3. Mesures des caractères de hauteur et de biomasse

Des mesures relatives à la hauteur et à la biomasse ont également été réalisées, sur l'ensemble des 9000 plants, dans l'objectif d'analyser la variation phénotypique entre provenances et leur relation avec le caractère $\delta^{13}C$ sur le sous échantillon des 3200 plants.

Les hauteurs totales des plants ont été mesurées avec une précision de +/- 1 cm, à la fin de la seconde (*HT02*) et de la troisième (*HT03*) année de croissance (Figure 40 et Tableau 17). La mesure a été faite depuis la base du plant jusqu'à la base du bourgeon terminal (Figure 40). Dans le cas d'un avortement accidentel du bourgeon terminal, la hauteur a été mesurée jusqu'au sommet de la branche la plus verticale, qui prend le relais dans la croissance de l'arbre. Lors de la troisième année de croissance, le nombre de cycles annuel (*Poly*) a été comptabilisé pour avoir une idée du caractère polycyclique ou monocyclique des différentes provenances, ce caractère pouvant être relié à la fabrication d'aiguilles, de branches ou de cônes, ou à l'accroissement total (Kremer et Roussel 1982).

A la fin de la troisième année de croissance, les 3200 plants ont été coupés au collet, et le diamètre au collet (*DC*) mesuré (Figure 40). Les plants ont été par la suite mis à l'étuve à 55°C pendant plus de 72 heures jusqu'à séchage complet (poids sec stable). La biomasse

Tableau 17. Caractères quantitatifs mesurés au sein du test de provenances / descendances.

Caractère	Unité	Description
<i>HT02</i>	cm	Hauteur totale des plants à la fin de la seconde année de croissance.
<i>HT03</i>	cm	Hauteur totale des plants à la fin de la troisième année de croissance.
<i>DC</i>	mm	Diamètre au collet à la fin de la troisième année de croissance.
<i>Poly ou PolyT</i>	nb	Nombre de cycle durant la troisième année de croissance (Poly), transformé en polyT selon $\sqrt{(n+0.5)}$ pour normaliser les données.
<i>TB</i>	gr	Biomasse sèche totale à la fin de la troisième année de croissance.
<i>NB</i>	gr	Biomasse sèche des aiguilles à la fin de la troisième année de croissance.
$\delta^{13}C$	‰	Composition isotopique du carbone sur des aiguilles d'arbres à leur troisième année de croissance.

aérienne sèche totale (*TB*), ainsi que la biomasse aérienne sèche du bois (tronc et branches) ont été pesées avec une précision de +/-5g, pour en déduire la biomasse sèche des aiguilles (*NB*).

B.4. Analyses statistiques

B.4.1. Modèle d'analyse de variance et estimation des composantes de variance

L'analyse présentée ici porte sur le sous échantillon sur lequel le $\delta^{13}C$ a été mesuré (3200 arbres appartenant à 18 populations). Un modèle mixte comprenant l'ensemble des effets population, famille (hiérarchisé à la population), bloc, et interaction entre population et bloc a été déclaré pour décrire la répartition de la variation des caractères phénotypiques en composantes génétiques et environnementales. L'effet interaction entre population et bloc n'étant pas significatif, nous avons retenu un modèle déclarant les effets populations, familles et résiduels comme effets aléatoires, et l'effet bloc comme effet fixe tel que :

$$Y_{ijkl} = \mu + P_i + F_{j(i)} + b_k + \varepsilon_{ijkl} \quad (1)$$

Avec,

- Y_{ijkl} , valeur phénotypique du caractère quantitatif du $l^{\text{ème}}$ arbre appartenant à la $j^{\text{ème}}$ famille au sein de la $i^{\text{ème}}$ population, localisé dans le $k^{\text{ème}}$ bloc,
- μ , moyenne générale du caractère,
- P_i , effet aléatoire de la $i^{\text{ème}}$ population (i variant de 1 à 18) de variance V_{pop}
- $F_{j(i)}$, effet aléatoire de la $j^{\text{ème}}$ famille de demi-frères au sein de la $i^{\text{ème}}$ population (j variant de 1 à 30) de variance V_{fam} ,
- b_k , effet fixe du $k^{\text{ème}}$ bloc (5 premiers blocs complets parmi les 15 blocs),
- ε_{ijkl} , effet aléatoire résiduel, de variance V_r .

La normalité des résidus a été vérifiée pour chaque caractère, d'après le test de Shapiro-Wilks (Sokal et Rohlf 1995). Seule la distribution des résidus du caractère Poly (Tableau 17) ne suivait pas une loi normale, car il s'agit d'un caractère discret (1 à 5). Pour obtenir une distribution normale, ce caractère a été par la suite transformé en polyT (polyT= $\sqrt{\text{Poly}+0.5}$) qui a permis de normaliser les résidus (Sokal et Rohlf 1995) (Annexe 3).

Les différentes composantes de la variance (V_{pop} et V_{fam}), ont été estimées par une méthode de maximum de vraisemblance dite « restreinte » connue pour être moins biaisée par

rapport à une méthode classique de maximum de vraisemblance (Searle *et al.* 1992). Ces estimations ont été obtenues grâce à l'option REML de la procédure PROC VARCOMP de SAS.

B.4.2. Estimation du coefficient de variation phénotypique, de la variance génétique additive et de l'héritabilité au sens strict

Le coefficient de variation phénotypique, CV_p (correspondant à l'écart type standardisé par la moyenne générale du caractère, μ) a été estimé sur l'ensemble des données pour tous les caractères à l'exception du $\delta^{13}C$. En effet, le $\delta^{13}C$ est une valeur calculée par rapport à un standard. La moyenne du $\delta^{13}C$ est donc dépendante du standard utilisé. De la même façon, le CV_p est également dépendante du standard utilisé, et est donc difficilement comparable avec celui d'autres caractères phénotypiques.

Après avoir enlevé l'effet population estimé par le modèle complet, une estimation globale de la variance génétique additive, V_A a été déduite de la variance entre familles de demi-frères V_{fam} à partir de l'équation $V_{fam} = \frac{1}{4} V_A$ (Gallais, 1990). Enfin, l'héritabilité au sens strict, h^2_s (Falconer et McKay 1996) a été estimée par le rapport entre V_A et la variance phénotypique, V_p . Cette variance phénotypique est la variance totale du modèle d'ANOVA intra population, égale à la somme de la variance familiale, V_{fam} , et de la variance résiduelle, V_r et exclue la composante population. Cette héritabilité est traditionnellement utilisée dans « l'équation du sélectionneur » ($R = h^2_s \times S$, avec R , la réponse attendue du caractère sélectionné, S , l'intensité de sélection, et h^2_s , l'héritabilité) pour prédire la réponse d'une population à une pression de sélection directionnelle (Falconer et McKay 1996).

L'erreur standard de l'héritabilité stricte est obtenue selon l'approche d'Osborne et Paterson (1952) et exprimée comme suit (Visscher 1998) :

$$S.E._{h^2} = 4x \sqrt{\frac{2(1 - \frac{h^2}{4})^2 \left[(1 + (s-1) \frac{h^2}{4}) \right]^2}{s(s-1)(f-1)}}$$

Avec, s , le nombre moyen de descendants par famille, et f le nombre de familles.

B.4.3. Estimation des corrélations génétiques et environnementales entre caractères quantitatifs

Les corrélations génétiques entre caractères ont été estimées à trois niveaux : populationnel (r_{pop}), familial (r_{fam}) (correspondant à la corrélation génétique additive dans un

dispositif de familles de demi-frères, Lynch et Walsh 1997), et résiduel (r_{err}). Les corrélations génétiques entre caractères mesurent un degré de relation génétique entre caractères et sont dues au sens général à une association ou covariation entre informations génétiques. Elles peuvent avoir plusieurs origines (Gallais 1990) 1) la pléiotropie (action simultanée d'un gène sur plusieurs caractères), 2) une liaison physique étroite entre plusieurs gènes entraînant un déséquilibre de liaison qui serait maintenu par l'effet de la sélection naturelle, la combinaison d'allèles à ces gènes présentant un avantage sélectif. Ce serait par exemple le cas d'un même QTL pouvant correspondre à de nombreux gènes candidats (Flint et Mott 2001, Christians et Keightley 2002, Masle *et al.* 2005) et affectant la variation de plusieurs caractères (Ebert *et al.* 1993), 3) un déséquilibre de liaison qui serait dû à d'autres facteurs que la liaison physique, notamment à un effet de dérive si des allèles à différents gènes sont associés par hasard dans certains génotypes de la population, entraînant des associations sans signification biologique (Gallais 1990), ou 4) de la sélection épistatique faisant intervenir plusieurs gènes pas forcément liés physiquement et plusieurs caractères (Shook et Johnson 1999, Wiehe et Slatkin 1998). Dans les cas 2), 3) et 4), la recombinaison et la sélection pourront permettre de rompre des corrélations génétiques défavorables, ce qui ne sera pas possible dans le cas de pléiotropie.

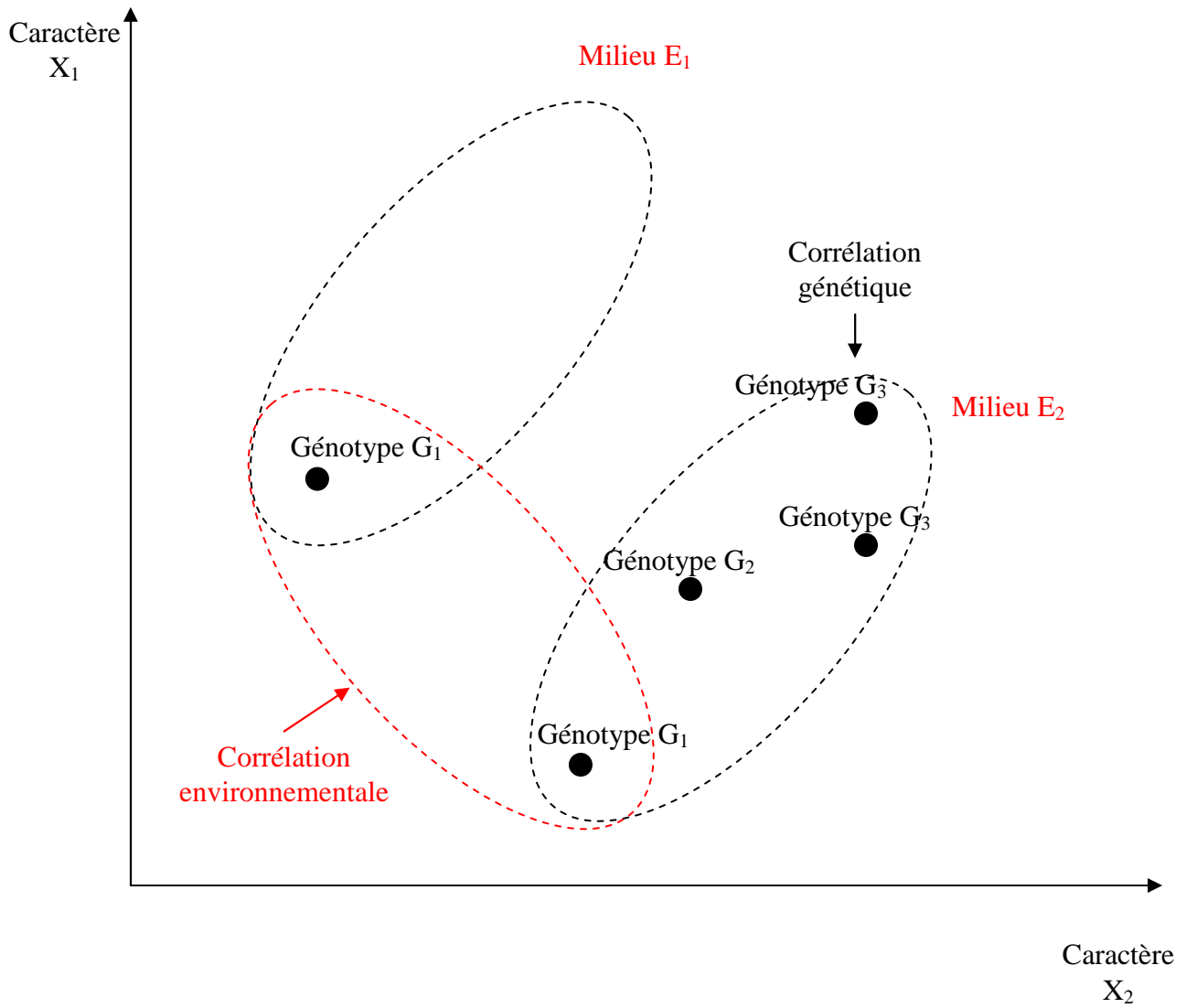
La corrélation résiduelle dans notre dispositif est basée sur la variation résiduelle comprenant un effet génétique intra famille et un effet environnemental. En faisant l'hypothèse que l'effet génétique non contrôlé est cependant peu important, la corrélation résiduelle pourrait donner une idée de la corrélation entre caractères lorsque le milieu varie pour un génotype donné, ce qui pourrait être interprété comme une corrélation due à la variation de l'expression des gènes selon le milieu (Gallais 1990) (Figure 41). Ce type de corrélation serait donc inhérent à la physiologie des génotypes ou des familles étudiées et ne pourrait pas être facilement modifiée, et correspondrait à une plasticité environnementale moyenne propres aux génotypes/familles d'une population.

Les corrélations génétiques et résiduelles entre deux caractères A et B ont été estimées à partir des matrices de variance – covariance, tel que :

$$r = \text{Cov}(A,B) / \sqrt{(\text{Var}(A) * \text{Var}(B))}$$

Ces corrélations ont été calculées uniquement lorsque les deux caractères présentait une variation significative pour le niveau de variation considéré. L'erreur standard de chaque corrélation est obtenue selon l'approche générale de Tallis (1959), qui est robuste lorsque le nombre de descendant est élevé (Visscher 1998) :

Figure 41. Illustration des corrélations génétiques et des corrélations environnementales entre caractères (d'après Gallais 1990).



$$S.E._r \approx \sqrt{\frac{(1-r^2)^2}{df}}$$

avec r , la corrélation (soit r_{pop} , r_{fam} ou r_{err}) et df , le nombre de degrés de liberté du niveau de variation considéré. Un intervalle de confiance à 95% a été calculé pour tester si les corrélations étaient significativement différentes de zéro, selon la méthode analytique basée sur le développement des carrés et produits moyens de Mode et Robinson (1959) (programme SAS-IML, com. pers. Garnier-Géré).

B.4.4. Etude de la structuration entre populations pour les caractères quantitatifs

Afin d'évaluer la différenciation entre populations pour des caractères phénotypiques, et déceler d'éventuelles signatures d'adaptation locale, l'indice Q_{ST} a été estimé à la fois sur l'ensemble des populations et par paires de populations. Cet indice est défini comme le ratio entre la variance entre populations (V_{pop}) et la variance totale estimée par $V_{pop} + V_A$, qui représente la variance génétique additive moyenne intra-population (Spitze, 1983) :

$$Q_{ST} = \frac{V_{pop}}{V_{pop} + 2(V_A)}$$

B.4.5. Analyse intra population

Afin de comparer la variation génétique au sein de chaque population, les différentes statistiques présentées (CV_p , V_A , h^2_S et corrélations) ont également été estimées dans chaque population avec un modèle simple déclarant un effet famille en utilisant les données corrigées au préalable par l'effet bloc estimé sur le modèle complet. Dans chaque population, nous avons approfondi l'étude de la variation et covariation de trois caractères représentatifs *a priori* de la performance des populations : hauteur ($HT03$), biomasse (TB) et efficience d'utilisation en eau ($\delta^{13}C$).

C. Problèmes méthodologiques et résultats préliminaires

C.1. Problème de l'estimation des variances entre familles pour le $\delta^{13}C$

Pour les données de composition isotopique ($\delta^{13}C$), certaines familles comportaient moins de 7 répétitions (arbres morts et mesures défectueuses). Par exemple, la famille 9 de la population 15 n'était représentée que par un seul arbre, ne permettant pas d'avoir une estimation de la moyenne familiale suffisamment précise. Une analyse en composantes principales (ACP) sur les moyennes familiales a d'ailleurs révélé que cette mesure s'écartait

très nettement des autres moyennes familiales de la population 15 et des autres populations (Figure 42). Elle a donc été exclue de l'analyse. Cet exemple soulève la question du nombre de répétitions au minimum nécessaires pour estimer ces moyennes familiales avec suffisamment de précision, et indirectement pour estimer la variance entre familles au sein des populations, ainsi que les indices qui en découlent comme l'héritabilité. Ce problème pourrait notamment être important pour la population 22, dans laquelle plus de la moitié des familles étaient représentée par moins de 5 répétitions.

Afin de répondre à cette question, nous avons exploré de façon très empirique l'impact du nombre de mesures sur la variation de l'estimateur d'héritabilité, en utilisant les données de 4 populations (29, 40, 42 et 44) qui comprenaient très peu de données manquantes (majorité de familles ayant 7 répétitions). Pour chaque population, cinq jeux de données ont été créés en conservant successivement 3, 4, 5, 6 et 7 répétitions choisies de façon aléatoire dans le lot de 7. Pour chaque jeu de données, la valeur d'héritabilité a été estimée sur les trois principaux caractères (*HT03*, *BT*, et $\delta^{13}C$), et l'évolution de l'héritabilité a été représentée en fonction du nombre de répétitions (Figure 43). Globalement, on observe une stabilisation des valeurs d'héritabilité à partir de 5 répétitions (valeurs variant au maximum de 0.3 point entre 5 et 7 répétitions, à l'exception de la population 29 pour *HT03*). Cette tendance a été vérifiée également sur d'autres populations représentatives de l'ensemble de la distribution. Sur cette base, nous avons décidé d'analyser par la suite uniquement les familles comprenant au moins 5 répétitions.

Un deuxième problème méthodologique était lié 1) au nombre de familles observées dans chaque population, limité seulement par le plan d'échantillonnage initial dans les populations d'origines (et variant de 20 à 30 par population), et 2) au nombre de répétitions par famille dont dépend la précision de l'estimation des moyennes familiales. Pour ce dernier, il a fallu réaliser un compromis avec le nombre de populations étudiées, car le nombre maximum de mesures était limité à 3200 dû à une contrainte économique (voir partie B).

Une analyse préliminaire indiquait un effet famille non significatif pour certaines populations, rendant difficile l'estimation d'une variance génétique entre familles et les indices qui en découlent (héritabilité ou corrélations génétiques). Donc afin d'augmenter la puissance du test de l'effet famille, et dans la mesure où certaines populations correspondaient à des échantillons très proches sur le plan géographique, nous avons essayé de voir si certains regroupements étaient justifiés.

Figure 42. Analyse en Composantes Principales sur les moyennes familiales, basée sur les 7 caractères *HT02*, *HT03*, *DC*, *PolyT*, *TB*, *NB* et $\delta^{13}C$. Les axes 1 et 2 expliquent 75,8% et 17,4% respectivement de la variation totale. Une famille de la population Pinia (159) n'est estimée que par une seule répétition et s'écarte du nuage de points des moyennes familiales.

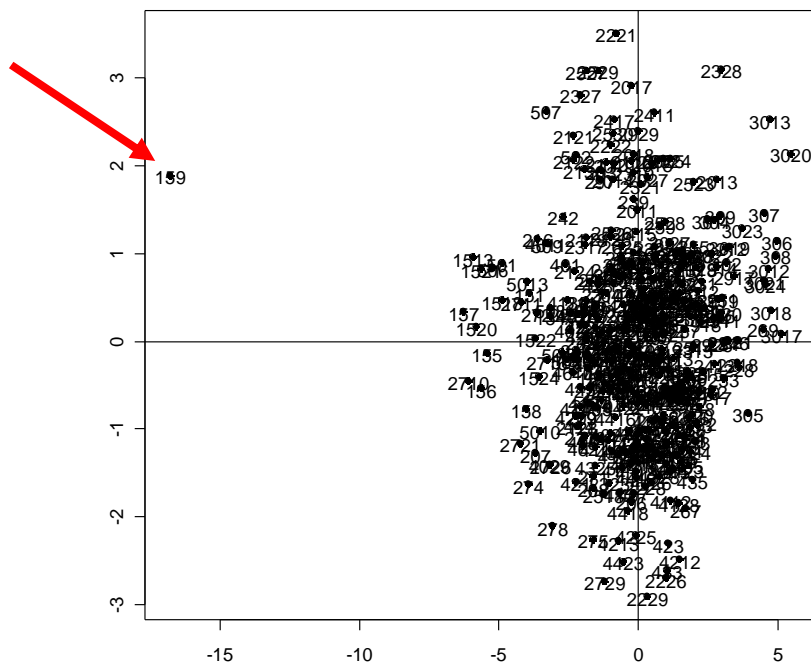
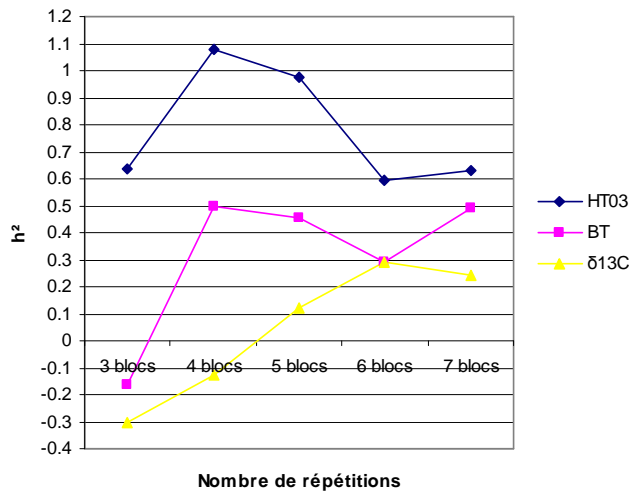
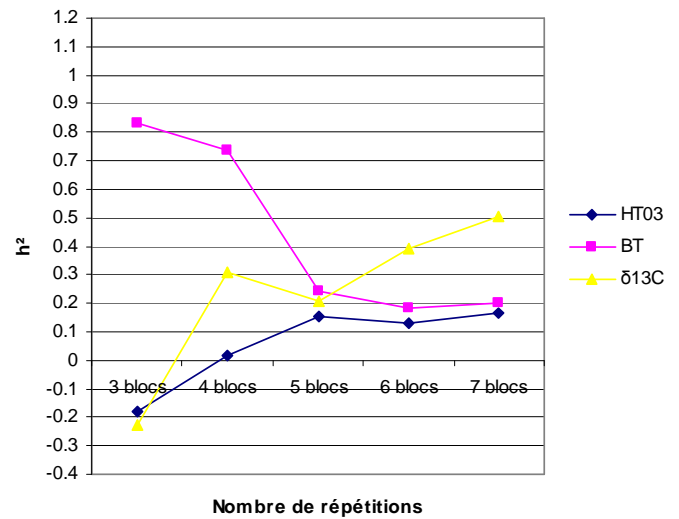


Figure 43. Evolution des valeurs d'héritabilité au sens strict (h^2) de 3 caractères (*HT03*, *BT* et $\delta^{13}C$) en fonction du nombre de répétitions (3 à 7) pour un sous échantillon des populations 29 (A), 40 (B), 42 (C) et 44 (D) comprenant 19 familles dans chaque population.

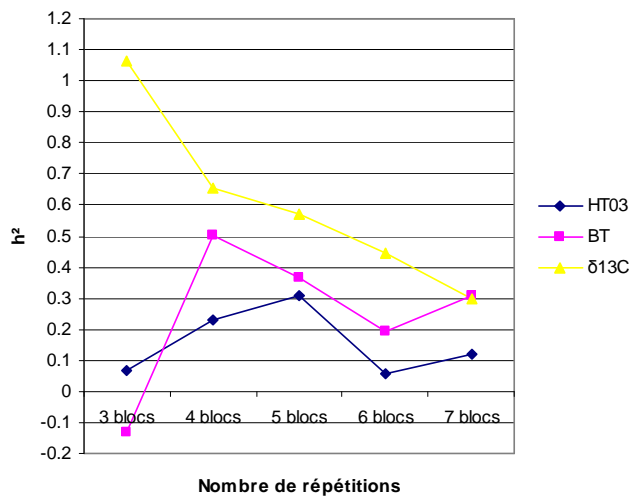
A)



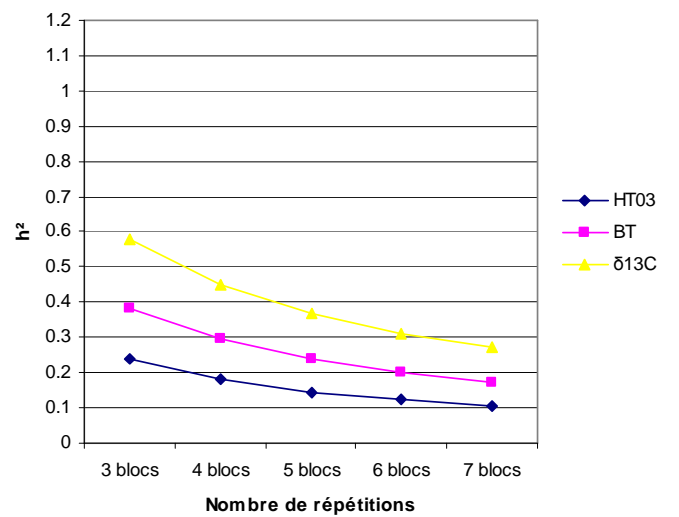
B)



C)



D)

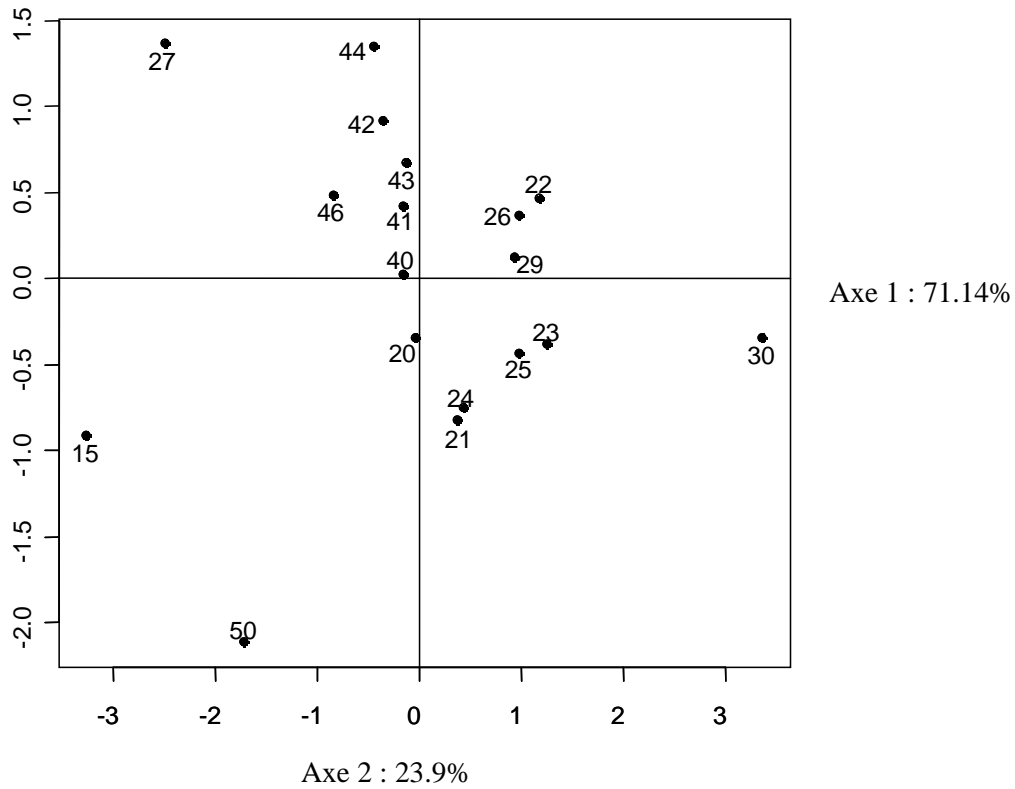


Dans un premier temps, nous avons réalisé une ACP sur les moyennes des populations des principaux caractères ($HT03$, BT et $\delta^{13}C$) (Figure 44). Les deux premières composantes principales expliquaient plus de 95% de la variation totale. L'axe 1 expliquant 71.14% de la variation totale était plutôt lié au caractère $HT03$, tandis que l'axe 2 (23.9% de la variation totale) était représentatif à la fois des caractères de biomasse (BT) et d'efficacité d'utilisation en eau ($\delta^{13}C$) (Figure 44B). Sur le plan principal, 4 populations (15, 27, 30 et 50) s'écartaient du nuage de points et ne pouvaient *a priori* pas être regroupées (Figure 44A). A l'inverse, les moyennes entre les populations de la côte atlantique française semblaient peu différentes, du moins sur la base de l'axe 1 (plutôt représentatif de la hauteur $HT03$). Sur la base de cette ACP, les populations 21 et 24 semblaient très proches (et proches aussi de la population 20 sur la base de l'axe 1), ainsi que les populations 23 et 25, et les populations 22 et 26 (et 29 sur la base de l'axe 1).

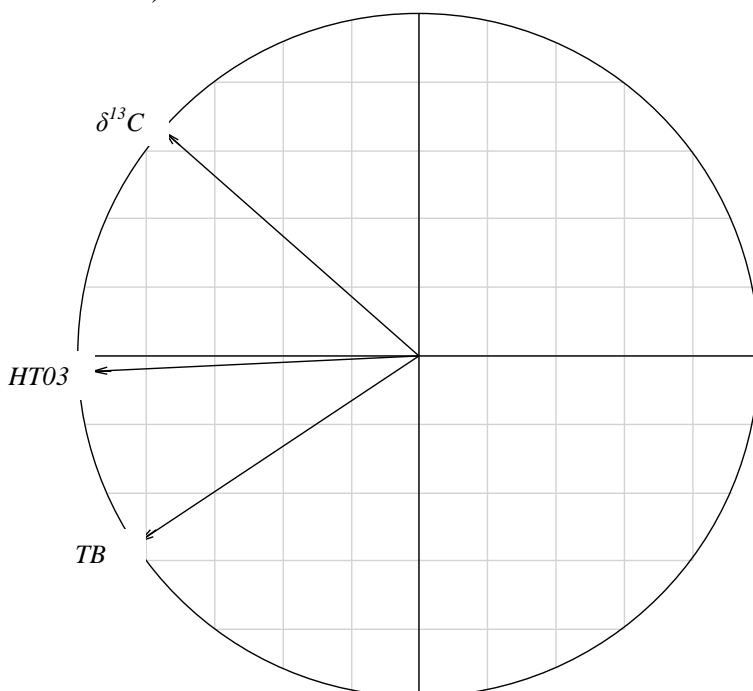
Nous avons ensuite effectué un test de Student de comparaison des moyennes entre paires de populations pour évaluer si les regroupements de populations étaient justifiés (Annexe 5). D'après ces tests, les populations 40, 41, 42 et 43 n'étaient pas significativement différentes les unes des autres pour les trois caractères $HT03$, TB et $\delta^{13}C$, sauf pour la paire de populations 40 et 42 pour TB et $\delta^{13}C$. Il a donc été décidé de regrouper d'une part, les populations 40 et 41 (groupe noté 4041), et d'autre part, les populations 42 et 43 (groupe noté 4243), d'autant plus que ces groupes concernent des populations proches géographiquement, situées dans la région Aquitaine (voir Figure 7 Chapitre II), et dont les conditions de température et de précipitations sont très similaires (Tableau 16). Un troisième regroupement constitué des populations 21 et 24 (noté 2124), était également justifié sur la base de ces tests, et du fait que ces populations soient toutes deux situées dans le même massif forestier (voir Figure 7 Chapitre II) au centre de l'Espagne avec des conditions climatiques proches (Tableau 16). La population 20 bien que proche n'a pas été regroupée avec celles-ci car on observait une différence entre les moyennes des populations 20 et 21 à la limite de la significativité à 5% ($p=0.07$) pour le $\delta^{13}C$ (Annexe 5). De plus les données de $\delta^{13}C$ pour cette population ne provenaient que du lot de mesures USA. Un dernier groupe aurait pu être constitué des populations 26 et 29 mais n'a pas été retenu compte tenu de la distance géographique les séparant (voir Figure 7 Chapitre II) et de conditions climatiques relativement différentes (Tableau 16). Enfin pour les populations 22 et 26 d'une part, et 23 et 25, d'autre part, même si les moyennes des populations n'étaient pas significativement différentes pour $HT03$ et $\delta^{13}C$, elles l'étaient pour TB ($p=0.04$ et 0.02 respectivement), excluant le regroupement.

Figure 44. Représentation graphique de l'analyse en composantes principales à partir des moyennes des populations, sur les 3 caractères *HT03*, *TB* et $\delta^{13}C$. Les axes principaux 1 et 2 expliquent 71.14% et 23.9% respectivement de la variation totale (A). Cercle de corrélation des caractères sur les axes 1 et 2 (B).

A)



B)



Par la suite, au sein de chacun de ces 3 groupes de populations, l'absence d'effet population significatif a été vérifiée. De plus, il a été vérifié que la dispersion entre familles n'était pas augmentée uniquement à cause de familles situées à l'extrémité inférieure ou supérieure de la distribution de chaque caractère, et qui appartenaient à des populations différentes. Pour cela, un test de comparaison des moyennes familiales de Student-Newman-Keuls (SNK) a été réalisé (Annexe 6). Ce test permet un regroupement de familles de proche en proche, et dans le cas de plusieurs groupes formés, l'appartenance des familles les plus extrêmes à non pas une seule population mais à un mélange de plusieurs populations a été confirmée.

C.2. Biais engendré par les analyses isotopiques

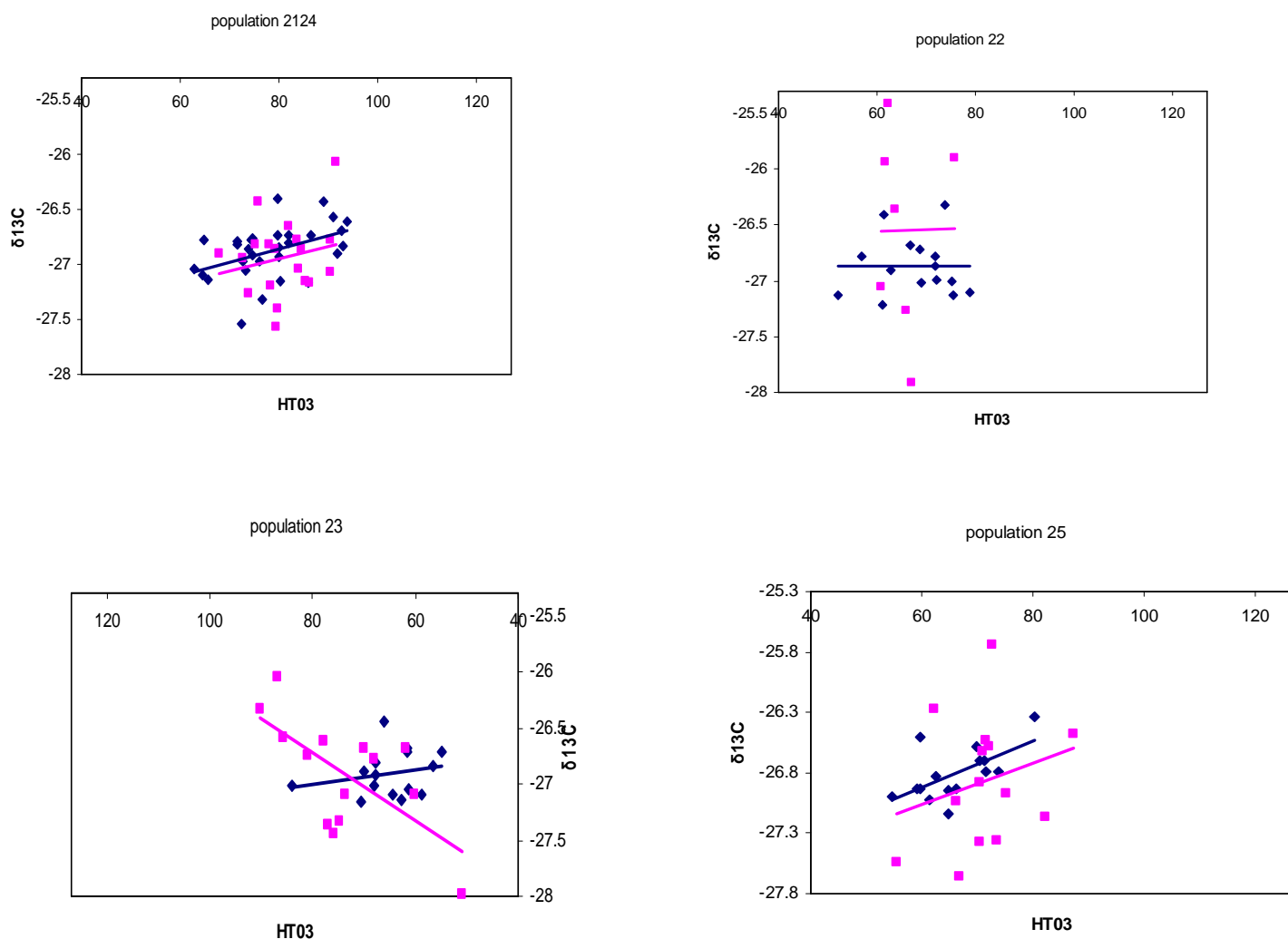
Une attention particulière a été portée sur les deux lots différents d'analyse du $\delta^{13}C$: le lot des 2300 mesures réalisées en France, noté *anFR*, et le lot des 900 mesures réalisées aux USA, noté *anUSA*. L'effet du lot d'analyse a été testé globalement sur les données incluant les 8 populations où l'on avait les deux types de mesures (la population marocaine ainsi que toutes les populations espagnoles à l'exception des populations 20 et 27). Cet effet était significatif mais expliquait moins de 1% de la variation totale en somme des carrés des écarts. L'effet lot d'analyse a également été testé au sein de chaque population comme un effet moyen. Pour la population 22 seulement, l'effet était à la limite de la significativité à 5% ($\text{Proba} < 0.08$) et expliquait 1,35% de la variation totale.

Nous avons également regardé l'effet possible du lot d'analyse sur la variance entre familles (σ^2_{fam}) et la variance résiduelle (σ^2_{err}). Pour cela, nous avons estimé ces variances séparément dans les deux jeux de données, et regardé si elles augmentaient ou diminuaient lorsque l'on passait du lot d'analyse français au lot d'analyse USA (Tableau 18, Figure 45). Ainsi σ^2_{fam} augmente dans la majorité des populations, et particulièrement pour les populations 21, 22, 23 et 25 (plus de 20 fois plus élevée). σ^2_{err} augmente (pour les populations 21, 22 et 23) ou diminue (pour les populations 24, 25, 26, 29 et 30) mais en moindre proportion. La conséquence générale de cette augmentation de la variance entre familles pourrait s'expliquer par le fait que les mesures du lot d'analyses *anUSA* ont été réalisées avec une quantité de poudre d'aiguilles deux fois moins importante, et seraient moins précises que celles réalisées avec plus de poudre (cas du lot *anFR*), à cause d'une dérive possible dans les mesures ou d'erreurs expérimentales plus importantes (O. Brendel, com. pers.).

Tableau 18. Tableau résumant l'augmentation (↑) ou la diminution (↓) des variances entre familles (σ^2_{fam}) et des variances résiduelles (σ^2_{err}) pour le $\delta^{13}C$, lorsque l'on passe du lot d'analyses français au lot d'analyses USA. Les valeurs entre parenthèses indiquent les facteurs d'augmentation ou de diminution.

Population	σ^2_{fam}	σ^2_{err}
21	↑ (>20x)	↑ (x1,7)
22	↑ (>20x)	↑ (x1,6)
23	↑ (>20x)	↑ (x1,4)
24	↑ (x2,5)	↓ (x2,4)
25	↑ (>20x)	↓ (x1,4)
26	↑ (x1,9)	↓ (x1,2)
29	↑ (1,9)	↓ (x1,2)
30	↓ (x2)	↓ (x1,1)

Figure 45. Quelques graphiques de régression (et courbes de régression associées) des moyennes familiales de $\delta^{13}C$ en fonction de *HT03* pour chaque lot d'analyse, *anFR* (bleu) et *anUSA* (rose). Les autres graphiques de régression sont présentés en Annexe 10.



En conclusion de ces résultats préliminaires, et dans la suite des analyses, les estimations ont été réalisées en excluant les familles avec moins de 5 répétitions (au total, 17 familles, soit 60 données ont été exclues de l'analyse, représentant environ 2% du nombre total de données), en regroupant certaines populations (4041, 4243 et 2124), et en prenant également en compte l'effet du lot d'analyse sur les différents estimateurs de variances génétique et résiduelle (voir résultats).

D. Résultats

Les résultats concernant les moyennes phénotypiques des différents caractères ont été considérés sur l'ensemble des populations ainsi qu'au sein de chaque population prise séparément. Pour l'estimation des héritabilités, des corrélations et des indices de différenciation génétique des caractères quantitatifs, nous avons décidé de considérer des groupes de populations quand cela était possible (voir partie C.1. ci-dessus) pour pouvoir augmenter la précision dans l'estimation de la variance entre familles qui est à la base de l'estimation de ces indices.

D.1. Une variation phénotypique répartie différemment selon les caractères entre effets génétiques et environnementaux

Un premier groupe de caractères (hauteurs (*HT02* et *HT03*) et biomasses (*TB*, *NB*)) montraient des coefficients de variation phénotypiques élevés, variant de 27% à 46%, alors qu'ils étaient plus faibles pour les caractères de diamètre (*DC*) et de polycyclisme (*PolyT*) (CV_p entre 11 et 22%) (Tableau 19). Cependant, pour le caractère *PolyT*, le faible CV_p s'explique par le fait que le polycyclisme est un caractère discret avec peu de possibilités (1, 2, 3, 4 ou 5).

Les résultats d'ANOVA sur l'ensemble des données indiquent que les effets population et famille sont significatifs pour tous les caractères (Tableau 20 et Annexe 9). L'effet bloc est également significatif, et explique entre 5.8 % et 12.8% de la variation totale en sommes de carrés des écarts totaux (SCEt), indiquant un relativement bon contrôle d'une partie de la variation environnementale, particulièrement pour les hauteurs (*HT02* et *HT03*) et le $\delta^{13}C$, et dans une moindre proportion pour les caractères de biomasse et le polycyclisme (R^2 aux alentours de 3%). Les observations de terrain montrent également que cette variation environnementale importante est en partie due à la différence entre les deux parcelles (Figure 38). Le modèle d'ANOVA explique entre 30% et 60% de la SCEt pour chacun des caractères.

Tableau 19. Moyenne, valeurs minimum et maximum, et coefficient de variation pour les 7 caractères *HT02*, *HT03*, *DC*, *PolyT*, *TB*, *NB* et $\delta^{13}C$.

Caractère	Moyenne	Ecart-type	Minimum	Maximum	CV_p (%)
<i>HT02</i>	30.99 cm	11.12 cm	6.00 cm	85.00 cm	36
<i>HT03</i>	78.42 cm	21.40 cm	16.00 cm	176.00 cm	27
<i>DC</i>	23.37 mm	5.07 mm	6.60 mm	42.94 mm	22
<i>PolyT</i>	1.65	0.19	1.22	2.35	11
<i>TB</i>	313.72 gr	143.71 gr	22.50 gr	999.30 gr	46
<i>NB</i>	183.58 gr	77.59 gr	14.30 gr	621.00 gr	42
$\delta^{13}C$	-26.69	0.69	-28.49	-24.50	-

Tableau 20. Coefficient de détermination de chaque effet déclaré dans le modèle d'ANOVA pour les 7 caractères *HT02*, *HT03*, *DC*, *PolyT*, *TB*, *NB* et $\delta^{13}C$. Tous les effets sont significatifs à 1 %.

Caractère	Sources de variation ^(a)			R^2 ^(b)
	Pop	Fam(pop)	Bloc	
<i>HT02</i>	28.26	16.60	12.80	58.10
<i>HT03</i>	32.85	15.59	5.84	56.04
<i>DC</i>	14.41	20.44	2.49	37.73
<i>PolyT</i>	11.79	15.44	3.06	31.61
<i>TB</i>	16.55	18.96	3.20	40.18
<i>NB</i>	15.36	18.95	2.87	39.70
$\delta^{13}C$	8.57	24.25	7.28	40.10

^(a) Proportion de la variation totale expliquée (en % de somme des carrés de type 1) pour chaque effet.

^(b) Coefficient de détermination du modèle.

La variation résiduelle d'environ 40 à 70% selon les caractères peut indiquer à la fois une sensibilité importante à la variation micro-environnementale et une composante génétique intra-famille non négligeable, notamment pour le caractère *PolyT* (avec plus de 68% de variation résiduelle). Quant à la variation contrôlée d'origine génétique, elle est répartie différemment selon les caractères. Pour les caractères de hauteur, elle est expliquée de façon plus importante par l'effet population (28 à 33% de la variation totale, représentant environ le double du pourcentage de variation totale expliquée par l'effet famille). Pour les caractères de biomasse et le polycyclisme, cette variation est répartie de façon équivalente entre populations et familles (chaque effet expliquant entre 15 et 19% de la variation totale pour les biomasses et 12% à 15% pour *PolyT*). Enfin, pour le $\delta^{13}C$, cette variation est davantage expliquée par la variation entre familles au sein des populations (>24% de la variation totale) (Tableau 20).

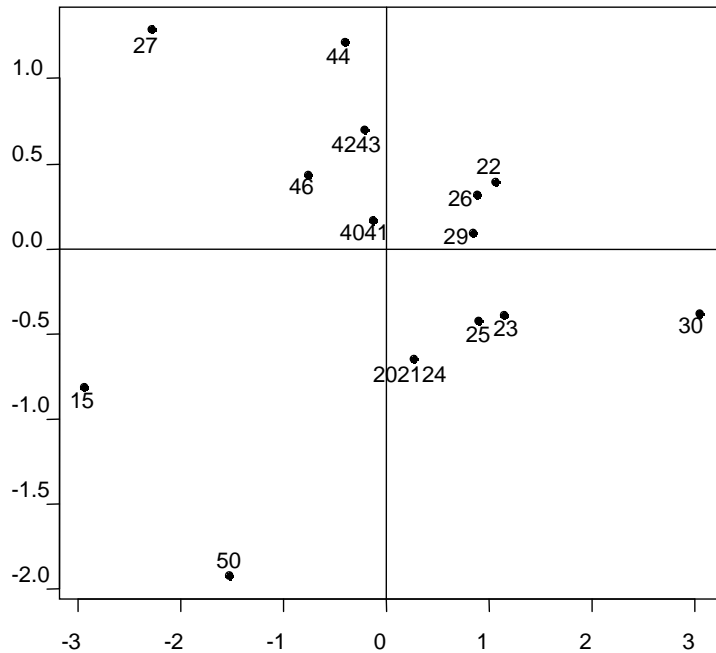
D.2. Comparaison des moyennes phénotypiques entre populations

D.2.1. Des populations de l'extrémité de l'aire de distribution divergentes pour l'efficience d'utilisation en eau et la hauteur.

La large variation observée sur l'ensemble des données est illustrée par la comparaison des moyennes de chaque population pour les 3 caractères *HT03*, *TB* et $\delta^{13}C$ (Figures 46 et 47), l'axe 1 de la figure 46 étant représentatif de la quasi-totalité de la variation de hauteur entre populations. La population Corse Pinia (15) présente des valeurs de hauteur (*HT03*) et de biomasse (*TB*) parmi les plus élevées, ainsi qu'une valeur de $\delta^{13}C$ parmi les plus importantes (-26.49 par rapport à une moyenne de -26.67) avec les populations 44, 46 et 27, faisant d'elle l'une des populations les plus performantes si l'on combine hauteur, biomasse et efficience d'utilisation en eau au sein de l'essai. A l'inverse, la population marocaine Tamrabta (30) montrent les valeurs les plus faibles pour ces caractères (Figures 47). Cette dernière fait partie de celles dont la température moyenne du lieu d'origine est maximale (17.2°C, Tableau 16). Les conditions climatiques du lieu d'origine de cette population *a priori* différentes de celles du milieu dans lequel a été implanté le test pourraient expliquer en partie le résultat observé. Entre ces deux extrêmes, différents comportements sont observés. La population tunisienne Tabarka (50) provenant d'un climat chaud mais avec un niveau de précipitations plus élevé que Tamrabta, et comparable à celui du milieu de test (Tableau 16), présente des valeurs de $\delta^{13}C$ et de *HT03* intermédiaires, mais une biomasse (*TB*) élevée (Figures 47), due à la fois à des biomasses de bois et d'aiguilles élevées.

Figure 46. Représentation graphique de l'analyse en composantes principales à partir des moyennes des populations et groupes de populations, sur les 3 caractères *HT03*, *TB* et $\delta^{13}C$. Les axes principaux 1 et 2 expliquent 74.03% et 22.3% respectivement de la variation totale (A). Cercle de corrélation des caractères sur les axes 1 et 2 (B).

A)



B)

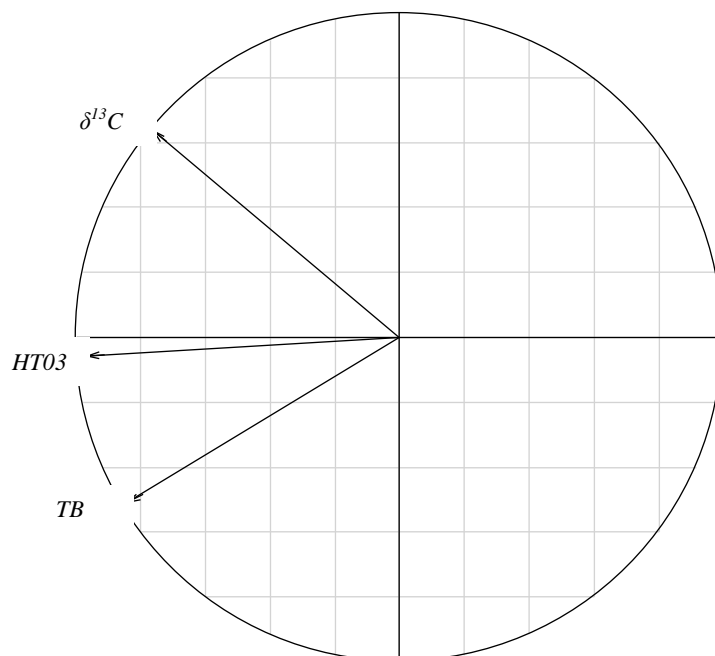
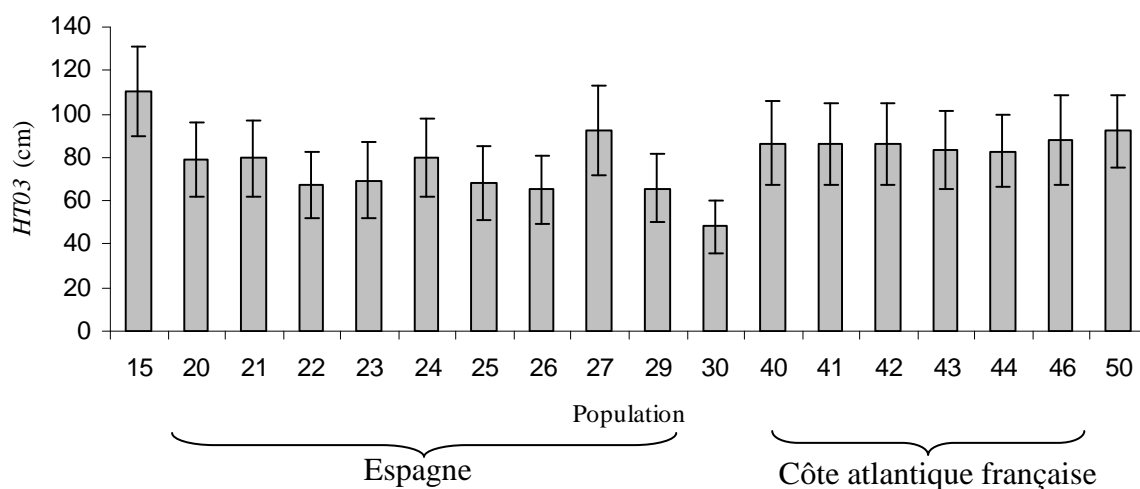
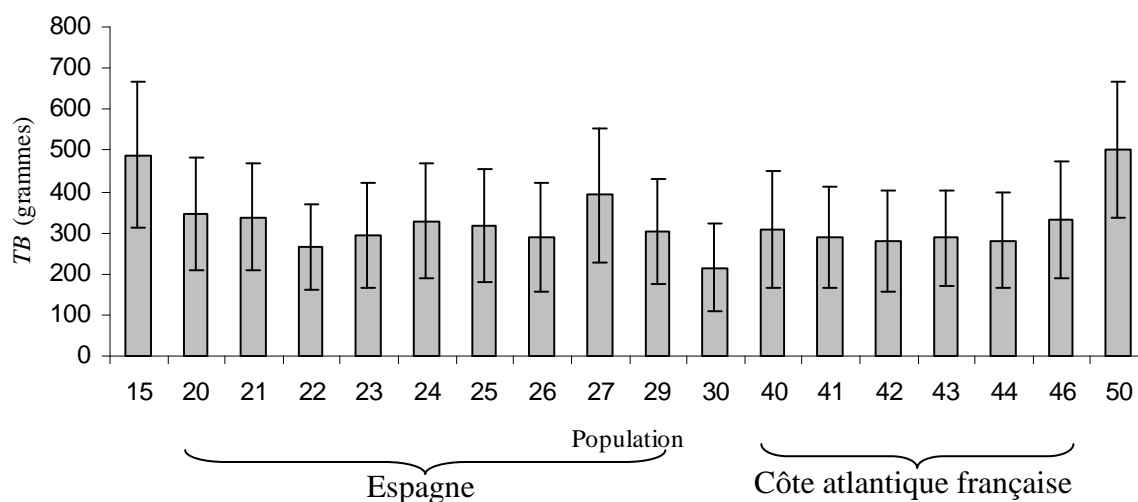


Figure 47. Moyennes (et erreurs standard) des 18 populations analysées pour *HTO3* (A), *TB* (B) et $\delta^{13}C$ (C).

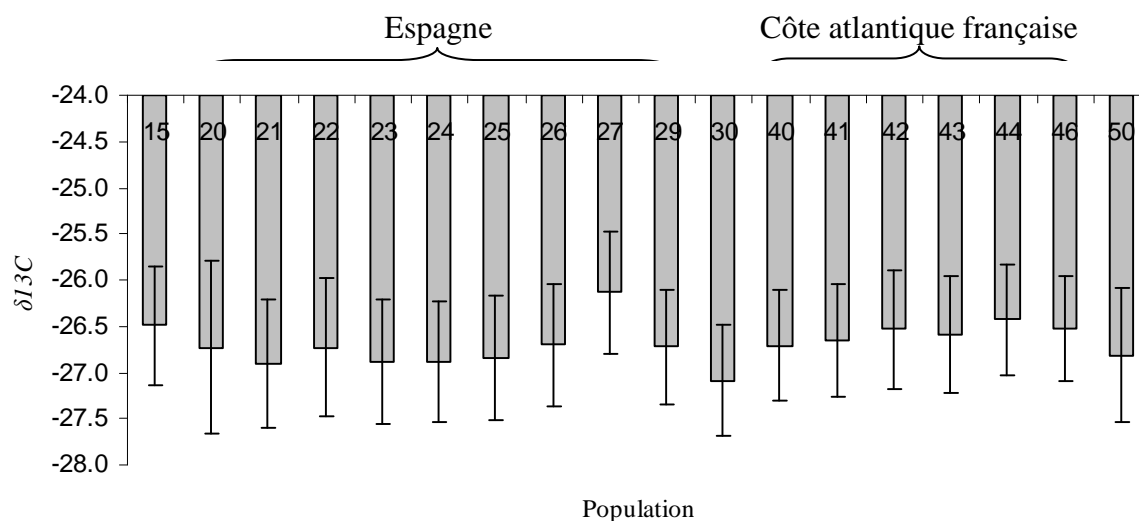
A)



B)



C)



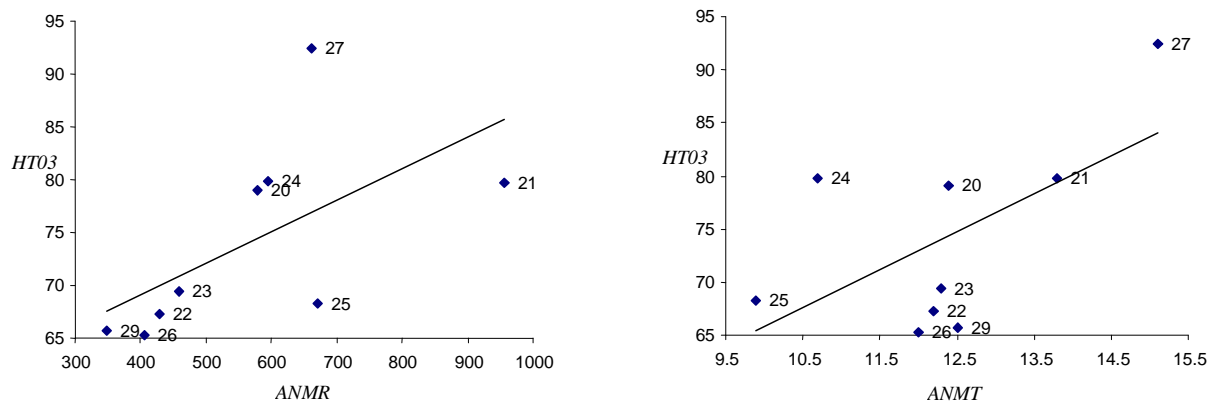
D.2.2. Des populations espagnoles moins efficaces que les populations françaises mais à biomasse similaire pour une hauteur plus faible

Parmi les populations espagnoles, une population (San Cipriano, 27) s'écarte très nettement des autres populations (Figure 46) avec la moyenne de $\delta^{13}C$ la plus élevée (-26.14 par rapport à une moyenne de -26.69 sur l'ensemble des populations espagnoles) (Figure 47A), et donc une EUE_{int} importante, ainsi que des moyennes de hauteur et de biomasse largement supérieures à celles des autres populations espagnoles. Cette population fait partie de celles dont les précipitations annuelles sont les plus élevées en Espagne (661 mm, Tableau 16), avec le niveau de précipitations du mois le plus sec le plus élevé (RDRYM=18 mm). Néanmoins, c'est aussi la population qui provient du milieu le plus chaud en moyenne après les populations du continent Africain. Si l'on exclut cette population particulière, on observe des corrélations significatives entre le cline de précipitations annuelles des lieux d'origine (ANMR) dans lesquels ont évolué les populations et les caractères $HT03$, TB et $\delta^{13}C$ (Figures 48). En effet, après exclusion de la population 27, la corrélation entre $\delta^{13}C$ et ANMR est négative ($r=-0.79$, $P<0.05$), tandis que les corrélations sont positives entre $HT03$ et ANMR et TB et ANMR ($r=0.84$, $P<0.01$ et $r=0.82$ $P<0.02$ respectivement). Au sein de ce groupe de populations, la corrélation est donc négative entre le $\delta^{13}C$ et les caractères de hauteur et biomasse ($r=-0.72$, $P<0.05$ et $r=-0.60$, $P>0.10$ respectivement). Donc il semblerait qu'au sein de la péninsule Ibérique, les populations ayant évoluées dans des régions plus sèches (par exemple, les populations Bayubas de Abajo (26), Oria (29) et Coca (22)) exprimeraient une meilleure efficacité d'utilisation en eau intrinsèque. Cependant, dans les conditions environnementales de notre essai qui peuvent être considérées comme moins stressantes d'après le niveau de précipitations annuelles, ces populations montrent à la fois une EUE_{int} plus faible et une hauteur réduite en comparaison des populations locales (populations 40 à 46) (Figure 47B). Néanmoins, leur biomasse est équivalente, voire légèrement supérieure (pour les populations 20 et 21) aux populations locales (Figure 47C) (également vrai pour le diamètre), ce qui voudrait dire que malgré une efficacité d'utilisation en eau plus faible dans l'essai (qui pourrait être due au changement de conditions drastique par rapport à leur milieu), ces populations espagnoles produisent autant de biomasse que les populations françaises.

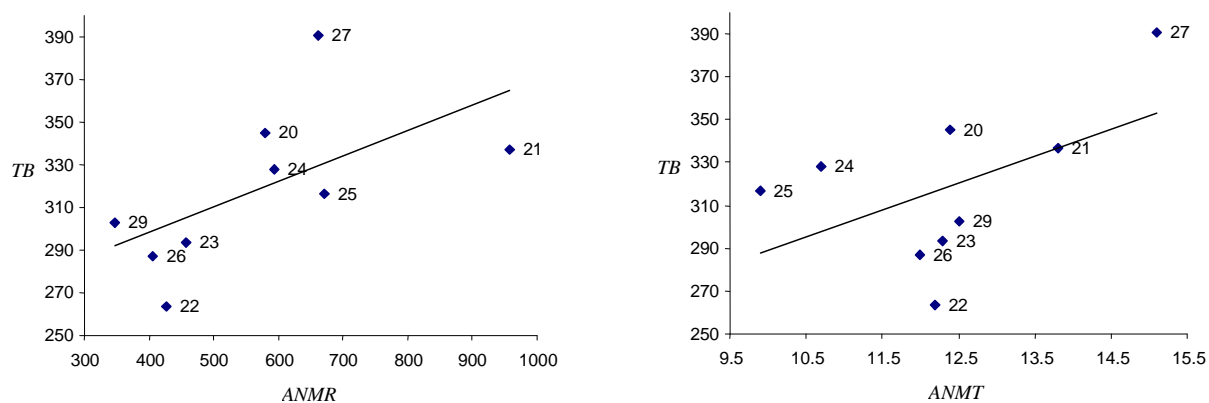
Les populations de la côte atlantique française présentent des moyennes de $\delta^{13}C$ du même ordre que celle de la population Corse Pinia (15), sauf pour les populations 40 et 41 à l'extrémité Sud de la côte française dont la valeur de $\delta^{13}C$ est davantage similaire à celles des populations espagnoles 22, 26 et 29 (Figure 47A). Les moyennes de hauteur pour les

Figure 48. Moyennes de $HT03$ (A), TB (B) et $\delta^{13}C$ (C) en fonction des précipitations moyennes annuelles ($ANMR$) et des températures moyennes annuelles ($ANMT$) des lieux d'origine pour les populations espagnoles. Les droites de régression ont été réalisées en excluant la population 27 pour le $\delta^{13}C$.

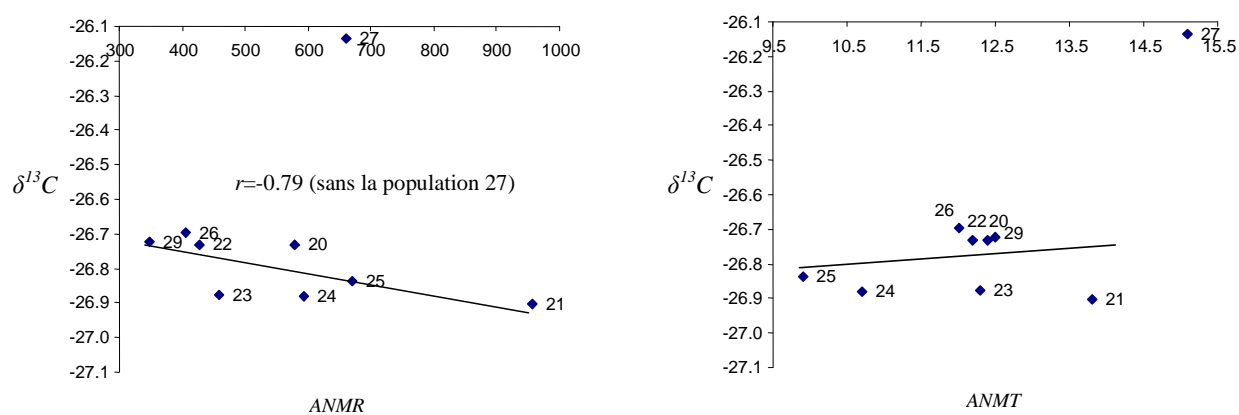
A)



B)



C)



populations françaises sont du même ordre de grandeur que celles de la population Tabarka et globalement plus élevées que celles des populations espagnoles (à l'exception de la population 27) (Figure 47B). A l'inverse, les moyennes de biomasse très similaires entre populations françaises (à l'exception d'une biomasse un peu plus élevée pour la population Pleucadeuc), restent néanmoins inférieures aux biomasses de certaines populations espagnoles (20, 21, 24 et 27) (Figure 47C). Il semble donc que dans le milieu de test, les populations françaises sont en moyenne plus efficaces pour l'utilisation de l'eau que les populations espagnoles (à l'exception de la population 27) avec une biomasse égale (ou légèrement inférieure) et une hauteur plus élevée. Ceci pourrait s'expliquer par le fait que l'environnement de test est proche de celui dans lequel elles ont évolué, tandis que les populations espagnoles proviennent dans l'ensemble de régions plus sèches et seraient moins bien adaptées (en termes d' EUE_{int}) que les populations locales.

D.3. Héritabilités

C.3.1. Des caractères à héritabilités globalement élevées toutes populations confondues

Sur l'ensemble des populations, les valeurs d'héritabilité sont élevées pour les hauteurs, diamètre et biomasse (entre 0.4 et 0.6). Pour les caractères *PolyT* et $\delta^{13}C$ (en considérant uniquement le lot d'analyse français), les héritabilités sont plus faibles (0.14 et 0.17 respectivement), et sont cohérentes avec celles estimées par Brendel et al. (2002) sur le $\delta^{13}C$ d'une population Aquitaine. Ces plus faibles valeurs d'héritabilité sont dues à des variances entre familles plus faibles en moyenne sur l'ensemble de populations, même si elles sont significatives (0.001 et 0.015 pour *PolyT* et $\delta^{13}C$ respectivement, c'est-à-dire 23 à 29 fois inférieures aux variances résiduelles, tandis que pour les autres caractères, cette variance est environ 8 fois plus faible que la variance résiduelle (Tableau 21)). L'héritabilité du $\delta^{13}C$ augmente beaucoup lorsque le lot d'analyse USA est inclus, ce qui est une conséquence du biais identifié en utilisant les données de ce lot d'analyse pour la mesure de $\delta^{13}C$ (voir C.2 et Tableau 3) et confirme qu'il vaut mieux contrôler le lot d'analyses pour l'interprétation des résultats.

D.3.2. Des valeurs d'héritabilité dans chaque population principalement liées à l'importance des variances génétiques additives

Au sein de chaque population, les valeurs d'héritabilité estimées sur les 3 caractères *HT03*, *TB* et $\delta^{13}C$, sont fortement corrélées aux valeurs de variance génétique additive V_A

Tableau 21. Estimation des variances entre populations, familles et résiduelles globalement sur l'ensemble des populations, et héritabilités au sens strict (h^2 , valeurs en gras significativement différentes de zéro), erreur standard des héritabilités (SE_{h^2}), et coefficients de différenciation quantitative entre populations (Q_{ST}) pour les 7 caractères. Pour le $\delta^{13}C$, les valeurs sont estimées sans le lot d'analyses USA (lot d'analyses français uniquement), et entre parenthèses à partir de toutes les données.

Effet	<i>HT02</i>	<i>HT03</i>	<i>DC</i>	<i>PolyT</i>	<i>TB</i>	<i>NB</i>	$\delta^{13}C$
Population	51.493	242.019	6.217	0.007	6954.000	1928.800	0.033 (0.049)
Famille	8.229	31.917	2.449	0.001	1726.400	502.452	0.015 (0.059)
Erreur	70.761	254.018	19.151	0.029	14906.100	4364.700	0.347 (0.349)
h^2	0.417	0.446	0.454	0.145	0.415	0.413	0.170 (0.579)
SE_{h^2}	0.063	0.064	0.064	0.051	0.063	0.063	0.064 (0.069)
Q_{ST}	0.439	0.487	0.241	0.458	0.335	0.324	0.212 (0.094)

(corrélation supérieure à 0.8), et beaucoup plus faiblement corrélées aux variances résiduelles V_r (corrélation entre 0.5 pour $HT03$ et 0.1 pour $\delta^{13}C$). Les niveaux d'héritabilités élevés sont donc plutôt associés à des variances génétiques additives élevées (et indirectement des variances entre familles élevées), tandis que les niveaux d'héritabilité faibles seraient causés par des variances additives faibles et non des variances résiduelles relativement plus importantes.

D.3.3. Variation de l'héritabilité entre populations

- Populations à l'extrémité de l'aire : des niveaux d'héritabilité négativement corrélées aux moyennes phénotypiques pour la hauteur et la biomasse

D'autre part, on observe qu'il existe globalement une corrélation négative entre les moyennes et les héritabilités pour les caractères de hauteur et de biomasse (Figure 50 A et B). Cette corrélation est essentiellement due aux populations corse Pinia (15) et tunisienne Tabarka (50) qui présentent des moyennes de hauteur et biomasse élevées tandis que les niveaux d'héritabilité sont nulles ou plus faibles que pour les autres populations. Ces populations dont la hauteur et la biomasse sont supérieures dans cet essai, présenteraient une variance génétique entre familles et donc une héritabilité non significatives. A l'inverse, pour la population marocaine Tamrabta (30), les moyennes de hauteur et de biomasse sont faibles pour des héritabilités plus élevées (Figure 50 A et B). Toutefois, cette corrélation négative entre moyennes et héritabilités n'est pas vérifiée pour le $\delta^{13}C$ sur l'ensemble des populations (Figure 50 C) et pour $HT03$ et TB au sein des groupes de populations espagnoles et françaises (Figure 50 A et B).

- Des caractères à héritabilité très variable parmi les populations espagnoles

Au sein de la péninsule Ibérique, on observe des populations avec des valeurs d'héritabilité pour $HT03$ de 0.27 (population 25) et 0.86 (population 23), et pour TB , de 0.068 (population 20) et 1 (population 23) (Figure 49 A et B), même si aucune héritabilité n'est significativement différente l'une de l'autre, probablement due à une forte imprécision dans les estimations. Les populations espagnoles présentant des héritabilités élevées, telles que les populations 20, 23 et 29, ont une variance entre familles supérieure, aucune famille extrême n'expliquant cette plus grande variance.

Pour le $\delta^{13}C$, les valeurs d'héritabilités sont directement dépendantes de la variance entre familles qui peut différer de façon importante entre lots d'analyse. Pour les populations

Figure 49. Héritabilités (et erreur standard) pour chaque population de *HT03* (A), *TB* (B) et $\delta^{13}C$ (C). Les signes * et # indiquent que l'effet famille est significatif à 5% et 10% respectivement.

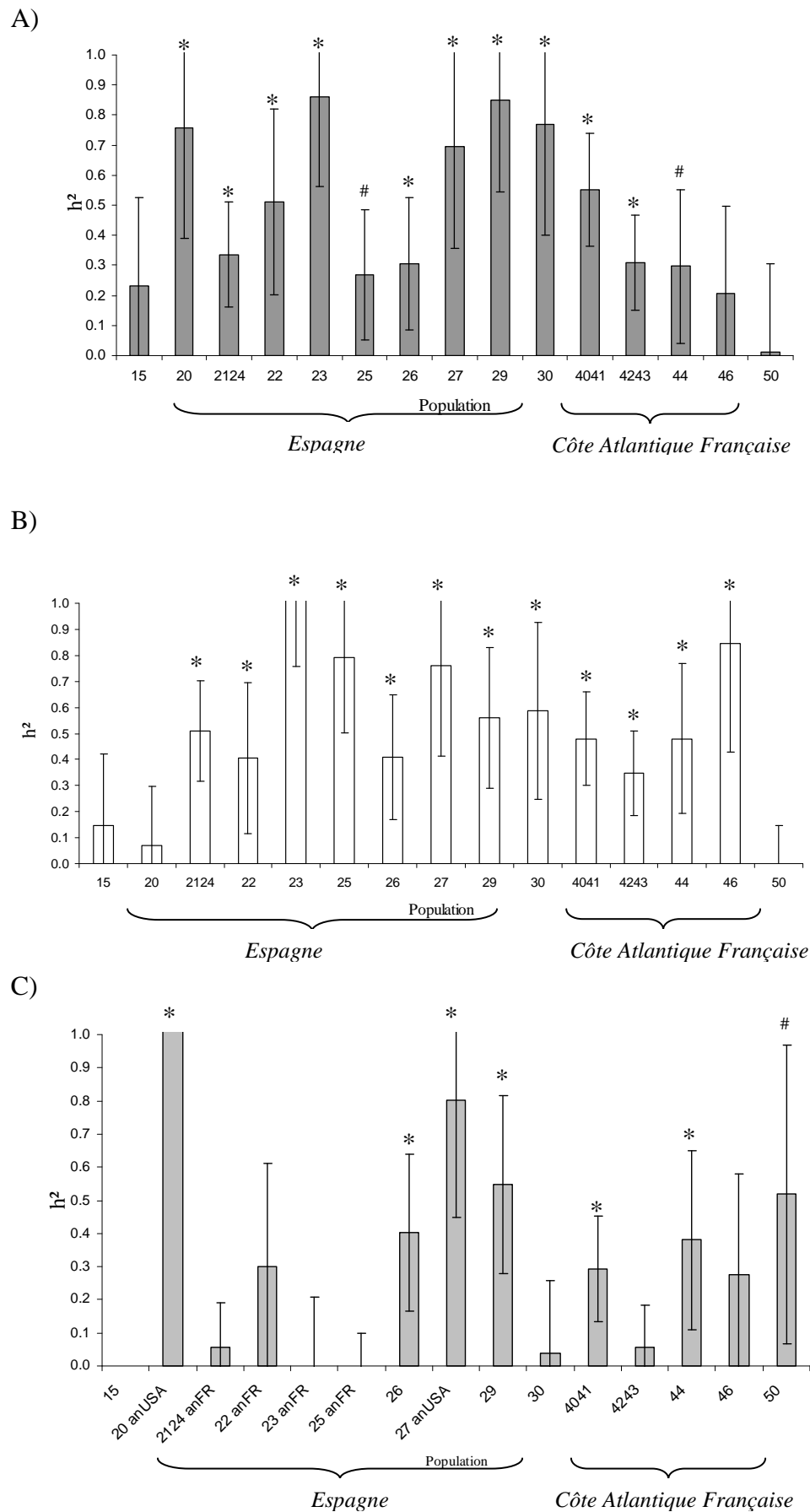
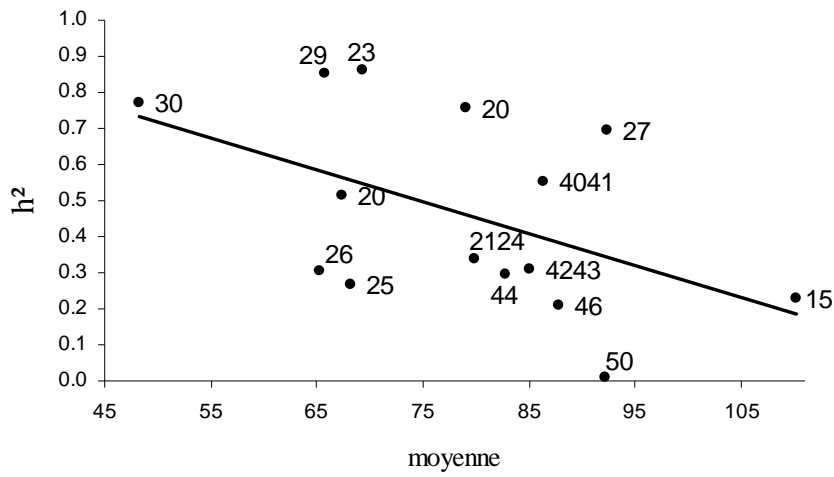
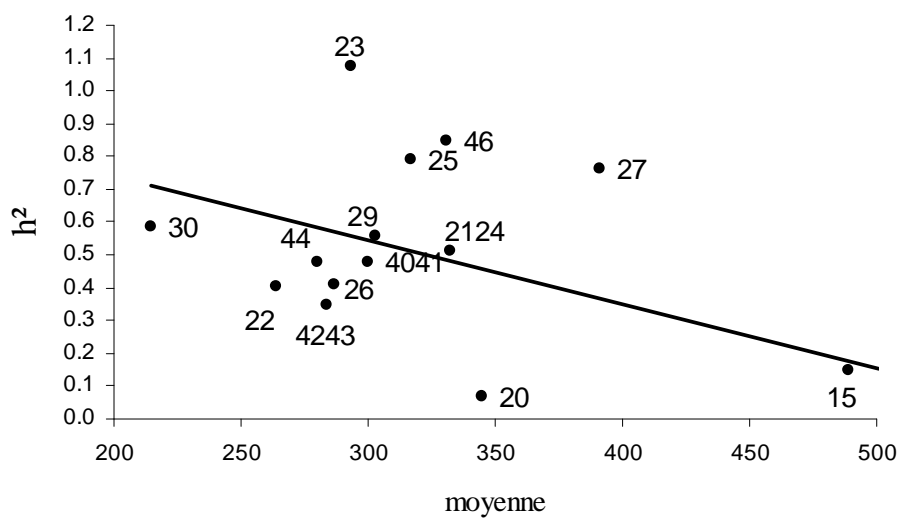


Figure 50. Héritabilités en fonction des moyennes par population pour les caractères *HT03* (A), *TB* (B) et $\delta^{13}C$ (C).

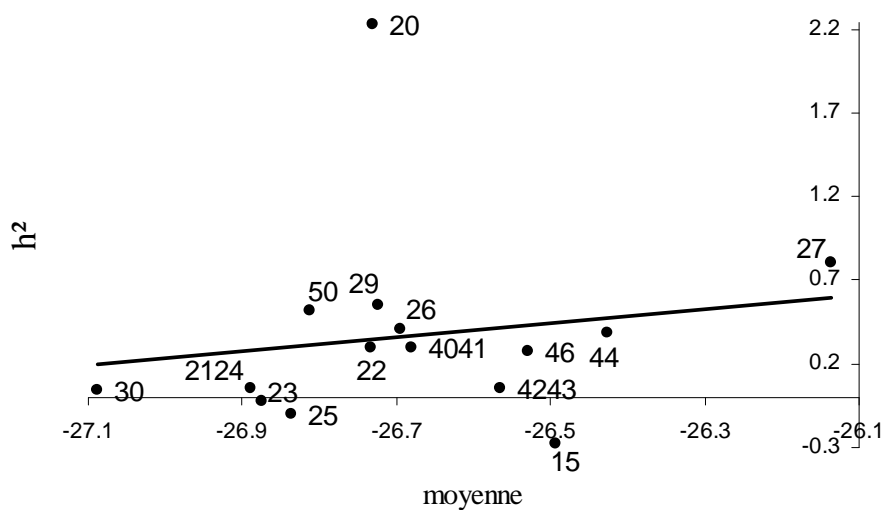
A)



B)



C)



21, 22, 23 et 25, nous avons vu que la variance entre famille était augmentée de plus de 20 fois lorsque l'on passait du lot d'analyse français au lot d'analyse USA (Tableau 18). Cette augmentation de la variance entraîne des héritabilités supérieures à 1 dans le lot d'analyse USA (Annexe 8). Nous avons donc décidé de ne considérer que le lot d'analyse français pour ces populations, ainsi que pour le groupe 2124. Pour les autres populations, la variance entre familles est relativement similaire dans les deux lots d'analyse (Tableau 18), et les héritabilités sont du même ordre de grandeur (Annexe 8). En ne considérant que les données où les lots d'analyses FR et USA pouvaient être comparés, une large variation dans les valeurs de h^2_s est néanmoins observée pour le $\delta^{13}C$ au sein des populations espagnoles (Figure 49C). Les héritabilités varient entre 0 (populations 23 et 25) et 0.55 (population 29). Les populations 20 et 27 pour lesquelles nous ne disposons que d'analyses USA pour le $\delta^{13}C$ ont l'héritabilité la plus élevée (1 et 0.80 respectivement) qui pourrait donc être en partie sur-estimée par une variance entre familles augmentée artificiellement par l'analyse USA. Cependant, on peut aussi se demander si ces populations présenteraient une héritabilité réellement plus élevée que pour les autres populations espagnoles, ce qui serait cohérent avec les héritabilités plus élevées observées pour la hauteur (Figure 49A) et pour la biomasse (Figure 49B uniquement pour la population 27).

- Des valeurs d'héritabilités moins variables au sein des populations françaises

Les populations françaises présentent une gamme de variation des valeurs d'héritabilité plus faible mais non négligeable pour *HT03* (h^2_s entre 0.21 et 0.55 pour les populations 46 et le groupe de populations 4041 respectivement) et *TB* (h^2_s entre 0.35 et 0.85 pour le groupe de populations 4243 et la population 46 respectivement (Figure 49 A et B). Comparativement aux autres populations de l'aire de distribution, ces populations ont en moyenne des valeurs d'héritabilité plus faibles pour ces deux caractères, avec néanmoins des moyennes plus élevées comparativement aux autres populations pour la hauteur (Figure 50A). Pour le $\delta^{13}C$, les héritabilités sont comparables à la gamme de variation de celles des autres populations (en considérant le lot d'analyses Français) (h^2_s entre 0.05 et 0.38 pour les populations 4243 et 44 respectivement), pour des moyennes plus élevées que les autres populations étudiées (Figure 50C).

D.4. Corrélations entre caractères

D.4.1. Des corrélations génétiques élevées entre caractères de hauteur et biomasse

- Corrélations génétiques positives entre hauteur et biomasse au niveau global

Les différents types de corrélations génétiques estimées sur l'ensemble des données, qu'elles soient au niveau des populations (r_{pop}) ou des familles (ces dernières correspondant aux corrélations génétiques additives) (r_{fam}) sont positives entre les caractères (Tableau 22, Figure 51). On retrouve les différents groupes de caractères, avec d'un côté ceux de biomasse (TB , NB et DC) qui ont des corrélations génétiques supérieures à 0.8 au niveau populations et familles, et ceux de hauteur ($HT02$ et $HT03$) avec des corrélations génétiques supérieures à 0.9 (Tableau 22, Figure 51). Ces deux groupes de caractères sont également fortement corrélés entre eux (r_{pop} et r_{fam} entre 0.5 et 0.9) (Tableau 22, Figure 51), ce qui est logique puisque toutes ces variables sont liées structurellement à la dimension et au niveau d'acquisitions de ressources de chaque arbre.

- Comparaison de la structure des corrélations génétiques entre populations

A l'échelle de chaque population et lorsque les effets familles sont significatifs, les r_{fam} estimées entre hauteur ($HT03$) et biomasse (TB) sont élevées, particulièrement pour les populations espagnoles 2124, 23, 27 et 29, les populations françaises 4243 et la population marocaine 30 ($r_{fam} > 0.7$) (Tableau 23). Seules les populations Bayubas de Abajo (26) ($r_{fam} = 0.16$) et le groupe de populations 4041 ($r_{fam} = 0$) présentent des r_{fam} inférieures à 0.5. Ces corrélations génétiques additives élevées entre hauteur et biomasse au sein d'une grande partie des populations, reflètent bien celle estimée sur l'ensemble des populations, ayant pour origine l'allométrie (c'est-à-dire les relations existant ici entre taille et volume qui conditionne la biomasse). Notons cependant le cas des populations 26 (Bayubas de Abajo), et du groupe 4041 (Aquitaine) pour lesquels la corrélation est nulle entre hauteur et biomasse (Tableau 23).

D.4.2. Corrélations génétiques entre efficacité d'utilisation en eau et hauteur ou biomasse

- Corrélations génétiques plus faibles avec le $\delta^{13}C$ au niveau global

Sur l'ensemble des données, les corrélations génétiques au niveau des populations (r_{pop}) entre le $\delta^{13}C$ et les autres caractères sont plus fortes au niveau des moyennes de

Tableau 22. Corrélations génétiques au niveau des populations (A), des familles (B) et résiduelles (C). Les valeurs en gras sont significativement différentes de zéro. Pour les corrélations entre $\delta^{13}C$ et les autres caractères, les valeurs sont estimées sans le lot d'analyses USA (lot d'analyses français uniquement), et entre parenthèses à partir de toutes les données.

A)

r_{pop}	HT02	HT03	DC	PolyT	TB	NB	$\delta^{13}C$
HT02							
HT03	0.964						
DC	0.809	0.699					
PolyT	0.850	0.909	0.592				
TB	0.747	0.684	0.930	0.645			
NB	0.624	0.551	0.876	0.555	0.983		
$\delta^{13}C$	0.61 (0.596)	0.74 (0.664)	0.25 (0.187)	0.66 (0.629)	0.32 (0.303)	0.26 (0.225)	

B)

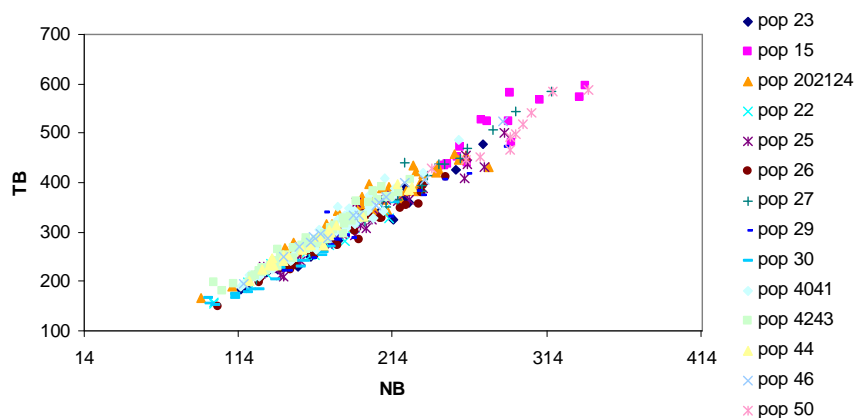
r_{fam}	HT02	HT03	DC	PolyT	TB	NB	$\delta^{13}C$
HT02							
HT03	0.946						
DC	0.795	0.796					
PolyT	0.446	0.533	0.252				
TB	0.713	0.701	0.933	0.204			
NB	0.615	0.600	0.870	0.215	0.976		
$\delta^{13}C$	0.23 (0.216)	0.34 (0.322)	0.06 (0.115)	0.63 (0.546)	0.05 (0.141)	-0.07 (0.109)	

C)

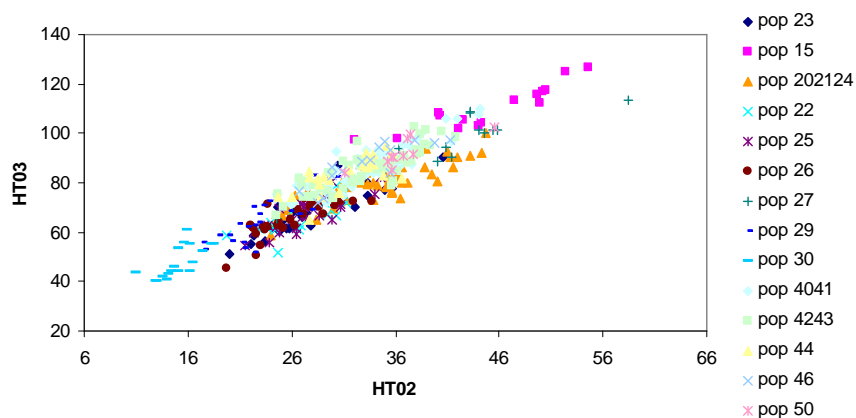
r_{err}	HT02	HT03	DC	PolyT	TB	NB	$\delta^{13}C$
HT02							
HT03	0.765						
DC	0.680	0.740					
PolyT	0.092	0.337	0.181				
TB	0.691	0.733	0.887	0.172			
NB	0.613	0.671	0.849	0.167	0.974		
$\delta^{13}C$	0.10 (0.103)	0.18 (0.188)	0.08 (0.076)	0.09 (0.108)	0.04 (0.053)	0.01 (0.042)	

Figure 51. Moyennes familiales par population de *TB* en fonction de celles de *NB* (A), de *HT03* en fonction de *HT02* (B), de *TB* en fonction de *HT03* (C), de *DC* en fonction de $\delta^{13}C$ (D), de *TB* en fonction de $\delta^{13}C$ (E).

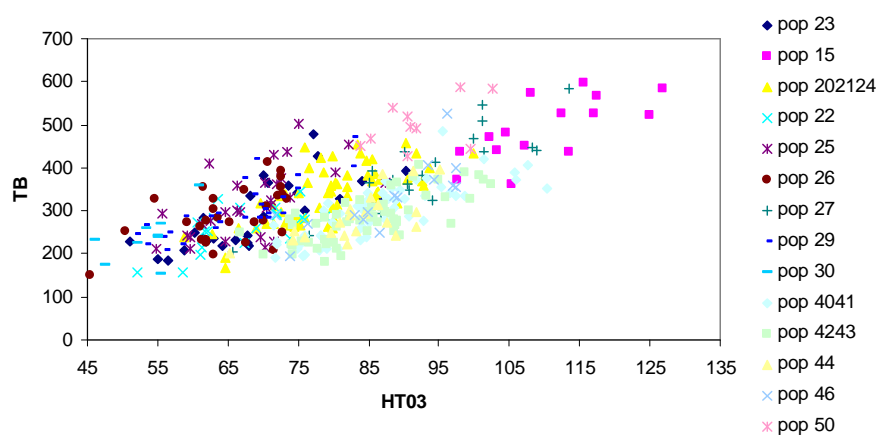
A)



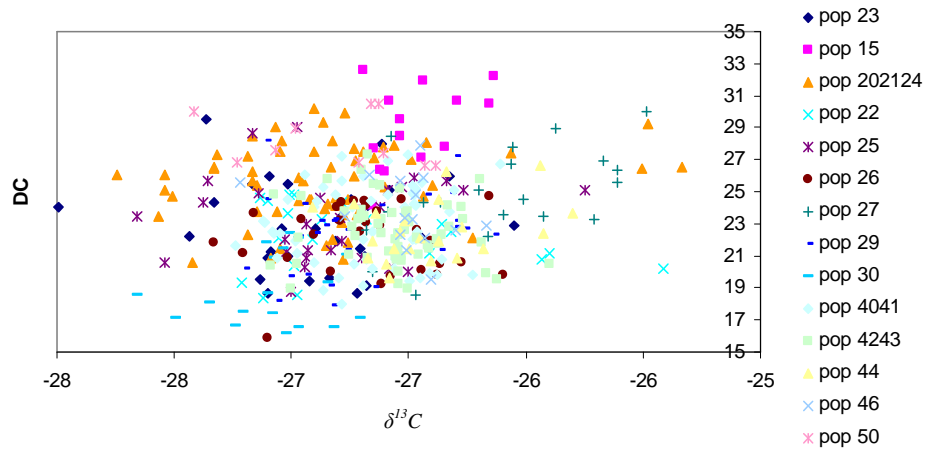
B)



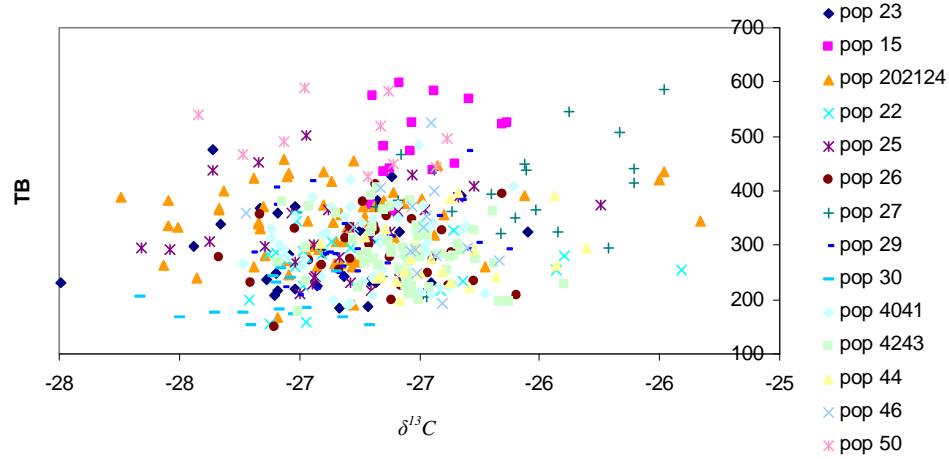
C)



D)



E)



F)

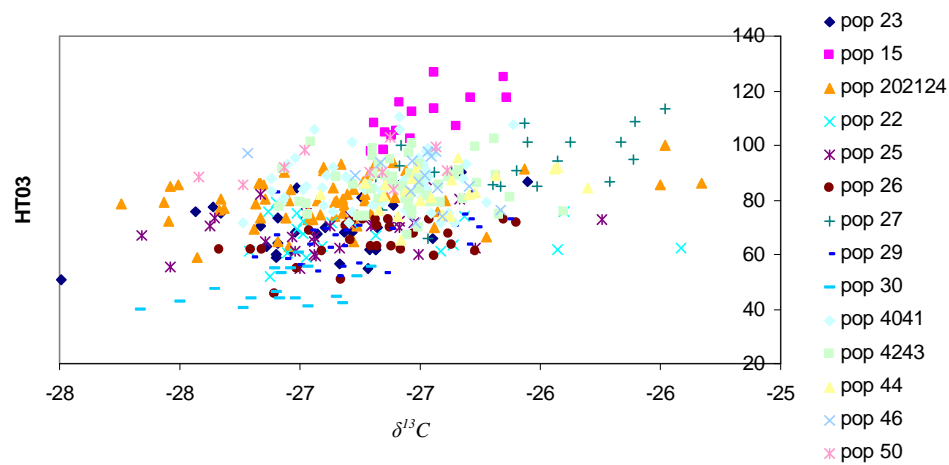


Tableau 23. Corrélations génétiques additives (=entre familles) et erreurs standards entre *HT03*, *TB* et $\delta^{13}C$ par population. Les erreurs standards des corrélations sont indiquées entre parenthèses. Les valeurs en gras sont significativement différentes de zéro.

Population		DC13/HT03		DC13/TB		HT03/TB	P- valeur effet famille HT03	P- valeur effet famille TB	P- valeur effet famille $\delta^{13}C$
	Df fam		Df fam		Df fam				
15		NA		NA		NA	NS	NS	NS
20	17	0.4264 (0.198) ^(a)		NA		NA	<0.003	NS	<0.0001
2124		NA		NA	46	0.7783 (0.058)	<0.0001	<0.001	NS
22	14	-0.251 (0.250) ^(b)	14	-0.307 (0.242) ^(b)	19	0.5887 (0.150)	<0.009	<0.037	<0.0001
23	14	0 (0.267) ^(b)	14	0 (0.267) ^(b)	28	0.7554 (0.081)	<0.0001	<0.0001	<0.001
25		NA		NA	28	0.5428 (0.133)	0.077	<0.0001	NS
26	28	0.8052 (0.066)	28	-0.155 (0.184)	28	0.1639 (0.184)	<0.046	<0.016	<0.023
27	19	0.769 (0.094) ^(a)	19	0.7333 (0.106) ^(a)	19	1 (0.000)	<0.006	<0.003	<0.002
29	26	0.5904 (0.128)	26	0.4752 (0.152)	26	0.8787 (0.045)	<0.0001	<0.003	<0.004
30		NA		NA	17	0.7318 (0.113)	<0.003	<0.024	NS
44	21	0.7667 (0.090)	21	0.4216 (0.179)	21	0.5602 (0.150)	0.063	<0.011	<0.043
46		NA	14	NA		NA	NS	<0.002	NS
4041	54	1 (0.000)	54	0 (0.136)	54	0 (0.136)	<0.0001	<0.0001	<0.012
4243		NA		NA	56	0.8281 (0.042)	<0.009	<0.004	NS
50		NA		NA		NA	NS	NS	0.063

^(a) Les données analysées proviennent uniquement d'un lot d'analyse USA.

^(b) Les corrélations sont estimées uniquement à partir du lot d'analyse français.

populations (variant de 0.2 à 0.6, tableau 7-A)) que au niveau des familles (corrélations génétiques additives r_{fam}) (entre 0.1 et 0.3, sauf pour la corrélation entre $\delta^{13}C$ et *PolyT* (Tableau 22-B)). Les valeurs sont comparables quand on exclut le lot d'analyse USA (Tableaux 22). Les corrélations les plus faibles sont en général entre $\delta^{13}C$ et diamètre (*DC*) ou biomasses (*TB*, *NB*), que ce soit au niveau des populations (corrélations non significatives) et des familles tandis que les plus élevées sont avec les hauteurs (*HT02* et *HT03*).

- Comparaison de la structure des corrélations génétiques additives entre populations

Les corrélations génétiques additives, tout comme les héritabilités, dépendent directement de la variance entre familles au sein de chaque population. C'est pourquoi, nous avons décidé, tout comme pour l'héritabilité, de ne considérer que le lot d'analyse FR pour les populations 2124, 22, 23 et 25 (voir la comparaison des corrélations entre moyennes phénotypiques des familles en Annexe 10). Pour le reste des populations, les corrélations génétiques additives ont été estimées sur l'ensemble des données, après avoir vérifié la consistance des estimations des corrélations dans les 2 lots.

Sur la base de ces estimations si l'on compare les différentes populations, les corrélations génétiques additives sont plus faibles (voire négatives mais non significativement différentes de zéro) et plus variables entre $\delta^{13}C$ et les autres caractères, qu'entre hauteur et biomasse (Tableau 23). Lorsque l'effet famille est significatif, les valeurs varient de -0.251 (population 22) à 0.80 (population 26) entre $\delta^{13}C$ et *HT03*, et de -0.307 (population 22) à 0.73 (population 27) entre $\delta^{13}C$ et *TB*. On peut distinguer différents groupes de populations au sein desquelles les corrélations génétiques additives entre $\delta^{13}C$ et les autres caractères semblent suivre le même patron. Pour quatre populations (20, 27, 29 et 44), les corrélations sont positives entre les trois couples de caractères (>0.4, sauf pour celles avec *TB* pour la population 20, Tableau 23), indiquant que l'efficacité d'utilisation en eau est aussi bien associée au niveau génétique à une croissance importante qu'à une biomasse élevée, et suggérant une possibilité importante de progrès génétique simultanément pour ces trois caractères. A l'inverse, pour la population 22 (et peut-être la population 46 si l'on avait plus de puissance pour détecter un effet famille significatif), on observe une tendance vers une corrélation génétique additive négative (mais non significativement différente de zéro) entre le $\delta^{13}C$ et *HT03* ou *TB*, tandis qu'elle est fortement positive entre *HT03* et *TB*. Pour cette population, ces valeurs de corrélation montrent une tendance à ce que les génotypes les plus efficaces soient également ceux qui aient une croissance en hauteur et une biomasse plus faibles. Pour les autres populations (26 et 4041) les corrélations quand elles sont estimables

sont en général non significatives et proches de zéro entre $\delta^{13}C$ et *TB*, alors qu'elles sont fortement positives entre $\delta^{13}C$ et *HT03*. Au sein de ces populations, l'efficacité d'utilisation en eau serait plus fortement associée au niveau génétique à la hauteur qu'à la biomasse. Enfin, la population 23 présente une corrélation nulle entre $\delta^{13}C$ et *HT03* ou *TB*.

D.4.3. Des corrélations résiduelles positives et élevées entre caractères de hauteur et biomasse mais plus faibles et variables avec le $\delta^{13}C$

- Au niveau global

Sur l'ensemble des données, les corrélations résiduelles sont positives et élevées entre caractères de hauteur et biomasse (Tableau 22). De même que pour les corrélations génétiques additives, les caractères de biomasse (*TB*, *NB* et *DC*) présentent des corrélations résiduelles supérieures à 0.8, et les caractères de hauteur (*HT02* et *HT03*), des corrélations supérieures à 0.7. Ces deux groupes de caractères sont également fortement corrélés entre eux (r_{err} entre 0.6 et 0.8). A l'inverse, le caractère *PolyT* a une corrélation résiduelle nettement inférieure (autour de 0.1), qui pourrait être en partie due à une plus faible corrélation génétique intra famille.

Les corrélations résiduelles entre $\delta^{13}C$ et les autres caractères sont plus faibles (r_{err} n'excédant pas 0.2, avec ou sans le lot d'analyse USA, Tableau 22). Les corrélations les plus faibles sont avec le diamètre (*DC*) et les biomasses (*TB*, *NB*), tout comme pour les corrélations génétiques. La corrélation la plus élevée est entre $\delta^{13}C$ et *HT03*, impliquant qu'il existe une corrélation génétique intra famille plus élevée entre ces caractères et/ou que l'expression des gènes impliqués à la fois dans les caractères de hauteur et d'efficacité d'utilisation en eau est sensible à la variation du milieu et qu'elle varie dans le même sens.

- Comparaison de la structure des corrélations résiduelles entre populations

Dans chaque population, les corrélations résiduelles sont similaires et élevées entre *HT03* et *TB* (r_{err} toutes significatives et en moyenne de 0.7, Tableau 22), et reflètent bien celles estimées sur l'ensemble des populations (également de 0.7) (Tableau 20C). Tout comme pour les corrélations génétiques additives, nous avons considéré uniquement le lot d'analyse français pour les populations 2124, 22, 23 et 25, pour estimer les valeurs de corrélations résiduelles entre $\delta^{13}C$ et les autres caractères.

Les corrélations résiduelles sont généralement beaucoup plus faibles entre $\delta^{13}C$ et *HT03* (r_{err} maximum de 0.33 pour la population 20) sauf pour la population 15 qui montre un

Tableau 24. Corrélations résiduelles et erreurs standards entre *HT03*, *TB* et $\delta^{13}C$ par population. Les erreurs standards des corrélations sont indiquées entre parenthèses. Les corrélations résiduelles ont été estimées sur le jeu de données complet car l'ajout du lot d'analyse USA n'entraîne pas de changements majeurs dans les valeurs de corrélations résiduelles. Les valeurs en gras sont significativement différentes de zéro.

Population	DC13/HT03	DC13/TB	HT03/TB
	Df tot	Df tot	Df tot
15	102 0.4911 (0.075)	102 0.2347 (0.094)	102 0.7188 (0.048)
20	108 0.3345 (0.085) ^(a)	108 0.1098 (0.095) ^(a)	108 0.7452 (0.043)
2124	203 0.1726 (0.068) ^(b)	203 0.0474 (0.070) ^(b)	415 0.7714 (0.020)
22	97 0.2584 (0.095) ^(b)	97 0.1212 (0.100) ^(b)	134 0.6825 (0.046)
23	104 0.2472 (0.092) ^(b)	104 0.0529 (0.098) ^(b)	194 0.7478 (0.032)
25	102 0.2104 (0.095) ^(b)	102 0.0788 (0.098) ^(b)	192 0.6648 (0.040)
26	190 0.171 (0.070)	190 0.0971 (0.072)	190 0.7468 (0.032)
27	123 0.1573 (0.088) ^(a)	123 -0.023 (0.090) ^(a)	123 0.6416 (0.053)
29	186 0.1994 (0.070)	186 0.061 (0.073)	186 0.7676 (0.030)
30	111 0.2012 (0.091)	111 0.1105 (0.094)	111 0.5856 (0.062)
4041	382 0.1215 (0.050)	382 -0.025 (0.051)	382 0.7539 (0.022)
4243	391 0.1949 (0.049)	391 0.0436 (0.050)	391 0.7376 (0.023)
44	150 0.1471 (0.080)	150 0.0457 (0.081)	150 0.7999 (0.029)
46	102 0.0016 (0.099)	102 -0.18 (0.096)	102 0.7731 (0.040)
50	69 0.192 (0.116)	69 0.0375 (0.120)	69 0.8115 (0.041)

^(a) Les données analysées proviennent uniquement d'un lot d'analyse USA.

^(b) Les corrélations sont estimées uniquement à partir du lot d'analyse français.

r_{err} de 0.49 entre ces 2 caractères (Tableau 24). Les corrélations sont encore plus faibles entre $\delta^{13}C$ et TB (r_{err} maximum de 0.23 pour la population 15) (Tableau 24), suggérant que les contraintes physiologiques entre $\delta^{13}C$ et $HT03$ ou TB sont faibles. La population 15 (Pinia) est un peu différente et montre les corrélations résiduelles les plus importantes entre les trois caractères ($r_{err}>0.4$ entre $\delta^{13}C$ et $HT03$, >0.2 entre $\delta^{13}C$ et TB , >0.7 entre $HT03$ et TB). Même si une partie de cette corrélation résiduelle peut être due à un effet génétique intra famille, elle peut également vouloir dire que si une variation du micro-environnement rend les génotypes de la population Pinia plus efficaces, ces génotypes seront également ceux qui produisent le plus de biomasse et ont la croissance la plus importante dans ce second environnement. Les populations 20, 22 et 30 présentent la même tendance de corrélations résiduelles positives mais avec des valeurs moins importantes (>0.2 entre $\delta^{13}C$ et $HT03$, et >0.10 mais non significativement différentes de zéro entre $\delta^{13}C$ et TB). Parmi ces trois populations, on note aussi que la population 30 est celle qui présente une corrélation résiduelle la plus faible entre $HT03$ et TB (<0.6). Parmi les populations, la 46 (située la plus au Nord) est la seule à présenter une corrélation résiduelle négative (mais non significative à 5%) entre $\delta^{13}C$ et TB d'environ 0.2 suggérant une contrainte physiologique possible entre production de biomasse et efficacité au sein des génotypes de cette population.

D.5. Résultats sur la différenciation entre populations

D.5.1. Une différenciation entre populations représentative des grands groupes géographiques

Sur l'ensemble des données, les valeurs de différenciation génétique entre populations aux caractères quantitatifs (Q_{ST}) sont moyennes à élevées, allant de 9% pour $\delta^{13}C$, de 24% à 33% pour les caractères de biomasse (TB , NB et DC), et de 43% à 48% pour les caractères de hauteur ($HT02$ et $HT03$) et le $PolyT$ (Tableau 21). Cependant, lorsque l'on considère les Q_{ST} par paires de populations pour les 3 caractères $HT03$, TB et $\delta^{13}C$, les valeurs sont beaucoup plus variables (de 0 à 1) (Annexe 10). Les populations qui contribuent le plus à la différenciation sont Pinia (15) et Tamrabta (30) pour le $\delta^{13}C$ et la hauteur, et Pinia et Tabarka (50) pour la biomasse (Annexe 10).

D'autre part, de manière générale, les valeurs de Q_{ST} par paires de populations sont plus élevées pour le caractère $HT03$ puis le caractère TB que pour le caractère $\delta^{13}C$ (valeurs environ 2 fois moins élevées que pour $HT03$), indiquant que la part de la variation génétique expliquée par la structure en populations est nettement plus importante pour la hauteur que

pour l'efficacité d'utilisation en eau au sein de cet essai, ce qui confirme la grande variation intra population (entre familles) pour ce caractère par rapport à la variation entre populations.

D.5.2. Maintien d'une différenciation élevée entre populations espagnoles

Dans le groupe des populations françaises de la côte atlantique, les valeurs de Q_{ST} sont inférieures à 0.1 sauf entre le groupe 4041 et la population 44 (Olonne/mer) pour le $\delta^{13}C$. A l'inverse, parmi les populations espagnoles, les valeurs de Q_{ST} peuvent être très variables (allant de 0 à 0.56 entre les populations 26 et 27 pour $HT03$, de 0 à 0.34 entre les populations 20 et 22 pour TB , et de 0 à 0.42 entre les populations 24 et 27 pour $\delta^{13}C$). Les populations espagnoles les plus différenciées des autres populations espagnoles sont la population 27 (San Cipriano) pour les trois caractères et la population 22 (Coca) pour le caractère de biomasse.

E. Discussion

Cette étude avait pour premier objectif d'évaluer la variabilité des caractères d'efficacité d'utilisation en eau (liée positivement au $\delta^{13}C$), de croissance et de biomasse au sein d'un large échantillon de populations de pin maritime. Globalement, une large variabilité a été démontrée entre populations au sein de cet essai situé sur un sol landais humide à mésophile, pour différents estimateurs (performances moyennes et relations avec la variation des lieux d'échantillonnage, héritabilités, corrélations génétiques et résiduelles et différenciation adaptative).

En effet, l'effet population est significatif pour la variation de la croissance, de la biomasse et du $\delta^{13}C$ (même si pour ce dernier, la variation est surtout présente entre familles au sein des populations). D'autre part, les héritabilités des populations peuvent être très différentes. Pour les populations présentant des héritabilités très faibles telle que la population Corse Pinia (correspondant à une variation très faible entre familles), on pourrait supposer qu'une forte pression de sélection sur ces caractères en serait la cause. On a également observé que les moyennes de $\delta^{13}C$ étaient corrélées à des variables écologiques, notamment les précipitations annuelles. Des différences entre populations ont également été observées dans les structures de corrélations génétiques et physiologiques (avec notamment la population Pinia qui présente des corrélations physiologiques fortes par rapport aux autres populations). Enfin, la différenciation génétique pour les caractères peut être variable entre paires de populations (différenciation importante entre la population San Cipriano et le reste pour les trois caractères, et entre la population Coca et le reste pour la biomasse). L'ensemble de ces observations suggère l'existence d'adaptations locales différentes en réponse à des

conditions environnementales contrastées. Nous essayons de discuter ici cette variabilité de réponse au sein de notre essai, en considérant les différentes stratégies possibles que ces populations auraient pu développer au cours de leur évolution dans leur aire d'origine.

E.1. Différents mécanismes de résistance à la sécheresse entre provenances de milieux humides et de milieux chauds ou secs

La large variabilité observée pour ces caractères est en grande partie due à quelques populations provenant des extrémités de l'aire de distribution du pin maritime (notamment les populations Corse et Tunisienne), comme en témoignent les matrices de Q_{ST} par paires de populations, et la distribution des moyennes phénotypiques entre populations. Une variabilité importante à une échelle géographique plus réduite est également présente au sein de la péninsule Ibérique, avec le cas particulier d'une population du Nord-Ouest de l'Espagne, San Cipriano (27), qui s'écarte très nettement des autres populations espagnoles, en termes de hauteur, biomasse et $\delta^{13}C$. Au sein de la côte atlantique française, les caractères quantitatifs mesurés sont beaucoup moins variables entre populations (valeurs de Q_{ST} plus faibles). La variabilité observée entre les populations pour ces caractères quantitatifs au sein de notre essai peut être reliée à l'hétérogénéité des conditions climatiques de l'aire d'origine des populations.

E.1.1. Stratégie basée sur le contrôle stomatal

Il semblerait que les populations ayant évolué dans des milieux plus humides, présentent en moyenne des valeurs de $\delta^{13}C$ (EUE_{int}) plus élevées et une croissance en hauteur et biomasse supérieures, dans notre milieu de test, à celles de populations provenant de régions plus chaudes ou plus sèches. En effet, les populations corse Pinia (705 mm par an) et espagnole San Cipriano (661 mm par an) présentent des moyennes de hauteur et biomasse ainsi que de $\delta^{13}C$ donc d' EUE_{int} parmi les plus élevées. Les populations françaises locales de Pleucadec à Petrock (de 716 à 995 mm par an) sont également parmi les plus efficaces mais avec des biomasses moins importantes.

Ces résultats rejoignent ceux obtenus sur la même espèce par Guehl *et al.* (1996), Nguyen-Queyrens *et al.* (1998), Tognetti *et al.* (2000) et Danjon (1994). Dans les deux premières études, trois provenances de pin maritime (Porto Vecchio (Corse), Landes (côte atlantique française) et Tamjoute (Maroc)) ont été testées pour leur croissance en hauteur et leur EUE_{int} grâce à une mesure de $\delta^{13}C$ foliaire, dans des essais implantés dans un milieu de Landes humide à mésophile pour des arbres de 3 ans (Nguyen-Queyrens *et al.* 1998), ou sous

serre dans différentes conditions de régimes hydriques pour des plantules (Guehl *et al.* 1996). La population Corse Porto Vecchio présentait également une croissance plus élevée ainsi que des valeurs de $\delta^{13}C$ plus élevées (donc une EUE_{int} plus élevée). La population Landaise atteignait une croissance moyenne et un $\delta^{13}C$ moyen plus faibles que la population Corse. Dans l'étude de Tognetti *et al.* (2000), des arbres âgés de 2 ans provenant de 5 populations du Portugal, de la Corse, de la Toscane et de la Sardaigne ont été testées au sein de 4 sites localisés en Sardaigne et caractérisés par un climat semi-humide (entre 1000 et 1100 mm de précipitations moyennes annuelles). La population du Portugal originaire du pool génétique Aquitain (Bucci *et al.* 2007) et ayant évolué dans l'environnement le plus humide, était également celle avec la croissance et des valeurs de $\delta^{13}C$ les plus élevées (EUE_{int} élevée). Enfin, Danjon (1994) montre une dynamique de croissance en hauteur différente notamment entre populations de la côte atlantique française, de l'Espagne et du Maroc, avec les populations de la côte atlantique françaises présentant une croissance en hauteur la plus élevée jusqu'à l'âge de 36 ans sur un milieu plus sec (dune littoral landaise) que celui de notre étude. Plusieurs études chez d'autres espèces forestières ont également montré une tendance à des valeurs de $\delta^{13}C$ plus élevées (EUE_{int} plus élevée) pour des génotypes ou écotypes ayant évolué dans des milieux plus humides (Aitken *et al.* 1995 chez le Douglas, Lauteri *et al.* 1997 chez le Châtaigner, Read et Farquhar 1991 chez le *Nothofagus*, Zhang et Marshall 1995 chez le Douglas et le Pin ponderosa, Ponton *et al.* 2001 chez le Chêne, Hubick et Gibson 1993 chez l'Eucalyptus).

Dans une étude chez le *Nothofagus*, Read et Farquhar (1991) suggèrent l'existence d'une stratégie adaptative particulière pour ce type de population. Ces populations mésiques (évoluant au sein de milieux humides) posséderaient une stratégie plutôt basée sur le contrôle stomatique (régissant l'ouverture et la fermeture des stomates) afin de réduire les pertes d'eau dues à la transpiration durant les épisodes courts et peu fréquents de sécheresse modérée. La sensibilité stomatale de ces génotypes conduirait à des valeurs élevées de $\delta^{13}C$ (EUE_{int} élevée), particulièrement dans des environnements semi-humides où des précipitations relativement élevées assure une croissance aérienne importante. Ce type d'environnement correspondrait relativement bien à celui de notre essai où le sol est de type humide à mésophile (semi-humide) et les précipitations moyennes annuelles sont d'environ 900 mm.

Nous serions donc en présence de quelques populations (Corse, San Cipriano, et populations de la côte atlantique française) qui auraient développé des stratégies caractéristiques de populations mésiques consistant à fermer plus rapidement leurs stomates lors de sécheresse momentanée et à les ouvrir plus rapidement lors de précipitations.

E.1.2. Stratégie basée sur le développement racinaire

Au sein de notre étude, la population présentant les moyennes de croissance (hauteur et biomasse aérienne) et de $\delta^{13}C$ les plus faibles est celle du Maroc (Tamrabta), dont la température moyenne annuelle de l'aire d'origine est l'une des plus élevées (17,2°C par rapport à une moyenne de 13°C dans l'essai). De même dans Guehl *et al.* (1996) et Nguyen-Queyrens *et al.* (1998), c'est la population Tamjoute du Maroc, population la plus résistante à la sécheresse (Guyon et Kremer 1982, Sarrauste 1982, Nguyen et Lamant 1989) qui présente la croissance moyenne (hauteur) la plus faible et le $\delta^{13}C$ moyen le plus bas, quelque soit le régime hydrique.

Read et Farquhar (1991) suggèrent que les populations plus xériques (évoluant dans des milieux secs) auraient développé des mécanismes d'évitement à la sécheresse, certes plus coûteux, mais plus adaptés à des conditions de sécheresse intense, répétée et longue. Ces mécanismes correspondraient à une allocation préférentielle du carbone dans la croissance racinaire plutôt que dans la croissance des tiges, une activité photosynthétique plus élevée, une consommation en eau plus forte au niveau des feuilles, et un contrôle stomatique plus limité, avec pour conséquence une discrimination isotopique du carbone (Δ) plus élevée (correspondant à une EUE_{int} plus faible).

Cette hypothèse a été vérifiée sur le plan physiologique notamment chez le pin maritime. Il a été montré que des populations xériques, telle que Tamjoute, était capable de maintenir une plus grande conductance stomatique (g_s) pour une même activité photosynthétique (A), et donc une plus faible EUE_{int} , par rapport à des populations mésiques (Guyon et Kremer 1982) et ce quel que soit le régime hydrique. Le ratio de biomasse racinaire sur biomasse aérienne plus élevée chez cette population (Sarrauste 1982) est également cohérent avec une conductance stomatique plus élevée, selon Meintzer et Grantz (1990). Danjon (1994) montre également qu'en dune littorale landaise (sol sec), la provenance marocaine Tamrabta aurait tendance à combler son retard de croissance en hauteur par rapport aux provenances locales à partir de l'âge de 20 ans, qui pourrait éventuellement s'expliquer par un meilleur enracinement et une meilleure prospection pour l'eau. Dans notre essai, la population marocaine serait moins réactive par rapport à l'ouverture/fermeture des stomates et donc moins performante, mais compenserait par un système racinaire plus développé. Quant à la population tunisienne, malgré des températures moyennes élevées, présente une hauteur et une efficacité de l'eau intermédiaires, accompagnées d'une des 2 valeurs de biomasse les

plus élevées (reliée à une production plus importante d'aiguilles observée de terrain), elle aurait développé une stratégie sans doute différente de la population marocaine.

E.1.3 Performances des populations locales versus non-locales

Le fait que des populations non-locales (même si les conditions climatiques du lieu de l'essai sont proches de celles de leur aire d'origine), telles que Pinia et San Cipriano, aient une biomasse plus importante que des populations locales (populations de la côte atlantique française) seraient cohérentes avec des observations sur les transferts de populations chez *Pinus sylvestris* (Savolainen *et al.* 2007). Dans ces transferts, les populations du Nord présentent un optimum de survie relative (survie de la population transférée sur survie de la population locale) supérieure non pas dans leur milieu d'origine, mais à quelques degrés de latitude plus au sud. Dans notre cas, les populations Pinia et San Cipriano auraient un optimum de biomasse plus important lorsqu'elles sont transférées dans un milieu un peu moins chaud (température moyenne par an de 15.1°C pour San Cipriano à environ 14°C pour le milieu du test).

E.2. Le cas particulier des populations espagnoles

Cependant, au sein de la péninsule Ibérique si l'on exclut la population San Cipriano au Nord de l'Espagne, une tendance différente est observée. Un petit gradient environnemental (précipitations allant de 957 mm à 348 mm) est également présent du centre jusqu'au sud de l'Espagne. Or le long de ce gradient, les populations ayant évolué dans des régions plus sèches (Bayubas, Oria et Coca), comparativement aux autres populations espagnoles (Cuellar et San Leonardo) présentent une meilleure EUE_{int} mais une moins bonne croissance en hauteur et surtout une biomasse moins importante. Ce comportement n'a encore jamais été rapporté dans les quelques études réalisées sur le pin maritime. Néanmoins, chez l'Eucalyptus, Li (1999) a observé le même comportement chez 3 populations d'Eucalyptus soumis en serre à des traitements sans et avec un léger stress hydrique (seule une quantité d'eau correspondant aux pertes hydriques durant la transpiration est maintenue). La population la plus adaptée à la sécheresse (provenant du Nord de l'Australie) présentait une meilleure efficacité d'utilisation en eau au niveau de la plante ($\delta^{13}C$ également plus élevée) mais une matière sèche totale plus faible. Des résultats similaires ont également été rencontrés dans d'autres tests de provenances chez l'Eucalyptus. L'hypothèse avancée serait que les populations provenant de régions plus sèches (telles que Bayubas, Oria et Coca) auraient développé une stratégie conservative d'utilisation de l'eau (selon Passioura 1982). Cette

stratégie consisterait à augmenter leur efficacité en fermant partiellement les stomates tout en réduisant la photosynthèse, avec pour conséquence une corrélation négative entre EUE_{int} et biomasse. Cette stratégie serait particulièrement efficace lors de longues périodes de sécheresse. Il est intéressant de noter qu'au vu des résultats de cette étude, les différences de stratégies entre populations espagnoles ont été maintenues suite probablement à de fortes adaptations locales, et ce malgré une proximité géographique des populations, impliquant des flux de gènes importants (par exemple, entre les populations Bayubas de Abajo avec 405 mm par an et San Leonardo avec 671 mm par an, et entre Coca et Cuellar dont les précipitations annuelles sont proches).

E.3. Différentes intensités de corrélation génétique et résiduelle entre populations

Au niveau des moyennes phénotypiques, nous avons vu qu'une plus forte EUE_{int} ($\delta^{13}C$ plus élevé) était associée à une croissance et biomasse plus élevée chez la population corse (considérée comme mésique) et qu'une plus faible EUE_{int} était associée à une plus faible croissance chez la population marocaine Tamjoute (population xérique), cette association s'inversant lorsque l'on comparait les populations espagnoles entre elles. Ces corrélations phénotypiques pour les populations corse et marocaine vont dans le même sens que les corrélations au niveau génétique (à l'échelle des populations et des familles) et résiduel, sur l'ensemble des données, même si ces dernières restent faibles à modérées (entre 0.04 et 0.32). Par contre, en considérant chaque population, les intensités et le sens des corrélations peuvent être très différents même s'il existe un problème de puissance pour l'estimation des variances entre familles pour certaines populations où le nombre de familles était limité, ne permettant pas d'avoir une estimation dans toutes les populations. Cependant, dans celles où l'estimation était possible, plusieurs présentaient des corrélations entre $\delta^{13}C$ et les autres caractères nulles ou avec une tendance négative (cas des populations espagnoles Coca, Cuellar et Bayubas, des populations françaises Petrock, Mimizan et Pleucadeuc). Une corrélation génétique nulle entre $\delta^{13}C$ et croissance a également été observée au sein de la population des Landes par Brendel *et al.* (2002). Il suggère que cette absence de corrélation peut s'expliquer par différents facteurs : 1) le contrôle génétique de la croissance et de l'efficacité d'utilisation en eau n'est pas le même avec une croissance relativement indépendante du taux d'assimilation en CO_2 (A), tandis que l'efficacité d'utilisation en eau serait plus contrôlée par A (assimilation nette du CO_2) que par g_s (conductance stomatique à diffusion de vapeur d'eau) au sein de cette population (Guehl *et al.* 1996), 2) les héritabilités étant faibles pour les deux types de caractères, la corrélation génétique était mal estimée.

A l'inverse, pour d'autres populations (populations espagnoles Cencimientos, San Cipriano et Oria, et la population française Olonne/Mer), les corrélations génétiques sont fortes et positives. D'autres études mettent en avant des corrélations génétiques positives et élevées entre $\delta^{13}C$ et croissance chez des arbres forestiers (Jonhsen *et al.* 1999 chez l'épicéa, Lauteri *et al.* 2004 chez le châtaignier), mais au niveau inter-populations seulement. La question est de savoir s'il existe un lien causal entre $\delta^{13}C$ (et donc EUE_{int}) et la croissance, ou si elle serait plutôt due à un facteur commun contrôlant ces deux caractères. Il est également intéressant de constater que les corrélations génétiques les plus élevées ne sont pas forcément associées aux populations provenant de milieux plus stressés et qui auraient subi des pressions de sélection *a priori* plus importantes.

La corrélation résiduelle est également intéressante car même si elle comprend une composante génétique (due à la variation intra famille), elle possède également une composante environnementale (corrélation physiologique), qui permettrait de faire des prédictions sur le comportement des individus de ces populations dans des environnements différents. Au sein de notre essai, nous avons mis en évidence que la population corse Pinia était celle où la corrélation résiduelle positive était la plus élevée entre le $\delta^{13}C$ et les autres caractères. Cette corrélation résiduelle est du même ordre de grandeur que celle estimée au sein de la provenance des Landes par Brendel *et al.* (2002) (0.52). On a également montré que cette population avait des moyennes de $\delta^{13}C$, hauteur et biomasse élevées au sein de notre essai dont les conditions climatiques étaient proches de celles de son aire d'origine. Cela voudrait dire que les génotypes de cette population sont dans l'ensemble adaptés au milieu du test. Néanmoins, il n'est pas sûr que ces génotypes le soient également dans un milieu plus sec. Si ce n'est pas le cas, tous les génotypes seraient moins efficaces, et donc produiraient moins de biomasse et auraient une croissance plus faible dans un milieu plus sec. A l'inverse, on observe pour les autres populations une corrélation non significativement différente de zéro, qui indiquerait qu'il n'y ait *a priori* pas de contraintes particulières dans ces populations si l'on veut sélectionner pour l'efficacité. Ces populations étant généralement celles présentant la variabilité la plus large, un progrès génétique serait possible.

E.4. Conservation d'un potentiel adaptatif (variabilité génétique) plus élevé au sein de quelques populations (tel qu'il a été exprimé dans le milieu de test)

Cette étude met également en évidence l'existence d'une variabilité intra population importante pour certaines populations. La valeur d'héritabilité au sens strict reste modérée pour le $\delta^{13}C$ ($h^2=0.17 \pm 0.06$), comparativement à ce qui a été observé pour le Δ chez

d'autres espèces (entre 0,6 et 0,9, Farquhar *et al.* 1989). Cependant, ces niveaux d'héritabilité se révèlent très différents selon les populations (de 0 à 0.8). Cette variation dans les valeurs d'héritabilité est avant tout liée à une différence entre variances génétiques entre familles (et beaucoup moins à la variance résiduelle), cette tendance étant vraie pour toutes les populations.

Il semblerait que les populations présentant les valeurs d'héritabilité les plus faibles (avec une variabilité génétique entre familles réduite), soient en partie celles qui ont une croissance en hauteur, une biomasse et une EUE_{int} ($\delta^{13}C$) les plus élevées au sein de l'essai. C'est le cas de la population corse Pinia et des populations françaises (pour le $\delta^{13}C$, la hauteur et la biomasse) et de la population tunisienne (uniquement pour la croissance et la biomasse). La seule étude estimant l'héritabilité du $\delta^{13}C$, et qui existe chez le pin maritime est celle de Brendel *et al.* (2002). L'héritabilité pour le $\delta^{13}C$ a été estimée à partir d'un demi-diallèle de 12 arbres mère (sélectionnés pour leur vigueur) provenant des Landes de Gascogne. L'héritabilité dans cet essai, estimée sur un échantillon limité d'individus est plus faible, bien que du même ordre de grandeur que nos estimations: héritabilité de 0.17 pour Brendel *et al.* (2002), par rapport à une héritabilité de 0.29 +/- 0.2 au sein du groupe de population Mimizan-Petrock pour notre essai. Chez d'autres espèces d'arbres forestiers, des héritabilités plus élevées pour le $\delta^{13}C$ ont été constatées ($h^2=0.31$ +/- 0.07 chez 6 populations de châtaignier, Lauteri *et al.* 2004, $h^2=0.54$ +/- 0.26 dans un semi-diallèle chez l'épicéa, Johnsen *et al.* 1999). Les valeurs d'héritabilités estimées, bien que dépendantes du milieu de test, sont comparables quantitativement entre les populations observées dans notre essai.

Par exemple, au sein de certaines populations espagnoles (Coca, Bayubas et Oria), correspondant aux populations les moins performantes qui auraient évolué dans des régions plus sèches, les héritabilités sont plus élevées notamment pour le $\delta^{13}C$. C'est également vrai pour la population tunisienne Tabarka pour le $\delta^{13}C$. Comment expliquer que ces populations présentent des héritabilités plus élevées ? Une explication possible serait l'existence de familles mal-adaptées au milieu de test dans ces populations venant d'un milieu d'origine différent, et qui pourrait expliquer les variances entre familles plus importantes. Une autre hypothèse possible est que des pressions de sélection hétérogènes au cours du temps (due à une grande variation inter-annuelle des conditions climatiques) auraient pu s'exercer au sein de ces populations et maintenir une variation entre génotypes plus importante (Via et Lande 1985, Hedrick 2006a).

E.5. Conclusion et perspectives

Cette étude est la plus importante jusqu'à présent ayant permis d'estimer la composition isotopique du carbone, en relation avec croissance en hauteur et biomasse au sein d'un large échantillon de populations naturelles représentatif de la distribution géographique du pin maritime. Nous avons montré que la grande variabilité observée entre populations pour ces caractères pouvait être interprétée en termes de stratégies d'adaptations locales différentes pour la résistance à la sécheresse. Cette étude montre aussi la présence d'une large variabilité génétique dans certaines populations pour ces caractères, ce qui pourrait être un avantage adaptatif pour leur adaptation aux changements climatiques futurs. Parmi les populations les plus performantes, seule la population San Cipriano possède des héritabilités élevées pour le $\delta^{13}C$, la croissance et la biomasse, mais également des moyennes phénotypiques élevées. Cette population pourrait être particulièrement intéressante dans un programme de sélection, d'autant plus qu'elle a été très peu étudiée jusqu'à présent. Il pourrait notamment être intéressant de croiser cette population avec la population corse Pinia, qui malgré des moyennes phénotypiques élevées, montre *a priori* une variabilité génétique très faible, et donc un potentiel adaptatif limité. La connaissance de la distribution de cette variabilité entre populations naturelles peut également servir de point de départ à des programmes de conservation génétique. Néanmoins, cette étude reste limitée à un seul environnement, et la généralisation à d'autres environnements est difficile, malgré quelques tentatives de prédictions. Ces prédictions doivent être vérifiées dans un environnement plus stressé pour la disponibilité en eau, de façon à savoir 1) si les populations les moins résistantes à la sécheresse, malgré une EUE_{int} plus élevée, survivraient sans problèmes, 2) si les populations les plus résistantes se montreraient plus performantes, 3) et enfin si les corrélations entre $\delta^{13}C$ et croissance ou biomasse seraient plus élevées.

Chapitre VI

Recherche d'associations entre diversité de gènes candidats et variabilité génétique de caractères liés à la résistance au stress hydrique

A. Introduction

Les approches permettant de disséquer les bases génétiques des caractères complexes consistaient traditionnellement à cartographier des QTLs (loci à effets quantitatifs) sur la base d'associations statistiques significatives entre marqueurs moléculaires et variation quantitative. Les études d'associations (« LD mapping » ou « association mapping » en anglais) sont de nouvelles méthodes d'étude du déterminisme génétique des caractères complexes, qui peuvent être appliquées en populations naturelles. La principale différence avec les approches QTLs concerne la résolution possible du marquage autour d'un gène ou d'un site potentiellement intéressant, qui est *a priori* beaucoup plus fine dans une étude d'association, à cause des fenêtres de déséquilibre de liaison beaucoup plus petites qui les encadrent (Cardon et Bell 2001). En effet, comme ces études concernent du matériel issu de populations naturelles ayant évolué sur de nombreuses générations, de nombreux évènements de recombinaison ont pu se produire, diminuant ainsi les déséquilibres de liaison (*DL*) entre marqueurs le long du génome. Dans ce cas, seuls les loci ou mutations étroitement liés physiquement à un locus impliqué dans la variation adaptative présenteront des associations statistiques, permettant une cartographie plus fine des QTLs ou des mutations causales au sein de populations naturelles qu'au sein de pédigrés (Walsh 2002, Gupta et al. 2005).

L'efficacité des études d'associations dépendra donc fortement des patrons et de l'étendue du *DL* dans le matériel utilisé. Cependant, leur faiblesse sera aussi que tous les facteurs évolutifs autres que les variations du taux de recombinaison, et affectant les patrons de *DL* ou générant des *DL* pourront augmenter le nombre de faux-positifs et diminuer la puissance de détection des associations (Abecassis *et al.* 2005).

Chez de nombreuses espèces forestières, les niveaux élevés de diversité génétique, estimés en général au sein de populations de grande taille efficace, et une baisse rapide des déséquilibres de liaison le long des chromosomes rendraient les approches de génétique d'association particulièrement favorables, notamment chez les conifères (Neale et Savolainen 2004, Gonzalez-Martinez *et al.* 2006). En effet, ces caractéristiques vont permettre *a priori* d'éviter les problèmes dus à un apparemment trop grand ou à une structuration cachée dans ces populations, qui peuvent être des sources d'erreur importantes dans les études d'association. Elles peuvent notamment induire des déficits en hétérozygotes (effet Wahlund, voir encadré 1 Chapitre I), ou bien des *DL* entre variants alléliques et variation phénotypique qui n'ont pas de signification biologique. Ceci causera l'observation de faux-positifs (Marchini *et al.* 2004, Hirschhorn et Daly 2005), et il sera important soit d'essayer de détecter

une structure cachée si elle est inconnue, soit de prendre en compte la structure en populations dans les modèles quand elle est connue. De nombreuses méthodes le permettant sont disponibles chez les humains (Pritchard *et al.* 2000, Price *et al.* 2006) et plus récemment chez les plantes (Yu *et al.* 2006, Zhao *et al.* 2007, Mackay et Powell 2007). Par exemple, Yu *et al.* (2006) développent un modèle mixte permettant d'intégrer à la fois la structure géographique entre populations et l'apparentement connu entre individus comme covariables dans le modèle.

Ce dernier chapitre très prospectif avait pour but de réaliser une étude d'association préliminaire visant à tester les relations entre la diversité moléculaire au sein des gènes candidats et la variation des trois principaux caractères quantitatifs étudiés dans le Chapitre V (hauteur, *HT03*, biomasse totale, *TB* et $\delta^{13}C$), sur les 18 populations naturelles pour lesquelles nous disposons de données sur le $\delta^{13}C$. Dans notre étude, la structuration en provenances d'origines différentes était connue et a été prise en compte dans les modèles de tests d'association. Une corrélation entre variation moléculaire et variation phénotypique entre populations (qui peut représenter des adaptations locales différentes), peut elle-même être une source d'information intéressante, et ne doit pas être seulement considérée comme une covariation à éliminer (Zhao *et al.* 2007). Nous avons également testé ici les corrélations entre moyennes des caractères phénotypiques et fréquences alléliques de différents polymorphismes. Initialement, ces analyses devaient constituer la phase finale du projet TREESNIPs, suite aux étapes précédentes d'étude des patrons de diversité nucléotidique et de *DL* intra-gènes (chapitre IV), et de celle de la variation des caractères liés à la tolérance au stress hydrique en populations naturelles (chapitre V).

Malheureusement, le génotypage à grande échelle sur les arbres-mères des plants observés au champ n'a été réalisé que de façon très incomplète, à cause de problèmes techniques rencontrés par le laboratoire qui devait fournir les données avec une technique de pyroséquençage (Syvänen 2001). Nous avons cependant pu obtenir des données moléculaires sur 7 marqueurs SNPs, seuls « rescapés » du projet de génotypage (voir méthodes ci-dessous).

Nous présentons donc ici quelques résultats sur les relations entre variations moléculaire et phénotypique qui ont été testées à deux niveaux :

- 1) au niveau inter-populations, en testant les corrélations entre moyennes phénotypiques et fréquences alléliques aux marqueurs SNPs sur un ensemble de populations représentatives de l'espèce,

- 2) au niveau individuel, en testant les associations entre variants alléliques aux SNPs et variation aux caractères quantitatifs avec différents modèles linéaires prenant en compte la structuration en populations.

B. Matériels et méthodes

B.1. Choix des marqueurs SNPs

Initialement, 27 SNPs parmi les 203 identifiés sur 5 gènes avaient été choisis selon plusieurs critères permettant *a priori* de penser qu'ils avaient une plus grande probabilité d'avoir été soumis à la sélection, et aussi selon leur niveau d'information :

- a) pour chaque gène, le sous-ensemble de SNPs devait permettre de reconstituer les haplotypes principaux,
- b) ces SNPs avaient une fréquence de l'allèle le plus rare supérieure à 5%,
- c) ils étaient indépendants ou en faible déséquilibre de liaison,
- d) pour certains, ils présentaient une forte différenciation génétique entre les 10 populations et/ou étaient situés dans des régions codantes et éventuellement non-synonymes.

Quelques SNPs de « contrôle » avaient également été choisis avec des critères différents (faible différenciation génétique, dans des régions non-codantes) pour avoir une hypothèse alternative permettant notamment de comparer les résultats entre SNPs montrant un fort ou un faible F_{ST} . Cela aurait permis sur un nombre suffisamment important de SNPs, de vérifier si le critère « F_{ST} fort » pouvait être un bon critère de choix des marqueurs à utiliser dans les études d'association pour des espèces structurées en populations.

Les SNPs ont ensuite été génotypés à partir d'aiguilles des arbres-mères sur chacune des familles des 24 populations, par pyroséquençage (Sylvänen 2001). Le taux de réussite initial des SNPs génotypés était assez important, mais nous avons rencontré de nombreuses incohérences entre le jeu de données des génotypes SNPs et celui des SNPs identifiés dans les séquences précédemment obtenues sur les mégagamétophytes d'un certain nombre d'arbres-mères en commun. Le premier type d'incohérence était une incompatibilité entre l'allèle détecté chez un des gamètes de la mère (données de séquences) et les deux allèles détectés dans le génotype de la même mère par la technique de pyroséquençage (par exemple, un génotype maternel de type T/T ne peut donner un mégagamétophyte avec l'allèle C). Ces observations nous ont permis d'estimer un premier taux d'erreur. Nous avons également

comparé différents paramètres entre les 2 jeux de données, et observé des différences trop importantes entre fréquences alléliques et valeurs de F_{ST} pour qu'elles puissent être dues seulement à des différences de taille des échantillons entre jeux de données. De plus les valeurs de F_{IS} (indice de consanguinité) dans certaines populations et pour certains SNPs étaient fortement négatives pour certains SNPs, traduisant l'observation exclusive d'hétérozygotes, ce qui est très peu probable dans une espèce à reproduction sexuée et allogame pour les échantillons suffisamment importants qui étaient génotypés (30 arbres-mères, donc 60 gamètes).

Une des causes possibles de ces incohérences peut être l'amplification en mélange de fragments de gènes paralogues d'une même famille multigénique. Une autre cause possible d'erreur de génotypage peut être le seuil arbitraire à partir duquel les génotypes dans la technique de pyroséquençage étaient considérés comme homozygotes ou hétérozygotes, qui serait mal adapté. Nous avons essayé de changer ce seuil sans amélioration des résultats. Au final, 7 SNPs parmi les 27 initiaux ont été retenus, dont le taux d'incohérence ne dépassait pas 5%, et qui présentaient de bonnes corrélations entre fréquences alléliques (et les F_{ST}) estimées à partir des 2 jeux de données de séquences et de génotypes (Tableau 25).

B.2. Méthodes et modèles d'analyses de la variation

B.2.1. Corrélations entre fréquences alléliques et moyennes phénotypiques

Les fréquences alléliques ont été estimées dans chaque population pour chacun des 7 SNPs et les corrélations ont été testées par un test de Pearson (à n-2 degrés de liberté) entre fréquence allélique à chaque polymorphisme et moyennes phénotypiques des 3 caractères étudiés sur les 18 populations sur lesquelles les données étaient disponibles (Tableau 26).

B.2.2. Modèles d'analyse de la variation

Nous avons testé les associations entre ces 7 polymorphismes SNPs et la variation phénotypique des caractères quantitatifs, par des modèles linéaires combinant analyses de variance (ANOVA) et régression en utilisant la procédure PROC GLM de SAS (1999).

Tableau 25. Caractéristiques des 7 SNPs retenus parmi les 27 initialement choisis. FAR : fréquence de l'allèle le plus rare, F_{ST} : différenciation génétique entre les 10 populations contrastées, F_{IS} : coefficient de consanguinité moyenné sur l'ensemble des 18 populations.

SNP	Type de polymorphisme	FAR (données de séquences)	FAR (données génotypiques)	F_{ST} (données de séquences)	F_{ST} (données génotypiques)	F_{IS} (données génotypiques)	Nature de la mutation
<i>PR-AGP4 S536</i>	T/C	0.054	0.090	0.087	0.238	-0.069	intron
<i>PR-AGP4 S1128</i>	T/G	0.090	0.065	0.061	0.081	0.002	ns ^(a)
<i>PR-AGP4 S1135</i>	C/T	0.090	0.043	0.061	0.070	0.109	ns ^(a)
<i>Glucan S2590</i>	C/T	0.388	0.258	0.111	0.139	-0.359	intron
<i>Dhn1 S187</i>	A/G	0.308	0.254	0.063	0.076	0.105	s ^(b)
<i>Dhn1 S242</i>	G/T	0.083	0.081	0.048	0.035	0.051	ns ^(a)
<i>CCoAOMT 1229</i>	T/C	NA ^(c)	0.192	NA ^(c)	0.083	-0.103	intron

^(a) ns : mutation non-synonyme

^(b) s : mutation synonyme

^(c) NA : données manquantes

Tableau 26. Fréquences alléliques des 7 SNPs et moyennes phénotypiques de *HT03* (en m), *TB* (en g), $\delta^{13}C$ au sein de chaque population. Le nombre de familles (N_{fam}) ayant servi à estimer la moyenne phénotypique est également indiqué.

Population	N_{fam}	<i>HT03</i>	<i>TB</i>	$\delta^{13}C$	<i>PR-AGP4</i> <i>S536</i>	<i>PR-AGP4</i> <i>S1128</i>	<i>PR-AGP4</i> <i>S1135</i>	<i>Glucan</i> <i>S2590</i>	<i>dhn1 S187</i>	<i>dhn1 S242</i>	<i>CCoAOMT</i> <i>S1229</i>
15	15	109.242	478.309	-26.502	0.0%	0.0%	100.0%	80.6%	94.4%	0.0%	0.0%
20	18	77.796	335.241	-26.761	7.5%	12.5%	87.5%	57.5%	67.5%	2.5%	25.0%
21	28	78.637	328.076	-26.907	6.7%	10.3%	89.7%	62.1%	65.0%	5.0%	40.0%
22	21	66.082	250.684	-26.729	6.7%	7.1%	92.9%	91.7%	66.7%	5.2%	30.0%
23	30	68.203	282.638	-26.904	6.7%	8.3%	91.7%	80.0%	61.7%	8.6%	20.0%
24	19	78.747	316.547	-26.884	5.0%	10.0%	90.0%	72.5%	57.5%	15.0%	40.0%
25	29	67.351	308.259	-26.842	0.0%	1.9%	98.1%	64.5%	50.0%	11.3%	21.0%
26	30	63.962	275.673	-26.697	4.8%	6.7%	93.3%	76.7%	46.8%	11.7%	31.7%
27	20	91.99	384.747	-26.115	7.1%	5.0%	95.0%	64.7%	57.1%	4.8%	22.5%
29	27	64.738	292.276	-26.722	20.0%	13.8%	100.0%	69.2%	91.7%	13.3%	18.3%
30	18	47.004	201.562	-27.080	50.0%	26.5%	79.4%	55.0%	75.0%	10.0%	5.0%
40	27	85.241	298.228	-26.716	3.7%	3.6%	96.4%	62.5%	57.4%	16.1%	8.9%
41	28	85.441	279.468	-26.655	0.0%	1.8%	98.2%	94.8%	63.8%	17.2%	20.7%
42	28	85.412	269.283	-26.543	1.8%	1.8%	98.2%	83.9%	60.3%	17.2%	16.1%
43	29	82.426	275.44	-26.599	3.6%	0.0%	100.0%	68.3%	65.0%	16.7%	20.0%
44	22	82.077	270.398	-26.425	0.0%	0.0%	100.0%	72.4%	67.2%	22.4%	12.1%
46	15	86.581	318.862	-26.530	0.0%	0.0%	100.0%	85.5%	76.7%	16.1%	19.4%
50	10	91.073	489.705	-26.818	0.0%	0.0%	100.0%	34.6%	64.3%	0.0%	14.3%

- *Prise en compte de la structuration en populations*

Suivant les recommandations de la littérature sur les études d'associations, nous avons de façon préliminaire essayé d'incorporer dans un modèle mixte une matrice de structure Q inférée à partir du logiciel Structure ver 2.0 (Pritchard *et al.* 2000) sur la base de marqueurs neutres (Yu *et al.* 2006). Cependant, la structuration neutre ne reflète pas de façon suffisamment efficace la forte structure populationnelle existant chez cette espèce pour les caractères quantitatifs (voir la comparaison entre valeurs de Q_{ST} (Annexe 10) et de F_{ST} (Tableau 9 Chapitre II), donc la structure résiduelle induisait des tests d'associations significatifs uniquement dus à cette structure. Nous avons donc choisi d'intégrer la structuration entre populations connue *a priori*, en déclarant directement l'effet population comme facteur principal dans les modèles d'analyse de variance.

- *Tests de l'effet « classe génotypique » de la mère à chaque SNP*

L'effet des classes génotypiques de la mère pour chacun des SNPs a été testé de façon globale et au sein de chaque population, d'une part en utilisant les données individuelles des descendants et d'autre part sur les moyennes familiales ajustées des descendants pour chacun des trois caractères quantitatifs. Les données phénotypiques proviennent d'un ensemble de génotypes qui ne sont pas fixés pour le SNP, car ils vont dépendre à la fois du génotype de la mère et de celui du pool de pères potentiels (représentatifs des fréquences alléliques des SNPs dans chaque population). Cependant, les marqueurs étant bialléliques, le test de l'effet classe pourra donner une information partielle sur des éventuels effets de dominance au marqueur (par exemple, pour un SNP A/G avec A qui serait dominant sur G, tous les descendants d'une mère A/A donneront des phénotypes [A] pour le marqueur).

Donc l'effet global du SNP a été testé en le déclarant comme facteur principal dans le modèle suivant :

$$P_{ijk(l)} = \mu + P_i + G_j + E_{ijk(l)} \quad (a)$$

Avec $P_{ijk(l)}$, la valeur phénotypique de l'individu (l) de la famille k pour la population i et la classe génotypique j , de variance σ^2_p ,

μ , la moyenne générale du caractère,

P_i , l'effet population i aléatoire de variance σ^2_p ,

G_j , l'effet de la classe génotypique j aléatoire de variance σ^2_g ,

$E_{ijk(l)}$, l'erreur résiduelle de variance σ^2_e .

Dans ce qui précède, il suffit d'enlever le suffixe (l) pour avoir le modèle déclaré sur les moyennes phénotypiques des familles k .

L'effet SNP au sein de chaque population a été testé en hiérarchisant l'effet SNP à l'effet population, dans le modèle précédent, G_j devient $G_{j/i}$ (effet classe génotypique hiérarchisé à l'effet population i , toujours en déclarant l'effet population en facteur principal. L'objectif est de pouvoir détecter des effets SNPs qui seraient plus ou moins importants ou de sens opposé au sein de certaines populations, et donc qui ne serait pas significatifs en testant leurs effets globalement sur l'ensemble des populations (dans le modèle (a)).

Des modèles plus complets ont été déclarés comprenant les effets familles hiérarchisés aux effets populations, ce qui permettait d'expliquer une proportion plus importante de la variation totale. Cependant, les résultats sur la significativité des tests des effets SNPs étaient similaires et donc les modèles détaillés ne sont pas présentés ici.

- Test de l'effet « dose allélique » du génotype de la mère

Nous avons testé l'effet de la dose d'un allèle à chaque SNP par régression sur les valeurs individuelles des descendants (ou les moyennes familiales ajustées des descendants). Dans ce modèle, la dose d'un des deux allèles (ou de l'autre) est considérée comme une variable quantitative X_j qui prendra les valeurs 0, 1 ou 2 : par exemple, si l'on dose l'allèle G alors le génotype G/G sera codé 2, le génotype G/C sera codé 1, et le génotype C/C sera codé 0).

Puis on déclare le modèle suivant :

$$Y_{ijk(l)} = \mu + P_i + \beta \cdot X_j + R_{ijk(l)} \quad (b)$$

avec $Y_{ijk(l)}$, la valeur phénotypique de l'individu (l) de la famille k pour la population i ayant la dose allélique j , de variance σ^2_p ,

μ , la moyenne générale du caractère,

P_i , l'effet population i aléatoire de variance σ^2_p ,

$\beta \cdot X_j$, la valeur prédite par la régression de la valeur phénotypique sur la dose allélique, avec β , le coefficient de régression,

$R_{ijk(l)}$, le résidu de la régression de variance σ^2_R .

De même, on obtient le modèle utilisé pour analyser les données de moyennes phénotypiques familiales en enlevant l'indice (l).

Dans le cas où les effets des SNPs sont seulement additifs, les résultats entre les modèles (a) et (b) devraient être équivalents.

- Paramètres génétiques associés aux effets des SNPs

Nous avons calculé les coefficients de détermination phénotypique R^2_p de l'effet de chaque SNP, qui représentent la part de somme des carrés des écarts (SCE) totale (SCEt) expliquée par la SCE due à l'effet SNP. La variance génétique associée au SNP pour le test de l'effet classe, σ^2_g , a également été estimée à partir de l'espérance des carrés moyens (CM) du modèle d'ANOVA (a), selon la formule suivante, $\sigma^2_g = (\text{CM}_g - \text{CM}_e)/k$, avec $k=(N-1)/(C-1)$ qui tient compte du fait que le tirage du génotype dans chaque classe génotypique suit une loi binomiale qui dépend des fréquences alléliques (Charcosset et Gallais 1996). Nous avons ainsi pu aussi estimer le coefficient

$$R^2_G = \sigma^2_g / \sigma^2_G$$

qui représente la part de la variance génétique moyenne entre familles du caractère quantitatif (voir Tableau 21 Chapitre V) expliquée par l'effet génétique associé au SNP.

C. Résultats

C.1. Corrélations entre fréquences alléliques aux SNPs et moyennes phénotypiques

Les fréquences alléliques des SNPs *S1128* et *S1135* du gène *PR-AGP4* sont corrélées significativement à la hauteur (*HT03*) et à la composition isotopique ($\delta^{13}C$) qui est liée à l'efficacité d'utilisation de l'eau, sur l'ensemble des 18 populations considérées (Tableau 27). Ils expliquent respectivement 54% et 37% de la variation de *HT03*, et 34% et 33% de la variation de $\delta^{13}C$. Donc les populations les plus efficaces et avec une croissance importante possèdent un des deux allèles en fréquence majoritaire tandis que les populations moins efficaces et avec une croissance plus faible possèdent l'autre allèle en fréquence plus élevée. La représentation graphique de ces corrélations nous montre des nuages de points relativement homogènes, indiquant que cette corrélation n'est pas due seulement à des populations extrêmes (Figure 52). Pour le SNP *S536* de *PR-AGP4*, la corrélation est également significative avec *HT03* (Tableau 27), mais est due en grande partie à la population Tamrabta (Maroc) qui s'écarte nettement du nuage de points (Figure 52). Cependant, elle reste significative ($r=-0.51$) sans cette population. Enfin, le SNP *S242* du gène *dhn1* présente une corrélation significative avec la biomasse (Tableau 27), et explique 41% de la variation du caractère (Figure 52).

C.2. Résultats des tests d'associations entre SNPs et variation phénotypique

Dans tous les modèles utilisés, l'effet population explique pour chaque caractère la même proportion de la variation totale, avec de très petites différences dues aux différents effets déclarés et aux calculs des SCE ajustés correspondants (Tableau 28).

C.2.1. Effets globaux des SNPs

Parmi les 7 SNPs, 3 présentent des associations significatives avec l'un des trois caractères quantitatifs lorsque l'on teste l'effet global des classes génotypiques des SNPs (en se basant sur les données individuelles phénotypiques) (Tableau 28). Les SNPs *S1128* et *S1135* de *PR-AGP4* montrent des associations significatives principalement avec les caractères de croissance et biomasse, avec des R^2_P entre 0.1% et 1.2% (cas du SNP *S1128* avec *HT03*, Tableau 28). Les R^2_G (décrivant la part de variance génétique entre familles du caractère expliquée par le SNP) varient entre 2.3% et 3.3%. Pour le gène *PR-AGP4*,

Tableau 27. Corrélations entre les fréquences alléliques des 7 marqueurs SNPs, les moyennes phénotypiques des caractères *HT03*, *TB* et $\delta^{13}C$, et quelques données climatiques sur les 18 populations.

SNP	<i>HT03</i>	<i>TB</i>	$\delta^{13}C$	<i>ANMR</i> ^(a)	<i>ANMT</i> ^(b)	<i>RDRYM</i> ^(c)
<i>PR-AGP4 S536</i>	-0.68681** ^(d)	-0.43824	-0.45629	-0.19810	0.45273	-0.48484
<i>PR-AGP4 S1128</i>	-0.73602***	-0.43318	-0.58589*	-0.34945	0.27170	-0.53388
<i>PR-AGP4 S1135</i>	0.60983**	0.38513	0.57799*	0.21569	-0.23754	0.22186
<i>Glucan S2590</i>	0.06918	-0.37657	0.28260	-0.21856	-0.57733*	0.11402
<i>Dhn1 S187</i>	0.19277	0.23609	0.05726	0.29293	0.11133	-0.36120
<i>Dhn1 S242</i>	-0.15734	-0.63886**	0.13122	-0.33918	-0.38470	0.19863
<i>CCoAOMT 1229</i>	-0.23324	-0.17983	-0.18444	-0.37571	-0.39527	-0.08148

^(a) *ANMR* : Précipitations totales annuelles moyennes (voir Chapitre II).

^(b) *RDRYM* : Moyenne des précipitations totales du mois le plus sec (voir Chapitre II).

^(c) *ANMT* : Températures annuelles moyennes (voir Chapitre II).

^(d) Corrélations significatives à 0.05, *, 0.01, **, 0.001, ***, avec un test de Pearson à n-2 degrés de liberté.

Figure 52. Représentation graphique des corrélations entre les fréquences alléliques de quelques marqueurs SNP et les moyennes des caractères *HT03*, *TB* et $\delta^{13}C$ sur l'ensemble des populations.

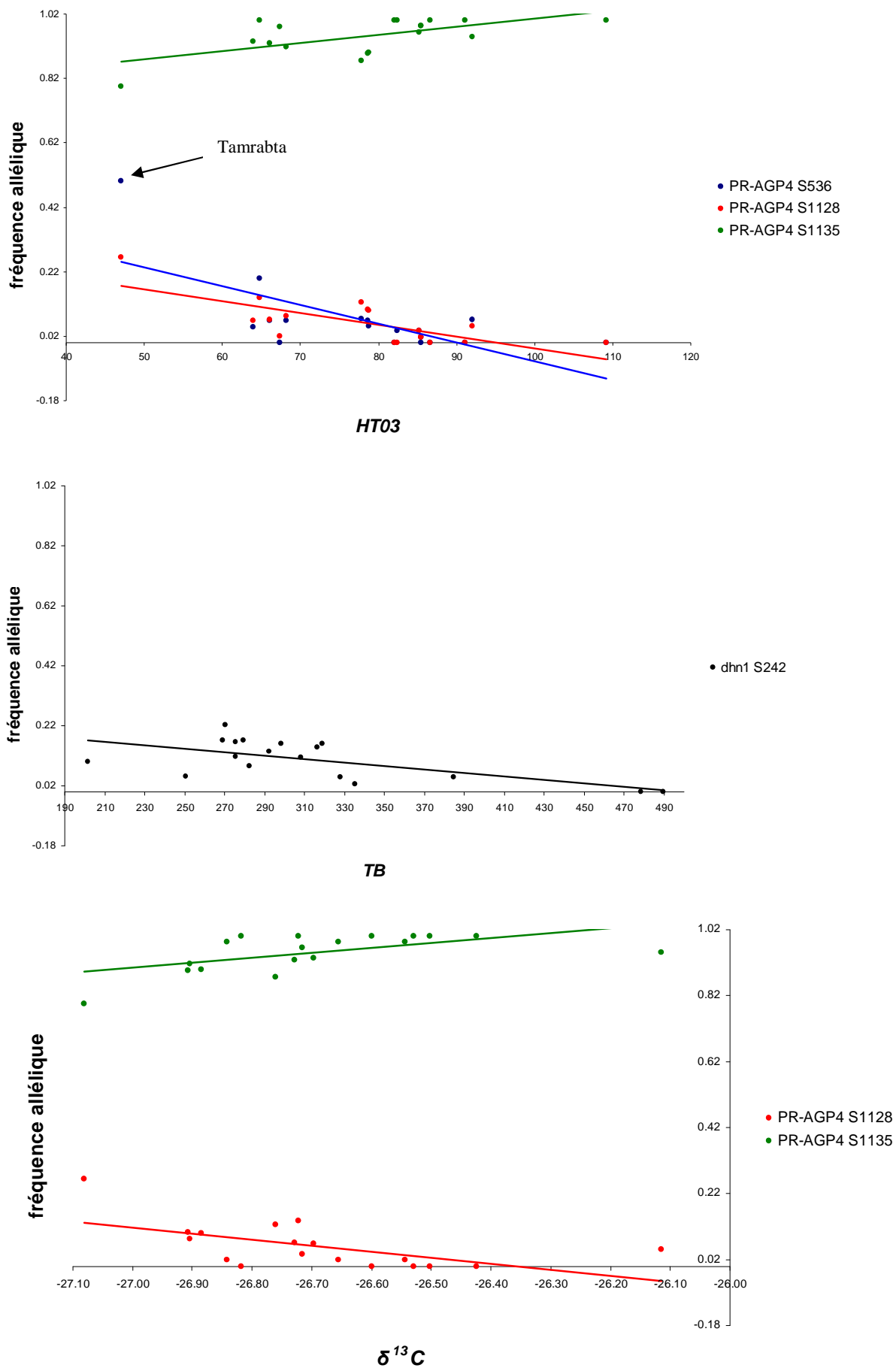


Tableau 28. Résultats des tests des effets « classe génotypique » et « dose allélique » au niveau global et en hiérarchisant l'effet du SNP au sein de la population. Sont indiqués les coefficients de détermination phénotypiques des effets des SNPs (R^2_P) et des effets des populations (R^2_{pop}), ainsi que le coefficient R^2_G représentant la part de la variance génétique du caractère quantitatif expliquée par le SNP. Seules les associations significatives sont indiquées.

SNP	Caractère	Type de données (a)	Effet global du SNP						Effet SNP hiérarchisé à la population				Effet population (ddl=14)
			Test classe génotypique (ddl=2)			Test dose allélique (ddl=1)			Test classe génotypique (12<ddl<28)		Test dose allélique (10<ddl<15)		
			R^2_P	R^2_G	R^2_P	R^2_P	R^2_P	R^2_P	R^2_P	R^2_P	R^2_{pop}		
<i>PR-AGP4</i> <i>S1128</i>	<i>HT03</i>	indiv	1.22%	** ^(b)	3.01%	1.06%	*	0.82%	**	0.63%	**	~34%	
	<i>TB</i>	indiv	0.08%	**	3.33%	0.0006%	**	1.02%	**	0.83%	**	~16%	
	$\delta^{13}C$	indiv											
<i>PR-AGP4</i> <i>S1135</i>	<i>HT03</i>	indiv	0.44%	**	2.56%	0.38%	*	0.74%	**	0.55%	*	~34%	
	<i>TB</i>	indiv	0.09%	*	2.34%	0.0005%	*	0.83%	*	0.63%	*	~16%	
	$\delta^{13}C$	indiv	0.81%	*	5.33%								
<i>PR-AGP4</i> <i>S536</i>	<i>HT03</i>	indiv						0.63%	*	0.51%	*	~35%	
	<i>TB</i>	indiv											
	$\delta^{13}C$	indiv											
<i>CCoAOMT</i> <i>1229</i>	<i>HT03</i>	indiv						1.26%	***	0.81%	**	~34%	
	<i>TB</i>	indiv						1.71%	***	1.22%	**	~8%	
	$\delta^{13}C$	indiv											
<i>Dhn1</i> <i>S187</i>	<i>HT03</i>	indiv						1.36%	**	0.91%	**	~34%	
	<i>TB</i>	indiv						1.60%	**	0.90%	*	~15%	
	$\delta^{13}C$	indiv moy	0.25%	*	5.87%	0.25%	**	2.83%	***	1.21%	**	~8%	
<i>Dhn1</i> <i>S242</i>	<i>HT03</i>	indiv											
	<i>TB</i>	indiv											
	$\delta^{13}C$	indiv						1.05%	*			~8%	
<i>Glucan</i> <i>S2590</i>	<i>HT03</i>	indiv						0.93%	*			~34%	
	<i>TB</i>	indiv						1.53%	***	1.22%	***	~15%	
	$\delta^{13}C$	indiv											

(a) Les résultats des ANOVA et régression sont sur les données individuelles (indiv) et sur les moyennes familiales (moy).

(b) *, **, *** : tests significatifs à 5%, 1% et 1% respectivement.

l'association significative la plus forte est entre le SNP *S1135* et le caractère $\delta^{13}C$ avec un R^2_G de 5.3%. Le SNP *S187* de *dhn1* présente une association significative avec le $\delta^{13}C$ avec une part de variance génétique entre familles du $\delta^{13}C$ expliquée de 5.9%, qui est la valeur de R^2_G la plus élevée parmi les associations significatives.

Les résultats des tests des effets de la dose allélique sont très similaires à ceux des tests des effets classes génotypiques aux SNPs pour les 2 associations entre *HT03* et SNPs de *PR-AGP4* (*S1128* et *S1135*), les valeurs de R^2_p étant légèrement inférieures pour les doses alléliques. Pour 3 autres associations, des différences importantes sont observées : l'association entre le SNP *S1135* de *PR-AGP4* et la variation du caractère $\delta^{13}C$ est significative (à 5%) uniquement sur les classes génotypiques, alors qu'elle ne l'est pas ($Pr < 0.07$) pour le test de la dose allélique (Tableau 28). D'autre part, les associations avec la biomasse totale (*TB*) pour les 2 SNPs *S1128* et *S1135* de *PR-AGP4*, bien que significatives expliquent beaucoup moins de variation phénotypique avec les modèles de tests de la dose allélique (Tableau 28).

On pourrait penser que ces différences seraient en partie dues à des effets de dominance qui ne sont pas modélisés dans les tests sur la dose allélique (effets additifs uniquement). De plus, on peut penser que cet effet possible de dominance sera testé de façon plus efficace si le pool des pères comprend une fréquence plus importante de l'allèle récessif. Par exemple, dans le cas d'un SNP A/G avec un allèle A dominant, les mères A/A auront toutes des descendants de phénotype A, quel que soit le génotype du père. Par contre, on prédit moins bien les phénotypes A des descendants pour les mères A/G et G/G, même si l'on suppose qu'il y aura moins de phénotypes A dans ce cas. Le nombre de phénotypes A dépendra alors de la fréquence de l'allèle A dans le pool des pères. Si ce pool a majoritairement l'allèle A dominant, le test sur les classes génotypiques des mères ne sera par informatif pour détecter les effets de dominance. Dans le cas contraire, le test sur les classes génotypiques des mères pourra donner sans doute une information supplémentaire comprenant à la fois des effets additifs mais également une partie des effets de dominance.

Enfin, les tests des effets classes génotypiques et doses alléliques lorsque l'on travaille sur les moyennes familiales ajustées des caractères phénotypiques ne montrent aucune association significative. Cette observation est en accord avec les simulations de Long et Langley (1999), et illustrerait le fait que même avec environ 500 données phénotypiques, la puissance pour tester les associations peut être très limitée.

C.2.1. Effets des SNPs hiérarchisés à l'effet population

Lorsque l'on teste les effets des SNPs hiérarchisés à l'effet population (que ce soit pour les classes génotypiques ou les doses alléliques), les associations qui sont significatives dans le cas du test de l'effet global du SNP, le sont aussi pour les effets hiérarchisés, avec cependant un pourcentage de variation expliquée par le SNP (R^2_P) en général supérieur (Tableau 28). Il est également intéressant de noter que l'on détecte des associations significatives supplémentaires dont le pourcentage de variation phénotypique expliquée varie entre 0.6 et 1.7%. C'est le cas des associations entre :

- le SNP *S536* (*PR-AGP4*) et *HT03* (R^2_P de 0.6%)
- le SNP *S1229* (*CCoAOMT*) et *HT03* (R^2_P de 1.3%) et $\delta^{13}C$ (R^2_P de 1.7%)
- le SNP *S187* (*dhn1*) et *HT03* (R^2_P de 1.4%) et *TB* (R^2_P de 1.6%)
- le SNP *S242* (*dhn1*) et $\delta^{13}C$ (R^2_P de 1.1%)
- le SNP *S2590* (*Glucan*) et *HT03* (R^2_P de 0.9%) et *TB* (R^2_P de 1.5%)

L'effet de ces SNPs serait donc dépendant du contexte génétique propre à chaque population. Cela signifierait que les effets des SNPs vont être sensibles aux interactions avec les autres locus potentiellement impliqués dans l'expression et la variation des caractères étudiés, et que leurs effets vont donc varier et pourront être plus ou moins importants dans chaque population. Par exemple, pour le SNP *S187* de *dhn1* qui est celui qui explique le plus de variation sur l'ensemble des tests (2.83% sur les données individuelles, et plus de 8% sur les moyennes phénotypiques de la biomasse), les classes génotypiques sont significativement différentes les unes des autres au sein des populations 20, 22, 25 et 46, mais avec des effets d'intensité différente au sein de chaque population (résultats non montrés).

Pour les modèles déclarant des effets hiérarchisés, les effets de la dose allélique sont en général significatifs, quand ils l'étaient pour les effets classes génotypiques, indiquant une majorité d'effets additifs au sein de chaque population. Cependant, les R^2_P sont en général inférieurs (jusqu'à 2 fois). Une illustration intéressante est celle du SNP *S242* de *dhn1*, dont l'effet classe génotypique est significatif alors que celui testant la dose allélique ne l'est pas (Tableau 28). Si l'on regarde plus en détail les tests des droites de régression au sein de chaque population, on observe qu'ils sont significatifs pour *TB* uniquement au sein des groupes de populations 2124 et 4041, mais avec un changement de signe de la pente entre ces

deux groupes de populations, ce qui explique que sur l'ensemble des populations, l'effet dose allélique ne soit pas significatif.

D. Conclusion

En conclusion sur les 7 SNPs retenus, seulement 3 (*S1128*, *S1135* de *PR-AGP4* et *S187* de *dhn1*) présentent des associations significatives au niveau global avec l'un des 3 caractères, les deux R^2_G les plus forts concernant le caractère d'efficacité d'utilisation de l'eau. Avec les modèles testant les effets SNPs au sein de chaque population, tous les SNPs présentent des associations significatives avec au moins l'un des trois caractères, l'ajustement intra-population permettant en général d'expliquer un pourcentage de variation phénotypique supérieur. Dans ces analyses préliminaires, les associations significatives n'expliquent cependant que très peu de la variation phénotypique du caractère (autour de 1%), même si la part de la variation génétique entre familles du caractère est supérieure (entre 1% et 6%). Les 3 associations les plus fortes sont observées avec le $\delta^{13}C$ entre un SNP de *dhn1* (*S187*, R^2_G de ~6%), un SNP de *CCoAOMT* (*S1229*, R^2_P de ~2%) et un SNP de *PR-AGP4* (*S1135*, *non-synonyme*, R^2_G de ~5%).

Chapitre VII

Discussion générale

Ce travail a essayé de répondre à trois questions principales :

1) Le potentiel adaptatif des populations naturelles de pin maritime est-il suffisant pour leur permettre de s'adapter aux changements climatiques annoncés, notamment à des sécheresses plus intenses ?

2) Quelle est l'importance de la sélection naturelle sur la variation de caractères liés à la réponse à la sécheresse, en particulier un caractère lié à l'efficacité d'utilisation en eau des arbres ?

3) Quelles sont les bases moléculaires de cette variation adaptative ?

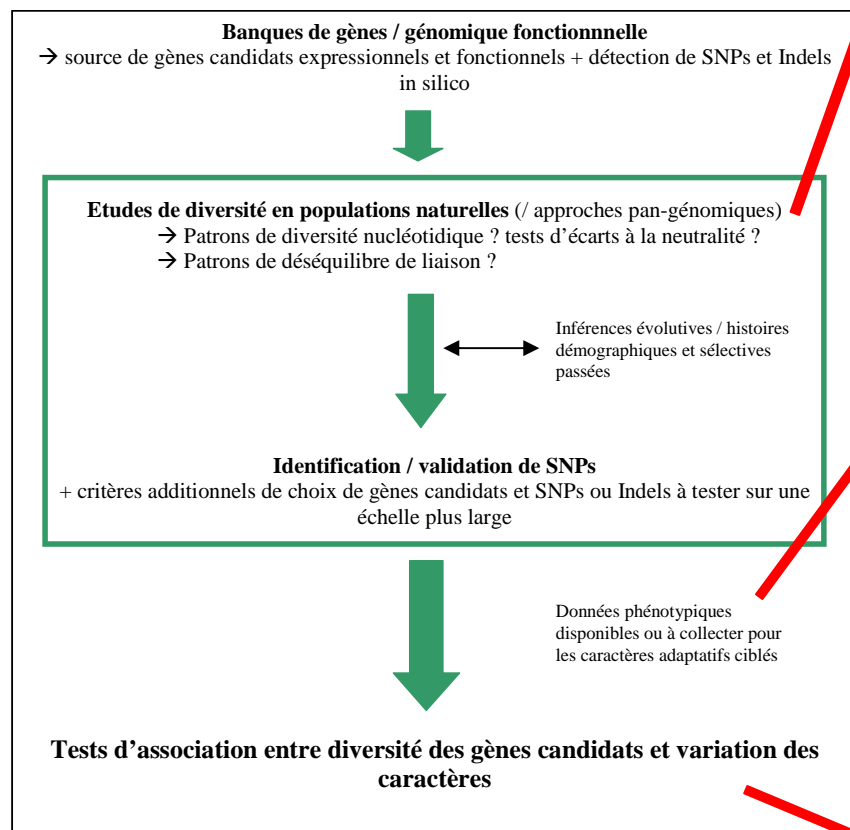
Pour répondre à ces questions, nous avons fait appel à plusieurs disciplines : la génomique des populations, la génétique quantitative et la génétique d'association (Figure 53), pour lesquelles nous avons collecté à la fois des données de diversité nucléotidique de gènes candidats et des données quantitatives pour des caractères potentiellement liés à l'adaptation chez le pin maritime. Nous discutons des résultats obtenus en vue d'apporter des réponses aux questions posées.

A. Le pin maritime face aux changements climatiques ?

Au cours de leur histoire évolutive et écologique, les espèces d'arbres forestiers ont déjà fait face à de nombreux changements environnementaux, sans perte apparente de diversité pour certaines malgré la recolonisation qui a suivi les dernières glaciations et les changements d'aires de distribution (Hamrick 2004). Le pin maritime a notamment évolué dans des conditions écologiques extrêmement variables : de sols sableux à calcaire, et sous des climats méditerranéens à atlantiques, avec pour conséquences l'évolution d'adaptations locales au sein de l'espèce pour la croissance et la survie (Danjon 1994, Alia *et al.* 1995, 1997, González-Martínez *et al.* 2004). Cependant, dans un contexte où l'amplitude des changements climatiques futurs prédits serait sans précédent, la question du potentiel adaptatif des populations de pin maritime pour répondre à ces changements reste posée.

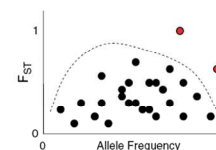
Nos résultats confirment la grande variabilité phénotypique et génétique d'un ensemble de populations représentatives du pin maritime pour des caractères liés à l'adaptation au déficit en eau (notamment un caractère de composition isotopique du carbone lié à l'efficacité d'utilisation en eau (*EUE*)). Nous avons observé cette grande variabilité génétique tout d'abord au niveau des héritabilités, qui indiquent un potentiel adaptatif

Figure 53. Stratégie utilisée au sein de cette étude faisant intervenir plusieurs disciplines.



Génomique des populations

- Détection d'outlier basée sur la différenciation génétique entre 10 populations contrastées
- Écarts à la neutralité basés sur la diversité génétique au sein de la métapopulation Aquitaine

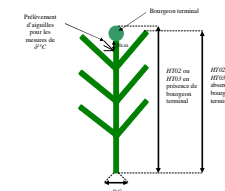


Importance du rôle de la sélection dans la variation moléculaire?

Identification de loci soumis à la sélection

Génétique quantitative

- Etude de la variabilité phénotypique et génétique entre populations et au sein des populations
- Études des structures des corrélations génétiques et environnementales



Potentiel adaptatif suffisant?

Etude d'association (prospectif)

Tests des effets des classes génotypiques et des doses alléliques des SNPs globalement et au sein des populations

$$P_{ijk} = \mu + p_i + g_j + e_{ijk}$$

Loci impliqués dans la variation adaptative?

important pour environ la moitié des populations atlantiques et espagnoles, quel que soit le caractère considéré (voir chapitre V), et par conséquent un potentiel important au niveau de l'espèce dans le contexte des changements climatiques futurs. Ce potentiel important a pu aussi être observé grâce aux indices de différenciation adaptative estimés entre populations, confirmant l'adaptation locale de certaines à des milieux très hétérogènes, que ce soit à l'échelle de l'espèce entre des milieux extrêmes (Maroc et Landes), ou bien à une échelle régionale (centre de l'Espagne).

Enfin, cette grande variation a été observée grâce aux corrélations génétiques ou environnementales estimées dans chaque population qui montrent en général l'absence de contraintes entre une meilleure *EUE* et une production de biomasse plus importante.

Ces résultats sont basés cependant sur des arbres jeunes, et il serait important de les confirmer dans de nouveaux essais multilocaux qui seraient conservés à plus long terme. Ceci est prévu dans le cadre du projet Novel Tree Breeding coordonné par C. Bastien de l'INRA d'Orléans et dont le laboratoire INRA de Pierroton est partenaire.

A.1 Variation des stratégies d'adaptation à un déficit en eau

Ces résultats ont pu être interprétés comme l'expression de stratégies d'adaptation différentes entre populations pour la réponse au stress hydrique (Chapitre V). Les populations ayant évolué dans des milieux plus humides (Corse, Landes, San Cipriano) sont les plus efficaces et auraient une stratégie basée sur la réactivité de la fermeture/ouverture des stomates. L'observation de meilleures performances pour les populations Pinia (Corse) et San Cipriano (Espagne de l'Ouest) en comparaison des populations landaises locales pourrait s'interpréter comme la conséquence de leur transfert dans un milieu légèrement moins stressant (moins chaud) mais proche de leur milieu d'origine (voir aussi discussion du chapitre V). A l'inverse, la population marocaine d'un milieu plus chaud qui est la moins efficace dans le lieu d'étude compenserait par un système racinaire plus développé. Des mesures de biomasse racinaire sont en cours de collecte pour ces populations et permettraient de confirmer cette stratégie. La population tunisienne aurait développé une stratégie différente avec une biomasse aérienne de bois élevée, et des aiguilles plus longues et plus fines. Enfin, les populations espagnoles du centre moins performantes que les populations atlantiques, montrent néanmoins une corrélation négative entre *EUE* moyennes et niveaux de précipitations, indiquant une adaptation à une meilleure *EUE* en conditions de stress hydrique.

A.2 Niveaux de plasticité phénotypique au sein des populations ?

Même si cette étude est limitée à un seul lieu de test, la comparaison des niveaux de variation résiduelle entre populations indique qu'il existe une plasticité différente des individus au sein des populations à l'échelle du micro-environnement. La population corse Pinia présente notamment une variance résiduelle parmi les plus élevées pour *HT03*, *TB* et $\delta^{13}C$, tandis que l'héritabilité de ses caractères est non-significativement différente de zéro, ce qui pourrait être la conséquence d'une forte pression de sélection. Cette population comprendrait donc des génotypes peu différents mais relativement plastiques. De plus, les corrélations résiduelles positives entre *EUE* et *HT03* ou *TB* au sein de cette population confirmeraient l'existence d'une forte similarité de réponses entre génotypes. On peut supposer que cette population située sur un sol sableux en bord de mer, aurait évolué dans un milieu plus drastique (stress salin important) avec des conditions plus hétérogènes au cours du temps que dans d'autres populations. La variation des précipitations au cours des années d'autre part parmi les plus élevées avec celle de la population Landaise. Cela aurait conduit à la sélection de génotypes plastiques en milieu hétérogène sous l'action de fortes pressions de sélection.

A l'inverse, les populations espagnoles seraient moins efficaces dans notre lieu d'étude plus humide que leur milieu d'origine, tout en restant globalement performantes (valeurs de *TB* intermédiaires), mais conserveraient un fort potentiel adaptatif qui pourrait être soit la conséquence de l'expression de familles mal-adaptées dans un milieu moins drastique et donc d'une variabilité de la plasticité dans ces populations, soit celle de maintien d'une forte hétérogénéité issue de la sélection dans leur milieu d'origine. L'observation de ces populations dans des milieux proches de leurs milieux d'origine permettrait de valider cette hypothèse.

Les 2 types de comportements que nous venons de décrire auraient pu être engendrés par de la sélection en milieux hétérogènes qui conduirait sur l'ensemble de l'espèce à une variabilité de la plasticité individuelle entre populations, cette plasticité étant elle-même variable entre familles des populations espagnoles (forte hétérogénéité, faibles corrélations additives entre caractères), ou bien montrant une similarité de réponses des génotypes au sein de la population Corse.

A.3 Prédiction des performances des populations dans un milieu plus aride ?

Au-delà de l'évaluation de la plasticité phénotypique, la question est de savoir si l'on peut faire quelques prédictions sur les performances des populations dans un environnement plus sec. Dans le cas de la population corse, on peut imaginer que sa plasticité pourrait s'extrapoler à une variation plus importante entre milieux. La plasticité phénotypique en général est d'ailleurs une caractéristique des arbres forestiers, et aurait contribué à leur survie lors des changements passés de conditions environnementales (Hamrick 2004, Petit et Hampe 2006). Chez le pin maritime, de fortes interactions génotypes x environnement ont été rapportées pour plusieurs provenances, avec notamment les provenances atlantiques ayant des performances beaucoup plus faibles en milieux secs, tandis que les génotypes des populations espagnoles sont beaucoup plus stables dans des environnements plus secs (Alia *et al.* 1997). D'autre part, des résultats récents visant à étudier plus finement la plasticité moléculaire de populations contrastées de pin maritime (Tamrabta et Mimizan) au niveau transcriptomique et protéomique, ont également mis en évidence l'existence d'une plasticité moléculaire importante (propriété d'un génotype à modifier l'expression de son génome en réponse à son environnement) (Chaumeil 2006).

A.4 Perspectives pour l'amélioration du pin maritime et la conservation de ses ressources génétiques

Si certaines populations montrent un potentiel adaptatif important, elles ne sont pas toutes performantes, en comparaison de la population Pinia peu variable mais plus performante et plus efficiente, et dont les génotypes seraient parmi les plus plastiques. Une perspective intéressante serait le croisement de cette population avec la population San Cipriano, qui présente à la fois des performances élevées et un fort potentiel adaptatif. Néanmoins, outre l'aspect résistance à la sécheresse, les changements climatiques pourraient provoquer une hétérogénéité des conditions provoquant des gels importants dans nos régions. Il faut donc rester prudent sur les propositions de transfert de populations du sud vers le nord, car elles pourraient être particulièrement sensibles à un gel rigoureux, comme cela avait été observé sur les populations portugaises qui avaient servi de sources de graines en Aquitaine pour replanter les forêts après les grands incendies de 1949 (Bonneau *et al.* 1969). Par ailleurs, la population corse Pinia est plus sensible à *Matsucoccus feytaudi* (Harfouche *et al.*

1995), et les données de résistance aux maladies et aux champignons doivent être prises en compte dans les stratégies d'amélioration et de conservation du pin maritime.

Finalement, l'existence d'adaptations locales et la possibilité d'interactions importantes entre populations et environnements pour l'efficacité d'utilisation de l'eau (déjà observées pour croissance et survie précédemment (Alia *et al.* 1997, González-Martínez *et al.* 2004) prône une stratégie d'amélioration intégrant les adaptations spécifiques à certaines régions.

B. Variation des niveaux et des patrons de diversité moléculaires

Nous avons mis en évidence des niveaux de diversité nucléotidique globalement comparables à ceux d'autres conifères et plus faibles que chez des espèces annuelles. Nous avons vu que ces niveaux de diversité nucléotidique, en comparaison à une diversité moléculaire plus élevée à d'autres marqueurs, ne pouvaient être interprétés uniquement comme la conséquence d'un taux de mutation faible ou d'évènements démographiques, mais que la sélection naturelle pouvait également avoir joué un rôle non négligeable. Nous avons trouvé à la fois des écarts à la neutralité significatifs sur les patrons de diversité nucléotidique au sein d'une métapopulation atlantique et sur les patrons de différenciation nucléotidique entre 10 populations contrastées. Ces résultats nous ont permis de proposer différents scénarios de sélection spécifique à chaque gène (voir chapitre IV).

B.1 Inférences évolutives : scénarios de sélection ou effets de l'histoire démographique des populations?

L'interprétation de ces écarts à la neutralité en termes de signatures de sélection semble justifiée pour plusieurs raisons. Tout d'abord, les résultats obtenus à partir des méthodes basés sur le F_{ST} sont relativement robustes à la démographie, notamment à la structure en populations, et l'observation d'« outliers » à la fois positifs et négatifs est difficilement compatible avec l'effet d'évènements démographiques seulement.

De plus, le sens des signaux observés sur les 10 populations sont cohérents avec ceux de la métapopulation atlantique : par exemple, le gène *PR-AGP4* présente une signature de hitch-hiking (auto-stop génétique) au sein de la métapopulation, cohérente avec une signature de sélection diversifiante entre populations. Pour *dhn1*, la signature de sélection balancée

observée au sein de cette métapopulation confirmerait la signature de sélection homogénéisante détectée entre populations.

Enfin, nos résultats montrent une grande hétérogénéité dans les signaux entre gènes sur l'ensemble des populations (diversifiante ou homogénéisante) mais également au sein de la métapopulation atlantique (directionnelle ou balancée), ce qui pourrait être le résultat de différentes formes de sélection affectant des loci en particulier.

Cependant, il est possible que l'histoire démographique des populations ait également agit en interaction avec ces évènements de sélection. La détection des signatures de sélection peut en effet dépendre de l'importance d'autres forces évolutives. Par exemple, Muirhead (2001) montre qu'il existe un niveau critique de taux de migration au-delà duquel la sélection ne pourra être détectée. Dans notre cas, les populations Marocaines, Tunisienne et Pleucadeuc montrent des effets spécifiques aux populations (significatifs dans l'analyse bayésienne), qui ont été interprétés comme la conséquence de flux de gènes plus limités avec les autres populations (Eveno *et al.* 2008). Cette situation, ainsi que des conditions environnementales plus arides ou plus chaudes dans le cas de Tamrabta et Tabarka ont sans doute permis des adaptations locales, et pourrait expliquer les observations de différenciation plus importantes entre ces populations et les autres. De plus, le fait d'avoir une structuration neutre relativement importante dans cette espèce, à cause de ces populations aux extrémités de l'aire de distribution, donne sans doute plus de poids à l'interprétation en termes de signatures de sélection homogénéisante entre populations (qui maintient des haplotypes similaires). En effet, ce type de sélection si elle n'est pas assez forte, risque facilement d'être masquée par la structuration, et elle a été interprétée comme la conséquence possible de conditions de milieu hétérogènes dans le temps et dans l'espace (partie C.4.2 chapitre IV).

Néanmoins, même si plusieurs évidences permettent de penser que des évènements démographiques n'ont pas affecté la diversité des populations de façon importante (conservation d'un fort potentiel adaptatif dans les populations, large diversité aux SSRs, peu de différenciation sur un axe de recolonisation Sud-Nord (Gonzalez-Martinez *et al.* 2002), il faudrait vérifier de façon plus approfondie les effets possibles d'une interaction complexe avec des évènements démographiques qui pourraient varier d'une population à l'autre. Pour cela, il faudrait obtenir plus de données de séquences (voir Conclusion et Perspectives) afin d'avoir une distribution empirique des statistiques de tests qui serait représentative de l'histoire évolutive de l'échantillon (voir Akey *et al.* 2002 chez les humains). Une autre alternative serait d'extraire par des méthodes de coalescence, une information contenue dans les marqueurs SSRs, permettant d'identifier d'éventuels effets de goulot d'étranglement ou

d'expansion (Beaumont 1999, Storz et Beaumont 2002, Beaumont et Rannala 2004). Pour l'instant, les données de 8 SSRs basées sur 20 à 30 individus par population n'ont été exploitées que pour estimer les niveaux moyens de différenciation neutre entre populations (Eveno *et al.* 2008, informations supplémentaires, Annexe 2). Une autre approche possible serait de simuler, à partir des données de séquences des gènes candidats, différents scénarios démographiques, par exemple en faisant varier l'intensité et la durée de goulots d'étranglement (Pyhäjärvi *et al.* 2007) pour vérifier si le type d'hétérogénéité observée au sein de cette étude pourrait être expliqué par de tels évènements démographiques.

B.2 Détection de signatures de sélection : quelles méthodes, quelles données et quels échantillonnages ?

B.2.1 Des méthodes complémentaires

Malgré la cohérence des interprétations entre les méthodes (voir ci-dessus), nous avons vu aussi que des signaux sur certains gènes mis en évidence par les méthodes basés sur le F_{ST} n'étaient pas détectables par celles basées sur la distribution des fréquences alléliques au sein de la métapopulation atlantique. Ceci est le cas par exemple de *GRP3*. Réciproquement pour d'autres gènes (par exemple, *CCoAOMT*, *Glucan*), des signatures de sélection balancée pouvaient être observées au sein de plusieurs populations, alors que la différenciation entre les populations n'était pas significativement différente de l'attendu neutre. Ce résultat était uniquement dû au fait que les allèles maintenus par la sélection balancée étaient différents d'une population à l'autre. Il semble donc qu'aucune méthode ne soit efficace pour détecter toutes les formes de sélection ou d'écarts à la neutralité et à l'équilibre démographique. Plusieurs auteurs recommandent en effet de combiner plusieurs méthodes pour détecter des signatures de sélection à des échelles de temps différentes (Storz 2005, Biswas et Akey 2006, et voir Chapitre I). Les résultats de notre étude (chapitre IV) confirment la nécessité d'une telle stratégie. D'autres méthodes testant conjointement plusieurs statistiques de tests et qui seraient moins sensibles à la démographie pourraient également être utilisées sur les mêmes jeux de données (Zeng *et al.* 2006).

B.2.2 Gènes candidats versus marqueurs anonymes et haplotypes versus SNPs ?

Nous avons vu également l'importance du choix des gènes candidats et du type d'information utilisée (haplotypique ou biallélique) pour la détection de ces signatures de sélection. En effet, le choix raisonné de gènes candidats fonctionnels et expressionnels a peut-être contribué à obtenir une proportion de gènes « outliers » (45%) plus importante que la proportion d'« outliers » détectée (au maximum 6%) dans plusieurs autres études basées sur le F_{ST} utilisant des marqueurs anonymes (Bonin *et al.* 2006, Wilding *et al.* 2001, Campbell et Bernatchez 2004, Acheré *et al.* 2005). Ceci rejoindrait les conclusions de deux autres études pan-génomiques privilégiant la détection d'« outliers » au sein de gènes (Vasemägi *et al.* 2005, Scotti-Saintagne *et al.* 2004) (voir discussion dans Eveno *et al.* 2008, Annexe 1).

D'autre part, la proportion d'« outliers » détectés à 5% sur l'ensemble des SNPs dans notre étude (11% de 94 SNPs) était inférieure à celle estimée au niveau haplotypique (45% de 11 gènes candidats), ce qui pourrait indiquer que des combinaisons particulières de SNPs au sein des gènes aient une signification fonctionnelle plus importante que les SNPs pris isolément. Ce résultat peut aussi être dû à une puissance plus importante des tests utilisant l'information haplotypique, à cause d'une meilleure précision de l'estimation du F_{ST} (Balding 2003). Le niveau haplotypique peut aussi être très informatif dans le cas de tests de neutralité basés sur le nombre attendu de ces haplotypes, comme dans le cas du test F_S de Fu dont les résultats se sont révélés positifs et significatifs dans différentes populations.

B.2.3. Perspectives d'amélioration des méthodes de détection d'« outliers »

Hormis leur robustesse à de nombreux scénarios démographiques (Beaumont 2005, Nielsen 2005), les approches basées sur les niveaux de différenciation utilisent jusqu'à présent un seul estimateur, le F_{ST} , qui considère tous les allèles comme équidistants, et qui semble peu approprié aux données de séquences. Nous avons au cours de cette étude estimé un autre indice, le N_{ST} , qui capture l'information des distances généalogiques entre allèles, et dont les valeurs étaient parfois le double de celles du F_{ST} (par exemple, pour *GRP3*) (Eveno *et al.* 2008). Il pourrait être judicieux de développer ces méthodes, en essayant de simuler non pas des F_{ST} mais des N_{ST} . Dans le même ordre d'idées, d'autres estimateurs utilisant l'information contenue dans les données de séquences pourraient être exploités. En particulier, la statistique Snn basée sur l'appartenance à une population des haplotypes les plus proches, serait plus puissante pour détecter de la différenciation que d'autres statistiques

adaptées aux données de séquences, pour une large gamme de valeurs de paramètres de mutation, de migration et de recombinaison (Hudson 1992, 2000).

B.2.4 Questions d'échantillonnage...

La question de l'échantillonnage se pose à la fois pour les méthodes basées sur le F_{ST} (voir la discussion dans Eveno *et al.* 2008, Annexe 1) et pour celles basées sur la distribution des fréquences alléliques. Il s'avère que les preuves de signatures sélectives chez les arbres forestiers restent encore limitées (voir chapitre IV, Tableau 2). Une hypothèse possible serait que l'absence de signaux soit due à un échantillonnage insuffisant pour effectuer par exemple un test de Tajima avec suffisamment de puissance (par exemple, 5 gamètes par population chez *Pinus sylvestris*, Pyhäjärvi *et al.* 2007). Dans cette étude, nous avons choisi de regrouper des populations de la côte atlantique française dans une métapopulation pour augmenter la taille d'échantillonnage. Le choix de ce regroupement était initialement basé sur l'absence de différenciation moléculaire moyenne entre populations au niveau des SSRs et des gènes candidats. *A posteriori*, connaissant les résultats de différenciation phénotypique, et de différenciation moléculaire par paires de populations sur les gènes candidats, nous aurions sans doute dû exclure la population Olonne/Mer de ce groupe, puisqu'elle présente une différenciation de 0.13 avec les autres populations pour le $\delta^{13}C$, ainsi que la population Pleucadeuc qui montre une différenciation de 9% avec Mimizan pour *GRP3* seulement. Ce point méthodologique soulève un problème plus général, puisque l'absence de structure en populations, condition requise pour appliquer un grand nombre de tests, est souvent vérifiée à l'aide de marqueurs moléculaires neutres, alors qu'il y a peu de chances qu'ils reflètent de façon efficace la différenciation à des caractères adaptatifs ou à des gènes impliqués dans l'adaptation (voir ci-dessous).

C. Associations entre variabilité phénotypique et diversité moléculaire

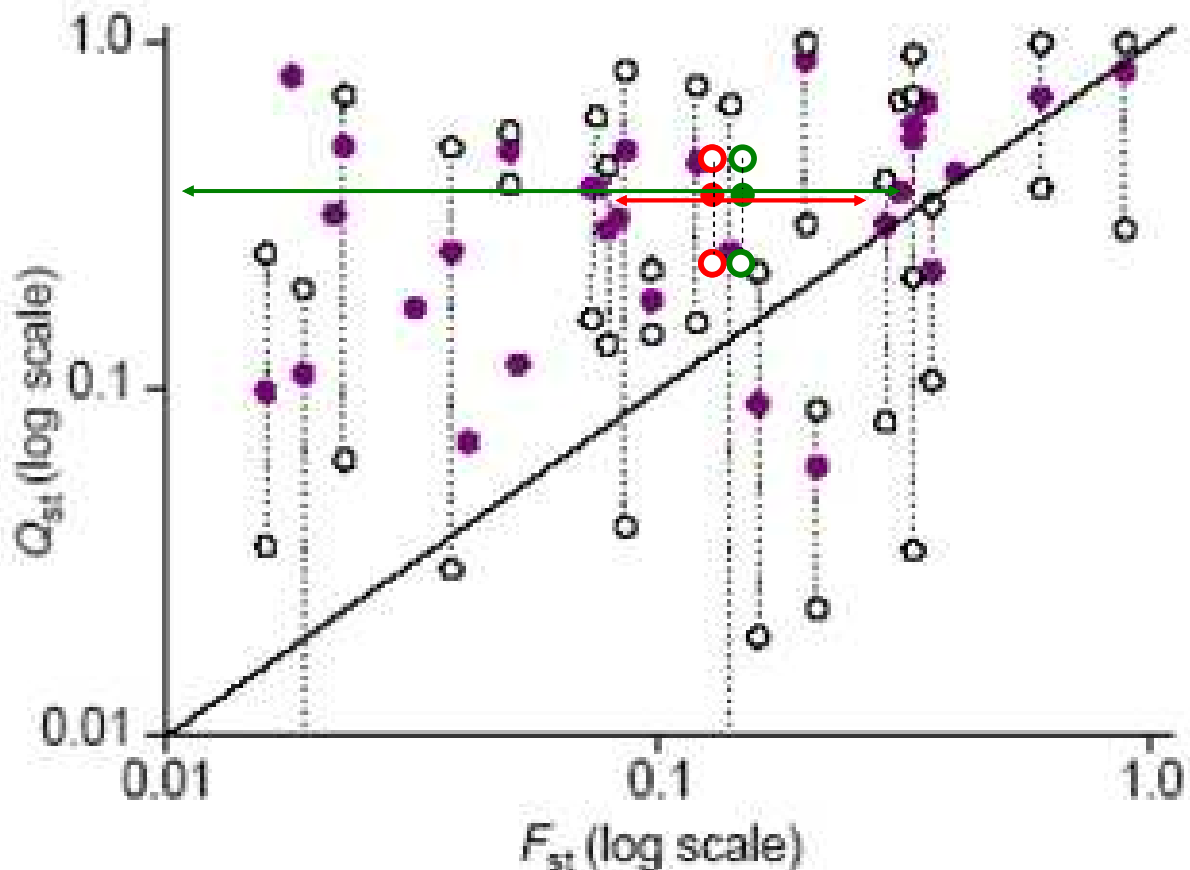
Avant l'ère de la génomique, la question des relations entre variation moléculaire et phénotypique était abordée à l'aide de marqueurs *a priori* neutres. Ces premières approches ont servi en réalité surtout à confirmer dans de nombreux cas le rôle de la sélection naturelle dans la variation et l'évolution de caractères quantitatifs, en montrant de façon empirique que les niveaux de différenciation génétique entre populations étaient généralement supérieurs à ceux estimés aux marqueurs moléculaires, et donc que ces caractères étaient impliqués dans

l'adaptation des populations à leur environnement local (McKay et Latta 2002, Pfrender *et al.* 2000, Beckeszy *et al.* 2003, Figure 54).

Cependant, dans la pratique, peu de relations simples ou répétables entre diversités neutre et adaptative ont été observées. Les quelques méta-analyses sur le sujet montrent que l'interprétation des résultats dépend de l'histoire des populations, de l'importance relative des différents mécanismes évolutifs en jeu, et du niveau auquel sont étudiées ces associations : un niveau individuel en comparant hétérozygotie individuelle et fitness (Britten 1996), un niveau populationnel en comparant variance moléculaire (H_S d'une population sur un ensemble de marqueurs) et variance génotypique intra-population des caractères (Reed et Frankham 2001), ou encore un niveau « espèce » en comparant différenciation génétique entre marqueurs et caractères quantitatifs (F_{ST} versus Q_{ST} (Spitze 1993)) (Merila et Crnokrak 2001). En effet, la diversité génétique neutre, est principalement affectée par les mécanismes de mutation, de migration et de dérive (Kimura 1968, 1983), et enregistre donc l'histoire démographique (au sens large) des populations (Godelle *et al.* 1998). Ces mécanismes affectant l'ensemble du génome, ils peuvent aussi en théorie agir de façon indirecte sur la diversité de gènes impliqués dans les caractères soumis à sélection: par exemple dans le cas d'un goulot d'étranglement ou d'un effet de fondation, ou bien dans le cas d'une forte structuration spatiale (Strauss *et al.* 1992, Godelle *et al.* 1998). Dans ce dernier cas, la diversité moléculaire a pu être valorisée pour proposer des stratégies de conservation de la variabilité adaptative (Schoen et Brown 1993, Bataillon *et al.* 1996).

Les faibles corrélations entre diversité neutre et adaptative pourraient donc s'expliquer également par des différences de taux de mutation (plus faible au niveau des marqueurs moléculaires), des effets non additifs (dominants et épistatiques) plus importants dans la variation des caractères quantitatifs ainsi que d'autres propriétés théoriques et statistiques (Lynch 1996). Bien que l'on reconnaisse désormais qu'il serait illusoire de ne s'intéresser qu'à la diversité neutre seulement si l'on veut conserver le potentiel adaptatif des populations (Crandall *et al.* 2000, Garnier-Géré *et al.* 2001), la diversité adaptative a souvent été beaucoup moins étudiée car elle est difficile et coûteuse à mesurer, notamment chez les espèces d'arbres forestiers dont le temps de générations est souvent de plusieurs dizaines d'années (10-15 ans chez le pin maritime pour une maturité sexuelle plus précoce que de nombreuses autres espèces où l'âge reproductif serait plutôt de l'ordre de 30 à 50 ans).

Figure 54. Valeurs de Q_{ST} et F_{ST} chez 29 espèces. Chaque point correspond à la valeur moyenne des F_{ST} et Q_{ST} pour une espèce, et les barres verticales et les points blancs représentent la variation des valeurs Q_{ST} entre les différents caractères au sein d'une même espèce. La diagonale correspond à la relation attendue sous l'hypothèse de neutralité ($F_{ST} = Q_{ST}$) (d'après McKay et Latta 2002). Nous avons rajouté les valeurs moyennes de F_{ST} à nos marqueurs neutres (points rouges) et nos gènes candidats (points verts) ainsi que leurs gammes de variation (flèches) par rapport aux valeurs moyennes de Q_{ST} aux caractères quantitatifs mesurés dans le chapitre V.



C.1. F_{ST} versus Q_{ST} chez le pin maritime ?

L'accès à un nombre croissant de séquences de gènes candidats, même chez des espèces non-modèles, pouvait laisser espérer que l'on observerait plus facilement des F_{ST} élevés pour des gènes impliqués dans la variation de caractères adaptatifs. C'est l'hypothèse de base des méthodes ayant pour but de détecter des « outliers » positifs, qui seraient soumis à de la sélection directionnelle, et ces méthodes nous ont permis de détecter 2 gènes rentrant dans cette catégorie (Eveno *et al.* 2008). Cependant, sur l'ensemble des gènes étudiés, les valeurs de F_{ST} sur 10 populations représentatives sont très variables (plus variables que celles des SSRs), et le F_{ST} moyen des gènes candidats se situe juste au-dessus du F_{ST} moyen aux SSRs mais bien en-dessous de la valeur moyenne pour le Q_{ST} comme la majorité des espèces représentées sur la figure 54. Même dans le cas théorique où l'on aurait pu étudier la différenciation de vrais gènes à effets quantitatifs (QTLs), on ne s'attend pas forcément à ce qu'ils montrent tous une différenciation élevée dans un scénario de sélection diversifiante entre populations avec des flux de gènes importants. En effet, il a été montré analytiquement et par simulations que dans ce cas, la différenciation du caractère sélectionné dépendait plus d'un terme exprimant les covariances entre fréquences et effets alléliques sur l'ensemble des populations que des valeurs individuelles de F_{ST} à chaque QTL (Le Corre et Kremer 2003, Latta 2003). De plus, le contraste entre Q_{ST} et F_{ST} sera d'autant plus fort que l'on se trouve dans un scénario où la sélection entre populations est forte (par exemple le long d'un cline), ce qui est couramment observé chez les espèces d'arbres forestiers, mais où la sélection intrapopulation optimisante est intermédiaire à faible, avec de forts flux de gènes entre populations. Dans leurs simulations, Le Corre et Kremer (2003) observent aussi que plus de la moitié des QTLs montrent des valeurs de F_{ST} similaires à celles de marqueurs neutres, qui ne seraient donc pas détectés par des méthodes basées sur le F_{ST} . Nos résultats sont cohérents avec ces prédictions théoriques, et illustrent l'une des limites des méthodes de détection d'outliers, indiquant aussi dans le futur la nécessité d'aborder ces approches sous l'angle multilocus, au fur et à mesure que la taille des jeux de données (en nombre de gènes et d'individus) le permettront (Holland 2007, Barton et Keightley 2002 et voir ci-dessous), et de mieux intégrer génomique des populations et génétique quantitative.

C.2. Bases génétiques de la variation pour la tolérance au stress hydrique en populations naturelles : utopie ou réalité proche ?

Cette meilleure intégration semble possible dans un futur très proche avec le développement des études d'associations (voir chapitre VI) chez des espèces non-modèles (chez lesquelles les ressources génomiques sont encore limitées). Ces études vont permettre de réconcilier les recherches de génétique évolutive en populations naturelles et celles visant à identifier des gènes impliqués dans la variation de caractères importants pour l'adaptation au sein d'une nouvelle discipline, la génomique fonctionnelle évolutive et écologique (Evolutionary and Ecological Functional Genomics (EEFG) en anglais) (Feder et Mitchell-Olds 2003, Mitchell-Olds *et al.* 2008). En effet, alors que les méthodes de détection de QTLs ne permettent pas de localiser ces gènes de façon précise, les études d'associations vont permettre de tester directement l'effet sur le phénotype des candidats, retenus par exemple selon des critères issus des analyses de diversité (Figure 53, et voir chapitre IV). L'observation de co-localisations avec des QTLs peut constituer (plus rarement) un argument supplémentaire de l'intérêt potentiel de certains de ces gènes, qui pourront être déterminés sur un grand nombre de génotypes et de populations (par exemple Krutovski et Neale 2005). Plus généralement, la combinaison d'informations de cartographie génétique (même si cela nécessite de gros investissements chez les arbres forestiers) et de génomique des populations en vue d'une validation dans des études d'associations peut constituer une approche particulièrement puissante pour révéler les mécanismes moléculaires responsables de l'adaptation (Stinchcombe et Hoekstra 2008, Figure 53). Le projet Treesnips et cette thèse en particulier s'inscrivent dans cette démarche, mais avec un petit nombre de gènes candidats et une partie de cartographie des gènes qui n'a pu être réalisée à cause des problèmes de génotypage (voir chapitre VI).

Les résultats du chapitre IV nous ont permis de proposer des critères de choix d'un certain nombre de gènes candidats et de polymorphismes pour un génotypage à plus grande échelle. Celui-ci a été réalisé de façon incomplète sur les arbres-mères dont nous avons étudié les descendants au niveau phénotypique dans le chapitre V, dans le but de valider le rôle éventuel des gènes dans la variation de caractères adaptatifs. Les résultats que nous avons présentés au chapitre VI sont très préliminaires mais cependant très encourageants pour le type d'approche présenté ci-dessus, et l'intérêt des inférences évolutives que l'on a pu faire avec les méthodes de génomique des populations.

C.2.1 Premières estimations sur l'organisation des DL chez le pin maritime

Les études d'association en populations naturelles permettent de tirer partie des nombreuses recombinaisons historiques qui ont eu lieu dans ces populations, pour cartographier des loci ayant un effet sur le caractère adaptatif (ou loci « adaptatifs ») plus finement que lors d'approches de détection de QTLs (Cardon et Bell 2001, et Walsh 2002, Gupta et al. 2005, voir Chapitre VI). La précision pour détecter des marqueurs proches de loci « adaptatifs » est donc conditionnée par l'intensité et les patrons de DL entre sites le long des chromosomes. Comparativement à l'espèce humaine chez laquelle les résultats du projet HapMap (couramment 10 millions de SNPs disponibles dans un panel d'une centaine d'individus) ont montré que les DL s'étendaient sur de longues distances (jusqu'à 50 kb, Reich *et al.* 2001), de même que chez d'autres espèces modèles telles que *Arabidopsis thaliana* (Nordborg *et al.* 2002), les premiers résultats publiés pour des espèces d'arbres forestiers font état d'une baisse rapide du DL par paires de sites intra-gène (Ingvarsson 2005a, Gonzalez-Martinez *et al.* 2006, Neale et Savolainen 2004). Cependant, ces estimations en plus d'être réalisées uniquement intra-gène, ne fournissent aucune information sur le DL entre gènes. De plus, plusieurs populations peuvent parfois être mélangées pour des espèces où l'on estime que la structuration n'est pas forte (par exemple, *Pinus sylvestris*). Enfin, Le nombre d'individus généralement échantillonnés par population (5-20) n'est pas suffisant pour estimer les DL avec précision (recommandation d'au moins 35 gamètes dans Brown 1975).

Nous avons observé en moyenne une baisse rapide des DL entre sites au sein des gènes étudiés, que ce soit dans la métapopulation atlantique *a priori* non structurée (Figure 30 du chapitre IV) ou sur l'ensemble des 10 populations (Informations supplémentaires à l'article de Eveno *et al.* 2008, Annexe 2). De même que chez d'autres espèces de conifères, ce résultat permet de penser que si l'on détecte des associations significatives, on se trouvera au niveau de gènes candidats intéressants. Nous avons également mis en évidence une absence de DL entre SNPs de différents gènes (Annexe 2). L'importance de ces DL varie cependant beaucoup entre gènes, et certains comme *Glucan* et *CcoAomt*, montrent un DL quasiment complet sur la longueur séquencée (métapopulation atlantique). Comprendre l'organisation des patrons de DL au niveau supérieur de différentes régions du génome de pin maritime nécessiterait des échantillonnages beaucoup plus importants sur de nombreux gènes.

C.2.2. Interprétation biologique de la fonction des gènes candidats

Les tests d'associations effectués, bien que sur un nombre limité de SNPs, mettent en évidence quelques associations remarquables avec des SNPs de *dhn1* (S187), *PR-AGP4* (S1135) et *CCoAOMT* (S1229). Ces résultats confirmeraient l'importance fonctionnelle de gènes sur lesquels nous avons précédemment détecté des effets de la sélection naturelle et la pertinence de l'approche suivie (voir partie B.1.). Ils sont cohérents aussi avec les observations faites chez *Pinus taeda* pour *CCoAOMT* (deux haplotypes majoritaires pouvant être maintenus par sélection balancée, Brown *et al.* 2004, Gonzalez-Martinez *et al.* 2007b), bien que ce gène n'ait pas été détecté dans l'étude d'association réalisée sur la même espèce dans un échantillon de 480 clones.

De plus, ces associations semblent cohérentes avec les informations *a priori* que l'on avait sur le rôle fonctionnel des gènes. Le gène *dhn1* qui serait impliqué dans la stabilisation des structures membranaires de la cellule ou la régulation du potentiel osmotique cellulaire, montre une sur-expression lors d'un stress hydrique (voir Chapitre II), ce qui pourrait expliquer l'association détectée entre un SNP de ce gène (S187) et le $\delta^{13}C$ (corrélé positivement à l'EUE). Ce SNP explique 8% de la variation phénotypique des classes génotypiques pour l'EUE dans un modèle ajustant l'effet du SNP au sein de chaque population. Il est également synonyme, et pourrait être en déséquilibre de liaison avec un SNP « fonctionnel » directement impliqué dans la variation de ce caractère. Le gène *PR-AGP4* interviendrait dans l'élongation de la paroi cellulaire (voir Chapitre II), et donc peut-être plus généralement dans la croissance des plantes. Ceci expliquerait que les associations soient surtout détectées avec la hauteur et la biomasse. Cependant l'association la plus forte est observée entre un SNP de ce gène (S1135) et le EUE. Une hypothèse possible est celle d'une stratégie visant à diminuer la croissance de la plante lors d'un stress hydrique (Chapitre II), ce qui expliquerait l'association détectée entre ce SNP et l'EUE. Il faudrait imaginer pour cela que ce SNP soit porté par un haplotype (ou soit lié à un gène) ayant un effet sur la régulation de l'expression de *PR-AGP4*. Pour *CCoAOMT* qui intervient dans le processus de lignification, on observe plutôt une sur-expression dans le cas d'un stress hydrique. Cette sur-expression serait liée à l'augmentation de la résistance des parois par fabrication de lignine, et pourrait expliquer l'association significative entre un SNP de ce gène (S1229) et l'efficacité d'utilisation en eau.

C.2.3 Importance des effets des SNPs ?

Le pourcentage de variation phénotypique d'un caractère expliqué par les quelques SNPs testés reste relativement faible (entre 2% et 8%), et du même ordre de grandeur que les associations détectées chez d'autres espèces forestières (Gonzalez-Martinez *et al.* 2007b, Thumma *et al.* 2005). Ces résultats sont également cohérents avec ceux de cartographie de QTLs chez ces espèces, où l'on constate que les caractères quantitatifs complexes sont généralement sous le contrôle génétique de nombreux loci à faibles effets (par exemple pour des caractères de propriété du bois, Brown *et al.* 2003, Jermstad *et al.* 2003, Pot *et al.* 2006), avec seulement quelques loci à effets majeurs, notamment pour la croissance et la discrimination isotopique du carbone (Prioul *et al.* 1997, Brendel *et al.* 2002). Le modèle de base de la génétique quantitative (nombreux loci à faibles effets) serait donc compatible avec ces résultats (Holland 2007). Mais ces approches sous-estimeraient l'importance du nombre de gènes à faibles effets et celle d'effets épistatiques impliqués dans l'architecture de caractères complexes (Barton et Keightley 2002, Malmberg *et al.* 2005, Holland 2007). Les études d'association en populations naturelles constitueraient une bonne alternative pour mieux comprendre cette architecture.

D'ailleurs, les résultats des modèles déclarant des effets hiérarchisés aux populations permettent d'approcher de façon élégante la complexité de l'expression des caractères. En effet, de nombreuses associations sont significatives seulement dans ces modèles alors qu'elles sont non-significatives quand les SNPs sont testés globalement. Ces résultats montrent à quel point ces associations dépendent du contexte génétique de chaque population, et donc des interactions épistatiques entre le SNP testé et tous les variants aux autres loci.

Même s'il est difficile de penser que les associations observées ne soient dues qu'au hasard (7 SNPs sur 7 montrant des tests significatifs avec au moins un caractère, et 3 sur 7 montrant des effets globaux), il serait important de tester l'hypothèse alternative des effets de SNPs « contrôle » dont on pourrait penser *a priori* qu'ils n'auront aucun effet sur la variation quantitative. Une confirmation de l'absence d'associations dans ce cas permettrait de confirmer la validité de nos critères de choix. C'est ce que nous avons proposé au sein du projet TREESNIPs, mais ceci n'a pu être réalisé, les 7 SNPs génotypés ne faisant pas partie du lô de contrôles. Il pourrait également être judicieux de confirmer ces associations ou d'identifier d'autres associations dans des milieux de tests où les stress hydriques seraient beaucoup plus élevés.

C.3. Quelles approches et quels échantillons pour les études d'associations chez des espèces structurées en populations?

C.3.1. Etudes d'association au sein de populations non structurée ou prise en compte de la structuration en populations ?

La structure pouvant être un facteur important de création de *DL* entre marqueurs non liés physiquement et par voie de conséquence entre marqueurs et phénotypes, on cherche soit à l'éviter en testant des associations au sein de métapopulations non-structurées, soit à l'intégrer dans les modèles d'analyses (Yu *et al.* 2006). Mais si l'objectif est d'identifier des polymorphismes impliqués dans l'adaptation locale et la différenciation entre populations comme dans le cas d'un cline climatique latitudinal ou altitudinal (Garcia-Gil *et al.* 2003), cette stratégie ne paraît pas optimale car au contraire échantillonner le long d'un cline pourra augmenter les chances d'élucider les bases moléculaires de la variation adaptative (Gonzalez-Martinez *et al.* 2006b). Un modèle composite de régression factorielle (voir Garnier-Géré *et al.* 2001) permettrait à la fois : 1) d'intégrer les fréquences alléliques des SNPs comme covariables pour expliquer des moyennes phénotypiques de chaque population, et 2) de tester les effets globaux ou hiérarchisés des SNPs à l'intérieur des populations dans le même modèle d'analyse. On pourrait aussi tenter de tester les effets combinés de plusieurs SNPs (par exemple un de *PR-AGP4* et l'autre de *dhn1*), pour voir s'ils s'additionnent ou non.

Dans le cas du pin maritime, il semble que la répartition des moyens de phénotypage et de génotypage dans plusieurs populations soit une stratégie efficace pour détecter des associations. Néanmoins, une stratégie mixte pourrait être développée où plusieurs populations contrastées sur l'ensemble de l'espèce seraient choisies, avec certaines échantillonnées plus largement au sein desquelles les associations pourraient être testés.

C.3.2. Fréquence minimum des allèles aux gènes candidats et taille des échantillons

Outre la répartition des moyens, deux autres facteurs clés doivent être pris en considération car ils risquent d'affecter la puissance des études d'associations : la fréquence minimale des allèles aux gènes candidats et la taille des échantillons (Wang *et al.* 2005). Un compromis dans le choix des polymorphismes à génotyper à plus grande échelle doit être trouvé entre une puissance *a priori* supérieure pour des allèles plus fréquents, et le fait que des allèles plus rares pourraient avoir des effets non négligeables sur le phénotype mais

nécessitent de très grands échantillons pour les tests. Par ailleurs, nos résultats illustrent bien la contrainte d'un nombre minimum d'au moins 500 individus (phénotypés) pour espérer détecter des effets alléliques relativement petits (Long et Langley 1999). La possibilité de réaliser des répétitions ou des clones de génotypes est un avantage chez les espèces végétales qui pourrait permettre d'augmenter la précision de l'estimation de la valeur phénotypique.

Conclusions et perspectives

En conclusion de cette thèse, les différentes approches utilisées (génomique des populations, génétique quantitative et génétique d'association) ont permis de mieux comprendre l'adaptation à la résistance au stress hydrique au sein de populations naturelles de pin maritime et de poser les bases d'expérimentations possibles dans ces domaines chez cette espèce. Les résultats que nous avons obtenus ont permis :

1) de montrer l'existence d'un potentiel adaptatif et d'une plasticité potentiellement importants pour des caractères liés à la résistance au stress hydrique dans de nombreuses populations naturelles, ce qui constitue des caractéristiques encourageantes pour l'adaptation de cette espèce aux changements climatiques futurs (voir chapitre VII).

2) de détecter des signatures de la sélection naturelle sur les patrons de diversité et de différenciation nucléotidiques d'un petit nombre de gènes candidats.

3) de valider par des tests d'associations le rôle potentiel de quelques-uns de ces gènes dans l'expression de caractères liés à l'adaptation. Malgré le petit nombre de SNPs testés, les tests d'associations réalisés sont très encourageants pour des études à plus grande échelle chez cette espèce.

On peut en effet se poser la question de savoir si l'on sera capable dans un futur proche de réaliser des études d'associations pan-génomiques chez le pin maritime et chez les espèces d'arbres forestiers en général. Nombre de ces espèces sont en effet de bons modèles d'espèces pérennes pour l'étude de l'adaptation par des approches de génomique des populations écologique et évolutive (Feder et Mitchell-Olds 2003), et elles peuvent profiter des ressources développées par l'analyse du génome séquencé du peuplier (Tuskan *et al.* 2003, *Populus trichocarpa*, en ligne sur <http://genome.jgi-psf.org/Poptr1/Poptr1.home.htm>).

Il est clair que l'on ne pourra améliorer les connaissances du génome des espèces d'arbres forestiers qu'en investissant dans des projets collaboratifs mettant en œuvre des moyens financiers suffisants, pour permettre tout d'abord d'augmenter le nombre de données de séquences disponibles de gènes candidats issus des banques d'ESTs (environ 30000 disponibles chez le pin maritime, et 65000 disponibles chez les chênes blancs). Cela pourrait être réalisé par exemple en constituant des bases de données de reséquençage allélique d'un grand nombre de fragments dessinés sur les unigènes disponibles dans les espèces étudiées.

Le reséquençage se ferait sur un échantillon soigneusement choisi de façon à disposer de données permettant d'avoir de meilleures estimations des patrons de diversité.

Dans les populations naturelles d'espèces d'arbres forestiers, nous avons vu que l'étendue des *DL* entre sites était *a priori* beaucoup plus faible que chez d'autres espèces de plantes ou chez les humains. Un effort important devrait cependant être effectué pour obtenir des données supplémentaires permettant de confirmer ces prédictions sur de larges régions génomiques et à l'extérieur des zones codantes des gènes. Les estimations disponibles ne concernent en effet que peu de gènes, souvent des fragments de ces gènes seulement (Neale et Savolainen 2004, Gonzalez-Martinez *et al.* 2006, Heuertz *et al.* 2006), et ne permettent *a priori* pas de savoir ce que l'on attend sur d'autres régions des génomes, même si les caractéristiques écologiques de ces espèces permettent de penser qu'une structure en blocs haplotypiques de grande taille est à exclure.

On pourrait donc investir aussi plus de moyens dans des techniques de marche sur le chromosome permettant d'avoir accès aux séquences des régions avoisinantes (5'UTR notamment), de gènes potentiellement intéressants et ainsi étudier les patrons de diversité le long de ces gènes « pleine-longueur ». En effet, il a été montré que la variation de séquences promotrices pouvait également faire partie de QTLs (Brem *et al.* 2002, Schadt *et al.* 2003). Dans notre travail, la proportion de SNPs « outliers » synonymes détectés par des méthodes basées sur le F_{ST} et qui étaient situés dans des régions non-codantes, était similaire à celle des SNPs « outliers » non-synonymes (22%), qui seraient pourtant *a priori* de meilleurs candidats à l'action de la sélection. Nous avons vu aussi dans les tests d'associations que 4 sur 7 des SNPs testés étaient synonymes ou situés dans des introns et montraient des associations significatives. Cela suggère que des SNPs situés dans les régions promotrices ou introniques pourraient aussi avoir une importance fonctionnelle, notamment pour l'expression des gènes, et donc être affectés par la sélection. Shabalina *et al.* (2001) montrent d'ailleurs chez les humains et les souris, que sur 1000 régions inter-géniques de 12000 pb en moyenne, une longueur d'au moins 2000 pb par région aurait été soumise à de fortes contraintes sélectives. D'autre part, des études sur les facteurs de transcription montrent également une importance fonctionnelle de ces régions (voir Patzlaff *et al.* 2003). Dans le cas particulier du génome des conifères qui contient de très nombreuses séquences répétées (rétrotransposons (Friesen *et al.* 2001), gènes ribosomiaux...), l'idée serait de se concentrer sur les régions les plus riches en gènes si elles existent.

Un autre besoin important concerne le génotypage à haut débit qui permettrait de valoriser les nombreux SNPs identifiés au sein d'approches gènes candidats, ainsi que ceux identifiés au sein de banques d'ESTs qui permettent maintenant de détecter un grand nombre de polymorphismes *in silico* (voir par exemple les bases de données « plantMarkers » (Rudd *et al.* 2005), ou MIPSPlantsDB (Spannagl *et al.* 2007)). Le génotypage à haut débit pourrait être réalisé sur le matériel installé dans des tests mis en place sur le long-terme, dans le but de réaliser des études d'associations. Techniquement, de telles méthodes de génotypage seraient *a priori* plus difficiles à mettre en oeuvre sur de gros génomes comme ceux des conifères. Nous avons vu en effet que les données qui devaient être initialement produites au sein du projet TREESNIPs n'ont pas pu être fournies, mais de gros efforts sont déployés au plan international pour développer ces méthodes, et elles devraient être accessibles en routine très prochainement, même chez les conifères (Pavy *et al.* 2008).

Plusieurs projets sont d'ailleurs en cours, notamment au sein du réseau Evoltree (<http://www.evoltree.org/>) pour des espèces feuillues (Peuplier et chêne) et des conifères (pin sylvestre, épicéa commun, pin maritime) concernant à la fois la détection *in silico* de polymorphismes SNPs, l'enrichissement de banques d'ESTs sur des tissus peu représentés, le développement des banques BAC, et le reséquençage allélique.

Si les besoins de génotypage sont honorés par le développement de techniques robustes et répétables, on peut espérer que la perspective d'études d'association non pas pan-génomiques mais « multilocus à haut débit » chez ces espèces n'est pas à exclure dans un futur proche chez les espèces d'arbres forestiers (Gonzalez-Martinez *et al.* 2007), en valorisant en parallèle les ressources génomiques déjà disponibles. De gros progrès restent cependant à faire, car malgré ces ressources importantes chez certaines espèces (peuplier par exemple), la génétique d'association a jusqu'à présent surtout été utilisée dans des approches ciblées permettant de confirmer le rôle de gènes candidats dans la variation de caractères d'intérêt (Thumma *et al.* 2005 chez l'eucalyptus, Ingvarsson *et al.* 2006 chez le peuplier et Gonzalez-Martinez *et al.* 2007b chez le pin taeda). Il y a donc eu encore peu de réalisations concrètes d'études d'associations intégrant du génotypage à grande échelle.

Dans un monde idéal, ayant à notre disposition de nombreux SNPs sur de nombreux génotypes clonés sur lesquels on aurait aussi de nombreuses données sur des caractères adaptatifs dans plusieurs environnements, des modèles complexes de détermination du

phénotype pourraient être testés, des SNPs à très faible effets identifiés, et on peut penser qu'on arriverait à expliquer une proportion de variation quantitative suffisamment importante à l'aide des génotypes, mais surtout, on améliorerait les connaissances concernant les interactions entre gènes, le choix des régions génomiques candidates à cibler, les meilleures méthodologies permettant de réaliser les études d'association, et celles sur les facteurs évolutifs ayant conduit aux patrons de diversité observés.

Références

- Abecasis GR, Ghosh D, Nichols TE. 2005. Linkage disequilibrium: ancient history drives the new genetics. *Human Heredity* 59:118-124.
- Acheré V, Favre JM, Besnard G, Jeandroz S. 2005. Genomic organization of molecular differentiation in Norway spruce (*Picea abies*). *Mol. Ecol.* 14: 3191-3201.
- Aguadé M, Miyashita N, Langley CH. 1989. Reduced variation in the *yellow-achaetescute* region in natural populations of *Drosophila melanogaster*. *Genetics* 122: 607-615.
- Aitken SN, Hannerz M. 2001. Genecology and gene resource management strategies for conifer cold hardiness. In: Bigras FJ, Columbo SJ, editors. *Conifer cold hardiness*. The Netherlands, Dordrecht: Kluwer Academic Publishers. p. 23–53.
- Aitken, S.N., Kavanagh, K.L., and Yoder, B.J. 1995. Genetic variation in seedling water-use efficiency as estimated by carbon isotope ratios and its relationship to sapling growth in Douglas-fir. *For. Genet.* 2: 199–206.
- Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA, Kruglyak L. 2004. Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biology.* 2: 15916-1599.
- Akey JM, Zhang G, Zhang K, Jin L, Shriver MD .2002. Interrogating a high density SNP map for signatures of natural selection. *Genome Res.* 12: 1805–1814.
- Alía R, Gil L, Pardos J. 1995. Performance of 43 *Pinus pinaster* Ait. provenances on 5 locations in central Spain. *Silvae Genetica.* 44:75-81.
- Alía R, Moro J, Denis JB. 1997. Performance of *Pinus pinaster* Ait. provenances in Spain: interpretation of the genotype-environment interaction. *Can. J. For. Res.* 27: 1548-1559.
- Allagulova C R, Gimalov FR, Shakirova FM, Vakhitov VA. 2003 The Plant Dehydrins: Structure and Putative Functions. *Biochemistry* 68: 945-951.
- Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P. 2005. A haplotype map of the human genome. *Nature* 437: 1299-1320.
- Andofalto P. 2001. Adaptive hitchhiking effects on genome variability. *Curr. Opin. Genet. Dev.* 11: 635-641.
- Araya YN. 2007. Ecology of water relations in plants. *Encyclopedia of Life Sciences* 5p.
- Ardlie KG, Kruglyak L, Seielstad M. 2002. Patterns of linkage disequilibrium in the human genome. *Nat. Rev. Genet.* 3: 299–309.
- Arumuganathan K, Earle ED. 1991. Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* 9: 208-218.
- Audigeos D. 2005. Patrons de diversité nucléotidique de deux gènes candidats à la réponse au stress hydrique chez le pin maritime : *PR-AGP4* et *CCoAOMT*. Rapport de Master 2 Recherche : Génétique et Développement des Plantes. 20p.
- Austerlitz F, Brachet S, Couvet D, Frascaria-Lacoste N, Jung-Muller B, Kremer A, Streiff R. 1999. Flux génétiques chez les arbres forestiers - Synthèse bibliographique. Commission Ressources Génétiques Forestières.
- Bacles CFE, Burczyk J, Lowe AJ, Ennos RA. 2005. Historical and contemporary mating patterns in remnant populations of the forest tree *Fraxinus excelsior* L. *Evolution* 54: 979-990.

- Baker J, Steele C, Leon D III. 1988. Sequence and characterization of 6 Lea proteins and their genes from cotton. *Plant Mol. Biol.* 11: 277-291.
- Balding DJ. 2003. Likelihood-based inference for genetic correlation coefficients. *Theor. Pop. Biol.* 63 : 221-230.
- Bamshad M, Wooding SP. 2003. Signature of natural selection in the human genome. *Nat. Rev. Genet.* 4: 99–111.
- Barba D, González-Martínez SC, Ribeiro MM, Agúndez D, Salvador L, Alía R, Gil L. 2001. Variación genética de *Pinus pinaster* Ait.: aplicación a la identificación y caracterización del material forestal de reproducción (MFR). In: Junta de Andalucía (ed.) *Montes para la Sociedad del Nuevo Milenio. Proceedings of the III Congreso Forestal Español* 5: 263-269.
- Barton NH, Bengsston BO. 1986. The barrier to genetic exchange between hybridizing populations. *Heredity.* 56: 357–376
- Barton NH, Keightley PD. 2002. Understanding quantitative genetic variation. *Nature Reviews Genetics.* 3:11-20.
- Barton NH. 2000. Genetic hitchhiking. *Philosophical Transactions of the Royal Society of London* 3555: 1553-1562.
- Bataillon TM, David JL, Schoen DJ. 1996. Neutral genetic markers and conservation genetics: simulated germplasm collections. *Genetics* 144: 409-447.
- Baudry E, Kerdelhue C, Innan H, Stephan W. 2001. Influence of recombination rates and mating system on DNA variability in the tomato genus. *Genetics* 158: 1725-1735.
- Bayes T. 1763. “An essay towards solving a problem in the doctrine of chances”. *Philosophical Transactions of the Royal Society of London* 53.
- Beaumont MA, Balding DJ. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Mol. Ecol.* 13: 969-980.
- Beaumont MA, Nichols RA. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond. Ser. Biol. Sci.* 263: 1619-1626.
- Beaumont MA, Rannala B. 2004. The Bayesian revolution in genetics. *Nature Reviews Genetics* 5: 251-261.
- Beaumont MA. 1999. Detecting population expansion and decline using microsatellites. *Genetics* 153: 2013-2029.
- Beaumont MA. 2005. Adaptation and speciation: What can F-st tell us? *Trends Ecol. Evol.* 20 : 435–440.
- Berli P. 2005. Comparison of Bayesian and maximum likelihood inference of population genetic parameters. *Bioinformatics* 22: 341-345.
- Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* 356: 519-520.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate – a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B.* 57: 289-300.
- Biswas S, Akey JM. 2006. Genomic insights into positive selection. *Trends in Genetics* 22: 437-446.
- Bohnert HJ, Sheveleva E. 1998. Plant stress adaptations – making metabolism move. *Current Opinion in Plant Biology.* 1: 267-274.
- Bonin A, Taberlet P, Miaud C, Pompanon F. 2006. Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Mol. Biol. Evol.* 23: 773-783.
- Bonneau M, Gelpe J, Le Tacon F. 1969. A propos du dépérissement du pin portugais dans les Landes. *Revue Forestière Française* 5: 343-350.

- Borevitz JO, Chory J. 2004. Genomic tools for QTL analysis and gene discovery. *Curr. Opin in Plant Biology* 7: 132-136.
- Bouillé M, Bousquet J. 2005. Trans-species shared polymorphisms at orthologous nuclear gene loci among distant species in the conifer *Picea* (Pinaceae): implications for the long-term maintenance of genetic diversity in trees. *Am. J. Bot.* 92: 63-73.
- Bowcock AM, Kidd JR, Mountain JL, Hebert JM, Carotenuto L, Kidd KK, Cavalli-Sforza LL. 1991. Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. *Proc. Natl. Acad. Sci. USA* 88: 839-843.
- Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W. 1995. The hitchhiking effect on the site frequency-spectrum of DNA polymorphisms. *Genetics* 140: 783-96.
- Bray EA. 1993. Molecular responses to water deficit. *Plant Physiol.* 103:1035-1040.
- Bray EA. 1997. Plant responses to water deficit. *Trends in Plant Science.* 2:48-54.
- Bray EA. 2001. Plant response to water-deficit stress. *Encyclopedia of Life Sciences* 5p.
- Brem RB, Yvert G, Clinton R, Kruglyak L. 2002. Genetic dissection of transcriptional regulation in budding yeast. *Science* 296: 752-755.
- Brendel O, Handley L, Griffiths H. 2003. The $\delta^{13}\text{C}$ of Scots pine (*Pinus sylvestris* L.) needles: spatial and temporal variations. *Ann For Sci* 60: 97-104.
- Brendel O, Pot D, Plomion C, Rozenberg P, Guehl JM. 2002. Genetic parameters and QTL analysis of delta C-13 and ring width in maritime pine. *Plant Cell Environ.* 25: 945-953.
- Britten HB. 1996. Meta-analyses of the association between multilocus heterozygosity and fitness. *Evolution* 50: 2158-2164.
- Brown ADH. 1975. Sample sizes required to detect linkage disequilibrium between two or three loci. *Theor. Popul. Biol.* 8: 184-201.
- Brown GR, Bassoni DL, Gill GP, Fontana JR, Wheeler NC, Megraw RA, Davis MF, Sewell MM, Tuskan GA, Neale DB. 2003. Identification of quantitative trait loci influencing wood property traits in loblolly pine (*Pinus taeda* L.). III. QTL verification and candidate gene mapping. *Genetics* 164: 1537-1546.
- Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB. 2004. Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proc. Natl. Acad. Sci. USA.* 101: 15255-15260.
- Bucci G, González-Martínez SC, Le Provost G, Plomion C, Ribeiro MM, Sebastiani F, Alía R, Vendramin GG. 2007. Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers. *Mol. Ecol.* 16: 2137-2153.
- Buntjer JB, Sorensen AP, Peleman JD. 2005. Haplotype diversity: the link between statistical and biological association. *Trends in Plant Science* 10: 466-471.
- Burban C, Petit RJ. 2003. Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. *Mol. Ecol.* 12: 1487-1495.
- Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD, Civello D, Adams MD, Cargill M, Clark AG. 2005. Natural selection on protein-coding genes in the human genome. *Nature* 437: 1153-1157.
- Campbell D, Bernatchez L. 2004. Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Mol. Biol. Evol.* 21: 945-956.

- Cardon LR, Bell JI. 2001. Association study designs for complex diseases. *Nature*. 2: 91-99.
- Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES. 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genetics* 22: 231-238.
- Carlson CS, Thomas DJ, Eberle MA, Swanson JE, Livingston RJ, Rieder MJ, Nickerson DA. 2005. Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Res.* 15: 1553–1565.
- Carrión JS, Navarro C, Navarro J, Munuera M. 2000. The distribution of cluster pine (*Pinus pinaster*) in Spain as derived from palaeoecological data: relationships with phytosociological classification. *The Holocene* 10 : 243-252.
- Carrión JS, Walker MJ, Legaz A, Chain C, Lopez A. 2003. Glacial refugia of temperate, Mediterranean and Ibero-North African flora in southeastern Spain: new evidence from cave pollen at two Neanderthal man sites. *Global Ecology and Biogeography* 12: 119-129.
- Carrión JS. 2002. Patterns and processes of Late Quaternary environmental change in a montane region of southwestern Europe. *Quaternary Science Reviews* 21 : 2047-2066.
- Caruso A, Morabito D, Delmotte F, Kahlem G, Carpin S. 2002. Dehydrin induction during drought and osmotic stress in *Populus*. *Plant Physiol. Biochem.* 40:1033–1042.
- Cassab GI, Varner JE. 1988. Cell wall proteins. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 39: 321–353.
- Cavalli-Sforza LL. 1966. Population structured and human evolution. *Proc. R. Soc. Lond. Ser. Biol. Sci.* 164: 362-379.
- Chagné D, Chaumeil P, Ramboer A, et al. (12 co-authors). 2004. Cross-species transferability and mapping of genomic and cDNA SSRs in pines. *Theor. Appl. Genet.* 109: 1204-1214.
- Chagné D, Lalanne C, Madur D, et al. (12 co-authors). 2002. A high density linkage map of *Pinus pinaster* based on AFLPs. *Ann. For. Sci.* 59: 627-636.
- Chang S, Puryear JD, Dias MADL, Funkhouser EA, Newton RJ, Cairney J. 1996. Gene expression under water deficit in loblolly pine (*Pinus taeda*): Isolation and characterization of cDNA clones. *Physiologia Plantarum.* 97:139-148.
- Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics.* 134: 1289–1303.
- Charlesworth B, Nordborg M, Charlesworth D. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res.* 70: 155-174.
- Charlesworth D, Awadalla P. 1998. Flowering plant self-incompatibility: the molecular population genetics of *Brassica S*-loci. *Heredity* 81: 1-9.
- Charlesworth D. 2004. Sex determination: balancing selection in the honey bee. *Current Biology* 14: R568-569.
- Charlesworth D. 2006. Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics* 2: 379-384.
- Chaumeil P. 2006. Plasticité moléculaire de deux ecotypes de pin maritime soumis à un stress osmotique. Thèse. Université de Bordeaux-I, France.

- Chikhi L, Garnier-Géré PH. 2005. Genetics of large populations and association studies. Encyclopedia of Life Sciences. 5p.
- Christians JK, Keightley PD. 2002. Genetic architecture : dissecting the genetic basis of phenotypic variation. *Curr. Biol.* 12:415-416.
- Clark AG, Hubisz MJ, Bustamente CD, Williamson SH, Nielsen R. 2005. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Res.* 15: 1496-1502.
- Close TJ. 1996. Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. *Physiol. plant* 97: 795-803.
- Close TJ. 1997. Dehydrins: a commonality in the response of plants to dehydration and low temperature. *Physiologia Plantarum* 100: 291–296.
- Cockerham CC, Weir BS. 1993. Estimation of gene flow from F-statistics. *Evolution* 47: 855-863.
- Cockerham CC. 1969. Variance of gene frequency. *Evolution* 23: 72-74.
- Cockerham CC. 1973. Analysis of gene frequencies. *Genetics* 74: 679-700.
- Cockerham W, Weir BS. 1987. Correlations, descent measures: Drift with migration and mutation. *PNAS* 84: 8512–8514.
- Condon AG, Richards RA, Rebetzke GJ, Farquhar GD. 2002. Improving intrinsic water-use efficiency and crop yield. *Crop Science* 42: 122–131.
- Condon AG, Richards RA. 1993. Exploiting genetic variation in transpiration efficiency in wheat: An agronomic view. In *Stable Isotopes and Plant Carbon-Water Relations*. J.R. Ehleringer et al. (ed.). Academic Press, San Diego, CA. p. 435–450.
- Conner JK, Hartl DL. 2004. A primer of ecological genetics. Sinauer Associates: Sunderland, MA.
- Conover DO, Van Voorhees DA, Ehtisham A. 1992. Sex ratio selection and changes in environmental sex determination in laboratory populations of *Menidia menidia*. *Evolution* 46: 1722-1730.
- Conover DO, Van Voorhees DA. 1990. Evolution of a balanced sex ratio by frequency-dependent selection in a fish. *Science* 250: 1556-1558.
- Costa P, Bahrman N, Frigerio JM, Kremer A, Plomion C. 1998. Water-deficit-responsive proteins in maritime pine. *Plant Mol. Biol.* 38:587-596.
- Craig H. 1954. Carbon 13 in Plants and the Relationship between carbon 13 and carbon 14 variations. *Nature* 62: 115.
- Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK. 2000. Considering evolutionary processes in conservation biology. *TREE* 15: 290-295.
- Cregg BM, Olivas-Garcia JM, Hennessey TC. 2000. Provenance variation in carbon isotope discrimination of mature ponderosa pine trees at two locations in the Great Plains. *Can. J. For.Res.* 30: 428-439.
- Crnokrak P, Merilä J. 2002. Genetic population divergence: markers and traits. *Trends in Ecology and Evolution.* 17: 501.
- Crow JF, Kimura M. 1970. An introduction to population genetics theory. Harper and Row (eds). New York.
- Cruz RT, Jordan WR, Drew MC. 1992. Structural changes and associated reduction of hydraulic conductance in roots of *Sorghum bicolor* L. following exposure to water deficit. *Plant Physiol.* 99: 203-212.

- Danjon F. 1994. Stand Features and Height Growth in a 36-Year-Old Maritime Pine (*Pinus pinaster* Ait.) Provenance Test. *Silvae Genetica* 43: 52-62.
- Danjon F. 1995. Observed section effects on height growth, diameter and stem form in maritime pine. *Silvae Genetica* 44: 10-18.
- Darwin C. 1865. On the sexual relations of the three forms of *Lythrum salicaria*. *Jour. Linn. Soc.* 8: 169-196.
- Davis MB, Shaw RG. 2001. Range shifts and adaptive responses to Quaternary climate change. *Science* 292: 673-679.
- De Mita S, Ronfort J, McKhann HI, Poncet C, El Malki R, Bataillon T. 2007. Investigation of the demographic and selective forces shaping the nucleotide diversity of genes involved in nod factor signaling in *Medicago truncatula*. *Genetics* 177: 2123-2133.
- Delden VW, Boerema AC, Kamping A. 1978. The alcohol dehydrogenase polymorphism in populations of *Drosophila melanogaster*. I. Selection in different environments. *Genetics* 90: 161-191.
- Derory J, Mariette S, González-Martínez SC, Chagné D, Madur D, Gerber S, Brach J, Persyn F, Ribeiro MM, Plomion C. 2002. What can nuclear microsatellites tell us about maritime pine genetic resources conservation and provenance certification strategies? *Ann. For. Sci.* 59: 699-708.
- Dreyer E, Bogeat-Triboulot MB, Le Thiec D, Guehl JM, Brignolas F, Villar M, Bastien C, Martin F, Kohler A. 2004. Drought tolerance of poplars: can we expect to improve it? *Biofutur* 247: 54-58.
- Dubos C, Le Provost G, Pot D, Salin F, Lalane C, Madur D, Frigerio JM, Plomion C. 2003. Identification and characterization of water-stress-responsive genes in hydroponically grown maritime pine (*Pinus pinaster*) seedlings. *Tree Physiol.* 23:169-179.
- Dubos C, Plomion C. 2003. Identification of water-deficit responsive genes in maritime pine (*Pinus pinaster* Ait.) roots. *Plant Mol. Biol.* 51: 249-262.
- Dure L III. 1993. Structural motifs in Lea proteins. In: Close TJ, Bray EA, editors. *Plant Responses to Cellular Dehydration during Environmental Stress. Current Topics in Plant Physiology, Vol 10.* American Society of Plant Physiologists, Rockville, MD, p. 91-103.
- Dvornyk V, Sirviö A, Mikkonen M, Savolainen O. 2002. Low Nucleotide Diversity at the *pal1* Locus in the Widely Distributed *Pinus sylvestris*. *Mol. Biol. Evol.* 19: 179-188.
- Ebert RH, Cherkasova VA, Dennis RA, Wu JH, Ruggles S, Perrin TE, Shmookler Reis RJ. 1993. Longevity-determining genes in *Caenorhabditis elegans*: chromosomal mapping of multiple noninteractive loci. *Genetics* 135: 1003-1010.
- Endler JA. 1986. *Natural selection in the wild.* Princeton, New Jersey, USA: Princeton University Press.
- Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611-2620.
- Eveno E, Collada C, Guevara MA, Léger V, Soto A, Diaz L, Léger P, Gonzalez-Martinez SC, Cervera MT, Plomion C, Garnier-Géré PH. 2008. Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate genes as revealed by genetic differentiation analyses. *Mol. Biol. Evol.* 25: 417-437.
- Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol. Bioinform Online* 1: 47-50.

- Excoffier L, Novembre J, Schneider S. 2000. SIMCOAL: A general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *J. Heredity* 91: 506-510.
- Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.
- Falconer DS, McKay TFC. 1996. *Introduction to Quantitative Genetics*. Ed 4. Longmans Green, Harlow, Essex, UK.
- Falconer DS. 1960. *Introduction to quantitative genetics*. New York, Ronald Press, Edinburgh, Oliver and Boyd.
- Farquhar GD, Ehleringer JR, Hubick KT. 1989. Carbon isotope discrimination and photosynthesis. *Ann. Rev. Plant Physiol.* 40: 503-37.
- Farquhar GD, O'Leary MH, Berry JA. 1982. On the relationship between carbon isotope discrimination and intercellular carbon dioxide concentration in the leaves. *Aust. J. Plant Physiol.* 9: 121-137.
- Farquhar GD, Richards DA. 1984. Isotopic composition of plant carbon correlates with water-use-efficiency of wheat genotypes. *Aust. J. Plant Physiol.* 11: 539-524.
- Fay JC, Wu C-I. 2000. Hitchhiking under positive Darwinian selection. *Genetics* 155: 1405-1413.
- Fay JC, Wu C-I. 2001. The neutral theory in the genomic era. *Current Opinion in Genetics and Development* 11: 642-646.
- Filatov DA. 2002. ProSeq: A software for preparation and evolutionary analysis of DNA sequence data sets. *Mol. Ecol. Notes* 2: 621-624.
- Flint J, Mott R. 2001. Finding the molecular basis of quantitative traits: successes and pitfalls. *Nature Rev. Genet.* 2: 437-445.
- Ford MJ. 2002. Applications of selective neutrality tests to molecular ecology. *Mol. Ecol.* 11:1245-1262.
- Fox CW, Roff DA, Fairbairn DJ. 2001. *Evolutionary ecology: concepts and case studies*. Oxford University Press.
- Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD. 2000. *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. *Science* 289: 85-88.
- Friesen N., Brandes A, Heslop-Harrison JSP. 2001. Diversity, Origin, and Distribution of Retrotransposons (gypsy and copia) in Conifers. *Mol. Biol. Evol.* 18: 1176-1188.
- Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics* 133: 693-709.
- Fu YX. 1997. Statistical tests of neutrality against population growth, hitchhiking and background selection. *Genetics* 147: 915-925.
- Futuyma DJ. 1998. *Evolutionary Biology*. 3rd edition, Sinauer Associates, Inc. Pp 681-685.
- Gallais A. 1990. *Théorie de la sélection en amélioration des plantes*. Collection des sciences agronomiques. Paris, Masson. 588 pp.
- García-Gil MR, Mikkonen M, Savolainen O. 2003. Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. *Mol. Ecol.* 12: 1195-1206.
- Garnier-Géré PH, Ades PK. 2001. Environmental surrogates for predicting and conserving adaptive genetic variability in tree species. *Conservation Biology.* 15: 1632-1644.
- Garnier-Géré PH, Ades PK. 2001. Environmental surrogates for predicting and conserving adaptive genetic variability in tree species. *Conservation Biology.* 15:1632-1644.
- Garrigan D, Hedrick PW. 2003. Perspective: detecting adaptive molecular evolution, lessons from the MHC. *Evolution* 57: 1707-1722.

- Garris AJ, McCouch SR, Kresovich S. 2003. Population structure and its effects on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.). *Genetics* 165: 759-769.
- Gillespie JH. 2000. Genetic drift in an infinite populations: the pseudohitchhiking model. *Genetics* 155: 909-919.
- Godelle B, Austerlitz F, Brachet S, Colas B, Cuguen J, Gandon S, Gouyon P-H, Lefranc M, Olivieri I, Reboud X, Vitalis R. 1998. Système génétique, polymorphisme neutre et sélectionné: implications en biologie de la conservation. *Génétique, Sélection, Evolution* 30: 515-528.
- Gómez J, Sánchez-Martínez D, Stiefel V, Rigau J, Puigdomènech P, Pagès M. 1988. A gene induced by the plant hormone abscisic acid in response to water stress encodes a glycine-rich protein. *Nature* 334: 262-264.
- González-Martínez SC, Alía R, Gil L. 2002. Population genetic structure in a Mediterranean pine (*Pinus pinaster* Ait.): a comparison of allozyme markers and quantitative traits. *Heredity*. 89: 199-206.
- González-Martínez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB. 2006a. DNA sequence Variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics* 172:1915-1926.
- González-Martínez SC, Gil L, Alía R. 2002. Population genetic structure in a Mediterranean pine (*Pinus pinaster* Ait.): a comparison of allozyme markers and quantitative traits. *Heredity*. 89: 199-206.
- González-Martínez SC, Gómez A, Carrión JS, Agúndez D, Alía R, Gil L. 2007a. Spatial genetic structure of an explicit glacial refugium of maritime pine (*Pinus pinaster* Aiton) in southeastern Spain. In: Weiss S, Ferrand N, editors. *Phylogeography in southern European refugia: evolutionary perspectives on the origins of European biodiversity*. London. p. 257-269.
- González-Martínez SC, Krutovsky KV, Neale DB. 2006b. Forest-tree population and genomics and adaptive evolution. *New Phytol.* 170: 227-238.
- González-Martínez SC, Mariette S, Ribeiro MM, Burban C, Raffin A, Chambel MR, Ribeiro C, Aguiar A, Plomion C, Alía R, Gil L, Vendramin GG, Kremer A. 2004. Genetic resources in maritime pine (*Pinus pinaster* Aiton): molecular and quantitative measures of genetic variation and differentiation among maternal lineages. *Forest Ecology and Management*. 197: 103-115.
- González-Martínez SC, Wheeler NC, Rezos E, Nelson CD, Neale DB. 2007b. Association Genetics in *Pinus taeda* L. I. Wood Property Traits. *Genetics*. 175:399-409.
- Graham J, Curran J, Weir BS. 2000. Conditional Genotypic Probabilities for Microsatellite Loci. *Genetics* 155: 1973–1980.
- Guehl, J-M, Nguyen-Queyrens A, Loustau D, Ferhi A. 1996. Genetic and environmental determinants of water-use efficiency and carbon isotope discrimination in forest trees. In: INRA éditeur. *EUROSILVA, contribution to forest tree physiology*, Dourdan, France, November 7.10.1994. Colloq. INRA. No. 76. pp. 297.321.
- Guevara MA, Chagné D, Almeida MH, et al. (16 co-auteurs). 2005. Isolation and characterization of nuclear microsatellite loci in *Pinus pinaster* Ait. *Mol. Ecol. Notes* 5: 57-59.
- Gupta PK, Rustgi S, Kulwal PL. 2005. Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Mol. Biol.* 57 : 461-485.

- Guyon JP, Kremer A. 1982. Stabilité phénotypique de la croissance en hauteur et cinétique journalière de la pression de sève et de la transpiration chez le pin maritime (*P. pinaster* Ait.). Can. J. For. Res. 12: 936-946.
- Haldane JBS. 1949. Disease and evolution. *Ricerca Sci.* 19: 3–10.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95-98.
- Hamrick J, Godt M, Sherman-Broyles S. 1992. Factors influencing levels of genetic diversity in woody plant species. *New Forests* 6: 95-124.
- Hamrick JL, Blanton HM, Hamrick KJ. 1989. The genetic structure of geographically marginal populations of ponderosa pine. *Am. J. Bot.* 76: 1559-1568.
- Hamrick JL, Godt MJ, Murawski DA, Loveless MD. 1991. Correlation between species traits and allozyme diversity: Implications for conservation biology. En: Falk, D.A. and K.E. Holsinger (eds.) *Genetics and conservation of rare plants (Genética y conservación de plantas raras)*. Oxford Univ. Press. p. 75-86.
- Hamrick JL. 2004. Response of forest trees to global environmental changes, *Forest Ecology and Management* 197: 323-335.
- Harfouche A, Baradat P, Durel CE. 1995. Variabilité intraspécifique chez le pin maritime (*Pinus pinaster* Ait.) dans le sud-est de la France. I. Variabilité des populations autochtones et des populations de l'ensemble de l'aire de l'espèce. *Annales des Sciences Forestières* 52 : 307-328.
- Harr B, Kauer M, Schlötterer C. 2002. Hitchhiking mapping — a population-based fine mapping strategy for adaptive mutations in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America* 99: 12949–12954.
- Hedrick PW, Ginevan M, Ewing E. 1976. Genetic polymorphism in heterogeneous environments. *Annu. Rev. Ecol. Syst.* 7: 1–32.
- Hedrick PW. 1986. Genetic polymorphism in heterogeneous environments: a decade later. *Annu Rev. Ecol. Syst.* 17: 535-566.
- Hedrick PW. 2006a. Genetic polymorphism in heterogeneous environments: the age of genomics. *Annu. Rev. Ecol. Syst.* 37: 67-93.
- Hedrick PW. 2006b. Balancing selection. *Current Biology.* 17: R230-231.
- Hedrick PW. 2007. Balancing selection. *Current Biology* 17: R230-231.
- Hein J, Schierup M, Wiuf C. 2005. *Gene genealogies, variation and evolution: a primer in coalescent theory*. New York: Oxford University Press.
- Hellmann I, Ebersberger I, Ptak SE, Pääbo S, Przeworski M. 2003. A neutral explanation for the correlation of diversity with recombination rates in humans. *Am. J. Hum. Genet.* 72: 1527-1535.
- Heuertz M, De Paoli E, Källman T, Larsson H, Jurman I, Morgante M, Lascoux M, Gyllenstrand N. 2006. Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of norway spruce [*Picea abies* (L.) Karst]. *Genetics* 174: 2095-2105.
- Hill AVS, Allsopp CEM, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennett S, Brewster D, McMichael AJ, Greenwood BM. 1991. Common west African HLA antigens are associated with protection from severe malaria. *Nature* 352: 595-600.

- Hill WG, Robertson A. 1968. Linkage disequilibrium in finite populations. *TAG* 38: 226-231.
- Hill WG, Weir BS. 1988. Variances and covariances of squared linkage disequilibria in finite populations. *Theor. Popul. Biol.* 33: 54-78.
- Hiorns RW, Harrison GA. 1977. The Combined Effects of Selection and Migration in Human Evolution. *Man New Series.* 12: 438-445.
- Hirschhorn JN, Daly MJ. 2005. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* 6: 95-108.
- Hoffman EA, Schueler FW, Jones AG, Blouin MS. 2006. An analysis of selection on a colour polymorphism in the northern leopard frog. *Mol. Ecol.* 15: 2627-2641.
- Holland JB. 2007. Genetic Architecture of Complex Traits in Plants. *Current Opinion in Plant Biology.* 10: 156-161.
- Houle D. 1992. Comparing Evolvability and Variability of Quantitative Traits. *Genetics.* 130: 195-204.
- Howe GT, Aitken SN, Neale DB, Jermstad KD, Wheeler N, Chen THH. 2003. From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. *Can. J. Bot.* 81: 1247-1266.
- Howe GT, Brunner AM. 2005. An evolving approach for understanding plant adaptation. *New Phytol.* 167:1-5.
- Hubick KT, Gibson A. 1993. Diversity in the relationship between carbon isotope discrimination and transpiration efficiency when water is limited. In: *Stable isotopes and plant carbon-water relations.* Ehleringer JR, Hall AE, Farquhar GD (eds). London: Academic Press. 311-325.
- Hudson RR, Boos DD, Kaplan NL. 1992. A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* 9: 138-151.
- Hudson RR, Kaplan NL. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111: 147-164.
- Hudson RR, Kaplan NL. 1988. The coalescent process in models with selection and recombination. *Genetics* 120: 831-840.
- Hudson RR, Kaplan NL. 1995. Deleterious background selection with recombination. *Genetics.* 141: 1605-1617.
- Hudson RR, Kreitman M, Aguadé M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116: 153-159.
- Hudson RR. 1990. Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* 7: 1-44.
- Hudson RR. 2000. A new statistic for detecting genetic differentiation. *Genetics* 155: 2011-2014.
- Huey RB, Gilchrist GW, Carlson ML, Berrigan D, Serra, L. 2000. Rapid evolution of a geographic cline in size in an introduced fly. *Science.* 287: 308-309.
- Hughes AL, Nei M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335: 167-170.
- Hughes L. 2000. Biological consequences of global warming: is the signal already apparent? *Trends in Ecology and Evolution* 15: 56-61.
- Hugouvieux V, Kwak JM, Schroeder JI. 2001. An mRNA cap binding protein, ABH1, modulates early abscisic acid signal transduction in *Arabidopsis*. *Cell* 106: 477-487.
- Hurme P, Repo T, Savolainen O, Pääkkönen T. 1997. Climatic adaptation of bud set and frost hardiness in Scots pine (*Pinus sylvestris*). *Canadian Journal of Forest Research.* 27:716-723.
- Ihaka R, Gentleman R. 1996. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5: 299-314.

- Ingram J, Bartels D. 1996. The molecular basis of dehydration tolerance in plants. *Annu. Rev. Plant Physiol. Plant. Mol. Biol.* 47: 377-403.
- Ingvarsson PK, Garci MV, Hall D, Luquez V, Jansson S. 2006. Clinal variation in *phyB2*, a candidate gene for day-length induced growth cessation and bud set, across a latitudinal gradient in European aspen (*Populus tremula* L., Salicaceae). *Genetics* 172: 1845-1853.
- Ingvarsson PK. 2005a. Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European Aspen (*Populus tremula* L., Salicaceae). *Genetics* 169: 945-953.
- Ingvarsson PK. 2005b. Molecular population genetics of herbivore-induced protease inhibitor genes in european aspen (*Populus tremula* L., Salicaceae). *Mol. Biol. Evol.* 22: 1802-1812.
- Innan H, Kim Y. 2004. Pattern of polymorphism after strong artificial selection in a domestication event. *PNAS* 101: 10667-10672.
- IPCC (Intergovernmental Panel on Climate Change). 2001. *Climate change 2001: The scientific basis*, Cambridge University Press, Cambridge, UK.
- IPCC (Intergovernmental Panel on Climate Change). 2007. *Climate change 2001: The scientific basis*, Cambridge University Press, Cambridge, UK.
- Jannink J-L, Walsh JB. 2002. Association mapping in plant populations. In *Quantitative Genetics, Genomics and Plant Breeding*. Kang SM (ed). CABI, Wallingford, UK. p. 59-68.
- Jarvinen P, Lemmetyinen J, Savolainen O, Sopanen T. 2003. DNA sequence variation in *BpMADS2* gene in two populations of *Betula pendula*. *Mol. Ecol.* 12: 369-384.
- JBIC / SCET Tunisie. 2005. Study of integrated water resources management in the North of Tunisia. Final report.
- Jermstad KD, Bassoni DL, Jech KS, Ritchie GA, Wheeler NC, Neale DB. 2003. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas fir. III. Quantitative trait loci-by-environment interactions. *Genetics* 165: 1489–1506.
- Johnsen KH, Flanagan LB, Huber DA, Major JE. 1999. Genetic variation in growth, carbon isotope discrimination, and foliar N concentration in *Picea mariana*: analyses from a half-diallel mating design using field-grown trees. *Can. J. For. Res.* 29: 1727-1735.
- Johnsen Ø, Skrøppa T. 2000. Provenances and families show different patterns of relationship between bud-set and frost hardiness in *Picea abies*. *Can. J. For. Res.* 30: 1858-1866.
- Joshi J, Schmid B, Caldeira MC, et al. (18 co-authors). 2001. Local adaptation enhances performance of common plant species. *Ecology Letters*. 4: 536-544.
- Joyce DC, Aspinall D, Edwards GR. 1983. Water deficit and the growth and anatomy of the radish fleshy axis. *New Phytol.* 93: 439-446.
- Jump AS, Hunt JM, Martinez-Izquierdo JA, Penuelas J. 2006. Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*. *Mol. Ecol.* 15: 3469-3480.
- Jump AS, Penuelas J. 2005. Running to stand still: adaptation and the response of plants to rapid climate change. *Ecology Letters* 8: 1010-1020.
- Kado T, Matsumoto A, Ujino-Ihara T, Tsumura Y. 2008. Amounts and patterns of nucleotide variation within and between two Japanese conifers, sugi (*Cryptomeria japonica*) and hinoki (*Chamaecyparis obtusa*) (Cupressaceae *sensu lato*). *Tree Genetics and Genomes*. 4: 133-141.
- Kado T, Ushio Y, Yoshimaru H, Tsumura Y, Tachida H. 2006. Search for nucleotide patterns of local adaptation in dehydrins and other cold related candidate genes in Scots pine (*Pinus sylvestris* L.). *Genes Genet. Syst.* 81: 103-113.

- Kado T, Yoshimaru H, Tsumura Y, Tachida H. 2003. DNA variation in a conifer, *Cryptomeria japonica* (Cupressaceae sensu lato). *Genetics* 164: 1547-1559.
- Kaplan NL, Darden T, Hudson RR. 1988. The coalescent process in models with selection. *Genetics* 120: 819-29.
- Kaplan NL, Hudson RR, Langley CH. 1989. The hitchhiking effect revisited. *Genetics* 123: 887-99.
- Kauer M, Dieringer D, Schlötterer C. 2003. A microsatellite variability screen for positive selection associated with the 'out of Africa' habitat expansion of *Drosophila melanogaster*. *Genetics* 165: 1137-1148.
- Kauer M, Zangerl B, Dieringer D, Schlötterer C. 2002. Chromosomal patterns of microsatellite variability contrast sharply in African and non-African populations of *Drosophila melanogaster*. *Genetics* 160: 247-256.
- Kaufman SR, Smouse PE, Alvarez-Buylla ER. 1998. Pollen-mediated gene flow and differential male reproductive success in a tropical pioneer tree, *Cecropia obtusifolia* Bertol. (Moraceae): a paternity analysis. *Heredity* 81: 164-173.
- Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. *Ecology Letters*. 7: 1225-1241.
- Kayser M, Brauer S, Stoneking M. 2003. A genome scan to detect candidate regions influenced by local natural selection in human populations. *Mol. Biol. Evol.* 20: 893-900.
- Kelley JL, Madeoy J, Calhoun JC, Swanson W, Akey AM. 2007. Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome Res.* 16: 980-989.
- Kelly JK. 1997. A test of neutrality based on interlocus associations. *Genetics* 146: 1197-1206.
- Kim Y, Nielsen R. 2004. Linkage disequilibrium as a signature of selective sweeps. *Genetics* 167: 1513-1524.
- Kim Y, Stephan W. 2002. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* 160: 765-777.
- Kimura M, Ohta T. 1978. Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proc. Natl. Acad. Sci. USA* 75: 2868-2872.
- Kimura M. 1968. Evolutionary rate at the molecular level. *Nature* 217: 624-626.
- Kimura M. 1983. *The neutral theory of molecular evolution*. Cambridge, UK: Cambridge Univ. Press. 367pp.
- Kimura R, Fujimoto A, Tokunaga K, Ohashi J. 2007. A practical genome scan for population-specific strong selective sweeps that have reached fixation. *PLoS* 3: 1-10.
- Kingman JFC. 1982. On the genealogy of large populations. *J Appl Prob* 19A: 27-43.
- Kittelson PM, Maron JL. 2001. Fine scale genetically-based differentiation of life-history traits in the perennial shrub, *Lupinus arboreus*. *Evolution*. 55: 2429-2438.
- Klein J. 1987. Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. *Hum. Immunol.* 19: 155-162.
- Klein Tank A, Wijngaard J, Van Engelen A. 2002. *Climate in Europe. Assessment of observed daily temperature and precipitation extremes*. European Climate Assessment, KNMI, the Bilt, the Netherlands.
- Koch M, Haubold B, Mitchell-Olds T. 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol Biol Evol.* 17: 1483-1498.

- Koelewijn HP, Koski V, Savolainen O. 1999. Magnitude and timing of inbreeding depression in Scots pine (*Pinus sylvestris* L.). *Evolution* 53: 758-768.
- Kraft T, Sall T, Magnusson-Rading I, Nilsson NO, Hallden C. 1998. Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (*Beta vulgaris* subsp. *maritima*). *Genetics* 150: 1239-1244.
- Kremer A, Le Corre V, Mariette S. 2000. Population differentiation for adaptive traits and their underlying loci in forest trees: theoretical predictions and experimental results. In: *Forest Genetics and Sustainability* (ed. Mátyás C.), pp. 59-74. Kluwer Academic Publisher, The Netherlands.
- Kremer A, Roussel G. 1982. Composantes de la croissance en hauteur chez le pin maritime (*Pinus pinaster* Ait). *Annales des Sciences Forestières* 38: 77-98.
- Kremer A. 1994. Diversité génétique et variabilité des caractères phénotypiques chez les arbres forestiers. *Genet. Sel. Evol.* 26: 105s-123s.
- Krutovsky KV, Neale DB. 2005. Nucleotide Diversity and Linkage Disequilibrium in Cold-Hardiness- and Wood Quality-Related Candidate Genes in Douglas Fir. *Genetics* 171: 2029-2041.
- Lande R, Arnold SJ. 1983. The measurement of selection on correlated characters. *Evolution.* 37: 1210-1226.
- Lande R, Schemske DW, Schultz ST. 1994. High inbreeding depression, selective interference among loci, and the threshold selfing rate for purging recessive lethal mutations. *Evolution* 48: 965-978.
- Langley CH. 1990. The molecular population genetics of *Drosophila*. In *Population Biology of Gaes and Molecules*. Takahata N and Crow JF (eds). Baifukan, Tokyo. pp. 75-91.
- Latta RG. 2003. Gene flow, adaptive population divergence and comparative population structure across loci. *New. Phytol.* 161: 51-58.
- Latta RG. 2004. Relating processes to patterns of genetic variation across landscapes. *Forest Ecology and Management.* 197: 91-102.
- Lauteri M, Pliura A, Monteverdi MC, Brugnoli E, Villani F, Eriksson G. 2004. Genetic variation in carbon isotope discrimination in six European populations of *Castanea sativa* Mill. originating from contrasting localities. *J. Evol. Biol.* 17: 1286-1296.
- Lauteri M, Scartazza A, Guido MC, Brugnoli E. 1997. Genetic variation in photosynthetic capacity, carbon isotope discrimination and mesophyll conductance in provenances of *Castanea sativa* adapted to different environments. *Functional Ecology* 11: 675-683.
- Laval G, Excoffier L. 2004. SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics* 20: 2485-2487.
- Lecorre V, Kremer A. 2003. Comparison of genetic variability at neutral markers, QTLs and adaptive trait in a subdivided population: effects of selection intensity, variance of local optima and migration. *Genetics.* 164:1205-1219.
- Leung J, Bouvier-Durand M, Morris P-C, Guerrier D, Cheddor F, Giraudat J. 1994. Arabidopsis ABA-response gene *ABI7*: Features of a calcium-modulated protein phosphatase. *Science* 264: 1448-1452.
- Leung J, Merlot S, Giraudat J. 1997. The arabidopsis *ABSCISIC ACID-INSENSITIVE2* (*ABI2*) and *ABI1* genes encode homologous protein phosphatases 2C involved in abscisic acid signal transduction. *Plant Cell* 9: 759-771.

- Leung J, Merlot S, Gosti F, Bertauche N, Blatt MR, Giraudat J. 1998. The role of ABI1 in abscisic acid signal transduction: from gene to cell. *Symp. Soc. Exp. Biol.* 51: 65-71.
- Lewontin RC, Krakauer JK. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74: 175-195.
- Lewontin RC. 1970. "The units of selection". *Annual Reviews of Ecology and Systematics* 1: 1-18.
- Li C, Berninger F, Koskela J, Sonninen E. 2000. Drought responses of *Eucalyptus microtheca* provenances depend on seasonality of rainfall in their place of origin. *Aus. J. Plant Physiol.* 27: 231-238.
- Li C. 1999. Carbon isotope composition, water-use efficiency and biomass productivity of *Eucalyptus microtheca* populations under different water supplies. *Plant and Soil* 214: 165-171.
- Li WH. 1976. Effect of migration on genetic distance. *Am. Nat.* 110: 841-847.
- Linhart YB, Grant MC. 1996. Evolutionary significance of local genetic differentiation in plants. *Ann. Rev. Ecol. Syst.* 27: 237-277.
- Liu A, Burke JM. 2006. Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics* 173: 321-330.
- Long AD, Langley CH. 1999. The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Res.* 9: 720-731.
- Long JC, Smouse PE, Wood JW. 1987. The allelic correlation structure of Gainj- and Kalam-speaking people. II. The genetic distance between population subdivisions. *Genetics* 117: 273-83.
- Long JC. 1986. The allelic correlation structure of Gainj- and Kalam-speaking people. I. The estimation and interpretation of Wright's F-statistics. *Genetics* 112: 629-647.
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nat. Rev. Genet.* 4: 981-994.
- Lynch M, Crease T. 1990. The analysis of population survey data on DNA sequence variation. *Mol. Biol. Evol.* 7 :377-394.
- Lynch M, Walsh B. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Assoc.
- Zhang JW, Marshall JD. Variation in carbon isotope discrimination and photosynthetic gas exchange among populations of *Pseudotsuga menziesii* and *Pinus ponderosa* in different environments. *Funct. Ecol.* 9 : 402-412.
- Lynch M. 1996. A quantitative-genetic perspective on conservation issues. In: *Conservation Genetics: case histories from nature*. Avise JC, Hamrick JL (eds). Springer-Verlag, New-York, pp471-501.
- Ma X-F, Szmidt AE, Wang X-R. 2006. Genetic structure and evolutionary history of a diploid hybrid pine *Pinus densata* inferred from the nucleotide variation at seven gene loci. *Mol. Biol. Evol.* 23: 807-816.
- Marchini J, Cardon LR, Phillips MS, Donnelly P. 2004. The effects of human population structure on large genetic association studies. *Nature Genetics* 36: 512-517.
- Mariette S, Chagné D, Decroocq S, Giovanni G, Lalanne C, Madur D, Plomion C. 2001. Microsatellite markers for *Pinus pinaster* Ait. *Ann. For. Sci.* 58: 203-206.
- Masle J, Gilmore SR, Farquhar GD. 2005. The *ERECTA* gene regulates plant transpiration efficiency in *Arabidopsis*. *Nature.* 436, 866-870
- Maynard Smith J, Haigh J. 1974. "The hitch-hiking effect of a favourable gene". *Genetical Research Cambridge.* 23: 23-35.

- Mayr S, Cochard H. 2003. A new method for vulnerability analysis of small xylem areas reveals that compression wood of Norway spruce has lower hydraulic safety than opposite wood. *Plant, Cell and Environment* 26: 1365–1371.
- McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *adh* locus in *Drosophila*. *Nature* 351: 652-654.
- McKay I, Powell W. 2007. Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci.* 12: 57-63.
- McKay JK, Latta RG. 2002. Adaptive population divergence: markers, QTL and traits. *Trends Ecol. Evol.* 17:285-291.
- McVean GAT, Awadalla P, Fearnhead P. 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 160: 1231–1241.
- McVean GAT, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. 2004. The fine-scale structure of recombination rate variation in the human genome. *Science* 304: 581-584.
- Merilä J, Crnokrak P. 2001. Comparison of genetic differentiation at marker loci and quantitative traits. *J. Evol. Biol.* 14: 892-903.
- Merlot S, Gosti F, Guerrier D, Vavasseur A, Giraudat J. 2001. The ABI1 and ABI2 protein phosphatases 2C act in a negative feedback regulatory loop of the abscisic acid signalling pathway. *Plant J.* 25: 295-303.
- Meyer K, Leube MP, Grill E. 1994. A Protein Phosphatase 2C involved in ABA signal transduction in *Arabidopsis thaliana*. *Science* 264: 1452-1455.
- Michalakis Y, Excoffier L. 1996. A generic estimation of population subdivision using distances between alleles with special reference to microsatellite loci. *Genetics* 142: 1061-1064.
- Mitchell-Olds T, Feder M, Wray G. 2008. Evolutionary and ecological functional genomics. *Heredity* 100: 101-2
- Mitchell-Olds T, Willis JH, Goldstein DB. 2007. Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature* 8: 845-856.
- Mitton JB. 2001. Adaptation and natural selection: overview. *Encyclopedia of Life Science.* 5p.
- Mode CJ, Robinson HF. 1959. Pleiotropism and the genetic variance. *Biometrics* 15: 518-537.
- Morton, N. E., and J. F. Crow, 1956 An estimate of the mutational damage in man from data on consanguineous marriages. *PNAS* 42: 855–863.
- Muirhead CA. 2001. Consequences of population structure on genes under balancing selection. *Evolution* 55: 1532-1541.
- Mundree SG, Baker B, Mowla S, et al. (11 co-auteurs). 2002. Physiological and molecular insights into drought tolerance. *African Journal of Biotechnology.* 1:28-38.
- Mundy J, Chua N-H. 1988. Abscisic acid and water-stress induce the expression of a novel rice gene. *EMBO J.* 7: 2279-2286.
- Murray B. 1998. Nuclear DNA amounts in gymnosperms. *Ann. Bot.* 83: 3-15.

- Nachman MW, Bauer VL, Crowell SL, Aquadro CF. 1998. DNA variability and recombination rates at X-linked loci in humans. *Genetics* 150: 1133-1141.
- Nair S, Williams J, Brockman A, Paiphun L, Mayxay M, Newton PN, Guthmann J-P, Smithuis FM, Hien TT, White NJ, Nosten F, Anderson TJC. 2003. A selective sweep driven by pyrimethamine treatment in southeast Asian malaria parasites. *Mol. Biol. Evol.* 20: 1526–1536.
- Nasrallah JB. 2002. Recognition and rejection of self in plant reproduction. *Science* 296: 305-308.
- Neale DB, Savolainen O. 2004. Association genetics of complex traits in conifers. *Trends Plant Sci.* 97: 325-330.
- Nei M, Chakravarti A, Tateno Y. 1977. Mean and variance of F_{ST} in a finite number of incompletely isolated populations. *Theor. Popul. Biol.* 11: 291–306.
- Nei M, Chakravarti A. 1977. Drift variance of F_{ST} and G_{ST} statistics obtained from a finite number of isolated populations. *Theor. Popul. Biol.* 11: 307–325.
- Nei M, Kumar S. 2000. *Molecular Evolution and Phylogenetics*. New York: Oxford University Press.
- Nei M, Maruyama T. 1975. Lewontin-Krakauer test for neutral genes. *Genetics* 80: 396.
- Nei M. 1973. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* 70: 3321-3323.
- Nei M. 1982. Evolution of human races at the gene level. In: Bonne-Tamir B, editeur. *Human Genetics, Part A: The Unfolding Genome*. Alan R. Liss, Inc., New York, p. 167-181.
- Nei M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York, NY.
- Nguyen A, Lamant A. 1989. Variation in growth and osmotic regulation of roots of water stressed maritime pine (*P. pinaster* Ait) provenances. *Tree Physiol.* 5: 123-133.
- Nguyen-Queyrens A, Fehri A, Loustau D, Guehl J-M. 1998. Within-ring $\delta^{13}C$ spatial variability and interannual variations in wood cellulose of two contrasting provenances of *Pinus pinaster*. *Can. J. For. Res.* 28: 766-773.
- Nielsen EE, Hansen MM, Meldrup D. 2006. Evidence of microsatellite hotch-hiking selection in Atlantic cod (*Gadus morhua* L.): implications for inferring population structure in nonmodel organisms. *Mol. Ecol.* 15: 3219-3229.
- Nielsen R. 2001. Statistical tests of selective neutrality in the age of genomics. *Heredity.* 86: 641-647.
- Nielsen R. 2005. Molecular signatures of natural selection. *Annu. Rev. Genet.* 39: 197-218.
- Nielsen R, Bustamente C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, Sninsky JJ, Adams MD, Cargill M. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* 3: 976-985.
- No E-G, Lupstra CA. 2000. Hormonal and developmental regulation of two arabinogalactan-proteins in xylem of loblolly pine (*Pinus taeda*). *Physiol. Plant.* 110: 524-529
- Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D. 2002. The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.* 30: 190-193.
- Nordborg M, Innan H. 2002. Molecular population genetics. *Curr. Opin. Plant Biol.* 5: 69–73.
- Nylander M, Svensson J, Palva ET, Welin BV. 2001. Stress-induced accumulation and tissue-specific localization of dehydrins in *Arabidopsis thaliana*. *Plant Mol. Biol.* 45: 263–279.
- O’Hagan A, Luce BR. 2003. *A Primer on Bayesian Statistics in Health Economics and Outcomes Research*. London: Medtap International Inc.

- Orr HA. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*. 52: 935-949.
- Osborne R, Paterson WSB. 1952. On the sampling variance of heritability estimates derived from variance analysis. *Proc. Roy. Soc. Edinburg* B64:456.
- Osório J, Osório M L, Chaves MM, Peirera JS. 1998. Effects of water deficits on ^{13}C discrimination and transpiration efficiency of *Eucalyptus globulus* clones. *Aust. J. Plant Physiol.* 25: 645-653.
- Padmanabhan V, Dias DMAL, Newton RJ. 1997. Expression analysis of a gene family in loblolly pine (*Pinus taeda* L.) induced by water deficit stress. *Plant Mol. Biol.* 35:801-807.
- Paquereau MM. 1964. Flores et climats post-glaciaires en Gironde. *Actes de la Société Linnéenne de Bordeaux* 101 : 1-156.
- Pariset L, Cappuccio I, Joost S, D'Andrea M, Marletta D, Marsan PA, Valentini A, Econogene Consortium. 2006. Characterization of single nucleotide polymorphisms in sheep and their variation as evidence of selection. *Animal Genetics* 37: 290-292.
- Parry ML. 2000. Assessment of potential effects and adaptation for climate change in Europe: The Europe Acacia Project, Jackson Environmental Institute, University of East Anglia, Norwich, UK.
- Passioura JB. 1977. Grain yield, harvest index, and water use of wheat. *J. Aust. Inst. Agric. Sci.* 43: 117-120.
- Passioura JB. 1982. Water in the soil-plant-atmosphere continuum. In: Lange, O.L., Nobel, P.S., Osmond, C.B. and Ziegler, H., Editors. *Physiological Plant Ecology 2. Water Relations and Carbon Assimilation. Encyclopedia of Plant Physiology, New Series Vol. 12 B*, Springer, Berlin, pp. 5-33.
- Patzlaff A, Newman LJ, Dubos C, Whetten RW, Smith C, McInnis S, Bevan M, Sederoff RR, Campbell MM. 2003. Characterisation of PtMYB1, an R2R3-MYB from pine xylem. *Plant Molecular Biology* 53: 597-608.
- Pavy N, Paule C, Parsons L, Crow J, Morency MJ, Cooke J, Johnson JE, Noumen E, Guillet-Claude C, Butterfield Y, Barber S, Yang G, Liu J, Stott J, Kirkpatrick R, Siddiqui A, Holt R, Marra M, Seguin A, Retzel E, Bousquet J, MacKay J. 2005. Generation, annotation, analysis and database integration of 16,500 white spruce EST clusters. *BMC Genomics* 6: 144.
- Perneger TV. 1998. What is wrong with Bonferroni adjustments. *British Medical Journal*. 136: 1236-1238.
- Petit RJ, Bahrman N, Baradat PH. 1995. Comparison of genetic differentiation in maritime pine (*Pinus pinaster* Ait.) estimated using isozyme, total protein and terpenic loci. *Heredity* 75: 382-389.
- Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG. 2005. Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Mol. Ecol.* 14: 689-701.
- Petit RJ, Hampe A. 2006. Some evolutionary consequences of being a tree. *Ann. Rev. Ecol. Evol. Syst.* 37: 187-214.
- Petry D. 1983. The effect on neutral gene flow of selection at a linked locus. *Theoretical Population Biology*. 23: 300-313.
- Pfrender ME, Spitze K, Hicks J, Morgan K, Latta L, Lynch M. 2000. Lack of concordance between genetic diversity estimates at the molecular and quantitative-trait levels. *Conservation Genetics* 1: 263-269.
- Pigliucci M, Murren CJ, Schlichting CD. 2006. Phenotypic plasticity and evolution by genetic assimilation. *Journal of Experimental Biology* 209: 2362-2367.
- Pigliucci M. 2008. Is evolvability evolvable? *Nature Reviews Genetics* 9: 75-82.

- Pita P, Soria F, Canas I, Toval G, Pardos JA. 2001. Carbon isotope discrimination and its relationship to drought resistance under field conditions in genotypes of *Eucalyptus globulus* Labill. For. Ecol. Manage. 141: 211-221.
- Plomion C, Bahrman N, Durel CE, O'Malley DM. 1995. Genomic analysis in *Pinus pinaster* (Maritime pine) using RAPD and protein markers. Heredity 74: 661-668.
- Plomion C, Pionneau C, Brach J, Costa P, Baillères H. 2000. Compression wood-responsive proteins in developing xylem of maritime pine (*Pinus pinaster* Ait.). Plant Physiol. 123:959-969.
- Pluzhnikov A, Donnelly P. 1996. Optimal sequencing strategies for surveying molecular genetic diversity. Genetics 144: 1247-1262.
- Pons O, Petit R. 1995. Estimation, variance and optimal sampling of gene diversity. 1: Haploid locus. Theor. Appl. Genet. 90: 462-470.
- Pons O, Petit R. 1996. Measuring and testing genetic differentiation with ordered versus unordered alleles. Genetics 144: 1237-1245.
- Ponton S, Dupouey J-L, Bréda N, Feuillat F, Bodénès C, Dreyer E. 2001. Carbon isotope discrimination and wood anatomy variations in mixed stands of *Quercus robur* and *Quercus petraea*. Plant, Cell and Environment 24: 861-868.
- Porter AH. 2003. A test for deviation from island-model population structure. Mol. Ecol. 12: 903-915.
- Posada D, Crandall KA. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. PNAS USA 98: 13757-13762.
- Pot D, McMillan L, Echt C, Le Provost G, Garnier-Géré P, Cato S, Plomion C. 2005. Nucleotide variation in genes involved in wood formation in two pine species. New Phytologist 167: 101-112.
- Pot D, Rodrigues J-C, Rozenberg P, Chantre G, Tibbits J, Cahalan C, Pichavant F, Plomion C. 2006. QTLs and candidate genes for wood properties in maritime pine (*Pinus pinaster* Ait.) Tree Genetics and Genomes 2: 10-24.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genetics 38: 904-9.
- Price TD, Qvarnstro A, Irwin DE. 2003. The role of phenotypic plasticity in driving genetic evolution. Proceedings of the Royal Society of London B 270: 1433-1440.
- Prioul JL, Quarrie SA, Causse M, de Vienne D. 1997. Dissecting complex physiological functions into elementary components through the use of molecular quantitative genetics. Journal of Experimental Botany 48 : 1151-1163.
- Pritchard JK, Przeworski M. 2001. Linkage disequilibrium in humans: models and data. Am. J. Hum. Genet. 69: 1-114.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959.
- Przeworski M, Coop G, Wall JD. 2005. The signature of positive selection on standing genetic variation. Evolution 59: 2312-2323.
- Przeworski M. 2002. The signature of positive selection at randomly chosen loci. Genetics 160: 1170-1189.
- Pyharjarvi T, Garcia-Gil MR, Knutt T, Mikkonen M, Wachowiak W, Savolainen O. 2007. Demographic history has influenced nucleotide diversity in European *Pinus sylvestris* populations. Genetics 177: 1713-1724.

- Ramanjalu S, Bartels D. 2002. Drought- and desiccation-induced modulation of gene expression in plants. *Plant, Cell and Environment* 25: 141-151.
- Read J, Farquhar G. 1991. Comparative studies in *Nothofagus* (Fagaceae). I. Leaf carbon discrimination. *Functional Ecology* 5: 684-695.
- Reed DH, Frankham R. 2001. How closely related are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution* 55: 1095-1103.
- Rehfeldt GE, Tschebakova NM, Parfenova YI, Wykoff WR, Kuzmina NA, Milyutin LI. 2002. Intraspecific responses to climate in *Pinus sylvestris*. *Glob. Change Biol.* 8: 912-29.
- Rehfeldt GE, Ying CC, Spittlehouse DL, Hamilton DA Jr. 1999. Genetic responses to climate in *Pinus contorta*: niche breadth, climate change and reforestation. *Ecological Monographs*. 69:375-407.
- Reich D, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES. 2001. Linkage disequilibrium in the human genome. *Nature* 411: 199-204.
- Remington DL, Thornsberry JM, Matuoka Y, Wilson L, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler IV ES. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *PNAS* 98: 11479-11484.
- Rhodes D, Nadolska-Orczyk A. 2002. Plant stress physiology. *Encyclopedia of Life Sciences* 7p.
- Riccardi F, Gazeau P, De Vienne D, Zivy M. 1998. Protein changes in response to progressive water deficit in maize. *Plant Physiol.* 117: 1253-1263.
- Richard, S., Morency MJ, Drevet C, Jouanin L, Séguin A. 2000. Isolation and characterization of a dehydrin gene from white spruce induced upon wounding, drought and cold stresses. *Plant Mol. Biol.* 43: 1-10.
- Robertson A. 1975a. Remarks on the Lewontin-Krakauer test. *Genetics* 80: 396.
- Robertson, A. 1975b. Gene frequency distribution as a test of selective neutrality. *Genetics* 81: 775-785.
- Robledo-Arnuncio JJ, Gil L. 2004. Patterns of pollen dispersal in a small population of *Pinus sylvestris* L. revealed by total-exclusion paternity analysis. *Heredity* 94: 13-22.
- Rodriguez PL. 1998. Protein phosphatase 2C (PP2C) function in higher plants. *Plant Mol. Biol.* 38: 919-927.
- Rousset F. 2001. Inferences from spatial population genetics. In: *Handbook of statistical genetics*. Balding D, Bishop M, Cannings C. (eds). Wiley. P239-269.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496-2497.
- Rudd S, Schoof H, Mayer K. 2005. PlantMarkers--a database of predicted molecular markers from plants *Nucleic Acids Res* 33(Database issue): D628-32.
- Rueffler C, Van Dooren TJM, Leimar O, Abrams PA. 2006. Disruptive selection then what? *Trends in Ecology and Evolution* 21: 238-45.
- Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooperk R, Kwiatkowski D, Ward R, Lander ES. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419: 832-837.
- Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES. 2006. Positive natural selection in the human lineage. *Science*. 12:1614-1620.
- Salvador L, Alía R, Agúndez D, Gil L. 2000. Genetic variation and migration pathways of maritime pine (*Pinus pinaster* Ait) in the Iberian peninsula. *Theor. Appl. Genetics* 100: 89-95.

- Sarrauste N. 1982. Photosynthèse, respiration et répartition de matière sèche de jeunes plants de pin maritime (*Pinus pinaster* Ait.) appartenant à sept provenances et conduits selon deux traitements hydriques. Diplôme d'études approfondies, Université Paris VII, Paris, France.
- Savolainen O, Pyhäjärvi T, Knurr T. 2007. Gene flow and local adaptation in trees. *Annu. Rev. Ecol. Evol. Syst.* 38: 595-619.
- Savolainen O, Pyhäjärvi T. 2007. Genomic diversity in forest trees. *Curr. Opin. Plant Biol.* 10:1-6.
- Savolainen V, Anstett M-C, Lexer C, Hutton I, Clarkson JJ, Norup MV, Powell MP, Springate D, Salamin N, Baker WJ. 2006. Sympatric speciation in palms on an oceanic island. *Nature* 441: 210-213.
- Schadt EE, Monks SA, Drake TA, Lusk AJ, Che N. 2003. Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422: 297-302.
- Schemske DW. 1984. Population structure and local selection in *Impatiens pallida* (Balsaminaceae), a selfing annual. *Evolution.* 38:817-832.
- Schlötterer C. 2002. A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics* 160: 753-763.
- Schmalhausen I I. 1949. *Factors of evolution*. Blakiston, Philadelphia, PA.
- Schmid KJ, Ramos-Onsins S, Ringys-Beckstein H, Weisshaar B, Mitchell-Olds T. 2005. A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. *Genetics* 169: 1601-1615.
- Schmid KJ, Ramos-Onsins S, Ringys-Beckstein H, Weisshaar B, Mitchell-Olds T. 2005. A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. *Genetics* 169: 1601-1615.
- Schoen DJ, Brown AHD. 1993. Conservation of Allelic richness in wild crop relatives is aided by assessment of genetic markers. *Proceedings of the National Academy of Science, USA.* 90: 10623-10627.
- Schöfl G, Schlötterer C. 2004. Patterns of microsatellite variability among X chromosomes and autosomes indicate a high frequency of beneficial mutations in non-African *D. simulans*. *Mol. Biol. Evol.* 21: 1384-1390.
- Scotti-Saintagne C, Mariette S, Porth I, Goicoechea PG, Barreneche T, Bodénès C, Burg K, Kremer A. 2004. Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.] *Genetics* 168: 1615-1626.
- Searle SR, Casella G, McCulloch CE. 1992. *Variance components*. J Wiley & sons, New-York.
- Shabalina SA, Ogurtsov AY, Kondrashov VA, Kondrashov AS. 2001. Selective constraint in intergenic regions of human and mouse genomes. *Trends in Genetics.* 17: 373-376.
- Shih F-L, Hwang S-Y, Cheng Y-P, Lee P-F, Lin T-P. 2007. Uniform genetic diversity, low differentiation, and neutral evolution characterize contemporary refuge populations of Taiwan fir (*Abies kawakamii*, Pinaceae). *Am. J. Bot.* 94: 194-202.
- Shinozaki K, Yamaguchi-Shinozaki K. 2000. Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signalling pathways. *Curr. Opin. Plant Biol.* 3:217- 233.
- Shoemaker J, Painter I, Weir BS. 1999. Bayesian statistics in genetics. *Trends Genet.* 15: 354-358.
- Shook DR, Johnson TE. 1999. Quantitative trait loci affecting survival and fertility-related traits in *Caenorhabditis elegans* show genotype-environment interactions, pleiotropy and epistasis. *Genetics* 153: 1233-1243.

- Showalter AM. 1993. Structure and function of plant cell wall proteins. *Plant Cell*. 5:9–23.
- Slate J. 2005. Quantitative trait locus mapping in natural populations: progress, caveats and future directions. *Mol. Ecol.* 14:363-379.
- Slatkin M, Voelm. 1991. *FST* in a Hierarchical Island Model. *Genetics* 127: 627-629.
- Slatkin M, Wiehe T. 1998. Genetic hitch-hiking in a subdivided population. *Genet. Res.* 71: 155-160.
- Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139: 457-462.
- Sneath PHA, Sokal RR. 1973. *Numerical Taxonomy — the principles and practice of numerical classification*. Freeman editeur, San Francisco.
- Sokal RR, Rohlf FJ. 1995. *Biometry : the principles and practice of statistics in biological research*. Freeman and Company (eds). p 887.
- Soulé ME. 1987. *Viable populations for conservation*. Cambridge University Press, Cambridge.
- Spannagl M, Noubibou O, Haase D, Yang L, Gundlach H, Hindemitt T, Klee K, Haberer G, Schoof H, Mayer KFX. 2007. MIPSPlantsDB—plant database resource for integrative and comparative plant genome research *Nucleic Acids Research*, Vol. 35 (Database issue) D834-D840.
- Spitze K. 1993. Population structure in *Daphnia obtusa*: quantitative genetics and allozyme variation. *Genetics* 135: 367–374.
- Spitze, K. 1993. Population structure in *Daphnia obtusa*: quantitative genetic and allozyme variation. *Genetics* 135 : 367–374
- Stephan W, Langley CH. 1998. DNA polymorphism in lycopersicon and crossing-over per physical length. *Genetics* 150: 1585-1593.
- Stephan W, Wiehe THE, Lenz MW. 1992. The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. *Theor. Popul. Biol.* 41: 237-254.
- Stinchcombe JR, Hoekstra HE. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100: 158-170.
- Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *PNAS* 100: 9440-9445.
- Storey JD. 2002. A direct approach to false discovery rates. *J. R. Stat. Soc. B.* 64: 479-498.
- Storz JF, Beaumont MA. 2002. Testing for genetic evidence of population expansion and contraction: an empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evolution* 56: 154–166.
- Storz JF, Dubach JM. 2004. Natural selection drives altitudinal divergence at the albumin locus in deer mice, *Peromyscus maniculatus*. *Evolution* 58: 1342-1352.
- Storz JF, Nachman MW. 2003. Natural selection on protein polymorphism in the rodent genus *Peromyscus*: evidence from interlocus contrasts. *Evolution* 57: 2628-2635.
- Storz JF, Payseur BA, Nachman MW. 2004. Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. *Mol. Biol. Evol* 21: 1800-1811.
- Storz JF. 2002. Contrasting patterns of divergence in quantitative traits and neutral DNA markers: analysis of clinal variation. *Mol. Ecol.* 11:2537-2551.

- Storz JF. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol. Ecol.* 14: 671-688.
- Strauss SH, Bousquet J, Hipkins VD, Hong YP. 1992. Biochemical and molecular genetic markers in biosystematic studies of forest trees. *New Forests* 6: 125-158.
- Stumpf MP, McVean GA. 2003 Estimating recombination rates from population-genetic data. *Nat. Rev. Genet.* 4: 959-968.
- Sykes MT, Prentice IC, Cramer W. 1996. A bioclimatic model for the potential distributions of north European tree species under present and future climates. *Journal of Biogeography* 23: 203-234.
- Sykes MT, Prentice IC. 1995. Boreal forest futures: Modelling the controls on tree species range limits and transient responses to climate change. *Water Air and Soil Pollution* 82: 415-428.
- Syvanen AC. 2001. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat. Rev. Genet.* 2: 930-942.
- Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437-460.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 123: 585-595.
- Tallis GM. 1959. Sampling errors of genetic correlation coefficients calculated from analyses of variance and covariance. *Aust. J. Stat.* 1: 35-43.
- Tenaillon M, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *PNAS* 98: 9161-9166.
- Tenaillon MI, Sawkins MC, Anderson LK, Stack SM, Doebley J, Gaut BS. 2002. Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Genetics* 162: 1401-1413.
- Thomas CD. 2005. Recent evolutionary effects of climate change. In: Lovejoy TE, Hannah L, editors. *Climate Change and Biodiversity*. New Haven, Connecticut: Yale University Press. p. 75-88.
- Thompson AJ, Corlett JE. 1995. mRNA levels of four tomato (*L. esculentum* Mill L.) genes related to fluctuating soil water status. *Plant Cell Environ.* 18: 773-780.
- Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler IV ES. 2001. Dwarf8 polymorphisms associate with variation in flowering time. *Nat. Genet.* 28: 286-289.
- Thumma BR, Nolan MF, Evans R, Moran GF. 2005. Polymorphisms in *Cinnamoyl CoA Reductase* (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics*. 171: 1257-1265.
- Tognetti R, Michelozzi M, Lauteri M, Brugnoli E, Giannini R. 2000. Geographic variation in growth, carbon isotope discrimination, and monoterpene composition in *Pinus pinaster* Ait. provenances. *Can. J. For. Res.* 30: 1682-1690.
- Tsumura Y, Kado T, Takahashi T, Tani N, Ujino-Ihara T, Iwata H. 2007. Genome scan to detect genetic structure and adaptive genes of natural populations of *Cryptomeria japonica*. *Genetics* 176: 2393-2403.
- Tuskan G, Difazio SP, Teichmann T. 2004. Poplar genomics is getting popular: The impact of the poplar genome project on tree research. *Plant Biology* 5: 1-3.

- Tuskan GA, DiFazio S, Jansson S, et al. (110 co-auteurs). 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313: 1596-1604.
- Valladares F, Sanchez-Gomez D, Zavala MA. 2006. Quantitative estimation of phenotypic plasticity: bridging the gap between the evolutionary concept and its ecological applications. *Journal of Ecology* 94: 1103–1116.
- Vasemägi A, Nilsson J, Primmer CR. 2005. Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in atlantic salmon (*Salmo salar* L.) *Mol. Biol. Evol.* 22: 1067-1076.
- Vasemägi A, Primmer CR. 2005. Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Mol. Ecol.* 14: 3623-3642.
- Vendramin GG, Anzidei M, Madaghiele A, Bucci G. 1998. Distribution of genetic diversity in *Pinus pinaster* Ait. as revealed by chloroplast microsatellites. *Theoretical and Applied Genetics* 97: 456-463.
- Verhoeven KJF, Simonsen KL, McIntyre LM. 2005. Implementing false discovery rate control: increasing your power. *OIKOS* 108: 643-647.
- Via S, Lande R. 1985. Genotype-by-environment interaction and the evolution of phenotypic plasticity. *Evolution* 39: 505-522.
- Via S. 2001. Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol. Evol.* 16: 381-390.
- Vigouroux Y, McMullen M, Hittinger CT, Houchins K, Schulz L, Kresovich, Matsuoka Y, Doebley J. 2002. Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. *Proceedings of the National Academy of Sciences of the United States of America.* 99: 9650–9655.
- Vitalis R, Dawson K, Boursot P. 2001. Interpretation of variation across marker loci as evidence of selection. *Genetics* 158: 1811-1823.
- Voight BF, Kudravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol.* 4: 446-458.
- Wachowiak W, Balk PA, Van Wordragen MF, Savolainen O. 2008. Search for nucleotide patterns of local adaptation in dehydrins and other cold related candidate genes in Scots pine (*Pinus sylvestris* L.). *Tree Genetics and Genome.* Submitted.
- Wagner GP, Altenberg L. 1996. Complex adaptations and the evolution of the evolvability. *Evolution* 50: 967-976.
- Wahid N, González-Martínez SC, El Hadrami I, Boulli A. 2004. Genetic Structure and Variability of Natural Populations of Maritime Pine (*Pinus pinaster* Aiton) in Morocco. *Silvae Genetica.* 53: 93-99.
- Wahlund S. 1928. Zusammensetzung von populationen und korrelation-serscheiunungen von standpunkt der vererbungslehre aus betrachtet. *Hereditas* 11: 65-106.
- Wakeley J. 1999. Nonequilibrium migration in human history. *Genetics* 153: 1863–1871.
- Wall JD. 2003. Estimating ancestral population sizes and divergence times. *Genetics* 163: 395-404.
- Walsh B. 2001. Quantitative genetics in the age of genomics. *TAG* 59: 175-184.
- Wang C-S, Liau YE, Huang JC, Wu TD, Su CC, Lin CH. 1998. Characterization of a desiccation-related protein in lily pollen during development and stress. *Plant Cell Physiol.* 39: 1307-1314.
- Wang ET, Kodama G, Baldi P, Moyzis RK. 2006. Global landscape of recent inferred Darwinian selection for *Homo sapiens*. *Proc. Natl. Acad. Sci. USA.* 103: 135–140.

- Wang WY, Barratt BJ, Clayton DG, Todd JA. 2005. Genome-wide association studies: theoretical and practical concerns. *Nat. Rev. Genet.* 6: 109–118.
- Ward JH. 1963. Hierarchical Grouping to optimize an objective function. *Journal of American Statistical Association* 58: 236-244.
- Watkinson JJ, Sioson AA, Vasquez-Robinet C, et al. (14 co-auteurs). 2003. Photosynthetic acclimation is reflected in specific patterns of gene expression in drought-stressed loblolly pine. *Plant Physiol.* 133: 1702-1716.
- Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Pop. Biol.* 7: 256-276.
- Weir BS, Cardon LR, Anderson AD, Nielsen DM, Hill WG. 2005. Measures of human population structure show heterogeneity among genomic regions. *Genome Res.* 15: 1468-1476.
- Weir BS, Cockerham CC. 1969. Group inbreeding with two linked loci. *Genetics* 63: 711–742.
- Weir BS, Cockerham CC. 1973. Mixed self and random mating at two loci. *Genetical Research* 21: 247-262.
- Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.
- Weir BS, Hill WG. 2002. Discussion on the meeting on 'Statistical modelling and analysis of genetic data'. *Journal of the Royal Statistical Society B* 64: 737-775.
- Weir BS. 1990. *Genetic Data Analysis*. Sinauer, Sunderland, MA.
- Wheeler NC, Jermstad KD, Krutovsky KV, Aitken SN, Howe GT, Krakowski J, Neale DB. 2005. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. IV. Cold-hardiness QTL verification and candidate gene mapping. *Molecular Breeding* 15: 145–156.
- Whitlock MC, McCauley DE. 1999. Indirect measures of gene flow and migration: F_{ST} doesn't equal $1/(4Nm+1)$. *Heredity* 82: 117-125.
- Wiehe T, Slatkin M. 1998. Epistatic selection in a multi-locus Levene model and implications for linkage disequilibrium. *Theor. Popul. Biol.* 53: 75-84.
- Wiehe T, Stephan W. 1993. Analysis of a genetic hitchhiking model and its application to DNA polymorphism data from *Drosophila melanogaster*. *Mol. Biol. Evol.* 10: 842-854.
- Wiehe T. 1998. The effect of selective sweeps on the variance of the allele distribution of a linked multiallele locus: hitchhiking of microsatellites. *Theoretical Population Biology.* 53: 272–283.
- Wilding CS, Butlin RK, Grahame J. 2001. Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *J. Evol. Biol.* 14: 611-619.
- Williard A, Syring J, Gernandt DS, Liston A, Cronn R. 2007. Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. *Mol. Biol. Evol.* 24: 90-101.
- Wiuf C, Christensen T, Hein J. 2001. A simulation study of the reliability of recombination detection methods. *Mol Biol. Evol.* 18: 1929-1939.
- Wootton JC, Feng X, Ferdig MT, Cooper RA, Mu J, Baruch DI, Magill AJ, Su X-Z. 2002. Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature* 418: 320–323.
- Wright S. 1931. Evolution in Mendelian populations. *Genetics* 16: 97-159.
- Wright S. 1943. Isolation by distance. *Genetics* 28: 114-138.
- Wright S. 1951. Evolution in Mendelian populations. *Genetics* 16: 97-159.
- Wright S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 19: 395-420.
- Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. 2005. The effects of artificial selection on the maize genome. *Science* 308: 1310–1314.

- Wright SI, Gaut BS. 2005. Molecular population genetics and the search for adaptive evolution in plants. *Mol. Biol. Evol.* 22: 506-519.
- Yamaguchi-Shinozaki K, Kasuga M, Liu Q, Nakashima K, Sakuma Y, Abe H, Shinwari ZK, Seki M, Shinozaki K. 2002. Development of drought-resistant and water stress-tolerant crops through biotechnology. In: JIRCAS, editors. *Proceedings of the 8th JIRCAS International Symposium, Water for Sustainable Agriculture in Developing Regions: More Crop for Every Scarce Drop*. Tsukuba, 27-28 November 2001. p. 23-34.
- Yang S-H, Wang H, Sathyan P, Stasolla C, Loopstra CA. 2005. Real-time RT-PCR analysis of loblolly pine (*Pinus taeda*) arabinogalactan-protein-like genes. *Physiol. Plant.* 124: 91-106.
- Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. *TREE* 15: 496-503.
- Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155: 431-449.
- Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* 19: 908-917.
- Yu JM, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler E. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* 38: 203-208.
- Yu LX, Chamberland H, Lafontain JG, Tabaeizadeh Z. 1996. Negative regulation of gene expression of a novel proline-, threonine-, and glycine-rich protein by water stress in *Lycopersicon chilense*. *Genome* 39: 1185-1193.
- Yu N, Chen F-C, Ota S, Jorde LB, Pamilo P, Patthy L, Ramsay M, Jenkins T, Shyue S-K, Li W-H. 2002. Larger genetic differences within Africans than between Africans and Eurasians. *Genetics* 161: 269-274.
- Zeng K, Shi S, Fu Y-X, Wu C-I. 2006. Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* 174: 1431-1439.
- Zhang JW, Fins L, Marshall JD. 1994. Stable isotope discrimination, photosynthetic gas exchange, and growth differences among western larch families. *Tree Physiology* 14: 531-539.
- Zhang JW, Marshall JD. 1994. Population differences in water-use efficiency of well-watered and water-stressed western larch seedlings. *Can. J. For. Res.* 24: 92-99.
- Zhao KY, Aranzana MJ, Kim S, Lister C, Shindo C, Tang CL, Toomajian C, Zheng HG, Dean C, Marjoram P, Nordborg M. 2007. An *Arabidopsis* example of association mapping in structured samples. *Plos Genetics* 3: 71-82.

ANNEXES

**Annexe 1. Article de Eveno et al. (2008)
publié dans *Molecular Biology and Evolution***

**Annexe 2. Information supplémentaire de
l'article Eveno *et al.* (2008)**

SUPPLEMENTARY INFORMATION

Material and Methods

Nuclear Simple Sequence Repeats (SSRs) analyses

Numbers of observed alleles (A), within-population diversity (H_s), fixation indices (Weir and Cockerham 1984) including inbreeding coefficient (F_{IS}) averaged across populations (with exact tests for F_{IS} performed within each population), and the between population component of variance in allelic frequencies (F_{ST}) were computed at each SSR using Genepop (http://wbiomed.curtin.edu.au/genepop/genepop_op6.html). The F_{ST} -analog for microsatellites, based on the variance in allele size (R_{ST} , Michalakis and Excoffier 1996), was also computed (Supp. Table S2).

SSR data were also used for defining groups of populations that could be considered as homogenous gene pools. Two clustering methods were combined to define homogeneous gene pools (detailed in methods supplementary material). The first one was based on the pairwise F_{ST} matrix from which a dendrogram was built with R ver. 2.0.1 (Ihaka and Gentleman 1996) using Ward's algorithm (Ward 1963), which allows populations to be clustered at each step by minimising the variation (in sums of squares deviations from the cluster mean) within cluster comparing to the variation between clusters. The Unweighted Pair-Groups Method with Arithmetic mean (UPGMA) (Sneath and Sokal 1973) gave very similar results and was thus not presented. An arbitrary threshold was fixed that identified groups of populations among which F_{ST} values were at a maximum of 0.06. The second method inferred the population clustering from a Bayesian algorithm implemented in Structure ver. 2.0 (Pritchard, Stephens and Donnelly 2000), considering a contact zone model with two groups (related to different lineages; see Burbán and Petit 2003; Bucci et al. 2007) analysed independently (western range and eastern range including Morocco), as suggested by Evanno, Regnaut, and Goudet (2005) when patterns of dispersal among populations are not homogeneous. Their statistic Δk was used to estimate the optimal number of clusters in each group. Admixture models with correlated allele frequencies, and both burn-in and run-length periods of 10^6 iterations were used for the Markov Chain Monte Carlo (MCMC) runs.

Since the "Structure" algorithm was unable to separate Moroccan and Tunisian populations (see suppl. Figure 2A), these populations were nevertheless considered as different gene pools due to their very high F_{ST} values (0.29 in supplementary Table S3B; see also Supp. Figure S2), and phylogeographic evidence from cpSSR data (Bucci et al. 2007). Also, Oria was not included in group S due to previous evidence on their high genetic differentiation (using allozymes and cpSSRs), and to its geographic isolation (González-Martínez et al. 2007; see also suppl. Figure S1B-C). Each group showed a range of within-group pairwise F_{ST} from 0 to 0.05 (supplementary Table S3A).

Litterature cited

- Michalakis Y, Excoffier L. 1996. A generic estimate of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics*. 142:1061-1064.
- Ihaka R, Gentleman R. 1996. R: a language for data analysis and graphics. *J. Comput Graph Statist*. 5:299-314.
- Ward JH. 1963. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*. 58:236-244.
- Sneath PHA, Sokal RR. 1973: Numerical taxonomy, the principles and practice of numerical classification. WH Freeman (eds), San Francisco, USA.
- Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*. 14:2611-2620.

Supplementary Table S1. Primer pairs, diversity and differentiation estimates across the eight SSRs.

SSR locus	Forward primer	Reverse primer	24 populations					10 populations					six gene pools including 10 populations			
			$A^{(d)}$	H_S	F_{IS}	F_{ST}	R_{ST}	$A^{(d)}$	H_S	F_{IS}	F_{ST}	R_{ST}	H_S	F_{IS}	F_{ST}	R_{ST}
<i>NZPR413</i> ^(a)	TGAACCTCGATGGAATAGCC	CCCGCCTTGCATCAATTA	7	0.453	-0.033	0.189	0.225	6	0.458	-0.046	0.285	0.318	0.467	-0.028	0.310	0.272
<i>NZPR1078</i> ^(a)	TGGTGATCAAGCCTTTTTCC	GTTGATGAGTGATGGCATGG	6	0.617	0.055	0.042	0.036	6	0.575	0.006	0.082	0.061	0.552	0.001	0.089	0.050
<i>FRPP91</i> ^(b)	GTACTCCACATAAAATGAGACTT	CCGAAATACATTGCAGGTTA	29	0.773	0.215*** ^(e)	0.130	0.128	27	0.721	0.196*** ^(e)	0.189	0.212	0.715	0.240*** ^(e)	0.195	0.150
<i>FRPP94</i> ^(b)	GGCAAACCTCTTTTAGAGTGC	TTTGTGATTTTTCTTGAAATCTAA	16	0.692	0.032	0.084	0.170	15	0.734	0.015	0.126	0.243	0.758	0.004	0.127	0.220
<i>ITPH4516</i> ^(b)	TGATGCAAACAAGTTCCATG	AGCACTCGCTAAACTATGAAGG	16	0.786	0.104*** ^(f)	0.104	0.087	16	0.753	0.055*** ^(f)	0.143	0.121	0.758	0.094*** ^(f)	0.143	0.109
<i>ctg64</i> ^(a)	GGAAGCTGTTACAAGTGCGG	ATCGAGAAGAGAGGAAGGGC	6	0.538	0.041	0.098	0.143	3	0.483	-0.043	0.144	0.182	0.475	-0.061	0.161	0.187
<i>A6F03</i> ^(c)	CCTGAAAATCGACGGATCG	ATGGTATTTGCGGGTTGC	12	0.616	0.038	0.122	0.170	10	0.561	0.097	0.179	0.273	0.554	0.085	0.199	0.229
<i>ctg275</i> ^(a)	ACGGAGATATATTGCTGGCG	AAAGAATAACGTGAAACAAACCC	23	0.727	0.027	0.089	0.101	16	0.709	0.005	0.102	0.159	0.681	0.047	0.106	0.079
Mean (all SSRs)			14.375	0.520	0.069	0.104	0.147	12.375	0.624	0.036	0.154	0.225	0.620	0.048	0.164	0.176
Mean (excluding NZPR413)			15.429	0.528	0.073	0.096	0.119	13.286	0.648	0.047	0.138	0.179	0.642	0.059	0.146	0.146

^(a) Chagné et al. 2004.

^(b) Mariette et al. 2001.

^(c) Guevara et al. 2005.

^(d) Number of observed alleles

^(e) Fisher exact test significant at $P < 0.001$ within populations Mimizan, Pleucadec, Arenas and Oria, and groups F (continental France), S (north-western and central Spain) and O (Oria).

^(f) Fisher exact test significant at $P < 0.001$ within populations Mimizan and Tamrabta, and groups F and M (Tamrabta).

Supplementary Table S2. Matrix of pairwise F_{ST} between the 6 homogeneous gene pools including the 24 populations (B).

Gene pools	C	F	S	O	M	T
C	*					
F	0.250	*				
S	0.162	0.046	*			
O	0.174	0.102	0.047	*		
M	0.300	0.264	0.196	0.156	*	
T	0.313	0.273	0.180	0.183	0.289	*

C: Corsica, F: continental France, S: north-western and central Spain, O: Oria, M: Tamrabta, T: Tabarka.

Supplementary Table S3. Molecular differentiation estimates for 11 candidate genes between the six gene pools (excluding missing data).

Candidate genes	N ^(a)	S _T ^(b)	N _{st}	N _H ^(c)	S _{NR} ^(d)	F _{ST}	Range of variation ^(e)	N _{HG} ^(f)	F _{STHG}
<i>PR-AGP4</i>	18.3	85	0.253*** ^(g)	25	19	0.266***	0.006 - 0.676	9	0.294***
<i>CCoAOMT</i>	14.3	31	0.151***	20	13	0.113***	-0.008 - 0.336	7	0.156**
<i>GRP3</i>	16.3	8	0.278***	17	7	0.123***	0.023 - 0.375	10	0.153***
<i>Glucan</i>	14.3	37	0.057	10	5	0.061*	0.004 - 0.073	4	0.058
<i>dhn-1</i>	17.2	20	0.084**	11	7	0.064	0.011 - 0.133	7	0.064
<i>erd3</i>	9.7	13	0.263***	12	4	0.288***	-0.050 - 0.536	10	0.263***
<i>dhn-2</i>	9.3	21	0.004	17	6	0.013	-0.031 - 0.141	7	-0.005
<i>lp3-1</i>	9.5	12	0.047	22	9	0.044*	-0.014 - 0.160	17	0.042
<i>lp3-3</i>	9	9	0.147**	8	6	0.263***	-0.081 - 0.292	6	0.167**
<i>pp2c</i>	8.8	12	0.045	10	4	-0.009	0.038 - 0.195	7	-0.044
<i>rd21A</i>	8.8	45	0.179***	15	12	0.130***	-0.045 - 0.362	5	0.158*
Total		293		167	92			89	
Mean			0.137	15.2		0.123		8.1	0.119

^(a) Mean number of haploid sequences analysed per population; ^(b) Total number of polymorphisms excluding missing data; ^(c) Number of haplotypes used for the estimation of F_{ST}; ^(d) Number of polymorphisms excluding rare ones (minor allele frequency < 5%) and those in complete LD; ^(e) Min-max values of F_{ST} at each SNP, excluding rare polymorphisms and those in complete LD. This range is slightly different from the range of F_{ST} values when including missing data; ^(f) Number of haplogroups used for the estimation of F_{STHG}; ^(g) Permutation tests: ***: P<0.001, **: P<0.01, *:P<0.05.

Supplementary Table S4. Outliers detected at the haplotype level (A), and at the SNP level for SNPs having a heterozygosity greater than 0.1 (B), comparing results from the 6 groups analysis, using either the “Fdist2”, “FstSNP” or “BayesFst” methods.

A)

Gene	F_{ST}	"Fdist2"		"BayesFst"		
		P-value ^(c)	"HG" ^(a)			"H" ^(b)
			P-value ^(c)	α_i ^(d)	SD ^(e)	
<i>PR-AGP4</i>	0.266		3%	0.77	0.47	
<i>erd3</i>	0.288		10%	0.77	0.56	3%
<i>dhn-2</i>	0.013	2%	6%	-0.73	0.61	7%
<i>lp3-1</i>	0.044	10%	1%	-1.38	0.64	1%
<i>pp2c</i>	-0.009	4%	4%	-1.21	0.7	6%
<i>dhn-1</i>	0.064		9%	-0.65	0.58	5%

B)

SNP	Gene region / type of mutation ^(f)	F_{ST}	"Fdist2"	"FstSNP"	"BayesFst"		
			P-value ^(c)	P-value ^(c)	P-value ^(c)	α_i ^(d)	SD ^(e)
<i>PR-AGP4</i> S182	intron / -	0.428	6%		4%	0.92	0.53
<i>PR-AGP4</i> S331	intron / -	0.437	6%		4%	0.94	0.55
<i>PR-AGP4</i> S384	intron / -	0.013		10%			
<i>PR-AGP4</i> S472	intron / -	0.414	6%				
<i>PR-AGP4</i> S525	intron / -	0.588	2%	4%			
<i>PR-AGP4</i> S666	intron / -	0.390	%		4%	1.12	0.55
<i>PR-AGP4</i> S762	intron / -	0.374	%		7%	0.86	0.58
<i>PR-AGP4</i> S1409	3'UTR / -	0.410	6%		2%	0.98	0.56
<i>GRP3</i> S164	exon / s	0.358	10%		10%	0.58	0.59
<i>erd3</i> S42	exon / ns	0.536	2%		4%	1.37	0.68
<i>rd21A</i> S3	5'UTR / -	-0.019		10%			
<i>rd21A</i> S494	exon / s	-0.039	2%	8%			
<i>CCoAOMT</i> I832	exon / ns	-0.033	2%	2%			
<i>CCoAOMT</i> S980	exon / ns	-0.031	2%	2%			
<i>dhn-2</i> S332	intron / -	-0.012					
<i>dhn-2</i> S406	exon / s	-0.007					
<i>lp3-1</i> S97	exon / s	-0.014					
<i>lp3-1</i> S324	3'UTR / -	-0.011		10%			
<i>lp3-3</i> S43	exon / s	-0.081	2%	2%			
<i>dhn-1</i> S725	exon / ns	-0.011	2%				

^(a) Groups of haplotypes including haplotypes with less than 10% of polymorphisms differences relative to the total number of polymorphisms along each gene.

^(b) All haplotypes are considered separately as independent alleles.

^(c) Significance level (P-value) for detected outliers .

^(d) Mode of the posterior distribution of α_i .

^(e) Standard deviation of the posterior distribution of α_i .

^(f) s: synonymous polymorphism, ns: non synonymous polymorphism.

Suppl. Table S5. Correlation coefficients between pairs of climatic variables, latitude and elevation for the 10 contrasted populations. Variables names are as in Table 1. Correlation values significant at P<5% are in bold.

	Latitude	Elevation	ANMR	RDRYM	ANMT variance	ANMT	MTCM	MTWM
Latitude	1							
Elevation	-0.62254	1						
ANMR	0.06995	-0.16544	1					
RDRYM	0.48598	-0.4897	0.40599	1				
ANMT	-0.67072	0.03863	0.17883	0.13224	1			
VarANMT	-0.08554	0.13336	0.82241	0.12105	-0.23264	1		
MTCM	0.21976	-0.45955	0.23995	0.75018	0.89348	-0.16279	1	
MTWM	-0.70622	0.60357	-0.18392	-0.17399	0.70109	-0.1483	0.31761	1

SUPPLEMENTARY FIGURES LEGENDS

Supplementary Figure S1. Dendrogram based on the pairwise F_{ST} matrix between the 10 populations using Ward (1963) algorithm, and either the whole set of SSRs (A), or excluding from the analysis 2 SSRs, *FRPP91* and *ITPH4516* (B) with significant heterozygote deficits. The dotted line is the threshold (fixed at a F_{ST} value of 0.06) used for identifying groups of populations. C: Corsica, F: continental France, S: north-western and central Spain, O: Oria, M: Tamrabta, T: Tabarka.

Supplementary Figure S2. Linkage disequilibrium between SNPs which are polymorphic ($MAF > 5\%$) and not in complete LD, in two sets of genes included in this study, considering gametes from all populations together. A) LD plot for genes *PR-AGP4*, *CCoAOMT*, *Glucan*, *GRP3*, *dhn1*; B) LD plot for genes *dhn2*, *erd3*, *lp3-1*, *lp3-3*, *pp2c*, *rd21*. In both plots, correlation between allelic frequencies to the square are represented above the diagonals, and probability for exact tests of independence below the diagonals.

Supplementary Figure S3. (A) Spatial distribution of maritime pine haplotypes at locus *GRP3*. (B) Ward's minimum-variance dendrogram based on the percentage of differences between pairs of haplotypes for *GRP3*.

Supplementary Figure S4. Distribution of observed F_{ST} values for each SSR, as a function of heterozygosities (H_S), considering six groups of populations with a sample size of 82 gametes (A), and the 10 populations separately with a sample size of 48 gametes (B). The 10% and the 90% quantiles of the simulated neutral envelopes were built using the infinite mutation site model implemented in Fdist2" (Beaumont and Nichols 1996). Considering the six groups of populations (A), the grey and black lines represent the envelopes based respectively on a mean F_{ST} of 0.164 (estimated from the eight SSRs) and a mean F_{ST} of 0.146 (estimated from seven SSRs excluding the outlier SSR *NZPR413*). Considering the 10 populations separately (B), the grey and black lines represent the envelopes based respectively on a mean F_{ST} of 0.154 (estimated from the eight SSRs) and a mean F_{ST} of 0.138 (estimated from seven SSRs excluding the SSR *NZPR413*). The outlier SSR *NZPR413* is indicated by an arrow.

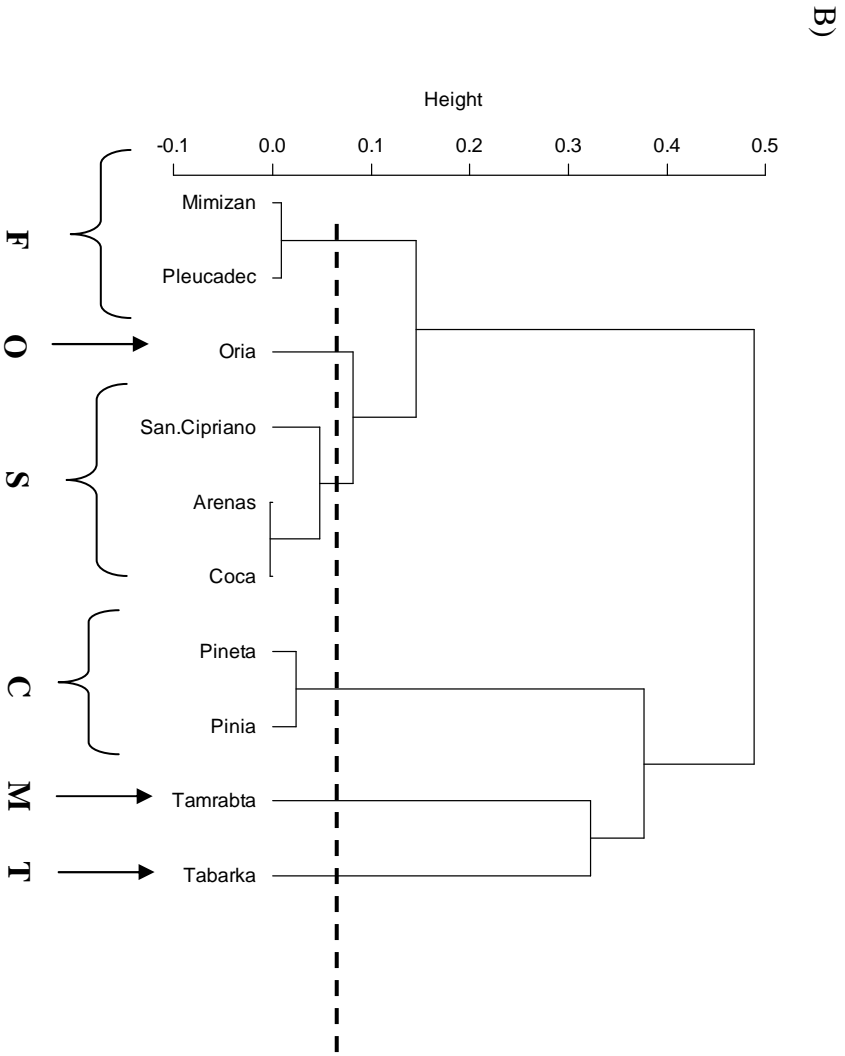
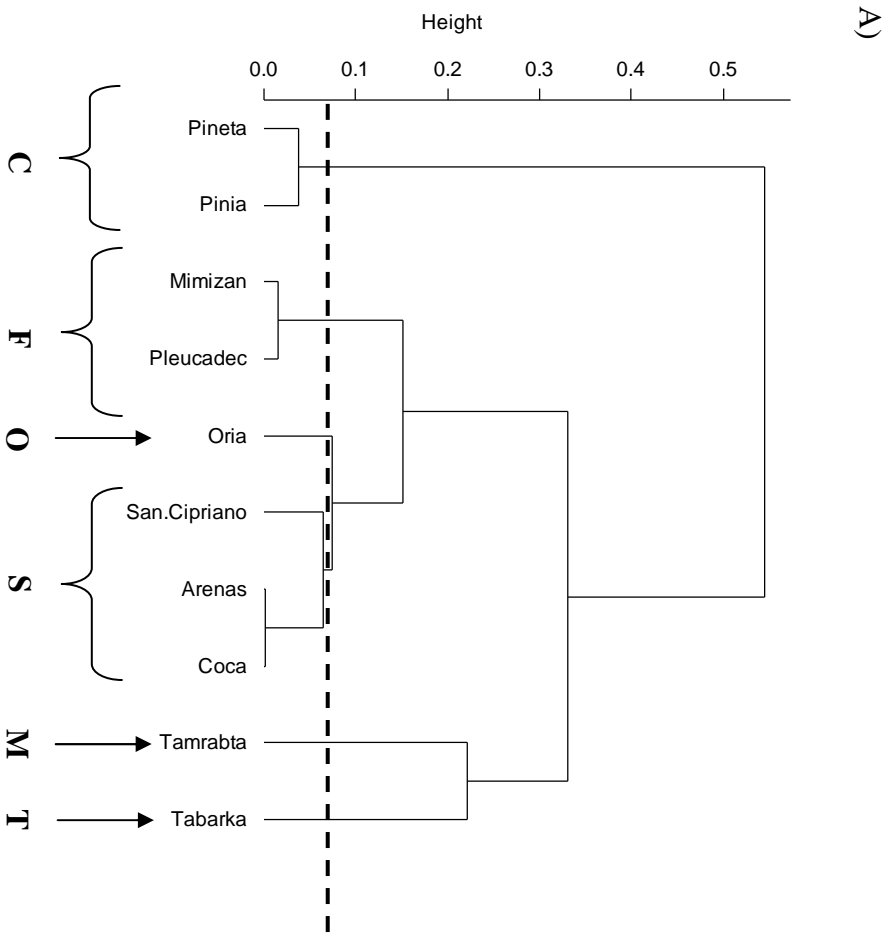
Supplementary Figure S5. Examples of posterior distributions of α_i for detected outlier genes (after grouping close haplotypes), and for a subsample of outlier SNPs in the 10 populations analysis.

Supplementary Figure S6. Distribution of observed F_{ST} values for each gene, for SNPs within genes, and for SSRs, as a function of the mean heterozygosities across groups (H_S), considering six groups of populations with a neutral F_{ST} of 0.146. The 5% and the 95% quantiles of the simulated neutral envelopes are represented by continuous and dotted lines (corresponding to a two-tailed global test at 10%), respectively for the "Fdist2" and "FstSNP" methods. Black lines represent the boundaries of envelopes for the first set of genes (with 24 gametes for SNPs, and 16 gametes for genes without missing data on average per group), and grey lines those for the second set of genes (with nine gametes for both SNPs and genes on average per group). Outlier genes detected with "Fdist2" are in bold face and outlier SNPs detected with "FstSNP" are indicated by arrows.

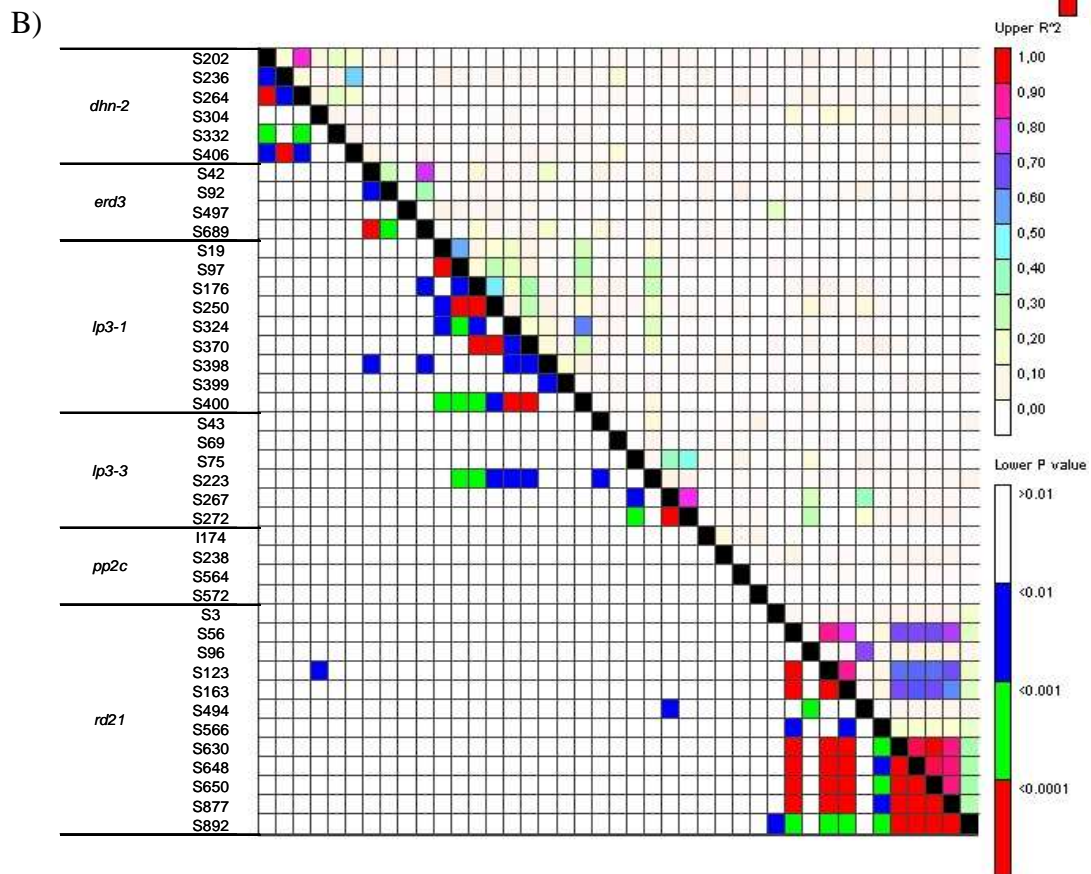
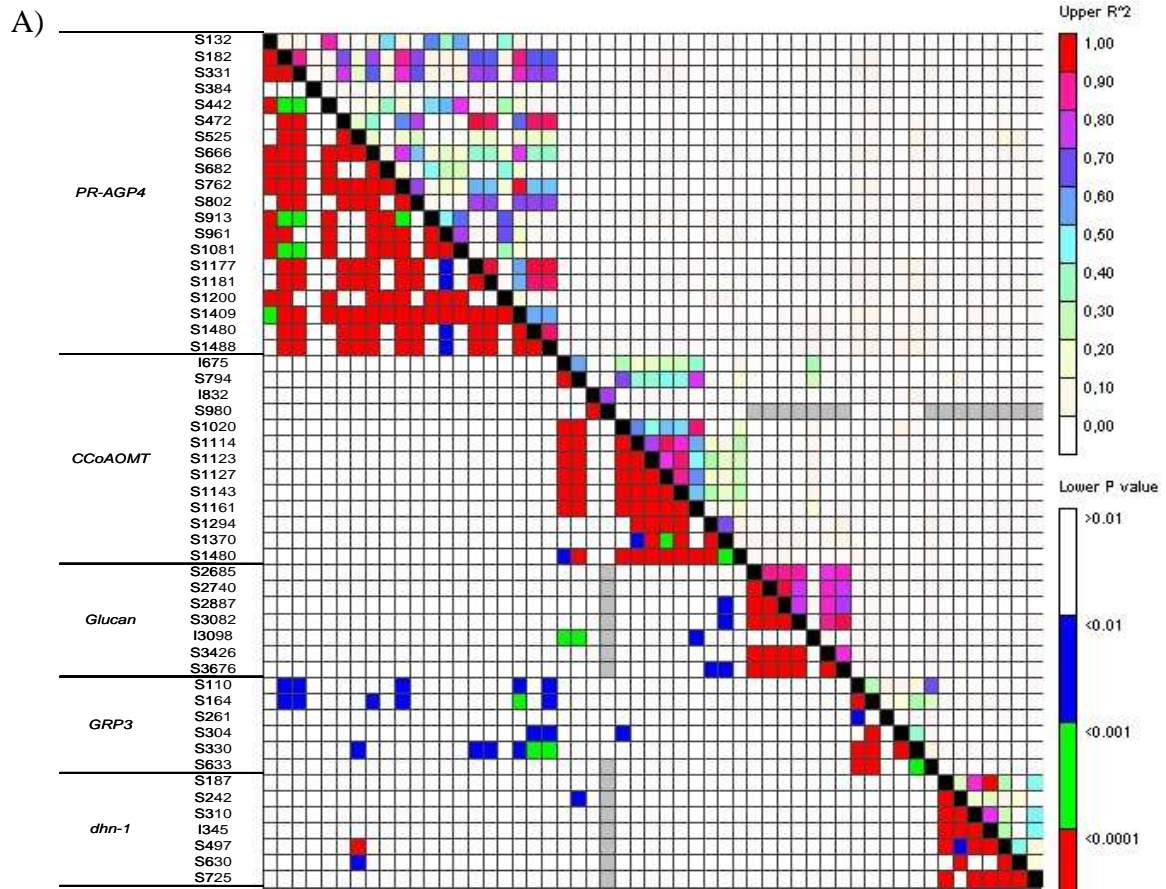
Supplementary Figure S7. Variation of allelic frequencies at SNPs detected as positive outliers with "FstSNP" or "BayesFst" within each gene pool (A), and within each population

(B) along a latitudinal gradient. C: Corsica, F : continental France, S : north-western and central Spain, O : Oria, M: Tamrabta, T: Tabarka.

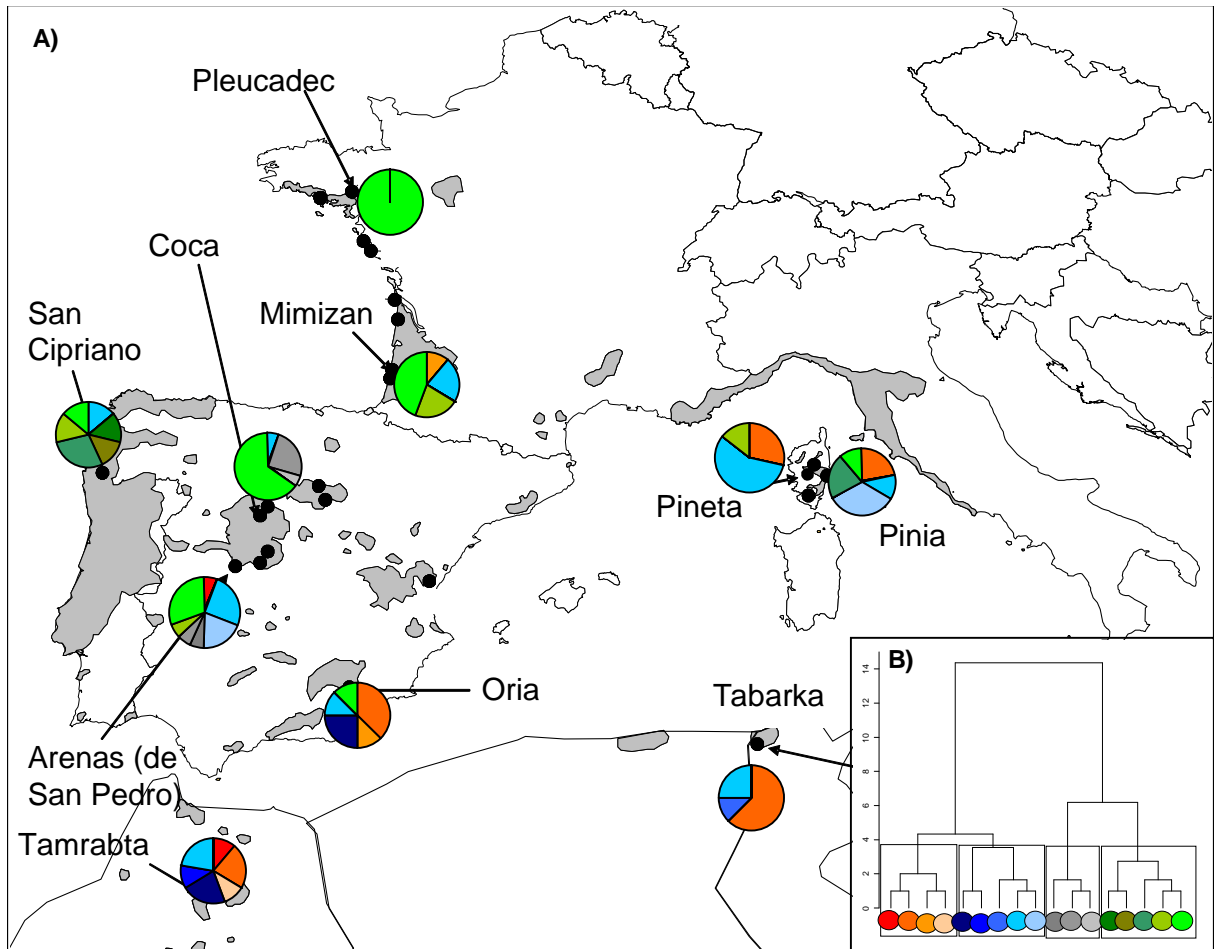
Supplementary Figure S1



Supplementary Figure S2

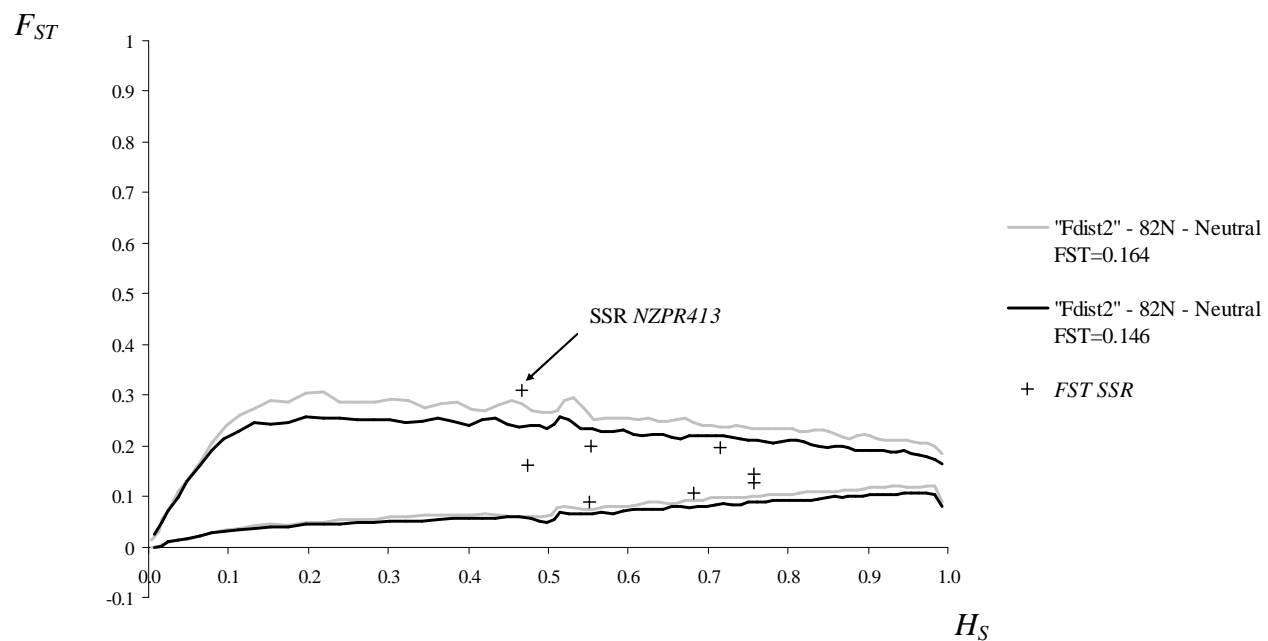


Supplementary Figure S3

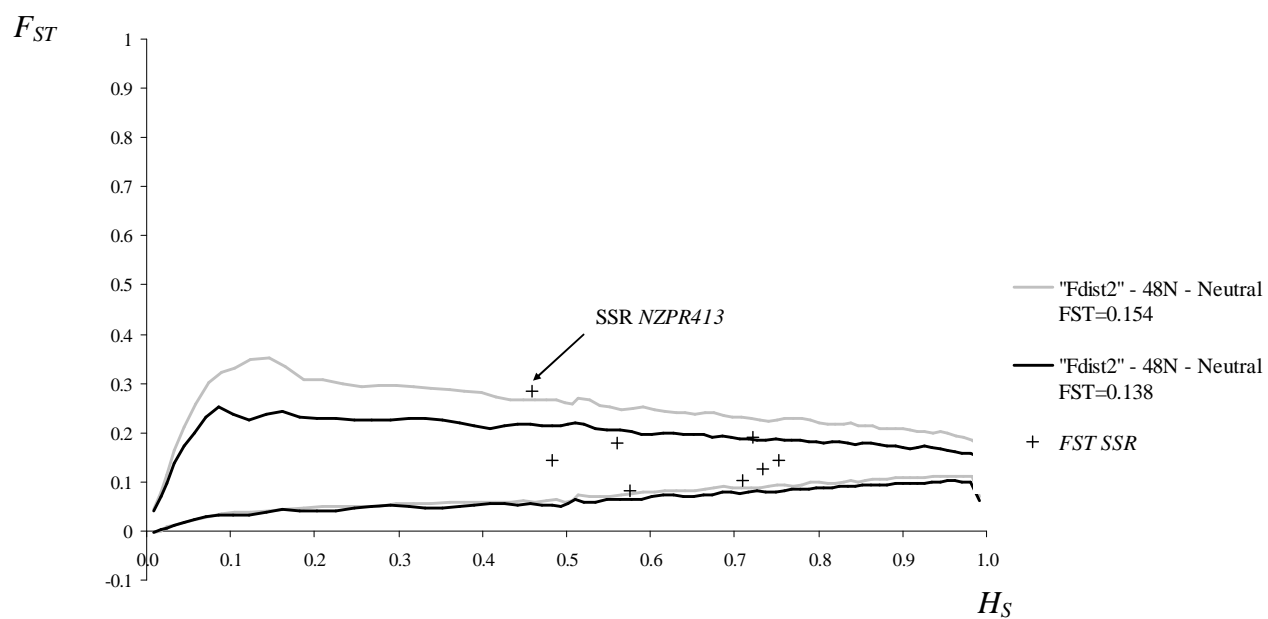


Supplementary Figure S4

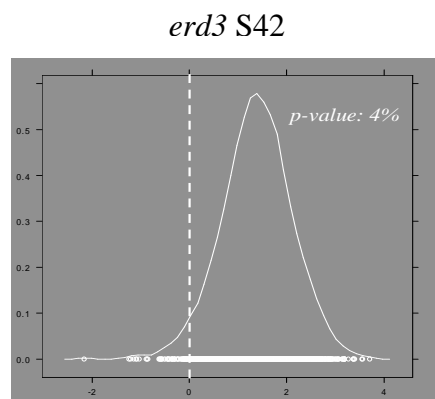
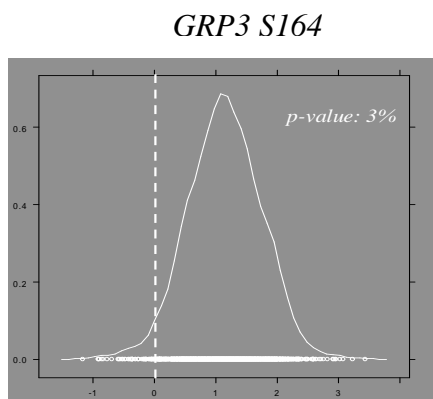
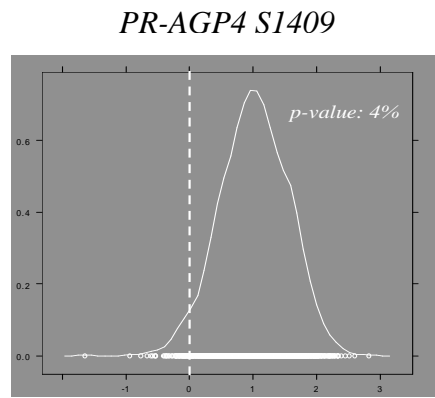
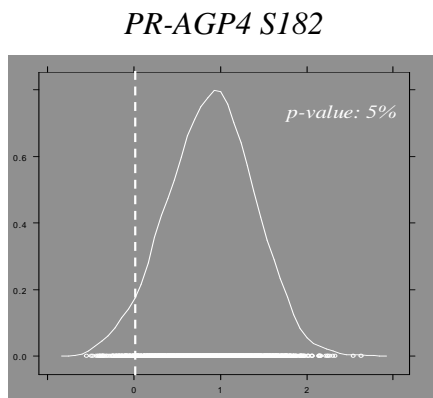
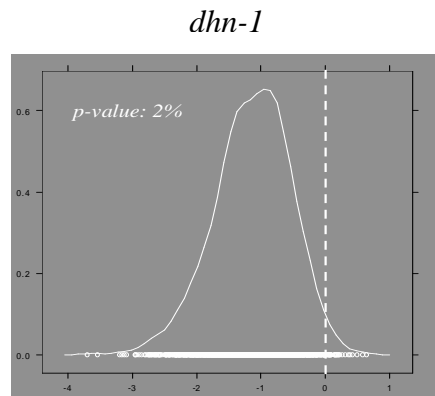
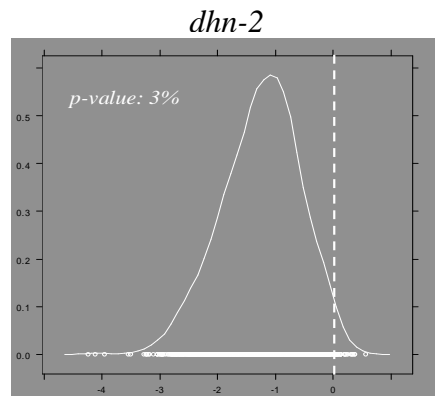
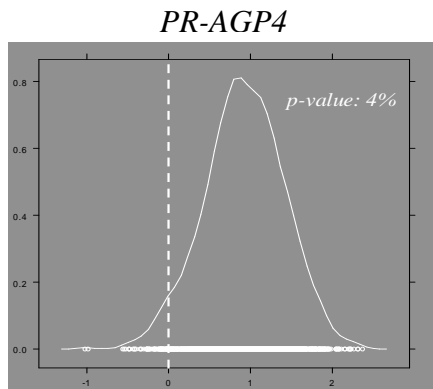
A)



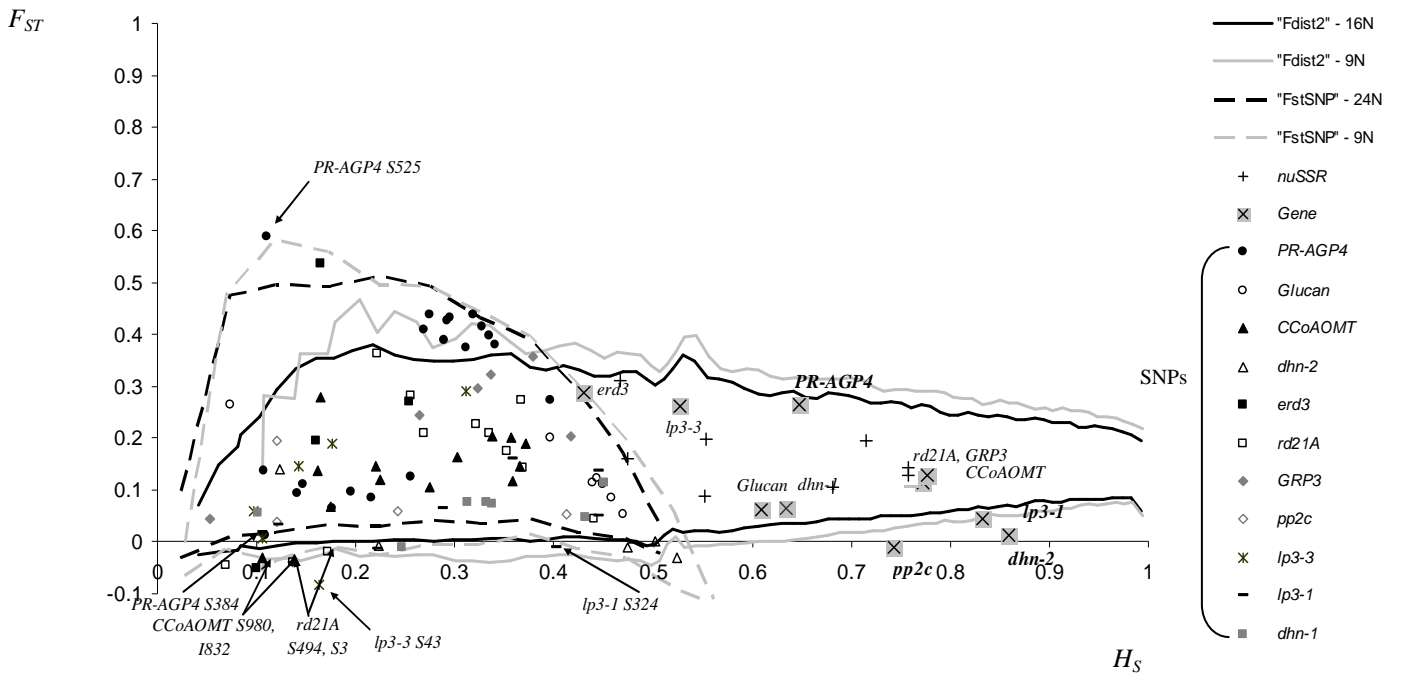
B)



Supplementary Figure S5

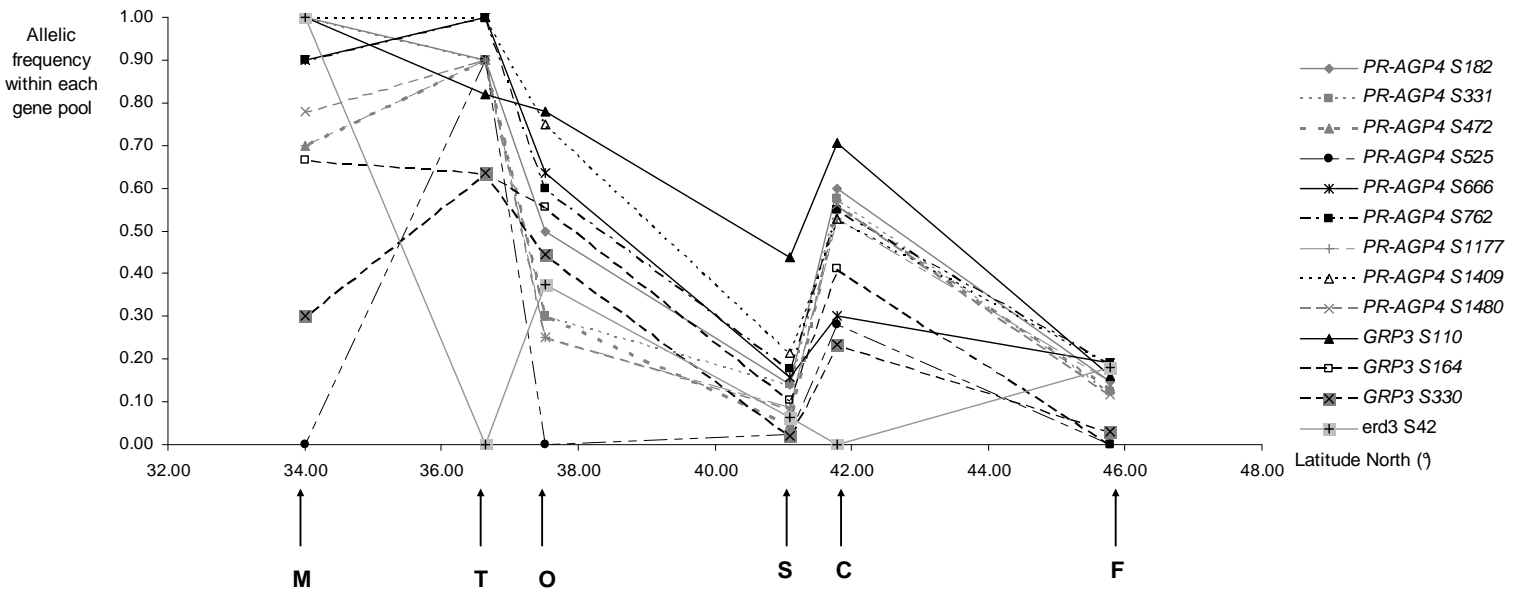


Supplementary Figure S6

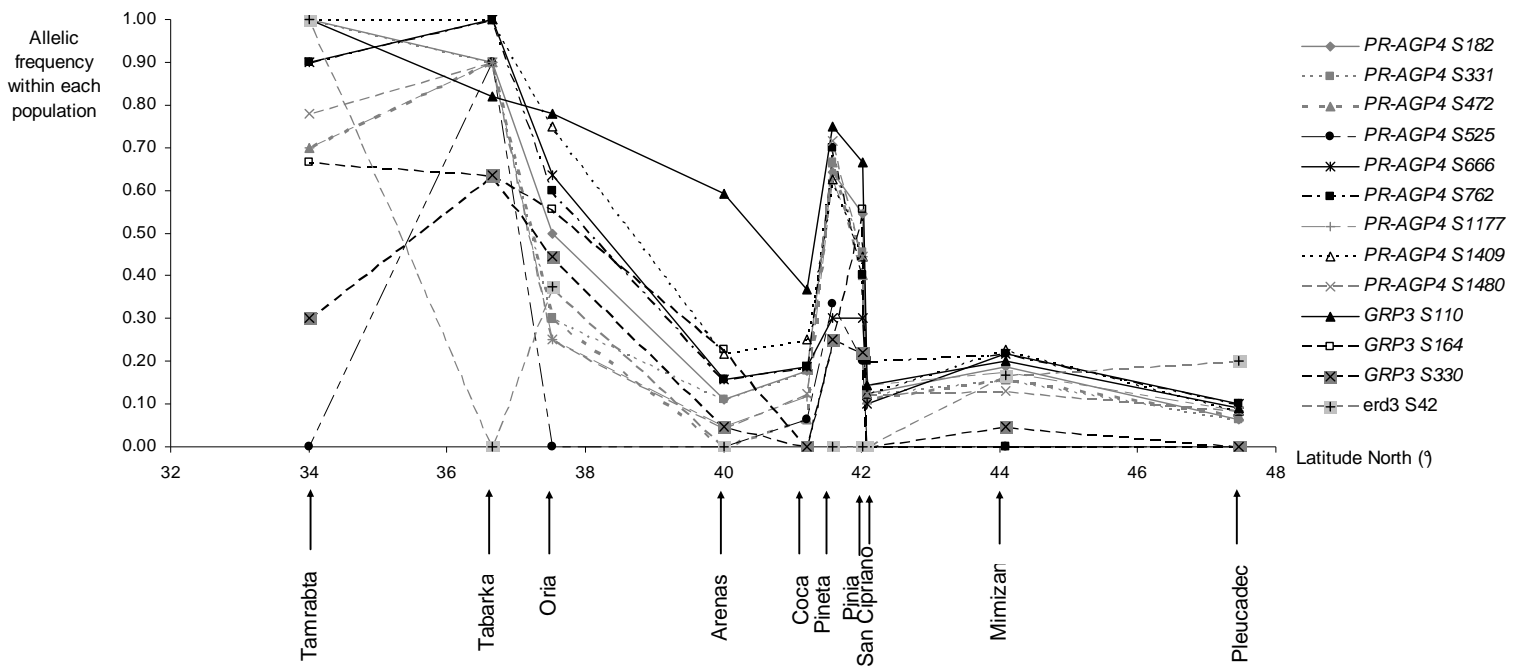


Supplementary FigureS7

A)



B)



**Annexe 3. Analyse de la distribution des
résidus pour *HT02, HT03, DC, PolyT, TB,*
*NB, $\delta^{13}C$***

Analyse de la distribution des résidus pour HT02

Moments

N	2815	Sum Weights	2815
Mean	0	Sum Observations	0
Std Deviation	7.7171825	Variance	59.5549057
Skewness	0.41616511	Kurtosis	1.10847972
Uncorrected SS	167587.505	Corrected SS	167587.505
Coeff Variation	.	Std Error Mean	0.14545196

Basic Statistical Measures

Location		Variability	
Mean	0.00000	Std Deviation	7.71718
Median	-0.38884	Variance	59.55491
Mode	-6.20865	Range	75.74414
		Interquartile Range	9.78342

NOTE: The mode displayed is the smallest of 7 modes with a count of 4.

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----
Student's t	t 0	Pr > t 1.0000
Sign	M -68.5	Pr >= M 0.0104
Signed Rank	S -67593.5	Pr >= S 0.1171

Tests for Normality

Test	--Statistic--	-----p Value-----
Kolmogorov-Smirnov	D 0.034123	Pr > D <0.0100
Cramer-von Mises	W-Sq 0.846817	Pr > W-Sq <0.0050
Anderson-Darling	A-Sq 5.128184	Pr > A-Sq <0.0050

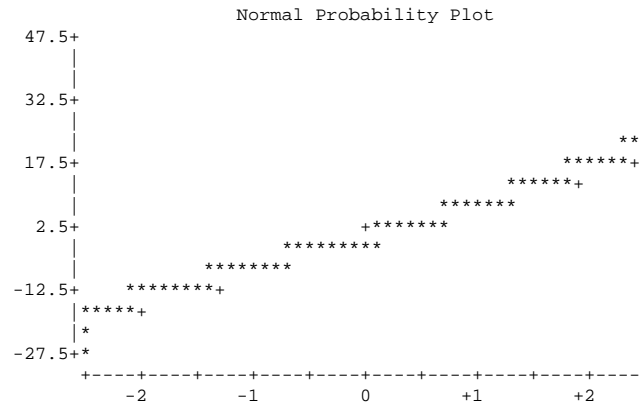
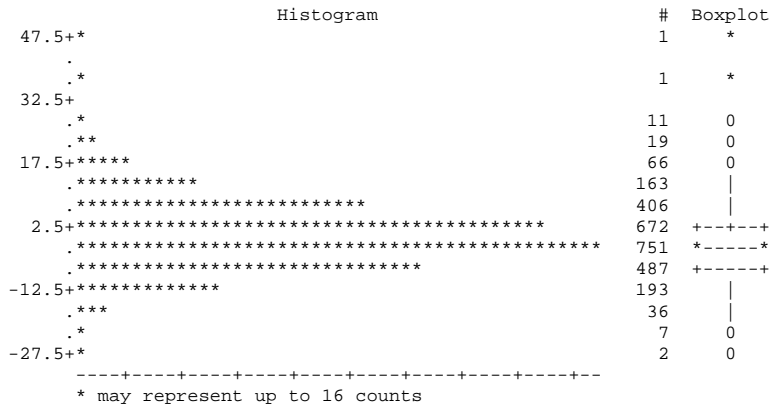
Quantiles (Definition 5)

Quantile	Estimate
100% Max	47.948347
99%	20.468298
95%	13.356414
90%	9.653094
75% Q3	4.648490
50% Median	-0.388845
25% Q1	-5.134933
10%	-9.480225
5%	-11.938647
1%	-17.480225

0% Min -27.795790

Extreme Observations

-----Lowest-----		-----Highest-----	
Value	Obs	Value	Obs
-27.7958	1829	28.6485	306
-25.2087	1319	28.7913	2011
-22.7899	21	29.8969	769
-22.7755	877	36.2341	2729
-21.9229	2479	47.9483	2365



Analyse de la distribution des résidus pour HT03

Moments

N	2815	Sum Weights	2815
Mean	0	Sum Observations	0
Std Deviation	14.7099778	Variance	216.383446
Skewness	-0.0302692	Kurtosis	0.42711752
Uncorrected SS	608903.018	Corrected SS	608903.018
Coeff Variation	.	Std Error Mean	0.27725081

Basic Statistical Measures

Location		Variability	
Mean	0.0000	Std Deviation	14.70998
Median	0.0155	Variance	216.38345
Mode	-19.1013	Range	128.00000
		Interquartile Range	18.76601

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----		
Student's t	t	0	Pr > t	1.0000
Sign	M	1.5	Pr >= M	0.9699
Signed Rank	S	7246.5	Pr >= S	0.8666

Tests for Normality

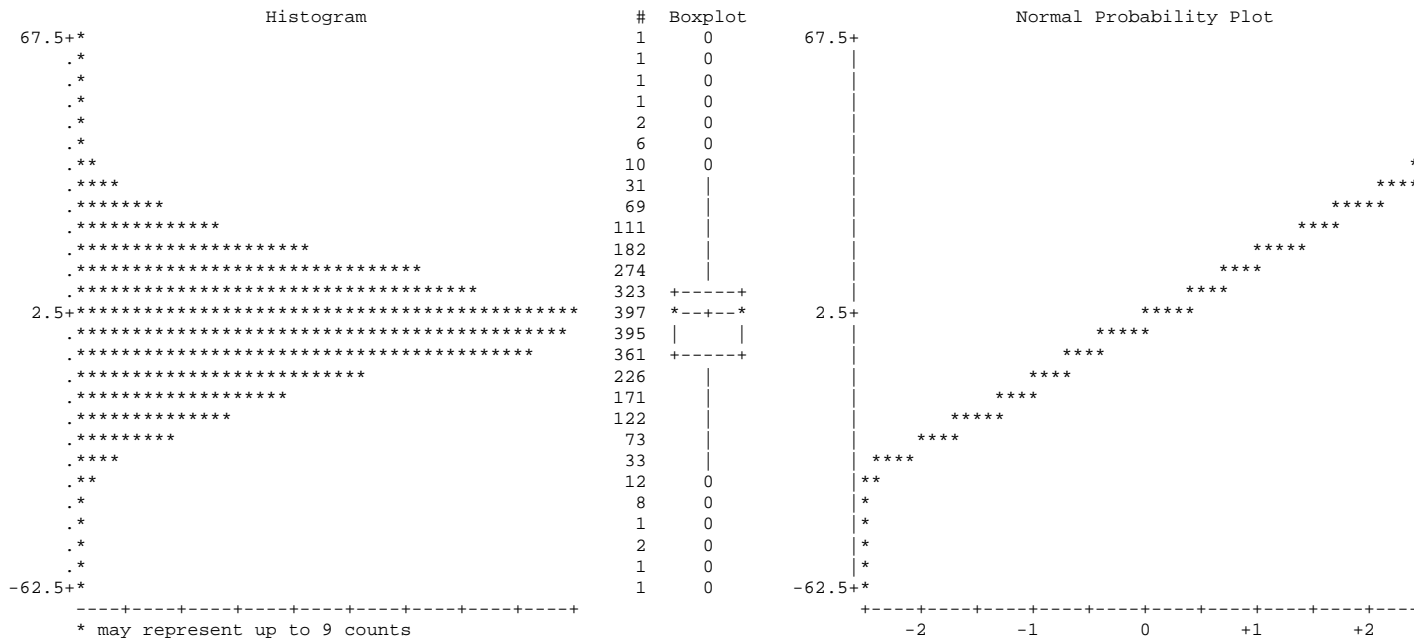
Test	--Statistic--	-----p Value-----		
Kolmogorov-Smirnov	D	0.020861	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.172283	Pr > W-Sq	0.0128
Anderson-Darling	A-Sq	0.962426	Pr > A-Sq	0.0167

Quantiles (Definition 5)

Quantile	Estimate
100% Max	65.5284718
99%	34.1221950
95%	23.8486008
90%	18.5582291
75% Q3	9.6936235
50% Median	0.0154755
25% Q1	-9.0723897
10%	-19.0206622
5%	-24.4731463
1%	-34.1635193
0% Min	-62.4715282

Extreme Observations

-----Lowest-----		-----Highest-----	
Value	Obs	Value	Obs
-62.4715	1829	48.0705	2128
-57.6438	2479	51.9276	2247
-54.7867	1319	56.3273	2102
-52.1514	1755	60.0705	306
-45.7867	51	65.5285	1025



Analyse de la distribution des résidus pour DC

Moments

N	2815	Sum Weights	2815
Mean	0	Sum Observations	0
Std Deviation	4.03407944	Variance	16.2737969
Skewness	-0.1379855	Kurtosis	0.19631676
Uncorrected SS	45794.4646	Corrected SS	45794.4646
Coeff Variation	.	Std Error Mean	0.07603355

Basic Statistical Measures

Location		Variability	
Mean	0.00000	Std Deviation	4.03408
Median	0.16986	Variance	16.27380
Mode	-6.06682	Range	30.19015
		Interquartile Range	5.30916

NOTE: The mode displayed is the smallest of 21 modes with a count of 2.

Tests for Location: Mu0=0

Test	-Statistic-		-----p Value-----
Student's t	t	0	Pr > t 1.0000
Sign	M	47	Pr >= M 0.0796
Signed Rank	S	27182	Pr >= S 0.5284

Tests for Normality

Test	--Statistic---		-----p Value-----
Kolmogorov-Smirnov	D	0.019009	Pr > D 0.0149
Cramer-von Mises	W-Sq	0.203981	Pr > W-Sq <0.0050
Anderson-Darling	A-Sq	1.110338	Pr > A-Sq 0.0070

Quantiles (Definition 5)

Quantile	Estimate
100% Max	14.344890
99%	9.217078
95%	6.526163
90%	5.038646
75% Q3	2.674605
50% Median	0.169857
25% Q1	-2.634550
10%	-5.170752
5%	-6.806823
1%	-9.716010
0% Min	-15.845264

Extreme Observations

-----Lowest-----		-----Highest-----	
Value	Obs	Value	Obs
-15.8453	877	11.7181	1324
-14.6836	1829	11.7362	1142
-14.5472	2479	12.2506	656
-14.3965	560	12.4439	98
-13.6107	1011	14.3449	559

Histogram

```

15+*
.*
.*
9+****
.*****
.*****

```

Boxplot

```

1 0
2 0
13 0
49 |
114 |
247 |

```

15+
|
|
9+
|
|

Normal Probability Plot

```

*
*
*
*****
*****

```

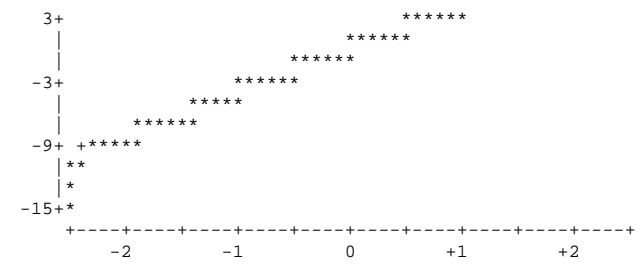


```

3+*****
.*****
.*****
-3+*****
.*****
.*****
-9+*****
.**
.*
-15+*
-----+
* may represent up to 13 counts

```

448	+-----+
581	*--*--*
507	---
398	+-----+
246	
133	
50	
17	0
5	0
4	0



Analyse de la distribution des résidus pour polyT (Poly transformé selon $\sqrt{(n+0.5)}$)

Moments

N	2815	Sum Weights	2815
Mean	0	Sum Observations	0
Std Deviation	0.1577028	Variance	0.02487017
Skewness	0.19470396	Kurtosis	0.64058839
Uncorrected SS	69.98467	Corrected SS	69.98467
Coeff Variation	.	Std Error Mean	0.00297235

Basic Statistical Measures

Location		Variability	
Mean	0.00000	Std Deviation	0.15770
Median	-0.01855	Variance	0.02487
Mode	-0.06267	Range	1.26817
		Interquartile Range	0.20335

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----
Student's t	t 0	Pr > t 1.0000
Sign	M -106.5	Pr >= M <.0001
Signed Rank	S -60043.5	Pr >= S 0.1639

Tests for Normality

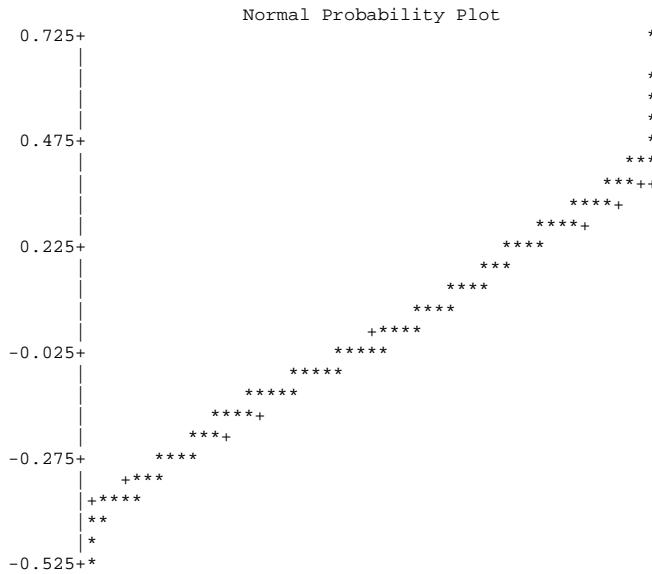
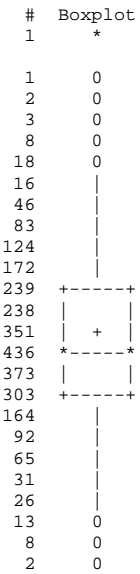
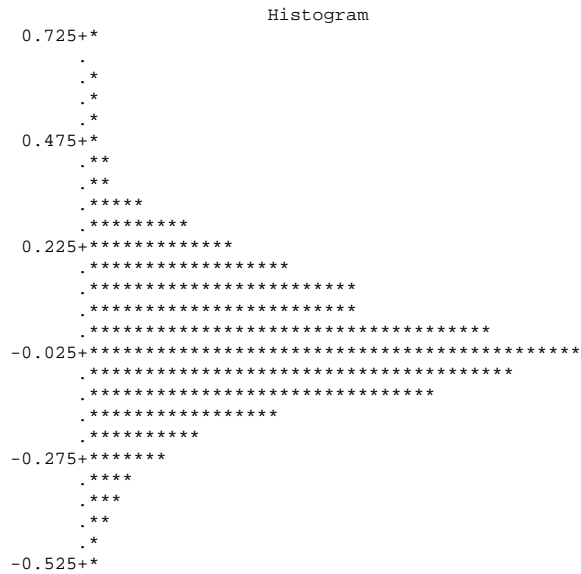
Test	--Statistic---	-----p Value-----
Kolmogorov-Smirnov	D 0.058282	Pr > D <0.0100
Cramer-von Mises	W-Sq 1.771248	Pr > W-Sq <0.0050
Anderson-Darling	A-Sq 8.874839	Pr > A-Sq <0.0050

Quantiles (Definition 5)

Quantile	Estimate
100% Max	0.7376759
99%	0.4030751
95%	0.2675885
90%	0.2098696
75% Q3	0.1028646
50% Median	-0.0185533
25% Q1	-0.1004904
10%	-0.1830699
5%	-0.2548865
1%	-0.3816550
0% Min	-0.5304955

Extreme Observations

-----Lowest-----		-----Highest-----	
Value	Obs	Value	Obs
-0.530495	1295	0.542957	2105
-0.519756	2268	0.563583	53
-0.492306	2776	0.571434	684
-0.491919	50	0.616480	391
-0.486347	393	0.737676	1636



-----+-----+-----+-----+-----+-----+-----
* may represent up to 10 counts

+-----+-----+-----+-----+-----+-----+-----+-----+
-2 -1 0 +1 +2

Analyse de la distribution des résidus pour TB

Moments

N	2815	Sum Weights	2815
Mean	0	Sum Observations	0
Std Deviation	112.385887	Variance	12630.5876
Skewness	0.29833565	Kurtosis	0.53187836
Uncorrected SS	35542473.5	Corrected SS	35542473.5
Coeff Variation	.	Std Error Mean	2.11822738

Basic Statistical Measures

Location		Variability	
Mean	0.000	Std Deviation	112.38589
Median	-5.008	Variance	12631
Mode	-130.335	Range	789.49128
		Interquartile Range	143.44636

NOTE: The mode displayed is the smallest of 9 modes with a count of 2.

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----
Student's t	t 0	Pr > t 1.0000
Sign	M -51.5	Pr >= M 0.0545
Signed Rank	S -58235.5	Pr >= S 0.1770

Tests for Normality

Test	--Statistic--	-----p Value-----
Kolmogorov-Smirnov	D 0.025797	Pr > D <0.0100
Cramer-von Mises	W-Sq 0.574668	Pr > W-Sq <0.0050
Anderson-Darling	A-Sq 4.040758	Pr > A-Sq <0.0050

Quantiles (Definition 5)

Quantile	Estimate
100% Max	429.08285
99%	292.63118
95%	196.17728
90%	138.52153

```

75% Q3      67.72921
50% Median  -5.00843
25% Q1     -75.71715
10%        -136.35468
5%         -173.48326
1%         -255.03635
0% Min     -360.40843

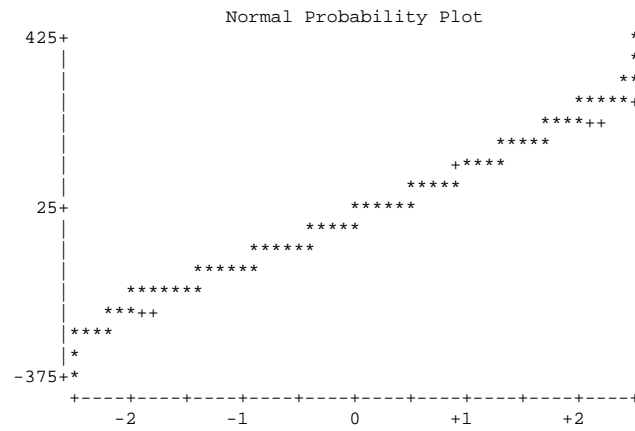
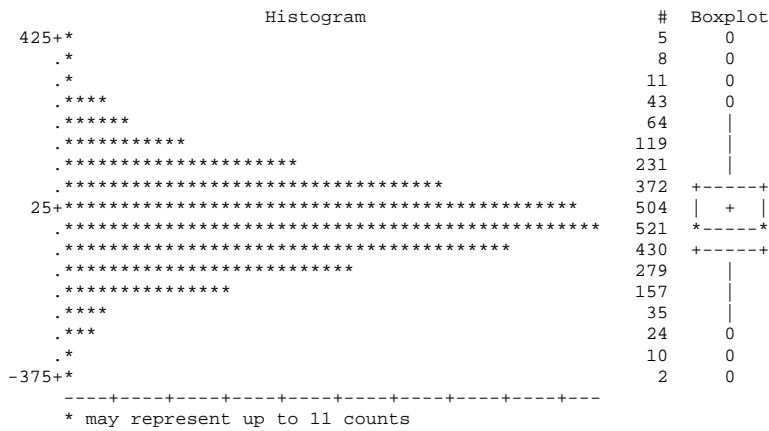
```

Extreme Observations

```

-----Lowest-----          -----Highest-----
Value      Obs          Value      Obs
-360.408   2789         400.013   100
-355.578   12           401.240  1324
-342.066   1319         407.736  1714
-338.087   1335         413.283   98
-337.693   560          429.083   91

```



Analyse de la distribution des résidus pour NB

Moments

N	2815	Sum Weights	2815
Mean	0	Sum Observations	0
Std Deviation	60.9098248	Variance	3710.00675
Skewness	0.22219291	Kurtosis	0.5589504
Uncorrected SS	10439959	Corrected SS	10439959
Coeff Variation	.	Std Error Mean	1.14801655

Basic Statistical Measures

Location		Variability	
Mean	0.0000	Std Deviation	60.90982
Median	-2.2739	Variance	3710
Mode	-75.0609	Range	489.24728
		Interquartile Range	78.30661

NOTE: The mode displayed is the smallest of 8 modes with a count of 2.

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----	
Student's t	t	0	Pr > t 1.0000
Sign	M	-51	Pr >= M 0.0569
Signed Rank	S	-39233	Pr >= S 0.3628

Tests for Normality

Test	--Statistic---	-----p Value-----	
Kolmogorov-Smirnov	D	0.022779	Pr > D <0.0100
Cramer-von Mises	W-Sq	0.346604	Pr > W-Sq <0.0050
Anderson-Darling	A-Sq	2.481739	Pr > A-Sq <0.0050

Quantiles (Definition 5)

Quantile	Estimate
100% Max	284.24558
99%	154.60529
95%	106.64558
90%	74.33970
75% Q3	37.36620
50% Median	-2.27388
25% Q1	-40.94041
10%	-76.79974
5%	-93.96502
1%	-141.72831
0% Min	-205.00170

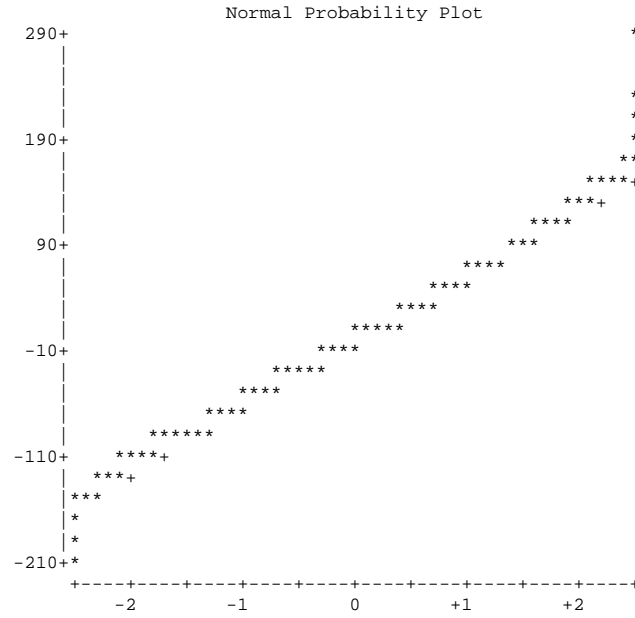
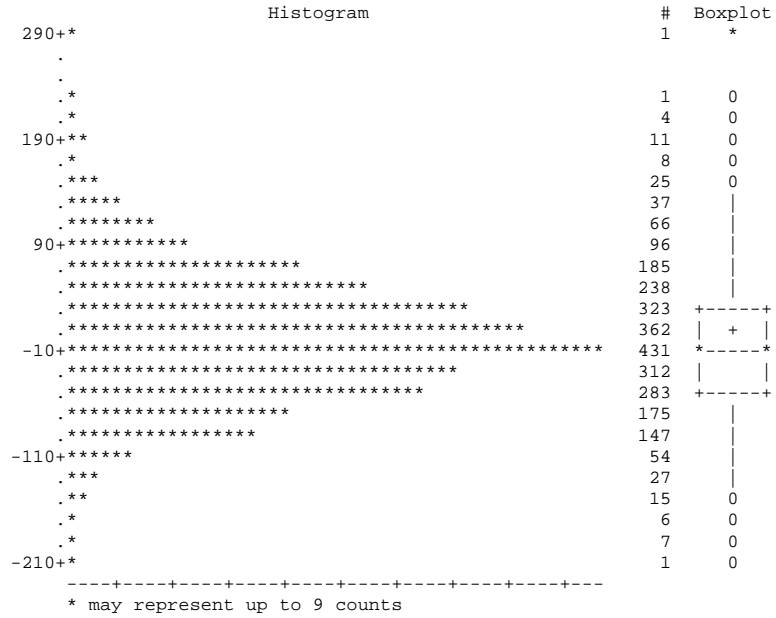
Extreme Observations

-----Lowest-----		-----Highest-----	
Value	Obs	Value	Obs
-205.002	1011	210.446	1324
-197.857	12	215.788	1716

```

-195.549    560    216.245    1447
-194.645    2789   220.055    100
-191.575    1235   284.246     91

```



Analyse de la distribution des résidus pour $\delta^{13}C$

Moments

Statistic	Value	Statistic	Value
N	2815	Sum Weights	2815
Mean	0	Sum Observations	0
Std Deviation	0.54607165	Variance	0.29819425
Skewness	0.03035442	Kurtosis	0.01527628
Uncorrected SS	839.118609	Corrected SS	839.118609
Coeff Variation	.	Std Error Mean	0.01029225

Basic Statistical Measures

Location		Variability	
Mean	0.00000	Std Deviation	0.54607
Median	-0.00834	Variance	0.29819

Mode -0.21943 Range 3.88445
 Interquartile Range 0.74033

NOTE: The mode displayed is the smallest of 2 modes with a count of 2.

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----
Student's t	t 0	Pr > t 1.0000
Sign	M -19	Pr >= M 0.4855
Signed Rank	S -4749	Pr >= S 0.9123

Tests for Normality

Test	--Statistic--	-----p Value-----
Kolmogorov-Smirnov	D 0.010446	Pr > D >0.1500
Cramer-von Mises	W-Sq 0.026629	Pr > W-Sq >0.2500
Anderson-Darling	A-Sq 0.184674	Pr > A-Sq >0.2500

Quantiles (Definition 5)

Quantile	Estimate
100% Max	2.07412263
99%	1.29396380
95%	0.89941911
90%	0.69044691
75% Q3	0.36805690
50% Median	-0.00833651
25% Q1	-0.37227017
10%	-0.71449697
5%	-0.89107386
1%	-1.22750934
0% Min	-1.81033226

Extreme Observations

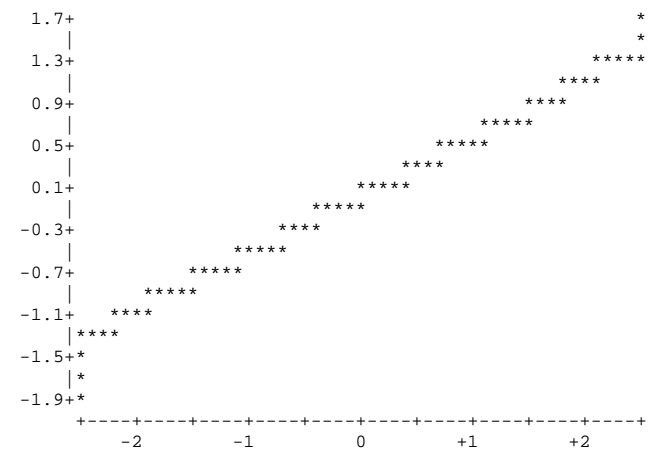
-----Lowest-----		-----Highest-----	
Value	Obs	Value	Obs
-1.81033	1892	1.62457	80
-1.72799	527	1.62535	1301
-1.69161	171	1.63844	862
-1.65372	993	1.75367	934
-1.64319	2808	2.07412	395

2.1+* Histogram # Boxplot Normal Probability Plot *



* may represent up to 10 counts

5	0
9	0
33	
57	
101	
165	
270	
367	+-----+
381	+
438	*-----*
326	+-----+
285	
176	
119	
47	
20	
8	0
6	0
1	0



**Annexe 4. Moyennes familiales par
population des caractères *HT02*, *HT03*, *DC*,
PolyT, *TB*, *NB* et $\delta^{13}C$**

Pop	Fam	N	HT02		HT03		DC		PolyT		TB		NB		$\delta^{13}C$	
			Moy	Ecart-type	Moy	Ecart-type	Moy	Ecart-type	Moy	Ecart-type	Moy	Ecart-type	Moy	Ecart-type	Moy	Ecart-type
15	1	7	44.286	6.264	104.429	12.040	27.732	2.460	1.747	0.155	481.529	87.158	291.086	51.458	-26.648	0.548
15	4	7	44.000	9.916	103.286	16.700	26.373	4.489	1.705	0.155	439.700	179.122	262.243	105.945	-26.619	0.576
15	5	7	50.143	16.787	117.143	20.530	32.247	4.438	1.747	0.155	526.343	194.414	271.429	82.490	-26.137	0.494
15	6	7	52.429	9.964	125.000	14.514	30.499	3.471	1.782	0.208	522.829	103.446	275.443	51.175	-26.153	0.880
15	7	6	54.667	15.384	126.833	27.802	31.921	4.254	1.774	0.150	582.650	184.314	290.267	68.225	-26.440	0.623
15	8	7	47.429	14.152	113.571	23.936	27.104	5.453	1.824	0.189	436.529	152.156	249.214	81.583	-26.442	0.505
15	9	1	95.000	.	184.000	.	44.965	.	1.871	.	1122.600	.	543.500	.	-25.692	.
15	10	7	36.143	7.058	98.143	15.027	24.178	6.411	1.782	0.208	435.743	191.340	248.043	95.329	-26.651	0.487
15	12	7	32.143	7.883	97.571	26.248	23.841	6.917	1.764	0.349	373.529	196.724	212.186	107.308	-26.702	1.169
15	13	7	49.714	14.648	115.571	16.762	30.665	3.605	1.788	0.141	597.214	171.610	338.871	89.578	-26.584	0.493
15	15	7	42.571	4.077	105.286	10.275	26.227	2.414	1.664	0.141	361.929	56.286	200.386	37.278	-26.602	0.459
15	18	7	50.000	16.422	112.429	22.984	29.476	4.211	1.747	0.155	524.700	159.489	289.071	66.831	-26.533	0.444
15	20	7	50.429	9.727	117.429	15.241	30.697	2.932	1.788	0.141	567.700	119.319	309.586	67.562	-26.291	0.921
15	21	7	40.143	10.074	108.143	20.820	32.604	7.326	1.907	0.095	573.343	239.676	335.300	143.928	-26.693	0.456
15	22	7	42.143	11.276	102.143	20.659	28.476	7.225	1.788	0.141	472.643	222.576	258.014	103.923	-26.539	0.353
15	24	6	40.333	14.679	107.167	29.499	27.744	6.436	1.774	0.150	450.683	243.187	256.317	132.856	-26.349	0.968
20	1	5	30.200	5.404	75.000	3.162	23.600	3.896	1.639	0.130	342.840	121.384	211.180	89.141	-26.707	0.581
20	2	7	35.286	5.589	89.000	9.000	23.571	2.992	1.613	0.219	315.914	93.505	171.086	46.934	-26.717	0.522
20	3	6	32.667	10.783	79.833	16.582	27.500	3.493	1.581	0.000	425.267	152.015	244.417	81.098	-27.048	0.442
20	4	5	32.800	10.780	80.400	16.365	28.200	3.511	1.697	0.159	360.460	77.258	196.340	44.154	-26.825	0.553
20	5	5	32.200	10.986	80.600	18.609	25.300	5.203	1.697	0.159	313.360	167.816	182.000	88.098	-26.602	0.996
20	6	4	31.000	4.082	77.000	24.590	25.250	6.702	1.654	0.145	277.600	183.351	157.725	97.568	-25.688	0.681
20	7	7	44.714	4.923	100.000	12.166	29.214	4.915	1.664	0.141	435.314	171.058	243.586	87.763	-25.479	0.617
20	8	7	38.857	5.786	86.286	14.115	26.571	4.439	1.572	0.187	345.900	104.600	191.900	57.203	-25.332	0.564
20	9	5	34.800	7.085	85.600	10.526	26.400	3.525	1.581	0.000	419.740	94.140	242.400	49.221	-25.503	0.384
20	10	7	25.143	6.149	66.571	14.570	22.143	5.566	1.623	0.109	260.043	165.348	149.100	68.690	-26.226	0.160
20	11	7	33.857	9.616	73.286	12.203	25.643	5.305	1.572	0.187	339.214	133.449	191.757	62.713	-27.168	0.579
20	12	7	27.571	6.803	70.571	15.841	21.786	3.661	1.479	0.174	270.929	76.858	169.029	43.751	-26.761	0.928
20	13	5	24.000	8.544	58.800	15.738	20.600	5.878	1.439	0.195	240.360	118.848	146.800	66.668	-27.425	0.688
20	14	6	40.000	10.198	80.333	13.880	26.500	2.665	1.581	0.000	366.300	63.368	207.867	32.998	-27.331	0.735
20	15	3	29.333	7.371	72.667	18.230	23.500	2.500	1.559	0.324	342.033	40.812	207.100	44.169	-27.458	1.186
20	16	5	33.400	11.610	85.400	25.056	24.700	5.880	1.755	0.159	332.020	170.334	185.500	86.237	-27.506	0.750
20	17	6	34.333	15.680	78.667	19.765	26.083	6.336	1.522	0.145	387.583	240.777	206.717	96.749	-27.746	0.530
20	18	6	27.500	8.264	76.333	17.455	26.417	4.984	1.570	0.205	366.800	124.719	219.400	81.288	-27.333	0.461
20	19	6	34.833	18.060	84.833	23.507	25.083	5.669	1.629	0.118	382.117	115.218	229.317	43.903	-27.542	0.871
20	20	7	30.000	9.557	69.714	16.670	24.714	7.751	1.581	0.000	317.229	158.463	174.600	77.482	-26.439	0.896
21	1	7	28.571	5.968	71.429	13.903	24.118	5.835	1.664	0.141	310.600	157.217	184.414	85.590	-26.811	0.709
21	2	7	30.857	8.572	74.714	21.305	24.567	6.636	1.530	0.135	268.686	130.675	144.100	60.693	-26.761	0.575
21	3	7	28.429	7.345	64.714	11.644	20.761	4.018	1.613	0.219	189.829	77.269	110.271	44.560	-26.777	0.482
21	4	7	38.000	9.967	92.571	18.192	27.456	4.303	1.741	0.216	369.871	87.263	193.886	31.038	-26.691	0.573
21	6	2	47.000	5.657	104.500	14.849	28.250	4.596	1.726	0.205	477.450	154.361	258.650	84.216	-27.087	0.231
21	7	7	30.857	9.191	79.857	7.151	24.516	3.882	1.705	0.155	260.957	72.337	151.129	42.698	-26.926	0.579
21	8	7	36.571	11.900	80.143	18.614	23.734	5.504	1.530	0.135	280.100	146.452	149.843	69.419	-27.145	0.550
21	9	7	35.429	12.394	82.000	23.930	26.931	5.563	1.603	0.289	354.371	168.864	194.900	83.253	-26.729	0.698
21	10	6	30.667	6.154	72.667	9.158	25.667	2.543	1.629	0.118	297.933	71.259	181.383	43.877	-26.948	0.634
21	11	7	30.286	5.619	76.000	8.042	24.391	2.149	1.572	0.187	284.157	56.257	164.371	34.231	-26.967	0.384
21	12	7	35.714	11.056	76.571	16.461	27.326	3.900	1.623	0.109	399.800	110.720	232.514	59.165	-27.315	0.789
21	13	7	26.286	4.855	74.571	12.739	24.535	2.619	1.623	0.109	291.443	67.845	173.871	39.085	-26.916	0.500
21	14	6	25.000	7.874	62.833	11.856	21.428	4.260	1.629	0.118	244.533	115.380	144.017	60.198	-27.046	0.290
21	15	7	39.000	17.078	93.714	34.750	27.769	5.841	1.613	0.219	398.229	159.803	198.571	74.365	-26.610	0.668

21	17	6	35.500	9.628	79.667	12.801	25.932	2.212	1.581	0.000	343.650	72.570	202.433	33.650	-26.731	0.612
21	18	6	35.667	12.275	75.833	22.085	28.083	4.934	1.570	0.205	445.950	167.024	257.200	86.643	-26.425	1.056
21	19	7	25.571	6.373	69.000	11.662	22.320	4.085	1.572	0.187	258.786	122.053	152.414	68.572	-27.189	0.520
21	20	7	36.000	9.000	86.000	18.157	26.989	4.762	1.623	0.109	386.457	149.592	218.257	71.886	-26.490	0.615
21	21	6	41.333	11.535	90.333	16.488	29.083	4.005	1.522	0.145	458.767	123.311	254.250	61.864	-27.067	0.637
21	22	6	39.500	7.120	83.500	11.327	29.917	2.396	1.522	0.145	453.967	63.123	261.450	40.129	-26.774	0.650
21	23	7	34.000	11.576	85.429	19.697	25.429	4.467	1.700	0.215	370.143	191.686	200.743	85.386	-27.149	0.680
21	24	5	36.000	13.266	84.600	13.315	29.300	3.546	1.747	0.244	416.760	130.835	234.120	47.586	-26.864	0.513
21	25	6	31.333	8.262	78.000	12.377	24.583	3.917	1.570	0.205	306.850	58.196	176.683	33.777	-26.813	0.509
21	26	6	36.667	10.764	86.167	13.467	28.417	6.240	1.629	0.118	342.683	90.060	178.767	43.351	-27.168	0.491
21	27	6	33.333	7.474	79.167	18.989	23.917	6.037	1.774	0.150	342.150	96.842	207.817	40.130	-26.858	0.769
21	28	7	35.000	8.869	91.429	18.591	27.429	3.220	1.741	0.216	390.086	107.658	211.571	50.263	-26.064	0.922
21	29	7	31.143	9.424	79.286	17.923	23.429	4.911	1.681	0.348	262.486	110.304	145.314	62.602	-27.567	0.539
21	30	5	33.200	6.834	83.800	10.354	28.200	4.945	1.626	0.267	432.280	126.925	276.700	108.562	-27.041	1.403
22	1	7	24.714	6.775	52.000	14.583	18.396	3.062	1.572	0.187	155.457	72.110	97.857	45.059	-27.124	0.752
22	2	7	30.000	9.469	73.857	10.123	22.471	2.435	1.613	0.219	234.471	66.266	134.986	37.338	-26.318	0.681
22	3	7	31.286	11.101	71.857	20.684	24.325	4.281	1.613	0.219	294.971	106.296	173.571	52.778	-26.785	0.574
22	4	7	31.000	7.572	75.571	15.054	24.609	3.382	1.581	0.000	284.786	103.640	164.386	53.612	-27.132	0.267
22	5	7	30.143	3.848	71.714	12.632	23.291	2.180	1.581	0.000	306.871	95.384	189.914	58.618	-26.868	0.546
22	8	4	19.250	6.850	51.000	12.623	16.875	4.211	1.564	0.264	226.875	121.869	102.725	49.991	-26.810	0.514
22	9	7	29.429	9.090	75.286	11.543	24.824	4.167	1.664	0.141	344.014	112.879	209.343	65.344	-27.011	0.453
22	10	7	26.857	8.415	62.857	16.252	21.989	5.633	1.530	0.135	284.800	137.926	167.243	87.213	-26.909	0.727
22	11	7	29.143	9.957	72.000	15.481	24.761	4.552	1.664	0.141	289.971	109.768	172.143	55.566	-26.987	0.423
22	12	7	26.714	8.200	61.143	13.692	19.299	4.233	1.428	0.191	198.657	101.469	120.014	55.044	-27.212	0.552
22	13	7	30.286	4.855	78.857	10.511	24.384	2.844	1.664	0.141	286.357	68.657	170.500	41.459	-27.099	0.590
22	14	7	26.143	6.939	69.000	13.241	23.634	4.006	1.705	0.155	259.871	81.836	157.771	48.152	-27.015	0.535
22	15	6	23.667	6.377	61.333	6.022	21.148	3.406	1.511	0.248	216.550	69.036	132.133	34.911	-26.412	0.709
22	16	4	20.250	4.031	59.750	18.355	18.375	3.146	1.654	0.145	176.650	62.637	110.375	35.188	-27.305	1.071
22	17	7	30.286	16.039	66.714	18.081	23.273	6.613	1.623	0.109	308.086	169.151	183.257	77.999	-26.681	0.354
22	18	6	19.667	4.761	58.667	10.132	18.555	4.367	1.570	0.205	157.100	77.719	98.533	44.055	-26.975	0.825
22	19	6	29.500	8.361	67.333	13.261	20.341	5.117	1.726	0.159	226.450	113.376	139.050	64.106	-26.988	0.637
22	20	4	28.250	12.010	68.750	24.473	20.625	2.428	1.654	0.145	233.400	64.763	142.925	27.401	-27.293	1.180
22	21	3	31.333	13.429	71.000	22.113	26.667	6.752	1.462	0.206	447.567	211.888	279.700	124.758	-27.429	0.746
22	22	4	29.250	2.062	79.000	18.019	25.625	3.924	1.654	0.145	402.125	126.583	270.525	93.591	-27.442	0.740
22	23	7	24.429	10.659	60.714	15.185	22.143	7.244	1.581	0.000	273.086	126.669	172.971	73.312	-27.053	0.850
22	24	7	22.714	5.589	61.714	11.427	20.786	2.984	1.581	0.000	253.286	81.737	160.629	45.239	-25.929	1.071
22	26	5	24.400	6.950	62.200	8.983	20.200	1.151	1.697	0.159	253.500	46.223	163.720	34.148	-25.412	0.559
22	27	7	24.000	4.203	63.571	12.634	23.000	3.536	1.572	0.187	327.200	107.184	211.414	66.555	-26.360	0.507
22	28	7	29.000	12.517	75.857	22.244	21.143	3.010	1.664	0.141	281.300	84.265	183.129	53.292	-25.900	0.551
22	29	4	21.750	7.455	68.250	14.431	22.125	3.816	1.789	0.260	268.675	108.591	169.725	59.344	-25.434	0.684
22	30	4	21.500	10.214	64.250	28.441	21.125	10.820	1.492	0.178	316.000	261.754	186.100	150.993	-26.584	0.865
23	1	7	27.286	10.323	68.000	15.308	20.945	2.090	1.479	0.174	218.029	66.842	132.529	37.445	-27.017	0.371
23	2	7	27.000	4.726	66.000	14.224	22.125	2.499	1.572	0.187	231.729	45.200	142.814	26.201	-26.447	0.757
23	3	7	22.000	10.661	55.000	13.796	18.686	4.707	1.572	0.187	186.500	103.203	119.386	61.547	-26.717	0.329
23	4	7	23.429	6.803	56.571	12.568	19.578	2.603	1.530	0.135	183.300	59.464	116.071	36.545	-26.835	0.826
23	5	7	22.286	8.655	58.714	16.560	18.690	2.597	1.572	0.187	208.629	53.253	129.029	27.514	-27.102	0.428
23	6	7	26.000	11.195	67.714	17.414	19.436	3.087	1.664	0.141	224.757	60.700	139.329	30.476	-26.925	0.585
23	7	7	26.714	11.884	67.857	15.192	21.903	5.987	1.572	0.187	243.386	114.877	143.886	54.931	-26.815	0.411
23	8	7	24.143	5.273	61.571	16.831	21.431	3.068	1.664	0.141	234.929	69.365	146.214	41.725	-26.708	0.387
23	9	7	28.000	5.802	70.714	8.674	25.513	3.716	1.623	0.109	366.743	96.781	222.271	54.484	-27.165	0.495
23	10	7	34.714	11.954	84.143	21.264	25.456	4.563	1.741	0.216	370.429	157.276	220.586	85.425	-27.016	0.676
23	11	7	25.143	9.335	64.286	18.962	21.271	6.012	1.581	0.000	217.757	151.146	126.786	77.335	-27.090	0.520
23	12	7	24.714	5.794	70.143	18.334	22.661	5.097	1.581	0.000	286.057	103.287	179.557	62.284	-26.896	0.710
23	13	7	27.857	8.133	62.857	15.994	19.561	3.315	1.623	0.109	235.971	114.582	147.157	60.717	-27.136	0.594

23	14	7	25.429	8.182	61.429	9.981	22.705	6.871	1.664	0.141	282.357	171.073	170.286	82.604	-27.042	0.513
23	15	7	25.714	6.576	61.571	10.014	20.961	4.650	1.572	0.187	236.629	101.712	144.043	55.992	-26.685	0.678
23	16	7	26.000	8.124	62.000	17.474	19.143	5.475	1.696	0.249	230.457	123.009	145.814	72.158	-26.683	0.496
23	17	6	35.667	7.230	77.833	19.167	28.000	4.970	1.629	0.118	425.650	146.314	255.383	79.341	-26.614	0.733
23	18	7	24.857	7.647	60.286	12.311	20.857	4.811	1.556	0.300	249.357	92.297	160.986	54.562	-27.098	0.752
23	19	7	29.429	3.910	73.571	12.765	26.000	4.752	1.530	0.135	359.471	139.301	224.614	88.070	-27.092	0.673
23	20	7	33.143	8.335	80.857	9.494	24.500	3.122	1.664	0.141	326.414	99.064	191.200	53.695	-26.744	0.554
23	21	7	33.286	9.793	74.857	18.542	24.357	5.328	1.562	0.264	340.057	120.237	204.929	63.896	-27.331	0.490
23	22	7	32.143	9.771	70.143	19.386	24.143	4.661	1.623	0.109	382.943	112.339	230.486	55.040	-26.677	0.678
23	23	6	30.500	10.691	86.833	24.629	22.917	3.878	1.774	0.150	323.717	105.248	214.050	84.488	-26.050	0.766
23	24	7	40.571	8.791	90.286	7.761	26.000	2.614	1.613	0.219	391.114	92.166	221.886	50.536	-26.329	0.657
23	25	7	27.429	5.996	68.143	14.588	23.500	3.028	1.530	0.135	334.186	85.123	204.214	45.622	-26.784	0.637
23	27	6	35.000	5.550	77.167	13.014	29.500	1.949	1.618	0.239	476.567	76.137	272.617	34.549	-27.362	0.650
23	28	3	20.000	7.000	51.000	13.077	24.000	1.500	1.462	0.206	230.167	91.526	152.533	65.472	-27.991	0.386
23	29	7	33.286	5.589	85.714	22.655	25.071	6.458	1.741	0.216	323.429	137.437	191.329	82.075	-26.585	0.970
23	30	6	28.833	6.113	75.833	14.442	22.250	3.460	1.678	0.150	299.233	55.922	181.217	29.384	-27.437	0.940
24	1	4	29.000	8.206	76.750	7.136	24.625	0.750	1.654	0.145	288.000	37.597	178.125	31.273	-26.938	0.609
24	2	7	44.286	13.853	91.857	20.203	30.216	5.830	1.613	0.219	433.757	164.123	226.986	75.915	-26.902	0.476
24	3	7	43.143	15.529	90.857	18.289	27.869	6.516	1.705	0.155	377.829	181.330	201.371	79.553	-26.562	0.663
24	4	7	33.714	11.412	85.857	12.877	26.217	4.324	1.649	0.280	329.586	139.385	189.086	76.819	-27.164	0.657
24	5	7	36.429	6.729	73.857	12.864	21.627	3.721	1.658	0.204	267.257	73.402	160.586	35.683	-26.854	0.559
24	6	7	30.429	10.147	71.429	17.615	23.104	5.315	1.788	0.141	271.200	134.984	164.129	75.335	-26.786	0.523
24	7	7	41.000	8.641	92.857	13.120	26.535	4.443	1.664	0.141	357.271	136.476	192.786	70.860	-26.829	0.489
24	8	7	36.286	8.976	81.857	16.777	24.321	3.824	1.664	0.141	333.414	114.304	186.114	55.691	-26.801	1.087
24	9	7	41.571	13.891	86.571	11.802	27.697	3.967	1.623	0.109	371.457	154.440	210.671	78.863	-26.734	0.601
24	10	7	33.857	8.355	79.857	14.416	24.259	5.195	1.782	0.208	305.700	132.040	170.800	68.242	-26.847	0.370
24	11	7	30.429	9.880	72.286	7.158	26.049	4.454	1.479	0.174	335.457	109.121	186.329	56.984	-27.543	0.738
24	12	7	29.000	9.129	74.429	20.719	23.604	4.333	1.664	0.141	261.314	94.036	151.400	47.867	-26.781	0.685
24	13	7	24.571	6.680	64.571	10.533	20.861	4.484	1.623	0.109	167.700	83.690	89.429	47.665	-27.091	0.520
24	14	7	31.286	6.775	73.143	17.131	23.702	3.564	1.572	0.187	271.843	99.596	156.400	53.308	-27.059	0.655
24	15	7	30.429	6.399	72.714	15.703	25.895	4.427	1.581	0.000	345.986	95.124	198.014	44.107	-26.967	0.891
24	16	7	37.143	10.869	79.714	19.661	25.418	5.938	1.623	0.109	357.186	137.432	193.943	68.865	-26.396	0.934
24	17	7	35.143	8.952	78.286	15.564	27.214	4.733	1.479	0.174	423.329	134.849	228.843	59.757	-27.187	0.599
24	18	7	35.857	17.199	81.857	22.049	27.071	8.253	1.623	0.109	382.657	186.095	223.186	101.514	-26.647	0.266
24	19	6	30.667	9.480	75.333	23.670	22.000	7.259	1.629	0.118	265.450	127.833	159.483	72.239	-26.823	0.522
24	20	6	42.000	18.942	90.500	24.139	27.500	4.980	1.671	0.221	373.267	174.021	201.550	88.076	-26.775	0.395
25	1	7	21.571	8.423	54.714	12.406	18.753	4.227	1.530	0.135	210.071	71.915	143.243	43.867	-27.000	0.583
25	2	7	26.429	10.097	59.286	14.500	21.317	5.912	1.581	0.000	242.800	112.196	151.500	62.925	-26.933	0.402
25	3	7	22.286	5.024	59.714	11.011	20.015	2.158	1.664	0.141	211.800	57.554	141.300	37.745	-26.507	0.531
25	4	7	27.429	6.949	66.286	12.500	23.016	3.120	1.521	0.228	300.186	65.461	182.771	41.643	-26.938	0.654
25	5	6	23.667	3.615	62.500	16.355	21.363	2.780	1.451	0.270	276.917	85.657	174.750	54.449	-26.834	0.659
25	6	7	28.571	7.435	73.857	18.398	23.318	5.285	1.623	0.109	332.871	160.438	205.500	104.542	-26.793	0.621
25	7	7	26.429	12.555	61.286	18.328	21.203	4.337	1.521	0.228	268.571	135.262	165.343	69.597	-27.020	0.743
25	8	7	30.714	5.851	70.286	14.080	20.905	4.996	1.623	0.109	215.114	116.182	128.129	70.704	-26.702	0.292
25	9	7	29.857	10.107	64.714	17.056	24.907	4.486	1.572	0.187	298.357	74.787	175.943	35.603	-27.142	0.370
25	10	6	29.833	8.085	80.333	17.558	25.667	4.645	1.823	0.118	389.167	168.263	234.067	91.841	-26.337	0.404
25	11	7	24.857	8.572	59.714	15.478	20.846	4.416	1.572	0.187	240.143	91.534	154.629	52.515	-26.938	0.559
25	12	7	24.857	5.815	69.714	11.629	20.193	4.503	1.623	0.109	237.529	70.134	138.571	32.974	-26.583	0.891
25	13	7	30.714	9.196	71.143	17.639	23.686	3.724	1.581	0.000	313.457	102.915	192.857	61.034	-26.700	0.920
25	14	7	26.000	8.851	64.714	11.828	20.335	4.582	1.623	0.109	228.086	128.182	141.014	76.807	-26.944	0.667
25	15	7	27.000	10.909	71.571	20.703	21.889	5.576	1.664	0.141	229.614	126.541	130.614	61.660	-26.786	0.875
25	16	6	30.333	12.987	72.000	27.662	25.083	5.783	1.522	0.145	362.200	200.789	217.383	108.353	-26.587	0.251
25	17	4	32.000	18.092	80.750	23.157	25.375	4.460	1.564	0.264	348.925	124.753	209.825	81.640	-26.665	0.554
25	18	7	27.571	9.607	72.571	17.106	25.143	6.149	1.809	0.315	374.971	171.838	213.557	97.740	-25.745	0.349

25	19	7	24.571	6.630	62.286	16.958	25.143	7.040	1.572	0.187	408.743	180.799	260.614	113.864	-26.270	0.415
25	20	7	27.571	7.231	70.286	6.993	24.643	1.909	1.658	0.204	365.471	54.821	225.000	32.752	-26.876	0.459
25	21	7	32.571	8.619	87.286	34.961	25.857	4.298	1.654	0.238	364.943	140.990	209.843	68.647	-26.477	0.607
25	22	6	28.333	7.815	71.000	14.873	24.083	4.364	1.522	0.145	325.050	124.359	201.083	77.684	-26.620	0.642
25	23	7	23.857	5.460	55.571	10.309	20.571	2.405	1.581	0.000	293.186	57.841	189.300	32.905	-27.537	0.636
25	24	7	28.714	6.102	66.714	11.800	23.500	1.607	1.572	0.187	296.043	88.286	181.157	58.435	-27.658	0.337
25	25	5	27.200	7.887	70.400	14.876	24.300	2.414	1.510	0.159	308.100	57.140	197.140	37.667	-27.375	0.414
25	26	6	26.833	7.910	71.500	15.070	25.250	5.383	1.581	0.000	430.400	192.523	273.917	98.948	-26.527	0.365
25	27	6	34.000	13.491	75.167	17.314	29.083	4.421	1.462	0.184	501.883	148.303	287.000	63.134	-26.974	0.651
25	28	7	27.000	5.292	66.286	7.181	22.000	3.894	1.530	0.135	359.800	141.333	190.171	91.894	-27.032	0.449
25	29	5	34.600	10.114	82.200	15.482	28.700	6.068	1.439	0.195	453.880	144.355	261.760	78.628	-27.168	0.813
25	30	7	29.286	9.178	73.429	16.582	25.643	4.488	1.623	0.109	436.629	126.713	262.657	71.034	-27.358	0.571
26	1	7	25.857	9.063	65.286	18.391	23.147	3.865	1.572	0.187	274.143	126.406	165.214	69.588	-26.789	0.594
26	2	7	22.000	5.385	62.857	17.639	19.259	3.140	1.572	0.187	198.714	70.948	127.914	40.840	-26.617	0.757
26	3	7	26.571	5.968	68.857	4.298	23.295	4.238	1.664	0.141	272.214	92.118	165.700	52.983	-26.959	0.902
26	4	7	32.000	11.165	72.714	19.483	24.406	6.022	1.705	0.155	336.957	169.629	195.429	89.273	-26.686	0.691
26	5	7	28.286	7.111	72.143	9.353	24.314	2.575	1.664	0.141	333.771	55.488	198.500	32.070	-26.785	0.739
26	6	7	28.000	4.726	72.714	12.175	22.596	3.187	1.747	0.155	249.900	83.816	146.371	41.623	-26.463	0.687
26	7	7	23.714	6.473	71.429	15.361	19.833	3.264	1.664	0.141	206.886	68.631	128.429	42.022	-26.097	0.776
26	8	7	22.286	7.064	61.286	14.952	20.575	6.990	1.623	0.109	233.571	144.847	150.629	83.493	-26.273	0.524
26	9	7	19.714	4.348	45.429	13.526	15.842	3.103	1.521	0.228	150.114	46.145	101.214	27.086	-27.103	0.585
26	10	5	24.800	8.438	63.600	8.678	20.500	2.449	1.639	0.130	285.040	51.732	191.880	39.898	-26.365	0.792
26	11	6	27.833	6.969	70.000	10.770	22.854	2.878	1.629	0.118	276.467	107.351	162.567	62.393	-26.624	0.511
26	12	6	23.667	9.158	61.833	23.319	21.195	2.871	1.618	0.239	231.683	109.008	145.600	57.260	-27.203	0.438
26	13	7	28.714	8.056	70.429	15.501	23.991	5.775	1.623	0.109	311.257	135.326	185.514	72.727	-26.808	0.312
26	14	7	27.143	10.024	67.571	21.007	19.846	4.109	1.623	0.109	226.486	125.915	141.700	71.175	-26.380	0.379
26	15	7	23.429	4.504	61.000	11.972	22.317	3.217	1.598	0.323	262.529	92.754	159.471	50.297	-26.906	0.698
26	16	7	26.286	6.726	62.857	17.199	22.483	3.802	1.623	0.109	302.243	96.130	189.614	47.823	-26.707	0.649
26	17	7	25.286	6.873	61.571	7.277	23.643	3.997	1.530	0.135	356.771	135.030	231.471	81.884	-27.161	1.098
26	18	5	23.000	7.314	54.600	16.196	20.900	7.877	1.626	0.267	329.160	215.316	206.500	123.344	-27.016	0.573
26	20	6	23.333	5.354	62.000	9.033	19.667	3.109	1.667	0.259	224.583	77.200	147.533	57.656	-26.578	0.514
26	21	6	27.333	10.053	73.500	17.456	21.917	3.121	1.726	0.159	328.183	86.047	207.567	39.954	-26.405	0.391
26	22	6	22.667	4.082	59.167	12.828	20.083	4.005	1.678	0.150	273.367	96.732	178.350	65.795	-26.442	0.486
26	23	7	29.000	7.810	67.429	14.673	24.500	5.276	1.623	0.109	347.800	136.025	219.529	80.322	-26.528	0.861
26	24	7	30.143	11.231	70.714	19.687	24.143	6.189	1.664	0.141	412.014	210.893	249.057	122.803	-26.680	0.301
26	25	5	33.800	6.380	72.600	8.325	24.400	4.980	1.568	0.229	379.540	144.378	232.720	85.292	-26.734	0.710
26	26	8	24.625	7.249	62.875	13.527	23.063	5.747	1.654	0.134	328.563	168.732	205.800	99.764	-26.680	0.851
26	27	6	28.667	8.116	72.500	18.053	24.667	6.377	1.629	0.118	393.033	205.480	233.467	117.873	-26.154	0.345
26	28	7	23.571	6.024	61.857	14.253	21.857	4.110	1.649	0.280	278.229	119.441	174.600	65.529	-27.332	0.602
26	29	7	30.714	11.557	72.571	16.742	23.786	4.338	1.664	0.141	354.300	130.052	222.629	79.253	-26.635	0.541
26	30	6	22.667	13.095	50.500	19.624	20.000	6.042	1.581	0.000	253.100	172.171	164.017	98.774	-26.830	0.470
27	1	6	36.500	9.203	90.833	9.152	23.583	2.108	1.774	0.150	349.533	84.737	209.617	55.566	-26.098	0.299
27	2	7	44.571	8.162	100.000	16.289	28.429	5.381	1.747	0.155	468.071	168.495	262.700	89.709	-26.578	0.190
27	3	7	38.286	13.022	92.714	23.852	23.714	3.684	1.737	0.250	383.757	107.090	233.300	65.649	-26.587	0.340
27	4	6	44.000	19.759	101.333	29.831	27.750	5.174	1.913	0.102	438.483	208.498	244.067	102.174	-26.055	1.006
27	5	6	36.333	10.596	94.000	21.420	23.500	3.550	1.816	0.206	323.417	103.065	192.717	52.283	-25.926	0.736
27	7	7	35.000	14.754	85.429	19.722	25.071	3.181	1.705	0.155	393.214	147.501	232.900	65.649	-26.200	0.679
27	8	7	40.857	11.711	94.571	22.795	25.571	7.328	1.818	0.244	414.086	225.431	237.100	122.433	-25.610	0.568
27	10	5	58.400	12.422	113.400	11.546	30.000	2.850	1.639	0.130	585.380	85.995	316.840	47.364	-25.483	0.296
27	11	6	45.833	12.303	101.167	19.793	28.917	5.643	1.570	0.205	545.150	200.641	293.733	106.188	-25.876	0.768
27	12	6	33.000	7.694	85.167	12.576	22.250	3.205	1.559	0.289	321.717	123.783	195.783	70.742	-26.164	0.510
27	13	6	40.000	15.900	88.500	23.594	22.583	6.822	1.678	0.150	372.183	187.003	207.633	93.806	-26.683	0.262
27	14	7	35.571	9.071	85.000	16.176	24.500	4.735	1.705	0.155	364.029	147.110	217.600	81.177	-26.014	0.788
27	15	7	29.429	2.070	76.429	5.127	20.000	3.253	1.613	0.219	241.857	90.574	145.171	49.534	-26.655	0.541

27	16	6	43.167	9.239	108.333	16.342	26.750	4.084	1.612	0.287	448.183	157.833	257.750	93.952	-26.063	0.652
27	17	6	24.500	10.540	65.667	25.508	18.583	5.962	1.511	0.248	203.883	137.059	125.517	72.104	-26.470	0.421
27	18	4	34.000	10.520	88.000	33.397	25.500	6.258	1.564	0.264	359.850	128.097	213.200	56.767	-25.909	0.867
27	19	7	41.429	7.913	90.429	12.246	24.357	3.761	1.782	0.208	361.957	92.854	207.200	49.082	-26.367	0.665
27	20	6	35.833	9.390	90.167	16.916	24.333	4.708	1.660	0.305	438.550	191.944	248.200	93.112	-26.439	0.724
27	21	5	45.400	12.321	101.200	19.447	26.900	3.190	1.755	0.159	508.040	133.426	279.120	58.203	-25.667	0.819
27	26	5	43.200	3.194	108.800	21.993	26.300	3.818	1.639	0.130	440.860	130.488	222.120	54.884	-25.608	0.264
27	29	6	36.333	9.266	86.667	19.876	23.250	6.525	1.816	0.206	294.333	129.930	177.067	71.061	-25.712	0.383
29	1	7	30.143	12.321	74.714	20.830	24.769	7.818	1.581	0.000	353.186	201.782	199.300	99.461	-26.422	0.618
29	2	7	23.143	5.928	70.571	12.286	22.576	3.805	1.658	0.204	293.543	99.066	182.900	57.755	-26.765	0.567
29	3	7	26.571	4.541	67.143	8.513	22.900	2.862	1.572	0.187	283.857	69.177	177.800	41.219	-26.860	0.629
29	4	7	21.429	4.429	53.429	9.163	19.810	2.711	1.581	0.000	220.757	66.542	141.986	39.236	-26.935	0.542
29	6	7	27.857	5.429	69.143	11.725	22.404	2.032	1.664	0.141	283.714	74.993	177.300	42.511	-26.900	0.441
29	7	7	30.286	10.889	82.571	14.351	28.194	3.371	1.623	0.109	404.543	103.811	246.886	99.734	-27.106	0.435
29	8	7	28.000	9.950	74.571	14.998	24.888	3.012	1.705	0.155	382.529	103.872	230.614	64.278	-26.322	0.653
29	9	7	21.714	10.012	62.286	20.031	23.183	6.314	1.530	0.135	285.200	140.141	171.800	72.651	-26.880	0.597
29	10	7	25.857	6.619	70.571	8.810	21.328	2.439	1.623	0.109	284.271	82.695	174.943	53.925	-26.364	0.516
29	11	7	21.286	5.155	56.000	10.970	19.746	5.763	1.572	0.187	207.371	111.836	126.743	62.053	-27.006	0.484
29	12	7	22.286	4.572	62.857	16.108	23.135	5.339	1.581	0.000	261.000	130.612	158.286	72.379	-26.824	0.359
29	13	7	22.714	5.589	63.571	8.886	22.237	3.324	1.623	0.109	273.043	75.996	172.586	42.389	-26.970	0.856
29	14	6	22.333	8.779	51.833	21.766	19.148	5.110	1.462	0.184	247.183	179.789	161.833	115.052	-26.841	0.642
29	15	7	22.143	6.283	70.143	24.162	22.704	4.993	1.581	0.000	302.986	140.463	182.557	76.295	-26.263	0.725
29	16	7	20.857	1.676	63.429	9.676	22.829	4.258	1.664	0.141	292.757	129.675	176.371	69.690	-26.302	0.845
29	17	7	17.429	5.412	55.571	20.032	17.857	5.610	1.428	0.191	240.429	147.174	144.071	89.985	-26.828	0.613
29	18	7	20.143	3.388	56.571	4.504	21.357	4.298	1.581	0.000	250.371	108.146	148.571	58.288	-26.766	0.667
29	19	3	23.667	7.095	71.667	19.009	18.667	0.577	1.462	0.206	201.933	13.639	125.700	4.689	-27.029	0.405
29	20	7	28.143	4.981	82.714	11.339	27.214	2.721	1.741	0.216	472.914	139.352	286.157	77.723	-26.299	0.542
29	21	6	23.667	5.164	72.500	9.333	22.333	1.941	1.629	0.118	293.383	40.413	181.117	33.112	-26.134	0.421
29	22	7	25.000	7.095	69.429	17.319	24.143	5.558	1.581	0.000	315.029	153.380	189.571	88.344	-26.623	0.629
29	23	7	25.714	5.469	68.571	17.934	23.286	5.282	1.613	0.219	339.729	167.661	170.586	109.698	-26.818	0.473
29	24	4	25.250	4.349	72.750	11.442	26.625	5.633	1.726	0.167	360.500	53.174	228.125	30.678	-26.191	0.829
29	25	8	17.500	3.703	53.125	9.992	19.063	4.075	1.581	0.000	264.775	60.677	154.550	52.336	-26.645	0.533
29	26	7	24.286	6.701	72.571	18.155	23.143	1.107	1.705	0.155	331.114	64.603	212.071	40.740	-26.299	0.889
29	27	7	22.714	6.130	67.143	9.974	24.500	2.466	1.572	0.187	374.229	90.553	233.314	43.627	-27.105	0.432
29	28	7	19.857	8.071	58.714	16.650	20.214	4.271	1.623	0.109	286.714	125.641	187.714	76.241	-27.198	0.575
29	29	6	25.000	5.550	68.833	14.919	24.250	3.934	1.462	0.184	418.583	154.210	262.200	87.416	-26.952	0.414
29	30	7	18.857	5.305	58.429	11.816	18.214	3.498	1.581	0.000	221.143	89.218	145.957	58.274	-27.061	0.412
30	1	7	16.143	2.911	55.429	16.298	22.119	2.889	1.479	0.174	269.443	106.877	174.229	67.843	-26.970	0.756
30	2	7	17.429	3.952	52.143	9.137	21.087	3.792	1.428	0.191	227.057	62.861	141.814	40.406	-26.763	0.753
30	4	7	18.286	4.112	55.000	8.679	21.834	3.953	1.479	0.174	242.671	55.280	156.786	35.133	-27.107	0.638
30	5	6	15.667	3.011	55.500	9.995	17.161	3.473	1.581	0.000	153.433	60.305	94.967	38.556	-26.706	0.280
30	6	6	14.333	3.670	44.000	11.662	17.519	2.713	1.392	0.275	152.150	52.341	99.183	32.805	-27.205	0.591
30	7	7	16.429	2.507	47.429	6.779	18.114	3.403	1.404	0.343	174.986	65.826	111.586	36.444	-27.352	0.646
30	8	5	15.000	2.828	44.000	10.124	16.182	2.555	1.367	0.195	172.820	66.945	111.680	42.551	-27.022	0.503
30	9	7	18.429	3.101	55.000	7.000	21.424	3.680	1.428	0.191	240.129	53.587	150.786	32.479	-27.038	0.479
30	12	6	12.833	2.858	40.333	16.158	16.667	2.443	1.511	0.248	176.867	47.014	120.233	26.878	-27.238	0.752
30	13	7	13.000	4.082	40.000	16.452	18.534	4.020	1.377	0.191	204.143	99.701	138.171	64.083	-27.661	0.414
30	17	4	14.250	1.708	39.750	5.123	15.500	1.291	1.403	0.206	147.075	46.656	97.700	28.320	-26.690	1.022
30	18	5	13.600	3.130	41.800	6.221	16.500	1.173	1.439	0.195	166.360	39.139	112.140	27.995	-26.819	0.557
30	19	6	16.167	2.229	44.333	9.004	19.333	3.125	1.462	0.184	251.667	77.505	168.250	43.343	-26.846	0.820
30	20	6	14.167	3.125	42.833	5.492	17.083	1.656	1.284	0.145	167.683	62.354	94.450	25.961	-27.497	0.601
30	21	5	11.000	1.414	43.800	8.289	17.400	2.679	1.497	0.275	182.860	55.728	121.840	38.640	-27.083	0.558
30	22	6	15.000	2.966	53.333	9.092	20.833	2.910	1.570	0.205	258.083	94.360	169.367	59.464	-27.077	0.486
30	23	6	14.667	3.011	46.000	11.679	18.667	4.750	1.462	0.184	230.500	108.966	154.583	67.960	-27.099	0.487

30	24	6	14.000	2.608	40.667	9.873	16.583	4.443	1.462	0.184	183.383	114.051	127.333	76.854	-26.969	0.463
30	27	7	15.857	2.968	60.857	13.018	22.429	6.630	1.649	0.280	359.686	259.165	215.057	154.421	-27.010	0.503
40	1	7	36.714	13.949	95.571	26.273	27.290	6.549	1.613	0.219	485.229	283.306	256.829	140.116	-26.504	0.462
40	2	7	42.000	9.309	105.714	18.016	26.034	4.015	1.747	0.155	388.843	127.306	215.029	69.310	-26.943	0.466
40	3	7	38.143	16.273	98.571	26.235	25.208	6.300	1.700	0.215	333.071	191.280	183.529	94.540	-26.468	0.661
40	4	7	34.286	4.680	88.857	9.459	24.436	2.671	1.581	0.000	324.243	84.124	178.200	40.094	-26.889	0.290
40	5	7	33.857	8.934	84.000	17.068	21.883	2.158	1.613	0.219	264.214	74.628	152.014	39.922	-26.984	0.231
40	7	7	34.286	12.816	87.286	27.741	20.194	6.138	1.818	0.244	279.529	136.407	158.371	70.845	-26.564	0.292
40	8	7	29.286	10.563	76.000	20.404	20.753	3.619	1.581	0.000	240.900	101.822	147.914	57.084	-26.556	0.556
40	9	7	34.429	7.976	84.000	13.077	24.101	3.197	1.664	0.141	320.914	114.147	187.886	66.884	-26.471	0.409
40	10	7	30.143	7.904	85.000	21.556	21.183	4.111	1.705	0.155	274.000	113.447	160.600	59.710	-26.977	0.414
40	11	7	33.143	8.821	83.429	8.848	23.121	2.636	1.581	0.000	311.214	109.146	184.229	67.416	-27.116	0.487
40	12	7	33.714	10.735	84.571	22.397	21.995	3.389	1.705	0.155	240.714	77.183	139.243	29.967	-26.693	0.462
40	13	7	31.143	5.460	80.857	15.071	20.890	4.518	1.664	0.141	293.029	187.158	173.714	106.507	-27.063	0.526
40	14	7	34.571	13.265	84.000	23.544	24.378	5.648	1.613	0.219	303.900	130.552	166.214	66.205	-26.661	0.413
40	15	7	25.857	4.981	71.571	13.649	21.631	1.619	1.613	0.219	265.371	47.087	164.443	24.497	-27.236	0.345
40	16	7	30.286	7.387	92.857	17.170	22.410	4.588	1.782	0.208	276.943	111.339	160.071	59.643	-27.050	0.767
40	17	7	34.286	12.619	93.000	22.745	25.109	5.258	1.747	0.155	334.871	142.157	188.971	67.139	-26.307	0.690
40	18	7	34.286	13.073	88.286	29.153	24.796	6.987	1.696	0.249	352.014	161.081	198.629	86.949	-26.811	0.950
40	20	7	29.714	6.370	84.857	11.423	24.816	3.769	1.623	0.109	323.771	145.856	184.600	77.461	-26.479	0.590
40	21	7	26.143	6.768	71.857	15.027	19.894	4.736	1.623	0.109	222.614	90.726	135.614	49.965	-26.413	0.572
40	23	7	29.857	5.429	75.857	14.713	18.822	4.841	1.705	0.155	198.729	114.759	119.486	69.910	-26.862	0.943
40	24	7	33.714	9.358	84.429	11.028	24.037	2.616	1.788	0.141	318.414	83.632	185.629	48.901	-26.906	0.590
40	25	7	36.429	6.876	87.000	8.869	24.936	3.845	1.705	0.155	342.957	87.496	198.214	43.549	-26.486	0.631
40	26	7	31.714	7.740	79.286	11.912	22.512	3.822	1.658	0.204	316.171	100.014	189.614	61.741	-26.657	0.430
40	27	7	32.429	13.202	90.000	21.229	24.016	4.675	1.788	0.141	326.943	130.426	185.100	60.594	-26.516	0.485
40	28	7	37.571	13.939	87.286	17.095	25.653	7.328	1.700	0.215	370.443	237.601	210.143	127.796	-26.787	0.529
40	29	7	44.000	13.964	107.571	15.576	26.769	5.405	1.782	0.208	407.557	187.100	209.214	86.244	-26.110	1.097
40	30	7	27.571	6.079	78.286	20.605	20.812	3.488	1.741	0.216	242.414	91.568	137.671	48.089	-26.683	0.440
41	1	7	37.286	13.524	97.429	18.716	24.266	4.957	1.788	0.141	335.857	117.958	184.529	56.433	-26.477	0.525
41	2	7	34.571	9.016	90.714	16.680	22.029	3.285	1.824	0.189	287.586	99.302	167.529	50.639	-26.472	0.465
41	3	7	28.714	9.340	75.714	13.200	17.937	3.498	1.664	0.141	193.800	76.159	117.171	43.892	-26.786	0.473
41	4	6	28.667	8.548	78.167	13.703	20.559	3.807	1.774	0.150	242.833	102.240	148.817	58.011	-26.903	0.481
41	5	7	41.429	12.177	96.571	25.954	25.547	3.707	1.788	0.141	342.229	107.570	189.314	52.026	-26.446	0.417
41	6	7	33.714	14.545	87.571	16.959	23.129	4.758	1.705	0.155	329.057	138.884	201.243	76.020	-27.120	0.241
41	7	7	29.571	8.638	78.714	15.987	22.453	3.559	1.782	0.208	300.686	122.443	177.414	65.733	-27.080	0.468
41	8	7	40.857	11.481	106.000	17.098	26.890	4.123	1.664	0.141	373.414	95.155	198.671	34.434	-26.586	0.761
41	9	7	35.143	6.694	91.714	18.750	25.321	3.553	1.741	0.216	371.814	140.374	213.814	71.414	-26.864	0.484
41	10	7	36.143	12.496	87.286	25.863	22.017	6.440	1.782	0.208	294.486	153.302	168.629	82.475	-26.481	0.504
41	11	7	35.000	4.123	87.714	14.511	23.792	3.811	1.747	0.155	326.686	114.050	193.071	64.435	-26.325	0.430
41	12	7	31.000	10.000	74.857	20.424	19.660	3.319	1.741	0.216	210.243	87.831	127.757	45.813	-26.451	0.875
41	13	7	29.857	9.873	83.429	23.151	20.229	4.696	1.521	0.228	254.014	132.784	149.514	72.956	-26.655	0.414
41	14	7	29.286	6.626	71.857	11.852	20.757	4.396	1.649	0.280	228.543	115.222	131.986	67.317	-26.358	0.481
41	15	7	31.429	5.912	80.714	9.534	19.449	2.404	1.623	0.109	213.114	55.621	121.171	29.145	-26.495	0.338
41	16	7	31.857	16.466	78.000	22.964	19.126	3.691	1.649	0.280	208.914	95.800	124.757	46.517	-26.695	0.854
41	18	6	33.667	13.110	89.500	20.246	25.945	4.267	1.726	0.159	348.667	132.838	186.217	57.170	-26.696	0.298
41	19	7	44.143	13.753	110.286	22.647	24.846	3.664	1.696	0.249	351.743	119.847	178.771	51.919	-26.586	0.724
41	20	7	37.286	9.656	95.286	14.829	24.485	2.538	1.818	0.244	353.500	124.656	200.971	75.475	-27.017	0.900
41	21	7	33.857	13.447	83.857	21.279	20.485	4.590	1.664	0.141	233.329	105.742	136.286	58.273	-26.733	0.438
41	22	7	26.143	8.454	71.714	10.950	19.654	3.308	1.664	0.141	189.800	58.971	115.357	32.158	-26.798	0.857
41	23	7	24.571	7.480	75.286	15.617	19.763	4.844	1.664	0.141	221.043	103.443	130.329	54.586	-26.379	0.448
41	24	7	33.143	8.764	93.286	13.313	25.414	3.666	1.788	0.141	374.014	133.864	213.529	69.213	-26.485	0.698
41	26	7	33.429	11.660	88.571	13.710	22.470	3.225	1.782	0.208	307.600	115.885	184.200	63.952	-26.467	0.707
41	27	7	33.714	7.319	84.857	15.700	24.004	5.943	1.747	0.155	341.914	190.324	208.129	111.125	-26.740	0.759

41	28	7	26.429	7.807	79.000	11.358	19.836	2.291	1.664	0.141	210.486	52.892	123.414	27.181	-26.226	0.685
41	29	7	28.286	10.766	79.857	17.267	22.442	2.565	1.747	0.155	273.957	63.315	159.286	29.926	-27.074	0.658
41	30	7	38.857	7.988	101.286	7.319	27.197	2.762	1.705	0.155	419.857	55.492	234.271	28.598	-26.786	0.742
42	1	7	37.857	9.720	102.429	27.946	25.746	5.409	1.809	0.315	361.943	138.917	190.900	62.510	-26.193	0.565
42	2	7	41.857	11.611	98.714	19.111	26.318	5.870	1.664	0.141	390.686	150.366	206.914	70.346	-26.699	0.650
42	3	7	24.714	4.923	75.429	12.354	20.520	3.165	1.705	0.155	228.629	113.391	139.514	78.804	-25.898	0.587
42	4	6	28.000	9.381	74.333	22.232	20.701	3.398	1.726	0.159	216.950	87.311	129.150	41.955	-26.531	0.529
42	5	7	31.857	16.668	81.714	29.330	22.097	5.168	1.747	0.155	317.500	191.866	182.900	99.972	-26.465	0.505
42	6	7	33.857	7.819	89.429	8.243	22.269	2.930	1.664	0.141	241.257	60.282	138.857	29.165	-26.541	0.542
42	8	7	38.000	14.457	91.286	21.531	24.803	6.438	1.705	0.155	378.771	218.631	218.029	114.905	-27.071	0.647
42	9	7	28.857	9.209	77.714	21.383	19.277	4.238	1.664	0.141	210.386	103.821	123.371	57.680	-26.545	0.752
42	10	7	39.000	9.363	101.429	10.907	24.709	2.369	1.747	0.155	370.586	118.960	212.657	72.092	-26.957	0.446
42	11	7	35.571	8.904	86.286	22.291	23.242	3.509	1.690	0.290	273.657	88.245	155.529	43.643	-26.264	0.587
42	12	7	27.857	4.220	74.286	7.158	19.860	2.524	1.747	0.155	197.229	47.091	97.914	41.942	-26.170	0.840
42	13	7	32.714	8.538	88.714	22.896	22.296	3.021	1.895	0.242	279.457	75.970	158.457	35.498	-26.356	0.676
42	14	7	33.143	6.939	85.286	11.842	21.677	3.952	1.664	0.141	283.243	88.928	164.457	46.704	-26.420	0.807
42	15	7	27.000	4.163	79.000	15.652	21.024	4.901	1.705	0.155	222.186	111.496	132.529	61.195	-26.455	0.523
42	16	7	35.857	6.568	80.571	8.960	20.005	3.223	1.664	0.141	211.429	86.543	124.571	48.299	-26.544	0.474
42	18	7	37.571	11.178	88.571	8.677	22.006	3.867	1.664	0.141	253.643	113.084	141.800	67.195	-26.382	0.733
42	19	7	34.286	9.250	85.286	10.404	23.251	2.919	1.664	0.141	286.157	66.879	158.543	37.323	-26.714	0.783
42	20	7	34.429	8.867	85.714	10.226	22.376	3.221	1.664	0.141	287.571	100.914	168.300	53.827	-26.618	0.587
42	21	7	38.857	4.180	95.857	11.187	23.843	2.520	1.747	0.155	331.871	70.686	186.486	39.766	-26.697	0.591
42	22	7	33.000	7.767	89.143	8.572	21.726	3.102	1.664	0.141	276.329	96.992	165.586	56.393	-26.369	0.714
42	23	7	27.143	8.071	74.000	21.158	21.023	5.014	1.581	0.000	222.029	100.957	127.614	51.018	-26.755	0.639
42	24	7	28.143	10.367	78.857	29.077	18.904	4.028	1.747	0.155	180.014	61.738	103.929	37.058	-27.002	0.405
42	25	7	31.000	8.869	88.857	15.126	21.853	3.844	1.782	0.208	265.486	118.336	139.771	66.160	-26.186	0.761
42	26	7	37.143	13.508	88.143	23.405	25.736	5.152	1.735	0.264	360.200	150.519	199.014	70.664	-26.848	0.779
42	27	7	34.714	9.793	87.286	18.581	22.598	5.255	1.767	0.326	296.000	131.064	172.029	74.378	-26.518	0.540
42	28	7	30.000	11.075	74.143	14.702	22.171	3.314	1.664	0.141	264.814	91.751	158.029	40.661	-26.729	0.382
42	29	7	40.429	9.199	100.857	18.425	25.162	4.590	1.788	0.141	382.443	134.815	202.014	62.536	-26.535	0.489
42	30	7	32.429	5.968	96.857	19.100	22.439	4.889	1.705	0.155	268.257	130.712	149.657	67.493	-26.545	0.580
43	1	6	35.667	13.277	78.833	21.977	20.453	5.539	1.629	0.118	232.617	117.145	134.600	61.160	-26.973	0.461
43	2	7	35.000	9.730	85.571	23.458	23.731	6.637	1.572	0.187	308.814	180.052	173.743	93.570	-26.775	0.589
43	3	7	32.000	15.133	81.143	22.431	19.528	4.455	1.741	0.216	196.286	90.260	110.814	42.109	-26.125	0.568
43	4	7	32.429	9.676	89.000	22.635	23.478	3.963	1.818	0.244	274.314	98.289	159.214	53.297	-26.689	0.378
43	5	7	25.286	5.282	70.143	11.539	18.940	3.633	1.705	0.155	197.829	80.722	121.371	47.533	-26.503	0.499
43	6	7	39.429	10.486	94.857	23.611	25.402	5.192	1.705	0.155	372.257	153.753	210.929	84.040	-26.624	0.837
43	7	7	28.286	6.264	74.571	13.538	22.089	6.615	1.664	0.141	295.071	189.830	181.771	107.139	-26.969	0.653
43	8	6	29.667	8.641	83.833	15.381	23.358	4.412	1.581	0.000	323.317	138.994	187.717	78.917	-26.703	0.665
43	9	6	36.833	13.497	90.833	17.244	24.150	6.364	1.726	0.159	299.600	147.595	173.433	75.026	-26.336	0.848
43	10	7	29.714	10.062	83.857	28.801	23.124	3.644	1.782	0.208	278.286	115.649	158.771	54.161	-26.499	0.710
43	11	7	32.714	19.889	82.429	12.674	22.066	3.801	1.654	0.238	279.229	93.379	161.171	49.539	-26.558	0.714
43	12	7	36.286	9.196	89.143	16.906	25.487	5.178	1.664	0.141	364.243	153.879	204.557	77.441	-26.648	0.512
43	13	7	32.571	7.068	87.143	14.848	23.792	4.584	1.664	0.141	314.086	124.316	176.843	55.910	-26.504	0.511
43	14	7	29.429	10.309	76.429	17.067	21.008	2.545	1.530	0.135	225.486	42.835	129.414	15.496	-26.531	0.868
43	15	5	24.600	6.950	66.800	10.354	20.722	3.425	1.581	0.000	248.000	79.017	150.100	44.619	-26.551	0.780
43	16	7	38.429	11.660	92.286	15.924	27.266	4.239	1.705	0.155	406.829	111.546	226.014	62.103	-26.673	0.508
43	17	7	31.000	11.075	77.571	19.857	20.402	5.082	1.700	0.215	268.557	162.028	159.914	86.935	-27.088	0.482
43	18	7	27.000	7.024	80.571	14.467	22.768	2.602	1.782	0.208	289.243	91.370	171.371	53.780	-26.323	0.713
43	19	7	35.571	8.810	88.286	10.735	23.239	3.395	1.572	0.187	294.071	81.334	172.043	44.392	-26.613	0.415
43	20	7	31.571	5.682	87.000	8.206	23.009	2.903	1.664	0.141	305.929	75.920	176.071	36.752	-26.629	0.627
43	21	7	30.000	8.737	73.857	8.971	22.819	3.039	1.664	0.141	280.543	81.967	167.129	49.602	-26.739	0.673
43	22	7	27.000	6.532	78.857	18.343	21.470	4.852	1.741	0.216	284.914	93.588	172.029	53.011	-26.412	0.909
43	23	7	26.714	7.135	82.000	10.231	21.495	2.588	1.747	0.155	272.114	68.107	158.914	34.828	-26.539	0.402

43	24	7	35.714	7.410	84.857	13.729	21.296	4.028	1.705	0.155	252.943	104.684	145.543	56.374	-26.580	0.611
43	25	7	38.429	10.486	99.429	14.943	24.069	4.525	1.741	0.216	326.671	93.007	181.871	48.737	-26.271	0.889
43	27	6	36.167	4.708	84.833	14.331	23.618	4.475	1.618	0.239	295.733	126.776	169.783	65.773	-26.570	0.704
43	28	7	32.000	8.563	80.143	20.465	20.977	4.442	1.741	0.216	225.771	105.058	131.029	54.073	-26.559	0.616
43	29	7	32.571	7.997	80.429	24.919	20.871	6.747	1.700	0.215	263.343	130.530	154.371	72.907	-26.671	0.521
43	30	7	38.000	5.132	93.286	12.338	23.619	2.450	1.664	0.141	334.971	75.615	194.743	43.075	-26.510	0.522
44	1	7	32.857	5.014	87.000	9.781	24.540	0.972	1.705	0.155	386.029	66.455	225.043	34.696	-26.372	0.492
44	3	6	35.167	14.034	95.167	17.759	26.218	3.570	1.726	0.159	397.650	120.124	217.517	45.843	-26.341	0.834
44	4	7	33.857	8.552	91.429	6.294	26.671	4.134	1.705	0.155	391.614	107.757	226.914	55.931	-25.939	0.468
44	6	7	28.571	8.182	81.571	15.153	21.895	4.816	1.664	0.141	241.971	101.711	134.671	51.345	-26.652	0.639
44	7	7	24.714	5.559	74.143	12.415	21.953	3.177	1.623	0.109	232.300	77.119	137.343	40.691	-26.372	0.604
44	8	7	27.714	9.050	65.286	17.500	19.616	2.627	1.581	0.000	200.471	48.168	121.529	26.350	-26.580	0.655
44	9	7	34.857	12.851	83.143	22.177	21.449	4.746	1.700	0.215	251.871	105.913	143.000	50.212	-26.562	0.395
44	10	7	36.571	14.774	82.000	18.358	20.879	3.806	1.581	0.000	222.800	82.100	129.129	41.542	-26.293	0.866
44	11	7	28.857	10.024	74.429	20.919	20.704	5.157	1.664	0.141	251.486	132.826	146.214	71.503	-26.407	0.375
44	12	7	33.857	11.711	87.571	16.672	22.524	4.466	1.623	0.109	274.100	105.770	160.843	61.799	-26.536	0.403
44	14	7	29.000	8.944	82.286	18.918	23.991	3.641	1.658	0.204	305.314	120.468	173.229	57.840	-26.912	0.744
44	15	7	31.571	8.039	85.286	13.450	23.615	4.073	1.705	0.155	331.314	123.005	194.829	64.708	-26.643	0.649
44	16	7	33.286	12.632	86.286	18.989	24.246	4.695	1.818	0.244	314.314	108.772	177.243	53.761	-26.730	0.703
44	17	7	29.571	4.237	81.000	13.454	22.071	3.344	1.700	0.215	273.457	83.918	169.271	51.054	-26.451	0.537
44	18	7	27.714	5.736	84.429	7.807	23.620	4.008	1.705	0.155	294.043	99.520	173.086	54.614	-25.802	0.592
44	19	6	27.667	7.528	82.333	14.222	22.772	2.778	1.768	0.224	291.767	106.910	166.883	58.384	-26.271	0.352
44	21	7	28.714	9.482	79.571	16.762	21.464	5.608	1.581	0.000	248.471	141.332	136.214	61.553	-26.499	0.454
44	22	7	35.714	11.398	91.857	21.729	23.146	4.838	1.664	0.141	298.371	158.710	173.771	88.243	-26.516	0.543
44	23	6	33.167	7.731	91.833	17.882	22.424	3.520	1.774	0.150	263.033	76.728	155.433	40.432	-25.928	0.526
44	24	7	30.571	13.963	76.000	21.954	20.825	5.115	1.623	0.109	240.529	124.705	143.157	70.377	-26.672	0.446
44	25	7	26.000	4.796	74.143	15.302	20.522	5.334	1.581	0.000	233.057	154.414	134.514	82.650	-26.594	0.376
44	26	7	33.286	8.712	88.857	12.089	21.451	3.654	1.623	0.109	241.629	89.683	135.214	41.864	-26.182	0.678
46	1	6	26.667	8.501	76.333	23.483	22.931	4.928	1.511	0.248	271.100	97.091	153.900	46.743	-26.165	0.611
46	2	7	37.857	6.619	97.429	23.387	24.818	3.310	1.829	0.109	354.671	98.366	203.814	48.051	-26.469	0.602
46	3	7	30.857	7.690	84.857	10.238	22.621	3.504	1.664	0.141	296.429	99.840	167.757	49.114	-26.297	0.652
46	4	7	29.857	6.842	86.571	18.840	21.312	2.767	1.705	0.155	248.843	85.047	143.643	34.153	-26.508	0.615
46	5	7	29.000	10.551	73.857	23.334	19.499	4.322	1.623	0.109	194.329	87.639	117.657	44.709	-26.406	0.405
46	6	7	35.000	7.703	97.000	24.180	25.599	5.126	1.782	0.208	359.457	110.028	205.657	55.286	-27.220	0.422
46	7	7	29.857	5.113	83.000	12.014	22.355	4.012	1.788	0.141	288.700	93.527	172.243	52.277	-26.539	0.466
46	8	7	32.714	14.705	88.714	17.007	23.446	4.046	1.623	0.109	333.914	133.081	192.214	62.224	-26.771	0.777
46	9	7	35.571	13.843	84.286	22.728	22.114	6.463	1.705	0.155	279.900	167.109	160.971	93.265	-26.431	0.545
46	10	7	41.286	10.719	97.429	24.535	25.846	4.627	1.705	0.155	400.200	151.429	221.829	71.155	-26.434	0.294
46	11	7	39.714	13.086	96.143	20.788	27.901	3.862	1.613	0.219	524.357	169.130	285.543	75.081	-26.453	0.562
46	12	7	35.571	9.914	93.429	20.582	26.032	4.264	1.696	0.249	405.400	139.669	235.086	70.940	-26.664	0.455
46	13	7	34.429	15.372	94.286	18.874	25.655	3.892	1.664	0.141	370.729	150.594	209.657	79.391	-26.533	0.333
46	14	7	33.571	11.443	89.000	13.723	23.289	3.124	1.623	0.109	332.214	90.455	189.057	49.579	-26.482	0.257
46	15	6	28.500	9.894	71.667	19.603	23.329	5.288	1.522	0.145	291.750	181.181	162.750	101.080	-26.517	1.016
50	1	7	37.143	10.574	98.000	14.248	28.934	7.145	1.978	0.134	588.900	173.501	341.386	96.133	-26.979	0.708
50	2	7	35.714	9.429	85.286	14.080	26.808	3.356	1.664	0.141	466.529	96.500	290.200	55.830	-27.234	0.468
50	4	7	35.857	7.581	90.429	13.265	26.829	6.305	1.860	0.221	427.614	182.489	239.814	94.766	-26.713	0.891
50	6	7	45.571	13.327	102.714	15.185	30.487	3.480	1.829	0.109	583.729	156.224	317.629	88.946	-26.628	0.601
50	7	7	35.286	11.131	88.429	19.940	30.016	5.268	1.737	0.250	540.386	246.849	304.414	125.695	-27.419	0.503
50	8	7	36.714	10.889	90.857	26.755	26.621	6.283	1.737	0.250	496.800	237.509	293.986	137.696	-26.385	0.851
50	9	7	37.714	8.770	91.714	19.440	27.581	4.932	1.824	0.189	491.229	187.074	290.429	108.503	-27.066	0.578
50	10	7	37.429	3.599	99.571	5.711	26.672	2.424	1.942	0.122	444.886	65.592	262.014	44.951	-26.433	0.870
50	12	7	31.143	6.914	83.857	17.276	27.372	4.996	1.901	0.184	450.929	166.689	270.800	91.942	-26.609	0.817
50	13	7	35.714	4.645	90.429	6.188	30.464	3.175	1.860	0.221	519.700	73.476	298.829	49.609	-26.659	0.402

**Annexe 5. Test de comparaison des
moyennes ajustées par paires de
populations pour *HT02, HT03, DC, PolyT,*
TB, NB et $\delta^{13}C$**

Probabilité $\Pr > |t|$ pour l'hypothèse nulle H_0 que la moyenne ajustée de la population i soit égale à la moyenne ajustée de la population j . Les populations sont codées de 1 à 16 :

Code population		Code dans les sorties du test
15	→	1
20	→	2
21	→	3
22	→	4
23	→	5
24	→	6
25	→	7
26	→	8
27	→	9
29	→	10
30	→	11
40	→	12
41	→	13
42	→	14
43	→	15
44	→	16
46	→	17
50	→	18

A) HT02

i/j	1	2	3	4	5	6	7	8	9
1		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
2	<.0001		0.6105	<.0001	<.0001	0.0409	<.0001	<.0001	<.0001
3	<.0001	0.6105		<.0001	<.0001	0.0782	<.0001	<.0001	<.0001
4	<.0001	<.0001	<.0001		0.2296	<.0001	0.6552	0.2062	<.0001
5	<.0001	<.0001	<.0001	0.2296		<.0001	0.4054	0.0068	<.0001
6	<.0001	0.0409	0.0782	<.0001	<.0001		<.0001	<.0001	<.0001
7	<.0001	<.0001	<.0001	0.6552	0.4054	<.0001		0.0600	<.0001
8	<.0001	<.0001	<.0001	0.2062	0.0068	<.0001	0.0600		<.0001
9	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	
10	<.0001	<.0001	<.0001	0.0001	<.0001	<.0001	<.0001	0.0043	<.0001
11	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
12	<.0001	0.6219	0.9774	<.0001	<.0001	0.0687	<.0001	<.0001	<.0001
13	<.0001	0.6991	0.8777	<.0001	<.0001	0.0518	<.0001	<.0001	<.0001
14	<.0001	0.5400	0.9153	<.0001	<.0001	0.0878	<.0001	<.0001	<.0001
15	<.0001	0.7482	0.3317	<.0001	<.0001	0.0069	<.0001	<.0001	<.0001
16	<.0001	0.1171	0.0193	<.0001	0.0020	0.0001	0.0001	<.0001	<.0001
17	<.0001	0.6242	0.9678	<.0001	<.0001	0.1311	<.0001	<.0001	<.0001
18	<.0001	0.0019	0.0035	<.0001	<.0001	0.1554	<.0001	<.0001	0.0366

i/j	10	11	12	13	14	15	16	17	18
1	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
2	<.0001	<.0001	0.6219	0.6991	0.5400	0.7482	0.1171	0.6242	0.0019
3	<.0001	<.0001	0.9774	0.8777	0.9153	0.3317	0.0193	0.9678	0.0035
4	0.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
5	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0020	<.0001	<.0001
6	<.0001	<.0001	0.0687	0.0518	0.0878	0.0069	0.0001	0.1311	0.1554
7	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0001	<.0001	<.0001
8	0.0043	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
9	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0366
10		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
11	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
12	<.0001	<.0001		0.8979	0.8902	0.3350	0.0183	0.9479	0.0029
13	<.0001	<.0001	0.8979		0.7886	0.4005	0.0243	0.8626	0.0020
14	<.0001	<.0001	0.8902	0.7886		0.2667	0.0122	0.9602	0.0038
15	<.0001	<.0001	0.3350	0.4005	0.2667		0.1408	0.3829	0.0002
16	<.0001	<.0001	0.0183	0.0243	0.0122	0.1408		0.0377	<.0001
17	<.0001	<.0001	0.9479	0.8626	0.9602	0.3829	0.0377		0.0083
18	<.0001	<.0001	0.0029	0.0020	0.0038	0.0002	<.0001	0.0083	

B) HT03

i/j	1	2	3	4	5	6	7	8	9
1		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
2	<.0001		0.6396	<.0001	<.0001	0.6488	<.0001	<.0001	<.0001
3	<.0001	0.6396		<.0001	<.0001	0.9877	<.0001	<.0001	<.0001
4	<.0001	<.0001	<.0001		0.2118	<.0001	0.4705	0.2734	<.0001
5	<.0001	<.0001	<.0001	0.2118		<.0001	0.5610	0.0101	<.0001
6	<.0001	0.6488	0.9877	<.0001	<.0001		<.0001	<.0001	<.0001
7	<.0001	<.0001	<.0001	0.4705	0.5610	<.0001		0.0459	<.0001

8	<.0001	<.0001	<.0001	0.2734	0.0101	<.0001	0.0459		<.0001
9	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
10	<.0001	<.0001	<.0001	0.4659	0.0303	<.0001	0.1114	0.6900	<.0001
11	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
12	<.0001	0.0001	0.0001	<.0001	<.0001	0.0003	<.0001	<.0001	0.0002
13	<.0001	<.0001	<.0001	<.0001	<.0001	0.0002	<.0001	<.0001	0.0003
14	<.0001	<.0001	<.0001	<.0001	<.0001	0.0002	<.0001	<.0001	0.0003
15	<.0001	0.0159	0.0259	<.0001	<.0001	0.0411	<.0001	<.0001	<.0001
16	<.0001	0.0341	0.0590	<.0001	<.0001	0.0808	<.0001	<.0001	<.0001
17	<.0001	<.0001	<.0001	<.0001	<.0001	0.0002	<.0001	<.0001	0.0098
18	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.6675

i/j	10	11	12	13	14	15	16	17	18
1	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
2	<.0001	<.0001	0.0001	<.0001	<.0001	0.0159	0.0341	<.0001	<.0001
3	<.0001	<.0001	0.0001	<.0001	<.0001	0.0259	0.0590	<.0001	<.0001
4	0.4659	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
5	0.0303	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
6	<.0001	<.0001	0.0003	0.0002	0.0002	0.0411	0.0808	0.0002	<.0001
7	0.1114	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
8	0.6900	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
9	<.0001	<.0001	0.0002	0.0003	0.0003	<.0001	<.0001	0.0098	0.6675
10	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
11	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
12	<.0001	<.0001		0.9024	0.9167	0.0833	0.0691	0.4927	0.0090
13	<.0001	<.0001	0.9024		0.9855	0.0620	0.0520	0.5577	0.0113
14	<.0001	<.0001	0.9167	0.9855		0.0642	0.0538	0.5472	0.0108
15	<.0001	<.0001	0.0833	0.0620	0.0642		0.8393	0.0324	0.0001
16	<.0001	<.0001	0.0691	0.0520	0.0538	0.8393		0.0272	<.0001
17	<.0001	<.0001	0.4927	0.5577	0.5472	0.0324	0.0272		0.0690
18	<.0001	<.0001	0.0090	0.0113	0.0108	0.0001	<.0001	0.0690	

C) DC

i/j	1	2	3	4	5	6	7	8	9
1		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
2	<.0001		0.2943	<.0001	<.0001	0.9039	0.0002	<.0001	0.5424
3	<.0001	0.2943		<.0001	<.0001	0.3282	<.0001	<.0001	0.0754
4	<.0001	<.0001	<.0001		0.1810	<.0001	0.0159	0.8857	<.0001
5	<.0001	<.0001	<.0001	0.1810		<.0001	0.2382	0.1041	<.0001
6	<.0001	0.9039	0.3282	<.0001	<.0001		<.0001	<.0001	0.4436
7	<.0001	0.0002	<.0001	0.0159	0.2382	<.0001		0.0051	0.0017
8	<.0001	<.0001	<.0001	0.8857	0.1041	<.0001	0.0051		<.0001
9	<.0001	0.5424	0.0754	<.0001	<.0001	0.4436	0.0017	<.0001	
10	<.0001	<.0001	<.0001	0.5232	0.4456	<.0001	0.0527	0.3915	<.0001
11	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
12	<.0001	0.0001	<.0001	0.0228	0.3013	<.0001	0.8865	0.0080	0.0011
13	<.0001	<.0001	<.0001	0.4511	0.5163	<.0001	0.0665	0.3242	<.0001
14	<.0001	<.0001	<.0001	0.5504	0.4105	<.0001	0.0443	0.4153	<.0001
15	<.0001	<.0001	<.0001	0.3668	0.6251	<.0001	0.0934	0.2500	<.0001
16	<.0001	<.0001	<.0001	0.3695	0.6843	<.0001	0.1295	0.2613	<.0001
17	<.0001	0.0151	0.0002	0.0037	0.0607	0.0077	0.3731	0.0012	0.0569
18	0.4871	<.0001	0.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001

i/j	10	11	12	13	14	15	16	17	18
1	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.4871
2	<.0001	<.0001	0.0001	<.0001	<.0001	<.0001	<.0001	0.0151	<.0001
3	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0002	0.0001
4	0.5232	<.0001	0.0228	0.4511	0.5504	0.3668	0.3695	0.0037	<.0001
5	0.4456	<.0001	0.3013	0.5163	0.4105	0.6251	0.6843	0.0607	<.0001
6	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0077	<.0001
7	0.0527	<.0001	0.8865	0.0665	0.0443	0.0934	0.1295	0.3731	<.0001
8	0.3915	<.0001	0.0080	0.3242	0.4153	0.2500	0.2613	0.0012	<.0001
9	<.0001	<.0001	0.0011	<.0001	<.0001	<.0001	<.0001	0.0569	<.0001
10		<.0001	0.0734	0.9041	0.9588	0.7777	0.7543	0.0121	<.0001
11	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
12	0.0734	<.0001		0.0916	0.0626	0.1262	0.1681	0.3133	<.0001
13	0.9041	<.0001	0.0916		0.8619	0.8705	0.8401	0.0153	<.0001
14	0.9588	<.0001	0.0626	0.8619		0.7356	0.7153	0.0101	<.0001
15	0.7777	<.0001	0.1262	0.8705	0.7356		0.9603	0.0218	<.0001
16	0.7543	<.0001	0.1681	0.8401	0.7153	0.9603		0.0322	<.0001
17	0.0121	<.0001	0.3133	0.0153	0.0101	0.0218	0.0322		<.0001
18	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	

D) PolyT

i/j	1	2	3	4	5	6	7	8	9
1		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0017
2	<.0001		0.2621	0.7967	0.6877	0.1078	0.4877	0.1365	<.0001
3	<.0001	0.2621		0.3634	0.3949	0.5357	0.0346	0.6869	0.0002
4	<.0001	0.7967	0.3634		0.8929	0.1515	0.2980	0.1935	<.0001
5	<.0001	0.6877	0.3949	0.8929		0.1562	0.1961	0.2001	<.0001
6	<.0001	0.1078	0.5357	0.1515	0.1562		0.0097	0.7963	0.0035
7	<.0001	0.4877	0.0346	0.2980	0.1961	0.0097		0.0101	<.0001
8	<.0001	0.1365	0.6869	0.1935	0.2001	0.7963	0.0101		0.0006
9	0.0017	<.0001	0.0002	<.0001	<.0001	0.0035	<.0001	0.0006	
10	<.0001	0.8781	0.1379	0.6459	0.5144	0.0453	0.5263	0.0542	<.0001
11	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
12	<.0001	<.0001	0.0006	<.0001	<.0001	0.0105	<.0001	<.0001	0.5062
13	0.0100	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.3606
14	0.0049	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.5054
15	<.0001	0.0001	0.0016	0.0001	<.0001	0.0222	<.0001	0.0048	0.3413
16	<.0001	0.0011	0.0139	0.0014	0.0009	0.0904	<.0001	0.0340	0.1736
17	<.0001	0.0034	0.0320	0.0046	0.0036	0.1382	<.0001	0.0668	0.1971
18	0.0196	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001

i/j	10	11	12	13	14	15	16	17	18
1	<.0001	<.0001	<.0001	0.0100	0.0049	<.0001	<.0001	<.0001	0.0196
2	0.8781	<.0001	<.0001	<.0001	<.0001	0.0001	0.0011	0.0034	<.0001
3	0.1379	<.0001	0.0006	<.0001	<.0001	0.0016	0.0139	0.0320	<.0001
4	0.6459	<.0001	<.0001	<.0001	<.0001	0.0001	0.0014	0.0046	<.0001
5	0.5144	<.0001	<.0001	<.0001	<.0001	<.0001	0.0009	0.0036	<.0001
6	0.0453	<.0001	0.0105	<.0001	<.0001	0.0222	0.0904	0.1382	<.0001
7	0.5263	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
8	0.0542	<.0001	0.0019	<.0001	<.0001	<.0001	0.0340	0.0668	<.0001
9	<.0001	<.0001	0.5062	0.3606	0.5054	0.3413	0.1736	0.1971	<.0001

10		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0006	<.0001
11	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
12	<.0001	<.0001	<.0001	0.0734	0.1308	0.7491	0.4171	0.4355	<.0001	<.0001
13	<.0001	<.0001	0.0734	0.7765	0.7765	0.0334	0.0124	0.0225	<.0001	<.0001
14	<.0001	<.0001	0.1308	0.7765	0.7765	0.0648	0.0252	0.0405	<.0001	<.0001
15	<.0001	<.0001	0.7491	0.0334	0.0648	0.6051	0.6051	0.6051	<.0001	<.0001
16	<.0001	<.0001	0.4171	0.0124	0.0252	0.6051	0.9567	0.9567	<.0001	<.0001
17	0.0006	<.0001	0.4355	0.0225	0.0405	0.6051	0.9567	0.9567	<.0001	<.0001
18	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001

E) TB

i/j	1	2	3	4	5	6	7	8	9
1		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
2	<.0001		0.6508	<.0001	<.0001	0.2394	0.0669	0.0001	0.0017
3	<.0001	0.6508		<.0001	0.0001	0.3989	0.1143	0.0001	<.0001
4	<.0001	<.0001	<.0001		0.0399	<.0001	<.0001	0.0389	<.0001
5	<.0001	0.0001	0.0001	0.0399		0.0062	0.0177	0.9876	<.0001
6	<.0001	0.2394	0.3989	<.0001	0.0062		0.5490	0.0066	<.0001
7	<.0001	0.0669	0.1143	<.0001	0.0177	0.5490		0.0187	<.0001
8	<.0001	0.0001	0.0001	0.0389	0.9876	0.0066	0.0187		<.0001
9	<.0001	0.0017	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	
10	<.0001	0.0037	0.0050	0.0024	0.2761	0.0814	0.2037	0.2841	<.0001
11	<.0001	<.0001	<.0001	0.0019	<.0001	<.0001	<.0001	<.0001	<.0001
12	<.0001	0.0122	0.0185	0.0005	0.1171	0.1869	0.4233	0.1217	<.0001
13	<.0001	0.0002	0.0001	0.0333	0.9432	0.0073	0.0208	0.9558	<.0001
14	<.0001	<.0001	<.0001	0.1650	0.4554	0.0006	0.0017	0.4470	<.0001
15	<.0001	<.0001	<.0001	0.0659	0.8010	0.0029	0.0082	0.7893	<.0001
16	<.0001	<.0001	<.0001	0.1650	0.5389	0.0016	0.0044	0.5300	<.0001
17	<.0001	0.3300	0.5264	<.0001	0.0071	0.8864	0.4777	0.0075	<.0001
18	0.5472	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001

i/j	10	11	12	13	14	15	16	17	18
1	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.5472
2	0.0037	<.0001	0.0122	0.0002	<.0001	<.0001	<.0001	0.3300	<.0001
3	0.0050	<.0001	0.0185	0.0001	<.0001	<.0001	<.0001	0.5264	<.0001
4	0.0024	0.0019	0.0005	0.0333	0.1650	0.0659	0.1650	<.0001	<.0001
5	0.2761	<.0001	0.1171	0.9432	0.4554	0.8010	0.5389	0.0071	<.0001
6	0.0814	<.0001	0.1869	0.0073	0.0006	0.0029	0.0016	0.8864	<.0001
7	0.2037	<.0001	0.4233	0.0208	0.0017	0.0082	0.0044	0.4777	<.0001
8	0.2841	<.0001	0.1217	0.9558	0.4470	0.7893	0.5300	0.0075	<.0001
9	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
10		<.0001	0.6369	0.3060	0.0660	0.1776	0.1017	0.0765	<.0001
11	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
12	0.6369	<.0001		0.1327	0.0203	0.0672	0.0369	0.1681	<.0001
13	0.3060	<.0001	0.1327		0.4112	0.4112	0.4939	0.0082	<.0001
14	0.0660	<.0001	0.0203	0.4112		0.6182	0.9335	0.0009	<.0001
15	0.1776	<.0001	0.0672	0.7451	0.6182		0.7026	0.0035	<.0001
16	0.1017	<.0001	0.0369	0.4939	0.9335	0.7026		0.0019	<.0001
17	0.0765	<.0001	0.1681	0.0082	0.0009	0.0035	0.0019		<.0001
18	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	

F) NB

i/j	1	2	3	4	5	6	7	8	9
1		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
2	<.0001		0.4790	<.0001	0.0061	0.0846	0.5685	0.0228	0.0015
3	<.0001	0.4790		<.0001	0.0198	0.2340	0.8621	0.0727	<.0001
4	<.0001	<.0001	<.0001		0.0250	0.0034	<.0001	0.0063	<.0001
5	<.0001	0.0061	0.0198	0.0250		0.3464	0.0103	0.5851	<.0001
6	<.0001	0.0846	0.2340	0.0034	0.3464		0.1690	0.6550	<.0001
7	<.0001	0.5685	0.8621	<.0001	0.0103	0.1690		0.0439	<.0001
8	<.0001	0.0228	0.0727	0.0063	0.5851	0.6550	0.0439		<.0001
9	<.0001	0.0015	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	
10	<.0001	0.0992	0.2844	0.0007	0.2008	0.8271	0.2024	0.4633	<.0001
11	<.0001	<.0001	<.0001	0.0075	<.0001	<.0001	<.0001	<.0001	<.0001
12	<.0001	0.0062	0.0201	0.0251	0.9983	0.3479	0.0104	0.5870	<.0001
13	<.0001	<.0001	0.0002	0.3202	0.1667	0.0284	<.0001	0.0539	<.0001
14	<.0001	<.0001	<.0001	0.8781	0.0081	0.0009	<.0001	0.0014	<.0001
15	<.0001	<.0001	0.0001	0.4099	0.1150	0.0179	<.0001	0.0339	<.0001
16	<.0001	<.0001	<.0001	0.7267	0.0528	0.0078	<.0001	0.0145	<.0001
17	<.0001	0.2501	0.5641	0.0012	0.1576	0.6162	0.4617	0.3404	<.0001
18	0.0705	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001

i/j	10	11	12	13	14	15	16	17	18
1	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0705
2	0.0992	<.0001	0.0062	<.0001	<.0001	<.0001	<.0001	0.2501	<.0001
3	0.2844	<.0001	0.0201	0.0002	<.0001	0.0001	<.0001	0.5641	<.0001
4	0.0007	0.0075	0.0251	0.3202	0.8781	0.4099	0.7267	0.0012	<.0001
5	0.2008	<.0001	0.9983	0.1667	0.0081	0.1150	0.0528	0.1576	<.0001
6	0.8271	<.0001	0.3479	0.0284	0.0009	0.0179	0.0078	0.6162	<.0001
7	0.2024	<.0001	0.0104	<.0001	<.0001	<.0001	<.0001	0.4617	<.0001
8	0.4633	<.0001	0.5870	0.0539	0.0014	0.0339	0.0145	0.3404	<.0001
9	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
10		<.0001	0.2020	0.0078	<.0001	0.0043	0.0017	0.7376	<.0001
11	<.0001		<.0001	0.0001	0.0061	0.0003	0.0021	<.0001	<.0001
12	0.2020	<.0001		0.1666	0.0081	0.1150	0.0528	0.1585	<.0001
13	0.0078	0.0001	0.1666		0.2039	0.8489	0.5182	0.0101	<.0001
14	<.0001	0.0061	0.0081	0.2039		0.2788	0.5884	0.0003	<.0001
15	0.0043	0.0003	0.1150	0.8489	0.2788		0.6387	0.0062	<.0001
16	0.0017	0.0021	0.0528	0.5182	0.5884	0.6387		0.0027	<.0001
17	0.7376	<.0001	0.1585	0.0101	0.0003	0.0062	0.0027		<.0001
18	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	

G) $\delta^{13}C$

i/j	1	2	3	4	5	6	7	8	9
1		0.0015	<.0001	0.0063	<.0001	<.0001	<.0001	0.0087	<.0001
2	0.0015		0.0699	0.5320	0.0292	0.1109	0.2545	0.3303	<.0001
3	<.0001	0.0699		0.0083	0.6980	0.8961	0.4154	0.0012	<.0001
4	0.0063	0.5320	0.0083		0.0023	0.0188	0.0523	0.7443	<.0001
5	<.0001	0.0292	0.6980	0.0023		0.6228	0.2179	0.0002	<.0001
6	<.0001	0.1109	0.8961	0.0188	0.6228		0.5355	0.0042	<.0001
7	<.0001	0.2545	0.4154	0.0523	0.2179	0.5355		0.0129	<.0001
8	0.0087	0.3303	0.0012	0.7443	0.0002	0.0042	0.0129		<.0001
9	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	
10	0.0024	0.5843	0.0061	0.8969	0.0014	0.0162	0.0475	0.6179	<.0001

11	<.0001	<.0001	0.0092	<.0001	0.0209	0.0102	0.0008	<.0001	<.0001
12	0.0031	0.5270	0.0044	0.9683	0.0009	0.0124	0.0368	0.6880	<.0001
13	0.0335	0.1351	0.0001	0.3780	<.0001	0.0006	0.0018	0.5434	<.0001
14	0.5638	0.0022	<.0001	0.0102	<.0001	<.0001	<.0001	0.0139	<.0001
15	0.1757	0.0224	<.0001	0.0841	<.0001	<.0001	<.0001	0.1245	<.0001
16	0.3077	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
17	0.7330	0.0046	<.0001	0.0177	<.0001	<.0001	<.0001	0.0254	<.0001
18	0.0006	0.5333	0.3687	0.2305	0.2293	0.4487	0.7644	0.1269	<.0001
i/j	10	11	12	13	14	15	16	17	18
1	0.0024	<.0001	0.0031	0.0335	0.5638	0.1757	0.3077	0.7330	0.0006
2	0.5843	<.0001	0.5270	0.1351	0.0022	0.0224	<.0001	0.0046	0.5333
3	0.0061	0.0092	0.0044	0.0001	<.0001	<.0001	<.0001	<.0001	0.3687
4	0.8969	<.0001	0.9683	0.3780	0.0102	0.0841	<.0001	0.0177	0.2305
5	0.0014	0.0209	0.0009	<.0001	<.0001	<.0001	<.0001	<.0001	0.2293
6	0.0162	0.0102	0.0124	0.0006	<.0001	<.0001	<.0001	<.0001	0.4487
7	0.0475	0.0008	0.0368	0.0018	<.0001	<.0001	<.0001	<.0001	0.7644
8	0.6179	<.0001	0.6880	0.5434	0.0139	0.1245	<.0001	0.0254	0.1269
9	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
10		<.0001	0.9213	0.2677	0.0031	0.0418	<.0001	0.0081	0.2473
11	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0036
12	0.9213	<.0001		0.3112	0.0042	0.0520	<.0001	0.0101	0.2180
13	0.2677	<.0001	0.3112		0.0622	0.3503	0.0003	0.0824	0.0479
14	0.0031	<.0001	0.0042	0.0622		0.3498	0.0641	0.8517	0.0009
15	0.0418	<.0001	0.0520	0.3503	0.3498		0.0064	0.3352	0.0078
16	<.0001	<.0001	<.0001	0.0003	0.0641	0.0064		0.1640	<.0001
17	0.0081	<.0001	0.0101	0.0824	0.8517	0.3352	0.1640		0.0017
18	0.2473	0.0036	0.2180	0.0479	0.0009	0.0078	<.0001	0.0017	

Annexe 6. Test SNK réalisées sur les moyennes familiales pour *HT02*, *HT03*, *DC*, *PolyT*, *TB*, *NB* et $\delta^{13}C$ dans les 3 groupes de populations 4041, 4243 et 2124, ainsi qu'au sein de chacune des populations 40, 41, 42, 43, 21 et 24

A) Groupe de populations

4041

1) HT02

	Mean	N	Newfam
A	44.143	7	1900
A	44.000	7	29
A	42.000	7	2
A	41.429	7	500
A	40.857	7	800
A	38.857	7	3000
A	38.143	7	3
A	37.571	7	28
A	37.286	7	100
A	37.286	7	2000
A	36.714	7	1
A	36.429	7	25
A	36.143	7	1000
A	35.143	7	900
A	35.000	7	1100
A	34.571	7	200
A	34.571	7	14
A	34.429	7	9
A	34.286	7	18
A	34.286	7	4
A	34.286	7	7
A	34.286	7	17
A	33.857	7	2100

A	33.857	7	5
A	33.714	7	12
A	33.714	7	2700
A	33.714	7	600
A	33.714	7	24
A	33.667	6	1800
A	33.429	7	2600
A	33.143	7	2400
A	33.143	7	11
A	32.429	7	27
A	31.857	7	1600
A	31.714	7	26
A	31.429	7	1500
A	31.143	7	13
A	31.000	7	1200
A	30.286	7	16
A	30.143	7	10
A	29.857	7	1300
A	29.857	7	23
A	29.714	7	20
A	29.571	7	700
A	29.286	7	8
A	29.286	7	1400
A	28.714	7	300
A	28.667	6	400
A	28.286	7	2900
A	27.571	7	30

A	26.429	7	2800
A	26.143	7	2200
A	26.143	7	21
A	25.857	7	15
A	24.571	7	2300

2) HT03

Newfam	Mean	N	
A	110.286	7	1900
A	107.571	7	29
B	106.000	7	800
B	105.714	7	2
B	101.286	7	3000
B	98.571	7	3
B	97.429	7	100
B	96.571	7	500
B	95.571	7	1
B	95.286	7	2000
B	93.286	7	2400
B	93.000	7	17
B	92.857	7	16
B	91.714	7	900
B	90.714	7	200
B	90.000	7	27
B	89.500	6	1800
B	88.857	7	4
B	88.571	7	2600

B	A	88.286	7	18
B	A	87.714	7	1100
B	A	87.571	7	600
B	A	87.286	7	1000
B	A	87.286	7	28
B	A	87.286	7	7
B	A	87.000	7	25
B	A	85.000	7	10
B	A	84.857	7	20
B	A	84.857	7	2700
B	A	84.571	7	12
B	A	84.429	7	24
B	A	84.000	7	5
B	A	84.000	7	14
B	A	84.000	7	9
B	A	83.857	7	2100
B	A	83.429	7	11
B	A	83.429	7	1300
B	A	80.857	7	13
B	A	80.714	7	1500
B	A	79.857	7	2900
B	A	79.286	7	26
B	A	79.000	7	2800
B	A	78.714	7	700
B	A	78.286	7	30
B	A	78.167	6	400
B	A	78.000	7	1600

B	A	76.000	7	8
B	A			
B	A	75.857	7	23
B	A			
B	A	75.714	7	300
B	A			
B	A	75.286	7	2300
B	A			
B	A	74.857	7	1200
B				
B		71.857	7	1400
B				
B		71.857	7	21
B				
B		71.714	7	2200
B				
B		71.571	7	15

3) DC

		Mean	N	
	Newfam			
	A	27.290	7	1
	A			
	A	27.197	7	3000
	A			
B	A	26.890	7	800
B	A			
B	A	26.769	7	29
B	A			
B	A	26.034	7	2
B	A			
B	A	25.945	6	1800
B	A			
B	A	25.653	7	28
B	A			
B	A	25.547	7	500
B	A			
B	A	25.414	7	2400
B	A			
B	A	25.321	7	900
B	A			
B	A	25.208	7	3
B	A			
B	A	25.109	7	17
B	A			
B	A	24.936	7	25
B	A			
B	A	24.846	7	1900
B	A			
B	A	24.816	7	20
B	A			
B	A	24.796	7	18

B	A	24.485	7	2000
B	A			
B	A	24.436	7	4
B	A			
B	A	24.378	7	14
B	A			
B	A	24.266	7	100
B	A			
B	A	24.101	7	9
B	A			
B	A	24.037	7	24
B	A			
B	A	24.016	7	27
B	A			
B	A	24.004	7	2700
B	A			
B	A	23.792	7	1100
B	A			
B	A	23.129	7	600
B	A			
B	A	23.121	7	11
B	A			
B	A	22.512	7	26
B	A			
B	A	22.470	7	2600
B	A			
B	A	22.453	7	700
B	A			
B	A	22.442	7	2900
B	A			
B	A	22.410	7	16
B	A			
B	A	22.029	7	200
B	A			
B	A	22.017	7	1000
B	A			
B	A	21.995	7	12
B	A			
B	A	21.883	7	5
B	A			
B	A	21.631	7	15
B	A			
B	A	21.183	7	10
B	A			
B	A	20.890	7	13
B	A			
B	A	20.812	7	30
B	A			
B	A	20.757	7	1400
B	A			
B	A	20.753	7	8
B	A			
B	A	20.559	6	400
B	A			
B	A	20.485	7	2100
B	A			

B	A	20.229	7	1300
B	A			
B	A	20.194	7	7
B	A			
B	A	19.894	7	21
B	A			
B	A	19.836	7	2800
B	A			
B	A	19.763	7	2300
B	A			
B	A	19.660	7	1200
B	A			
B	A	19.654	7	2200
B	A			
B	A	19.449	7	1500
B	A			
B	A	19.126	7	1600
B	A			
B	A	18.822	7	23
B				
B		17.937	7	300

4) PolyT

	Mean	N	Newfam
A	1.82384	7	200
A			
A	1.81824	7	7
A			
A	1.81824	7	2000
A			
A	1.78806	7	24
A			
A	1.78806	7	100
A			
A	1.78806	7	27
A			
A	1.78806	7	2400
A			
A	1.78806	7	500
A			
A	1.78246	7	2600
A			
A	1.78246	7	1000
A			
A	1.78246	7	16
A			
A	1.78246	7	29
A			
A	1.78246	7	700
A			
A	1.77427	6	400
A			
A	1.74668	7	17
A			
A	1.74668	7	2900

A	1.74668	7	2
A			
A	1.74668	7	1100
A			
A	1.74668	7	2700
A			
A	1.74108	7	1200
A			
A	1.74108	7	30
A			
A	1.74108	7	900
A			
A	1.72598	6	1800
A			
A	1.70529	7	600
A			
A	1.70529	7	10
A			
A	1.70529	7	12
A			
A	1.70529	7	3000
A			
A	1.70529	7	23
A			
A	1.70529	7	25
A			
A	1.69969	7	28
A			
A	1.69969	7	3
A			
A	1.69576	7	1900
A			
A	1.69576	7	18
A			
A	1.66391	7	2300
A			
A	1.66391	7	800
A			
A	1.66391	7	9
A			
A	1.66391	7	2800
A			
A	1.66391	7	300
A			
A	1.66391	7	2100
A			
A	1.66391	7	2200
A			
A	1.66391	7	13
A			
A	1.65831	7	26
A			
A	1.64878	7	1400
A			
A	1.64878	7	1600

A	1.62252	7	1500
A	1.62252	7	20
A	1.62252	7	21
A	1.61299	7	15
A	1.61299	7	14
A	1.61299	7	1
A	1.61299	7	5
A	1.58114	7	4
A	1.58114	7	11
A	1.58114	7	8
A	1.52070	7	1300

B	335.86	7	100
B	334.87	7	17
B	333.07	7	3
B	329.06	7	600
B	326.94	7	27
B	326.69	7	1100
B	324.24	7	4
B	323.77	7	20
B	320.91	7	9
B	318.41	7	24
B	316.17	7	26
B	311.21	7	11
B	307.60	7	2600
B	303.90	7	14
B	300.69	7	700
B	294.49	7	1000
B	293.03	7	13
B	287.59	7	200
B	279.53	7	7
B	276.94	7	16
B	274.00	7	10
B	273.96	7	2900
B	265.37	7	15
B	264.21	7	5
B	254.01	7	1300
B	242.83	6	400
B	242.41	7	30
B	240.90	7	8
B	240.71	7	12

B	233.33	7	2100
B	228.54	7	1400
B	222.61	7	21
B	221.04	7	2300
B	213.11	7	1500
B	210.49	7	2800
B	210.24	7	1200
B	208.91	7	1600
B	198.73	7	23
B	193.80	7	300
B	189.80	7	2200

B	189.31	7	500
B	188.97	7	17
B	187.89	7	9
B	186.22	6	1800
B	185.63	7	24
B	185.10	7	27
B	184.60	7	20
B	184.53	7	100
B	184.23	7	11
B	184.20	7	2600
B	183.53	7	3
B	178.77	7	1900
B	178.20	7	4
B	177.41	7	700
B	173.71	7	13
B	168.63	7	1000
B	167.53	7	200
B	166.21	7	14
B	164.44	7	15
B	160.60	7	10
B	160.07	7	16
B	159.29	7	2900
B	158.37	7	7
B	152.01	7	5
B	149.51	7	1300
B	148.82	6	400
B	147.91	7	8
B	139.24	7	12

5) TB

	Mean	N
Newfam		
A	485.23	7
A	419.86	7
A	407.56	7
A	388.84	7
A	374.01	7
A	373.41	7
A	371.81	7
A	370.44	7
A	353.50	7
A	352.01	7
A	351.74	7
A	348.67	6
A	342.96	7
A	342.23	7
A	341.91	7

A	3000
A	29
A	2
A	2400
A	800
A	900
A	28
A	2000
A	18
A	1900
A	1800
A	25
A	500
A	2700

6) NB

Mean	N	Newfam
A	256.83	7
A	234.27	7
A	215.03	7
A	213.81	7
A	213.53	7
A	210.14	7
A	209.21	7
A	208.13	7
A	201.24	7
A	200.97	7
A	198.67	7
A	198.63	7
A	198.21	7
A	193.07	7
A	189.61	7

A	1
A	3000
A	2
A	900
A	2400
A	28
A	29
A	2700
A	600
A	2000
A	800
A	18
A	25
A	1100
A	26

B	A	137.67	7	30
B	A			
B	A	136.29	7	2100
B	A			
B	A	135.61	7	21
B	A			
B	A	131.99	7	1400
B	A			
B	A	130.33	7	2300
B	A			
B	A	127.76	7	1200
B	A			
B	A	124.76	7	1600
B	A			
B	A	123.41	7	2800
B	A			
B	A	121.17	7	1500
B	A			
B	A	119.49	7	23
B	A			
B	A	117.17	7	300
B	A			
B		115.36	7	2200

7) $\delta^{13}C$

	Mean	N	Newfam
A	-26.1098	7	29
A			
A	-26.2260	7	2800
A			
A	-26.3071	7	17
A			
A	-26.3247	7	1100
A			
A	-26.3580	7	1400
A			
A	-26.3789	7	2300
A			
A	-26.4132	7	21
A			
A	-26.4460	7	500
A			
A	-26.4511	7	1200
A			
A	-26.4667	7	2600
A			
A	-26.4684	7	3
A			
A	-26.4711	7	9
A			
A	-26.4717	7	200
A			
A	-26.4765	7	100
A			
A	-26.4792	7	20

A				
A	-26.4814	7	1000	
A				
A	-26.4846	7	2400	
A				
A	-26.4855	7	25	
A				
A	-26.4949	7	1500	
A				
A	-26.5037	7	1	
A				
A	-26.5160	7	27	
A				
A	-26.5560	7	8	
A				
A	-26.5641	7	7	
A				
A	-26.5860	7	1900	
A				
A	-26.5863	7	800	
A				
A	-26.6551	7	1300	
A				
A	-26.6566	7	26	
A				
A	-26.6607	7	14	
A				
A	-26.6829	7	30	
A				
A	-26.6929	7	12	
A				
A	-26.6949	7	1600	
A				
A	-26.6962	6	1800	
A				
A	-26.7332	7	2100	
A				
A	-26.7398	7	2700	
A				
A	-26.7857	7	3000	
A				
A	-26.7862	7	300	
A				
A	-26.7873	7	28	
A				
A	-26.7977	7	2200	
A				
A	-26.8108	7	18	
A				
A	-26.8624	7	23	
A				
A	-26.8645	7	900	
A				
A	-26.8889	7	4	
A				
A	-26.9035	6	400	
A				

A	-26.9055	7	24	
A				
A	-26.9427	7	2	
A				
A	-26.9766	7	10	
A				
A	-26.9838	7	5	
A				
A	-27.0167	7	2000	
A				
A	-27.0497	7	16	
A				
A	-27.0634	7	13	
A				
A	-27.0742	7	2900	
A				
A	-27.0804	7	700	
A				
A	-27.1160	7	11	
A				
A	-27.1198	7	600	
A				
A	-27.2356	7	15	

B) Groupe de populations

4243

B1) HT02

g	Mean	N	Newfam
A	41.857	7	2
A			
A	40.429	7	29
A			
A	39.429	7	600
A			
A	39.000	7	10
A			
A	38.857	7	21
A			
A	38.429	7	1600
A			
A	38.429	7	2500
A			
A	38.000	7	3000
A			
A	38.000	7	8
A			
A	37.857	7	1

A	37.571	7	18	
A				
A	37.143	7	26	
A				
A	36.833	6	900	
A				
A	36.286	7	1200	
A				
A	36.167	6	2700	
A				
A	35.857	7	16	
A				
A	35.714	7	2400	
A				
A	35.667	6	100	
A				
A	35.571	7	11	
A				
A	35.571	7	1900	
A				
A	35.000	7	200	
A				
A	34.714	7	27	
A				
A	34.429	7	20	
A				
A	34.286	7	19	
A				
A	33.857	7	6	
A				
A	33.143	7	14	
A				
A	33.000	7	22	
A				
A	32.714	7	13	
A				
A	32.714	7	1100	
A				
A	32.571	7	2900	
A				
A	32.571	7	1300	
A				
A	32.429	7	30	
A				
A	32.429	7	400	
A				
A	32.000	7	2800	
A				
A	32.000	7	300	
A				
A	31.857	7	5	
A				
A	31.571	7	2000	
A				
A	31.000	7	25	
A				
A	31.000	7	1700	

A			
A	30.000	7	28
A			
A	30.000	7	2100
A			
A	29.714	7	1000
A			
A	29.667	6	800
A			
A	29.429	7	1400
A			
A	28.857	7	9
A			
A	28.286	7	700
A			
A	28.143	7	24
A			
A	28.000	6	4
A			
A	27.857	7	12
A			
A	27.143	7	23
A			
A	27.000	7	15
A			
A	27.000	7	2200
A			
A	27.000	7	1800
A			
A	26.714	7	2300
A			
A	25.286	7	500
A			
A	24.714	7	3
A			
A	24.600	5	1500

B2) HT03

g	Mean	N	Newfam
A	102.429	7	1
A			
A	101.429	7	10
A			
A	100.857	7	29
A			
A	99.429	7	2500
A			
A	98.714	7	2
A			
A	96.857	7	30
A			
A	95.857	7	21
A			
A	94.857	7	600

A			
A	93.286	7	3000
A			
A	92.286	7	1600
A			
A	91.286	7	8
A			
A	90.833	6	900
A			
A	89.429	7	6
A			
A	89.143	7	1200
A			
A	89.143	7	22
A			
A	89.000	7	400
A			
A	88.857	7	25
A			
A	88.714	7	13
A			
A	88.571	7	18
A			
A	88.286	7	1900
A			
A	88.143	7	26
A			
A	87.286	7	27
A			
A	87.143	7	1300
A			
A	87.000	7	2000
A			
A	86.286	7	11
A			
A	85.714	7	20
A			
A	85.571	7	200
A			
A	85.286	7	14
A			
A	85.286	7	19
A			
A	84.857	7	2400
A			
A	84.833	6	2700
A			
A	83.857	7	1000
A			
A	83.833	6	800
A			
A	82.429	7	1100
A			
A	82.000	7	2300
A			
A	81.714	7	5
A			

A	81.143	7	300
A			
A	80.571	7	16
A			
A	80.571	7	1800
A			
A	80.429	7	2900
A			
A	80.143	7	2800
A			
A	79.000	7	15
A			
A	78.857	7	2200
A			
A	78.857	7	24
A			
A	78.833	6	100
A			
A	77.714	7	9
A			
A	77.571	7	1700
A			
A	76.429	7	1400
A			
A	75.429	7	3
A			
A	74.571	7	700
A			
A	74.333	6	4
A			
A	74.286	7	12
A			
A	74.143	7	28
A			
A	74.000	7	23
A			
A	73.857	7	2100
A			
A	70.143	7	500
A			
A	66.800	5	1500

B3) DC

g	Mean	N	Newfam
A	27.266	7	1600
A			
A	26.318	7	2
A			
A	25.746	7	1
A			
A	25.736	7	26
A			
A	25.487	7	1200

A	25.402	7	600
A			
A	25.162	7	29
A			
A	24.803	7	8
A			
A	24.709	7	10
A			
A	24.150	6	900
A			
A	24.069	7	2500
A			
A	23.843	7	21
A			
A	23.792	7	1300
A			
A	23.731	7	200
A			
A	23.619	7	3000
A			
A	23.618	6	2700
A			
A	23.478	7	400
A			
A	23.358	6	800
A			
A	23.251	7	19
A			
A	23.242	7	11
A			
A	23.239	7	1900
A			
A	23.124	7	1000
A			
A	23.009	7	2000
A			
A	22.819	7	2100
A			
A	22.768	7	1800
A			
A	22.598	7	27
A			
A	22.439	7	30
A			
A	22.376	7	20
A			
A	22.296	7	13
A			
A	22.269	7	6
A			
A	22.171	7	28
A			
A	22.097	7	5
A			
A	22.089	7	700
A			
A	22.066	7	1100

A			
A	22.006	7	18
A			
A	21.853	7	25
A			
A	21.726	7	22
A			
A	21.677	7	14
A			
A	21.495	7	2300
A			
A	21.470	7	2200
A			
A	21.296	7	2400
A			
A	21.024	7	15
A			
A	21.023	7	23
A			
A	21.008	7	1400
A			
A	20.977	7	2800
A			
A	20.871	7	2900
A			
A	20.722	5	1500
A			
A	20.701	6	4
A			
A	20.520	7	3
A			
A	20.452	6	100
A			
A	20.402	7	1700
A			
A	20.005	7	16
A			
A	19.860	7	12
A			
A	19.528	7	300
A			
A	19.277	7	9
A			
A	18.940	7	500
A			
A	18.904	7	24

B4) PolyT

g	Mean	N	Newfam
A	1.89541	7	13
A			
A	1.81824	7	400
A			
A	1.80884	7	1

A			
A	1.78806	7	29
A			
A	1.78246	7	25
A			
A	1.78246	7	1000
A			
A	1.78246	7	1800
A			
A	1.76733	7	27
A			
A	1.74668	7	5
A			
A	1.74668	7	24
A			
A	1.74668	7	21
A			
A	1.74668	7	2300
A			
A	1.74668	7	12
A			
A	1.74668	7	10
A			
A	1.74108	7	2500
A			
A	1.74108	7	2800
A			
A	1.74108	7	2200
A			
A	1.74108	7	300
A			
A	1.73548	7	26
A			
A	1.72598	6	4
A			
A	1.72598	6	900
A			
A	1.70529	7	2400
A			
A	1.70529	7	600
A			
A	1.70529	7	15
A			
A	1.70529	7	30
A			
A	1.70529	7	1600
A			
A	1.70529	7	500
A			
A	1.70529	7	8
A			
A	1.70529	7	3
A			
A	1.69969	7	2900
A			
A	1.69969	7	1700

A	1.69016	7	11
A			
A	1.66391	7	28
A			
A	1.66391	7	2
A			
A	1.66391	7	19
A			
A	1.66391	7	18
A			
A	1.66391	7	14
A			
A	1.66391	7	6
A			
A	1.66391	7	16
A			
A	1.66391	7	9
A			
A	1.66391	7	1300
A			
A	1.66391	7	20
A			
A	1.66391	7	700
A			
A	1.66391	7	22
A			
A	1.66391	7	2100
A			
A	1.66391	7	2000
A			
A	1.66391	7	3000
A			
A	1.66391	7	1200
A			
A	1.65438	7	1100
A			
A	1.62942	6	100
A			
A	1.61830	6	2700
A			
A	1.58114	6	800
A			
A	1.58114	7	23
A			
A	1.58114	5	1500
A			
A	1.57161	7	200
A			
A	1.57161	7	1900
A			
A	1.53023	7	1400

B5) TB

g	Mean	N	Newfam
---	------	---	--------

A	406.83	7	1600
A			
A	390.69	7	2
A			
A	382.44	7	29
A			
A	378.77	7	8
A			
A	372.26	7	600
A			
A	370.59	7	10
A			
A	364.24	7	1200
A			
A	361.94	7	1
A			
A	360.20	7	26
A			
A	334.97	7	3000
A			
A	331.87	7	21
A			
A	326.67	7	2500
A			
A	323.32	6	800
A			
A	317.50	7	5
A			
A	314.09	7	1300
A			
A	308.81	7	200
A			
A	305.93	7	2000
A			
A	299.60	6	900
A			
A	296.00	7	27
A			
A	295.73	6	2700
A			
A	295.07	7	700
A			
A	294.07	7	1900
A			
A	289.24	7	1800
A			
A	287.57	7	20
A			
A	286.16	7	19
A			
A	284.91	7	2200
A			
A	283.24	7	14
A			
A	280.54	7	2100
A			
A	279.46	7	13

A			
A	279.23	7	1100
A			
A	278.29	7	1000
A			
A	276.33	7	22
A			
A	274.31	7	400
A			
A	273.66	7	11
A			
A	272.11	7	2300
A			
A	268.56	7	1700
A			
A	268.26	7	30
A			
A	265.49	7	25
A			
A	264.81	7	28
A			
A	263.34	7	2900
A			
A	253.64	7	18
A			
A	252.94	7	2400
A			
A	248.00	5	1500
A			
A	241.26	7	6
A			
A	232.62	6	100
A			
A	228.63	7	3
A			
A	225.77	7	2800
A			
A	225.49	7	1400
A			
A	222.19	7	15
A			
A	222.03	7	23
A			
A	216.95	6	4
A			
A	211.43	7	16
A			
A	210.39	7	9
A			
A	197.83	7	500
A			
A	197.23	7	12
A			
A	196.29	7	300
A			
A	180.01	7	24

B6) NB

g	Mean	N	Newfam
A	226.01	7	1600
A			
A	218.03	7	8
A			
A	212.66	7	10
A			
A	210.93	7	600
A			
A	206.91	7	2
A			
A	204.56	7	1200
A			
A	202.01	7	29
A			
A	199.01	7	26
A			
A	194.74	7	3000
A			
A	190.90	7	1
A			
A	187.72	6	800
A			
A	186.49	7	21
A			
A	182.90	7	5
A			
A	181.87	7	2500
A			
A	181.77	7	700
A			
A	176.84	7	1300
A			
A	176.07	7	2000
A			
A	173.74	7	200
A			
A	173.43	6	900
A			
A	172.04	7	1900
A			
A	172.03	7	2200
A			
A	172.03	7	27
A			
A	171.37	7	1800
A			
A	169.78	6	2700
A			
A	168.30	7	20
A			
A	167.13	7	2100

A			
A	165.59	7	22
A			
A	164.46	7	14
A			
A	161.17	7	1100
A			
A	159.91	7	1700
A			
A	159.21	7	400
A			
A	158.91	7	2300
A			
A	158.77	7	1000
A			
A	158.54	7	19
A			
A	158.46	7	13
A			
A	158.03	7	28
A			
A	155.53	7	11
A			
A	154.37	7	2900
A			
A	150.10	5	1500
A			
A	149.66	7	30
A			
A	145.54	7	2400
A			
A	141.80	7	18
A			
A	139.77	7	25
A			
A	139.51	7	3
A			
A	138.86	7	6
A			
A	134.60	6	100
A			
A	132.53	7	15
A			
A	131.03	7	2800
A			
A	129.41	7	1400
A			
A	129.15	6	4
A			
A	127.61	7	23
A			
A	124.57	7	16
A			
A	123.37	7	9
A			
A	121.37	7	500
A			

A	110.81	7	300
A			
A	103.93	7	24
A			
A	97.91	7	12

B7) $\delta^{13}C$

g	Mean	N	Newfam
A	-25.8982	7	3
A			
A	-26.1245	7	300
A			
A	-26.1700	7	12
A			
A	-26.1856	7	25
A			
A	-26.1935	7	1
A			
A	-26.2640	7	11
A			
A	-26.2707	7	2500
A			
A	-26.3230	7	1800
A			
A	-26.3356	6	900
A			
A	-26.3558	7	13
A			
A	-26.3693	7	22
A			
A	-26.3815	7	18
A			
A	-26.4119	7	2200
A			
A	-26.4198	7	14
A			
A	-26.4555	7	15
A			
A	-26.4654	7	5
A			
A	-26.4994	7	1000
A			
A	-26.5034	7	500
A			
A	-26.5039	7	1300
A			
A	-26.5096	7	3000
A			
A	-26.5176	7	27
A			
A	-26.5307	7	1400
A			
A	-26.5309	6	4
A			

A	-26.5349	7	29
A	-26.5394	7	2300
A	-26.5414	7	6
A	-26.5444	7	16
A	-26.5446	7	30
A	-26.5448	7	9
A	-26.5510	5	1500
A	-26.5583	7	1100
A	-26.5590	7	2800
A	-26.5700	6	2700
A	-26.5798	7	2400
A	-26.6129	7	1900
A	-26.6177	7	20
A	-26.6238	7	600
A	-26.6288	7	2000
A	-26.6476	7	1200
A	-26.6711	7	2900
A	-26.6727	7	1600
A	-26.6894	7	400
A	-26.6972	7	21
A	-26.6987	7	2
A	-26.7033	6	800
A	-26.7145	7	19
A	-26.7285	7	28
A	-26.7388	7	2100
A	-26.7553	7	23
A	-26.7752	7	200
A	-26.8478	7	26
A	-26.9567	7	10

A	-26.9689	7	700
A	-26.9729	6	100
A	-27.0020	7	24
A	-27.0712	7	8
A	-27.0885	7	1700

C) Groupe de populations

2124

C1) HT02

SNK Grouping Newfam	Mean	N
A	44.714	7 7
A	42.000	6 20000
A	41.571	7 9000
A	41.333	6 2100
A	41.000	7 7000
A	40.143	14 2000
A	40.000	6 14
A	39.500	6 2200

A	39.000	7	1500	A	35.429	7	900
A				A			
A	39.000	12	3000	A	35.286	7	2
A				A			
A	38.857	7	8	A	35.143	7	17000
A				A			
A	38.000	7	400	A	35.000	7	2800
A				A			
A	37.143	7	16000	A	34.833	6	19
A				A			
A	36.667	6	2600	A	34.800	5	9
A				A			
A	36.571	7	800	A	34.333	6	17
A				A			
A	36.429	7	5000	A	34.000	7	2300
A				A			
A	36.286	7	8000	A	33.857	7	10000
A				A			
A	36.000	5	2400	A	33.857	7	11
A				A			
A	35.857	7	18000	A	33.714	7	4000
A				A			
A	35.714	7	1200	A	33.400	5	16
A				A			
A	35.667	6	1800	A	33.333	6	2700
A				A			
A	35.500	6	1700	A	32.800	5	4
A				A			

A	32.667	6	3	A	30.000	7	20	B A	93.714	7	1500	B A	85.429	7	2300
A				A				B A				B A			
A	32.200	5	5	A	29.000	7	12000	B A	92.857	7	7000	B A	85.400	5	16
A				A				B A				B A			
A	31.333	6	2500	A	28.571	7	100	B A	92.571	7	400	B A	84.833	6	19
A				A				B A				B A			
A	31.286	7	14000	A	28.429	7	300	B A	91.429	7	2800	B A	84.600	5	2400
A				A				B A				B A			
A	31.143	7	2900	A	27.571	7	12	B A	90.500	6		B A	83.500	6	2200
A				A				20000				B A			
A	30.857	7	700	A	27.500	6	18	B A				B A	82.000	7	900
A				A				B A	90.333	6	2100	B A			
A	30.857	7	200	A	26.286	7	1300	B A				B A			
A				A				B A	89.000	7	2	B A	81.857	7	
A	30.667	6	19000	A	25.571	7	1900	B A				B A			
A				A				B A	88.929	14	2000	B A	81.857	7	8000
A	30.667	6	1000	A	25.143	7	10	B A				B A			
A				A				B A	87.917	12	3000	B A	80.600	5	5
A	30.429	7	15000	A	25.000	6	1400	B A				B A			
A				A				B A	86.571	7	9000	B A	80.400	5	4
A	30.429	7	11000	A	24.571	7	13000	B A				B A			
A				A				B A	86.286	7	8	B A	80.333	6	14
A	30.429	7	6000	A	24.000	5	13	B A				B A			
A								B A	86.167	6	2600	B A	80.143	7	800
A	30.286	7	1100					B A				B A			
A								B A	85.857	7	4000	B A	79.857	7	
A	30.200	5	1					B A				B A			
A								B A	85.600	5	9	B A	79.857	7	700
								B A							

C2) HT03

SNK Grouping
Newfam

Mean N

A 100.000 7 7

A

B	A				B	A	75.000	5	1	B	A	69.714	7	20	A				
B	A	79.833	6	3	B	A				B	A				A	28.200	5	4	
B	A				B	A	74.714	7	200	B	A	69.000	7	1900	A				
B	A	79.714	7		B	A				B	A				A	28.083	6	1800	
16000					B	A	74.571	7	1300	B	A	66.571	7	10	A				
B	A				B	A				B	A				A	28.007	12	3000	
B	A	79.667	6	1700	B	A	74.429	7		B	A	64.714	7	300	A				
B	A				12000					B	A				A	27.769	7	1500	
B	A	79.286	7	2900	B	A				B	A	64.571	7		A				
B	A				B	A	73.857	7	5000	B	A	13000			A	27.697	7	9000	
B	A	79.167	6	2700	B	A				B					A				
B	A				B	A	73.286	7	11	B		62.833	6	1400	A	27.500	6	20000	
B	A	78.667	6	17	B	A				B					A				
B	A				B	A	73.143	7		B		58.800	5	13	A	27.500	6	3	
B	A	78.286	7		14000					C3) DC									
17000					B	A				SNK Grouping	Mean	N			A	27.456	7	400	
B	A				15000					Newfam					A				
B	A	78.000	6	2500	B	A				A	29.917	6	2200	A	27.429	7	2800		
B	A				B	A	72.667	6	1000	A				A					
B	A	76.571	7	1200	B	A				A	29.300	5	2400	A	27.326	7	1200		
B	A				B	A	72.286	7		A				A					
B	A	76.333	6	18	11000					A	29.214	7	7	A	27.214	7	17000		
B	A				B	A				A				A					
B	A	76.000	7	1100	B	A	71.429	7	100	A	29.083	6	2100	A	27.071	7	18000		
B	A				B	A				A				A					
B	A	75.833	6	1800	B	A	71.429	7	6000	A	28.603	14	2000	A	26.931	7	900		
B	A				B	A				A				A					
B	A	75.333	6		B	A	70.571	7	12	A				A					
19000					B	A				A	28.417	6	2600	A	26.571	7	8		
B	A																		

												<i>C4) PolyT</i>			
												SNK Grouping	Mean	N	
												Newfam			
A				A				A				A	1.78806	7	6000
A	26.535	7	7000	A	25.083	6	19	A	23.604	7	12000	A			
A				A				A				A			
A	26.500	6	14	A	24.714	7	20	A	23.600	5	1	A			
A				A				A				A			
A	26.417	6	18	A	24.700	5	16	A	23.571	7	2	A			
A				A				A				A	1.78246	7	10000
A	26.400	5	9	A	24.583	6	2500	A	23.429	7	2900	A			
A				A				A				A			
A	26.217	7	4000	A	24.567	7	200	A	23.104	7	6000	A	1.77427	6	2700
A				A				A				A			
A	26.083	6	17	A	24.535	7	1300	A	22.320	7	1900	A	1.75495	5	16
A				A				A				A			
A	26.049	7	11000	A	24.516	7	700	A	22.143	7	10	A	1.74711	5	2400
A				A				A				A			
A	25.932	6	1700	A	24.391	7	1100	A	22.000	6	19000	A	1.74108	7	2800
A				A				A				A			
A	25.895	7	15000	A	24.321	7	8000	A	21.786	7	12	A	1.74108	7	400
A				A				A				A			
A	25.667	6	1000	A	24.259	7	10000	A	21.627	7	5000	A	1.70529	7	700
A				A				A				A			
A	25.643	7	11	A	24.118	7	100	A	21.428	6	1400	A	1.69969	7	2300
A				A				A				A			
A	25.429	7	2300	A	23.917	6	2700	A	20.861	7	13000	A	1.69701	5	5
A				A				A				A			
A	25.418	7	16000	A	23.734	7	800	A	20.761	7	300	A	1.69701	5	4
A				A				A				A			
A	25.300	5	5	A	23.702	7	14000	A	20.600	5	13	A	1.68076	7	2900

A	1.67214	12	3000	A	1.62942	6	1400	A	1.58114	6	3	A	1.53023	7	800
A				A				A				A			
A	1.67117	6	20000	A	1.62252	7	9000	A	1.58114	7	20	A	1.52174	6	17
A				A				A				A			
A	1.66391	7	7000	A	1.62252	7	13000	A	1.58114	6	14	A	1.52174	6	2200
A				A				A				A			
A	1.66391	7	100	A	1.62252	7	1300	A	1.58114	7	15000	A	1.52174	6	2100
A				A				A				A			
A	1.66391	7	8000	A	1.62252	7	10	A	1.58114	5	9	A	1.47931	7	12
A				A				A				A			
A	1.66391	7	7	A	1.62252	7	1200	A	1.57161	7	14000	A	1.47931	7	17000
A				A				A				A			
A	1.66391	7	12000	A	1.62252	7	18000	A	1.57161	7	1100	A	1.47931	7	11000
A				A				A				A			
A	1.65831	7	5000	A	1.62252	7	16000	A	1.57161	7	1900	A	1.43858	5	13
A				A				A							
A	1.64878	7	4000	A	1.61776	14	2000	A	1.57161	7	8				
A				A				A							
A	1.63908	5	1	A	1.61299	7	1500	A	1.57161	7	11				
A				A				A							
A	1.62942	6	2600	A	1.61299	7	300	A	1.57002	6	1800				
A				A				A							
A	1.62942	6	19000	A	1.61299	7	2	A	1.57002	6	2500				
A				A				A							
A	1.62942	6	19	A	1.60346	7	900	A	1.57002	6	18				
A				A				A							
A	1.62942	6	1000	A	1.58114	6	1700	A	1.53023	7	200				
A				A				A							

C5) TB																
												SNK Grouping	Mean	N		
												Newfam				
												A	458.77	6	2100	
												A				
												B	A	453.97	6	2200
												B	A			
												B	A	445.95	6	1800
												B	A			
												B	A	435.31	7	7
												B	A			
												B	A	425.27	6	3

B A				B A	370.14	7	2300	B A				B A	284.16	7	1100
B A	423.33	7		B A				B A	339.21	7	11	B A			
17000				B A	369.87	7	400	B A				B A	280.10	7	800
B A				B A				B A	335.46	7		B A			
B A	419.74	5	9	B A	366.80	6	18	11000				B A	271.84	7	
B A				B A				B A				14000			
B A	416.76	5	2400	B A	366.30	6	14	B A	333.41	7	8000	B A			
B A				B A				B A				B A	271.20	7	6000
B A	410.11	14	2000	B A				B A	332.02	5	16	B A			
B A				B A	360.46	5	4	B A				B A	270.93	7	12
B A	400.52	12	3000	B A				B A	329.59	7	4000	B A			
B A				B A	357.27	7	7000	B A				B A	268.69	7	200
B A	399.80	7	1200	B A				B A	317.23	7	20	B A			
B A				16000	357.19	7		B A				B A	267.26	7	5000
B A	398.23	7	1500	B A				B A	315.91	7	2	B A			
B A				B A	354.37	7	900	B A				19000	265.45	6	
B A	390.09	7	2800	B A				B A	313.36	5	5	B A			
B A				B A	345.99	7		B A				B A	262.49	7	2900
B A	387.58	6	17	15000				B A	310.60	7	100	B A			
B A				B A				B A				B A	261.31	7	
B A	382.66	7		B A	345.90	7	8	B A	306.85	6	2500	12000			
18000				B A				B A				B A			
B A				B A	343.65	6	1700	B A				B A	260.96	7	700
B A	382.12	6	19	B A				10000	305.70	7		B A			
B A				B A	342.84	5	1	B A				B A	260.04	7	10
B A	373.27	6		B A				B A	297.93	6	1000	B A			
20000				B A	342.68	6	2600	B A				B A	258.79	7	1900
B A				B A				B A	291.44	7	1300	B A			
B A	371.46	7	9000	B A	342.15	6	2700	B A				B A	244.53	6	1400
B A															

SNK Grouping	Mean	N	SNK Grouping	Mean	N	SNK Grouping	Mean	N	SNK Grouping	Mean	N
B A			B A C	232.76	12	B A C	207.82	6	B A C	192.79	7
B A	240.36	5	3000			2700			7000		
B A			B A C			B A C			B A C		
B A	189.83	7	B A C	232.51	7	B A C	206.72	6	B A C	191.90	7
B			1200			17			8		
B			B A C			B A C			B A C		
B	167.70	7	B A C	229.32	6	B A C	202.43	6	B A C	191.76	7
13000			19			1700			11		
			B A C			B A C			B A C		
C6) NB			B A C	228.84	7	B A C	201.55	6	B A C	189.09	7
			17000			20000			4000		
SNK Grouping	Mean	N	B A C			B A C			B A C		
Newfam			B A C	223.19	7	B A C	200.74	7	B A C	186.33	7
A	261.45	6	B A C			B A C			B A C		
			18000			2300			11000		
A			B A C			B A C			B A C		
B A	257.20	6	B A C	222.62	14	B A C	198.57	7	B A C	186.11	7
1800			2000			1500			8000		
B A			B A C			B A C			B A C		
B A	254.25	6	B A C	219.40	6	B A C	198.01	7	B A C	185.50	5
2100			18			15000			16		
B A			B A C			B A C			B A C		
B A	244.42	6	B A C	211.57	7	B A C	196.34	5	B A C	184.41	7
3			2800			4			100		
B A			B A C			B A C			B A C		
B A	243.59	7	B A C	211.18	5	B A C	194.90	7	B A C	182.00	5
7			1			900			5		
B A			B A C			B A C			B A C		
B A	242.40	5	B A C	210.67	7	B A C	193.94	7	B A C	181.38	6
9			9000			16000			1000		
B A			B A C			B A C			B A C		
B A C	234.12	5	B A C	207.87	6	B A C	193.89	7	B A C	178.77	6
2400			14			400			2600		
B A C			B A C			B A C			B A C		

								SNK	Grouping		Mean									
								N	Newfam											
B	A	C	176.68	6	B	A	C	152.41	7				B	D	F	C	-26.6911	E		
2500					1900								7	400						
B	A	C			B	A	C			A	-25.3320	7	8	B	D	F	C		E	
B	A	C	174.60	7	B	A	C	151.40	7	A				B	D	F	C	-26.6961	E	
20					12000					B	A	-25.4795		14	2000					
B	A	C			B	A	C			7	7			B	D	F	C		E	
B	A	C	173.87	7	B	A	C	151.13	7	B	A			B	D	F	C	-26.7067	E	
1300					700					B	A	C	-25.5030	5	1					
B	A	C			B	A	C			B	A	C		B	D	F	C		E	
B	A	C	171.09	7	B	A	C	149.84	7	B	A	C		B	D	F	C	-26.7171	E	
2					800					B	D	A	C	-26.0641	7	2800				
B	A	C			B	A	C			B	D	A	C		B	D	F	C		
B	A	C	170.80	7	B	A	C	149.10	7	B	D	A	C		B	D	F	C	-26.7291	E
10000					10					B	D	A	C	-26.2263	7	900				
B	A	C			B	A	C			B	D	A	C		B	D	F	C		
B	A	C	169.03	7	B	A	C	146.80	5	B	D	C		B	D	F	C	-26.7315	E	
12					13					B	D	F	C	-26.3956	7	16000				
B	A	C			B	A	C			B	D	F	C		B	D	F	C		
B	A	C	164.37	7	B	A	C	145.31	7	B	D	F	C		B	D	F	C	-26.7340	E
1100					2900					B	D	F	C	-26.4250	6	1800				
B	A	C			B	A	C			B	D	F	C		B	D	F	C		
B	A	C	164.13	7	B	A	C	144.10	7	B	D	F	C		B	D	F	C	-26.7607	E
6000					200					B	D	F	C	-26.4393	7	200				
B	A	C			B	A	C			B	D	F	C		B	D	F	C		
B	A	C	160.59	7	B	A	C	144.02	6	B	D	F	C		B	D	F	C	-26.7612	E
5000					1400					B	D	F	C	-26.6023	5	5				
B	A	C			B		C			B	D	F	C		B	D	F	C		
B	A	C	159.48	6	B		C	110.27	7	B	D	F	C		B	D	F	C	-26.7619	E
19000					300					B	D	F	C	-26.6105	12	3000				
B	A	C			C					B	D	F	C		B	D	F	C		
B	A	C	156.40	7	C	89.43	7	13000		B	D	F	C		B	D	F	C	-26.7739	E
14000										B	D	F	C	-26.6474	6	2200				
B	A	C								B	D	F	C		B	D	F	C		
										7	18000									
										B	D	F	C							

C7) $\delta^{13}C$

B 6	D 20000	F C	-26.7750	E	D 5000	F C	-26.8542	E 7	D 2100	F C	-27.0668	E 6	D 18	F C	-27.3334	E 6
B 7	D 300	F C	-26.7766	E	D 2700	F C	-26.8579	E 6	D 13000	F C	-27.0906	E 7	D 13	F C	-27.4252	E 5
B 7	D 12000	F C	-26.7812	E	D 2400	F C	-26.8641	E 5	D 800	F C	-27.1454	E 7	D 16	F C	-27.5057	E 5
B 7	D 6000	F C	-26.7860	E	D 1300	F C	-26.9158	E 7	D 2300	F C	-27.1494	E 7	F		-27.5420	6
D 8000	F C		-26.8010	E 7	D 700	F C	-26.9257	E 7	D 4000	F C	-27.1643	E 7	F 11000			
D 100	F C		-26.8112	E 7	D 1000	F C	-26.9482	E 6	D 11	F C	-27.1677	E 7	F		-27.5668	7
D 2500	F C		-26.8134	E 6	D 15000	F C	-26.9666	E 7	D 2600	F C	-27.1685	E 6	F		-27.7461	6
D 19000	F C		-26.8233	E 6	D 1100	F C	-26.9668	E 7	D 17000	F C	-27.1872	E 7				
D 4	F C		-26.8249	E 5	D 1400	F C	-27.0456	E 6	D 1900	F C	-27.1886	E 7				
D 7000	F C		-26.8285	E 7	D 3	F C	-27.0479	E 6	D 1200	F C	-27.3150	E 7				
D 10000	F C		-26.8475	E 7	D 14000	F C	-27.0589	E 7	D 14	F C	-27.3312	E 6				
D	F	C		E	D	F		E	D	F		E				

D) Population 40			
D1) HT02			
g	Mean	N	fam
A	44.000	7	29
A	42.000	7	2
A	38.143	7	3
A	37.571	7	28
A	36.714	7	1
A	36.429	7	25

A	34.571	7	14
A			
A	34.429	7	9
A			
A	34.286	7	18
A			
A	34.286	7	4
A			
A	34.286	7	17
A			
A	34.286	7	7
A			
A	33.857	7	5
A			
A	33.714	7	12
A			
A	33.714	7	24
A			
A	33.143	7	11
A			
A	32.429	7	27
A			
A	31.714	7	26
A			
A	31.143	7	13
A			
A	30.286	7	16
A			
A	30.143	7	10
A			
A	29.857	7	23
A			
A	29.714	7	20
A			
A	29.286	7	8
A			
A	27.571	7	30
A			
A	26.143	7	21
A			
A	25.857	7	15

D2) HT03

g	Mean	N	fam
A	107.571	7	29
A			

A	105.714	7	2
A			
A	98.571	7	3
A			
A	95.571	7	1
A			
A	93.000	7	17
A			
A	92.857	7	16
A			
A	90.000	7	27
A			
A	88.857	7	4
A			
A	88.286	7	18
A			
A	87.286	7	28
A			
A	87.286	7	7
A			
A	87.000	7	25
A			
A	85.000	7	10
A			
A	84.857	7	20
A			
A	84.571	7	12
A			
A	84.429	7	24
A			
A	84.000	7	14
A			
A	84.000	7	9
A			
A	84.000	7	5
A			
A	83.429	7	11
A			
A	80.857	7	13
A			
A	79.286	7	26
A			
A	78.286	7	30
A			
A	76.000	7	8
A			
A	75.857	7	23
A			
A	71.857	7	21

D3) DC

g	Mean	N	fam
A	71.571	7	15
A			
A	27.290	7	1
A			
A	26.769	7	29
A			
A	26.034	7	2
A			
A	25.653	7	28
A			
A	25.208	7	3
A			
A	25.109	7	17
A			
A	24.936	7	25
A			
A	24.816	7	20
A			
A	24.796	7	18
A			
A	24.436	7	4
A			
A	24.378	7	14
A			
A	24.101	7	9
A			
A	24.037	7	24
A			
A	24.016	7	27
A			
A	23.121	7	11
A			
A	22.512	7	26
A			
A	22.410	7	16
A			
A	21.995	7	12
A			
A	21.883	7	5
A			
A	21.631	7	15
A			
A	21.183	7	10

D4) PolyT

g	Mean	N	fam
A	20.890	7	13
A			
A	20.812	7	30
A			
A	20.753	7	8
A			
A	20.194	7	7
A			
A	19.894	7	21
A			
A	18.822	7	23
A			
A	1.81824	7	7
A			
A	1.78806	7	24
A			
A	1.78806	7	27
A			
A	1.78246	7	16
A			
A	1.78246	7	29
A			
A	1.74668	7	17
A			
A	1.74668	7	2
A			
A	1.74108	7	30
A			
A	1.70529	7	10
A			
A	1.70529	7	12
A			
A	1.70529	7	23
A			
A	1.70529	7	25
A			
A	1.69969	7	28
A			
A	1.69969	7	3
A			
A	1.69576	7	18
A			
A	1.66391	7	13

A	1.66391	7	9
A			
A	1.65831	7	26
A			
A	1.62252	7	21
A			
A	1.62252	7	20
A			
A	1.61299	7	14
A			
A	1.61299	7	15
A			
A	1.61299	7	5
A			
A	1.61299	7	1
A			
A	1.58114	7	4
A			
A	1.58114	7	11
A			
A	1.58114	7	8

D5) TB

ng fam	Mean	N
1 A	485.23	7
B A	407.56	7
B A	388.84	7
B A	370.44	7
B A	352.01	7
B A	342.96	7
B A	334.87	7

B A	333.07	7
3 B A		
B A	326.94	7
27 B A		
B A	324.24	7
4 B A		
B A	323.77	7
20 B A		
B A	320.91	7
9 B A		
B A	318.41	7
24 B A		
B A	316.17	7
26 B A		
B A	311.21	7
11 B A		
B A	303.90	7
14 B A		
B A	293.03	7
13 B A		
B A	279.53	7
7 B A		
B A	276.94	7
16 B A		
B A	274.00	7
10 B A		
B A	265.37	7
15 B A		
B A	264.21	7
5 B A		
B A	242.41	7
30 B A		
B A	240.90	7
8		

B A	240.71	7
12 B A		
B A	222.61	7
21 B		
B	198.73	7
23		

D6) NB

g	Mean	N	fam
A	256.83	7	1
A	215.03	7	2
A	210.14	7	28
A	209.21	7	29
A	198.63	7	18
A	198.21	7	25
A	189.61	7	26
A	188.97	7	17
A	187.89	7	9
A	185.63	7	24
A	185.10	7	27
A	184.60	7	20
A	184.23	7	11
A	183.53	7	3
A	178.20	7	4
A	173.71	7	13
A	166.21	7	14

A	164.44	7	15
A			
A	160.60	7	10
A			
A	160.07	7	16
A			
A	158.37	7	7
A			
A	152.01	7	5
A			
A	147.91	7	8
A			
A	139.24	7	12
A			
A	137.67	7	30
A			
A	135.61	7	21
A			
A	119.49	7	23

D7) δ¹³C

g	Mean	N	fam
A	-26.1098	7	29
A	-26.3071	7	17
A	-26.4132	7	21
A	-26.4684	7	3
A	-26.4711	7	9
A	-26.4792	7	20
A	-26.4855	7	25
A	-26.5037	7	1
A	-26.5160	7	27
A	-26.5560	7	8
A	-26.5641	7	7
A	-26.6566	7	26

A	-26.6607	7	14
A			
A	-26.6829	7	30
A			
A	-26.6929	7	12
A			
A	-26.7873	7	28
A			
A	-26.8108	7	18
A			
A	-26.8624	7	23
A			
A	-26.8889	7	4
A			
A	-26.9055	7	24
A			
A	-26.9427	7	2
A			
A	-26.9766	7	10
A			
A	-26.9838	7	5
A			
A	-27.0497	7	16
A			
A	-27.0634	7	13
A			
A	-27.1160	7	11
A			
A	-27.2356	7	15

E) Population 41

E1) HT02

ng fam	Mean	N
19 A	44.143	7
19 A		
5 B A	41.429	7
5 B A		
8 B A	40.857	7
8 B A		

B A			
B A			
30 B A	38.857	7	
30 B A			
20 B A	37.286	7	
20 B A			
1 B A	37.286	7	
1 B A			
B A	36.143	7	
B A			
10 B A	35.143	7	
10 B A			
9 B A	35.000	7	
9 B A			
11 B A	35.000	7	
11 B A			
B A	34.571	7	
B A			
2 B A			
2 B A			
21 B A	33.857	7	
21 B A			
B A	33.714	7	
B A			
27 B A	33.714	7	
27 B A			
B A	33.667	6	
B A			
18 B A	33.429	7	
18 B A			
26 B A			
26 B A	33.143	7	
26 B A			
24 B A			
24 B A	31.857	7	
24 B A			
16 B A			
16 B A	31.429	7	
16 B A			
15 B A			
15 B A	31.000	7	
15 B A			
12 B A			
12 B A			

B A	29.857	7	
B A			
13 B A			
13 B A	29.571	7	
13 B A			
7 B A			
7 B A	29.286	7	
7 B A			
14 B A			
14 B A	28.714	7	
14 B A			
3 B A			
3 B A	28.667	6	
3 B A			
4 B A			
4 B A	28.286	7	
4 B A			
29 B A			
29 B A	26.429	7	
29 B A			
28 B A			
28 B A	26.143	7	
28 B A			
22 B			
22 B	24.571	7	
22 B			
23 B			
23 B			

E2) HT03

Grouping fam	Mean	N
A	110.286	7
A		
B A	106.000	7
B A		
B A C	101.286	7
B A C		
B A C	97.429	7
B A C		
B A C	96.571	7
B A C		
B A C	95.286	7
B A C		
B A C		

B A C	93.286	7	
B A C			
24 B A C			
24 B A C	91.714	7	
24 B A C			
9 B A C			
9 B A C	90.714	7	
9 B A C			
2 B A C			
2 B A C	89.500	6	
2 B A C			
18 B A C			
18 B A C	88.571	7	
18 B A C			
26 B A C			
26 B A C	87.714	7	
26 B A C			
11 B A C			
11 B A C	87.571	7	
11 B A C			
6 B A C			
6 B A C	87.286	7	
6 B A C			
10 B A C			
10 B A C	84.857	7	
10 B A C			
27 B A C			
27 B A C	83.857	7	
27 B A C			
21 B A C			
21 B A C	83.429	7	
21 B A C			
13 B A C			
13 B A C	80.714	7	
13 B A C			
15 B A C			
15 B A C	79.857	7	
15 B A C			
29 B C			
29 B C	79.000	7	
29 B C			
28 B C			
28 B C	78.714	7	
28 B C			
7 B C			
7 B C	78.167	6	
7 B C			
4 B C			
4 B C	78.000	7	
4 B C			
16 B C			
16 B C	75.714	7	
16 B C			
3 B C			
3 B C	75.286	7	
3 B C			
23 B			
23 B			

B	C	74.857	7	
12				
	C			
	C	71.857	7	
14				
	C			
	C	71.714	7	
22				
E3) DC				
Grouping		Mean	N	
fam				
	A	27.197	7	
30				
	A			
B	A	26.890	7	
8				
B	A			
B	A	25.945	6	
18				
B	A			
B	A	C	25.547	7
5				
B	A	C		
B	A	C	25.414	7
24				
B	A	C		
B	A	C	25.321	7
9				
B	A	C		
B	A	C	24.846	7
19				
B	A	C		
B	A	C	24.485	7
20				
B	A	C		
B	A	C	24.266	7
1				
B	A	C		
B	A	C	24.004	7
27				
B	A	C		
B	A	C	23.792	7
11				
B	A	C		
B	A	C	23.129	7
6				
B	A	C		
B	A	C	22.470	7
26				
B	A	C		

B	A	C	22.453	7
7				
B	A	C		
B	A	C	22.442	7
29				
B	A	C		
B	A	C	22.029	7
2				
B	A	C		
B	A	C	22.017	7
10				
B	A	C		
B	A	C	20.757	7
14				
B	A	C		
B	A	C	20.559	6
4				
B	A	C		
B	A	C	20.485	7
21				
B	A	C		
B	A	C	20.229	7
13				
B	A	C		
B	A	C	19.836	7
28				
B	A	C		
B	A	C	19.763	7
23				
B	A	C		
B	A	C	19.660	7
12				
B	A	C		
B	A	C	19.654	7
22				
B	A	C		
B	A	C	19.449	7
15				
B		C		
B		C	19.126	7
16				
		C		
		C	17.937	7
3				
E4) PolyT				
g	Mean	N	fam	
A	1.82384	7	2	
A				
A	1.81824	7	20	
A				
A	1.78806	7	1	

A					
A	1.78806	7	5		
A					
A	1.78806	7	24		
A					
A	1.78246	7	26		
A					
A	1.78246	7	10		
A					
A	1.78246	7	7		
A					
A	1.77427	6	4		
A					
A	1.74668	7	11		
A					
A	1.74668	7	29		
A					
A	1.74668	7	27		
A					
A	1.74108	7	12		
A					
A	1.74108	7	9		
A					
A	1.72598	6	18		
A					
A	1.70529	7	6		
A					
A	1.70529	7	30		
A					
A	1.69576	7	19		
A					
A	1.66391	7	23		
A					
A	1.66391	7	8		
A					
A	1.66391	7	22		
A					
A	1.66391	7	28		
A					
A	1.66391	7	3		
A					
A	1.66391	7	21		
A					
A	1.64878	7	14		
A					
A	1.64878	7	16		
A					
A	1.62252	7	15		
A					
A	1.52070	7	13		

A	419.86	7	30	
A				
B	A	374.01	7	24
B	A			
B	A	373.41	7	8
B	A			
B	A	371.81	7	9
B	A			
B	A	353.50	7	20
B	A			
B	A	351.74	7	19
B	A			
B	A	348.67	6	18
B	A			
B	A	342.23	7	5
B	A			
B	A	341.91	7	27
B	A			
B	A	335.86	7	1
B	A			
B	A	329.06	7	6
B	A			
B	A	326.69	7	11
B	A			
B	A	307.60	7	26
B	A			
B	A	300.69	7	7
B	A			
B	A	294.49	7	10
B	A			
B	A	287.59	7	2
B	A			
B	A	273.96	7	29
B	A			
B	A	254.01	7	13
B	A			
B	A	242.83	6	4
B	A			
B	A	233.33	7	21
B	A			
B	A	228.54	7	14
B	A			
B	A	221.04	7	23
B	A			
B	A	213.11	7	15
B	A			
B	A	210.49	7	28
B	A			
B	A	210.24	7	12
B	A			
B	A	208.91	7	16
B				
B		193.80	7	3
B				
B		189.80	7	22

E5) TB				
ng	Mean	N	fam	

E6) NB

ng	Mean	N	fam
A	234.27	7	30
A			
B A	213.81	7	9
B A			
B A	213.53	7	24
B A			
B A	208.13	7	27
B A			
B A	201.24	7	6
B A			
B A	200.97	7	20
B A			
B A	198.67	7	8
B A			
B A	193.07	7	11
B A			
B A	189.31	7	5
B A			
B A	186.22	6	18
B A			
B A	184.53	7	1
B A			
B A	184.20	7	26
B A			
B A	178.77	7	19
B A			
B A	177.41	7	7
B A			
B A	168.63	7	10
B A			
B A	167.53	7	2
B A			
B A	159.29	7	29
B A			
B A	149.51	7	13
B A			
B A	148.82	6	4
B A			
B A	136.29	7	21
B A			
B A	131.99	7	14
B A			
B A	130.33	7	23
B A			
B A	127.76	7	12
B A			
B A	124.76	7	16
B A			
B A	123.41	7	28

B A			
B A	121.17	7	15
B			
B			
B	117.17	7	3
B			
B			
B	115.36	7	22

E7) $\delta^{13}C$

g	Mean	N	fam
A	-26.2260	7	28
A			
A	-26.3247	7	11
A			
A	-26.3580	7	14
A			
A	-26.3789	7	23
A			
A	-26.4460	7	5
A			
A	-26.4511	7	12
A			
A	-26.4667	7	26
A			
A	-26.4717	7	2
A			
A	-26.4765	7	1
A			
A	-26.4814	7	10
A			
A	-26.4846	7	24
A			
A	-26.4949	7	15
A			
A	-26.5860	7	19
A			
A	-26.5863	7	8
A			
A	-26.6551	7	13
A			
A	-26.6949	7	16
A			
A	-26.6962	6	18
A			
A	-26.7332	7	21
A			
A	-26.7398	7	27
A			
A	-26.7857	7	30
A			
A	-26.7862	7	3
A			
A	-26.7977	7	22

A			
A	-26.8645	7	9
A			
A			
A	-26.9035	6	4
A			
A	-27.0167	7	20
A			
A	-27.0742	7	29
A			
A	-27.0804	7	7
A			
A	-27.1198	7	6

F) Population 42

F1) HT02

g	Mean	N	fam
A	41.857	7	2
A			
A	40.429	7	29
A			
A	39.000	7	10
A			
A	38.857	7	21
A			
A	38.000	7	8
A			
A	37.857	7	1
A			
A	37.571	7	18
A			
A	37.143	7	26
A			
A	35.857	7	16
A			
A	35.571	7	11
A			
A	34.714	7	27
A			
A	34.429	7	20
A			
A	34.286	7	19
A			
A	33.857	7	6
A			
A	33.143	7	14
A			
A	33.000	7	22
A			
A	32.714	7	13
A			
A	32.429	7	30

A			
A	31.857	7	5
A			
A			
A	31.000	7	25
A			
A	30.000	7	28
A			
A	28.857	7	9
A			
A	28.143	7	24
A			
A	28.000	6	4
A			
A	27.857	7	12
A			
A	27.143	7	23
A			
A	27.000	7	15
A			
A	24.714	7	3

F2) HT03

g	Mean	N	fam
A	102.429	7	1
A			
A	101.429	7	10
A			
A	100.857	7	29
A			
A	98.714	7	2
A			
A	96.857	7	30
A			
A	95.857	7	21
A			
A	91.286	7	8
A			
A	89.429	7	6
A			
A	89.143	7	22
A			
A	88.857	7	25
A			
A	88.714	7	13
A			
A	88.571	7	18
A			
A	88.143	7	26
A			
A	87.286	7	27
A			
A	86.286	7	11

A			
A	85.714	7	20
A			
A	85.286	7	14
A			
A	85.286	7	19
A			
A	81.714	7	5
A			
A	80.571	7	16
A			
A	79.000	7	15
A			
A	78.857	7	24
A			
A	77.714	7	9
A			
A	75.429	7	3
A			
A	74.333	6	4
A			
A	74.286	7	12
A			
A	74.143	7	28
A			
A	74.000	7	23

F3) DC

g	Mean	N	fam
A	102.429	7	1
A			
A	101.429	7	10
A			
A	100.857	7	29
A			
A	98.714	7	2
A			
A	96.857	7	30
A			
A	95.857	7	21
A			
A	91.286	7	8
A			
A	89.429	7	6
A			
A	89.143	7	22
A			
A	88.857	7	25
A			
A	88.714	7	13
A			
A	88.571	7	18

A			
A	88.143	7	26
A			
A	87.286	7	27
A			
A	86.286	7	11
A			
A	85.714	7	20
A			
A	85.286	7	14
A			
A	85.286	7	19
A			
A	81.714	7	5
A			
A	80.571	7	16
A			
A	79.000	7	15
A			
A	78.857	7	24
A			
A	77.714	7	9
A			
A	75.429	7	3
A			
A	74.333	6	4
A			
A	74.286	7	12
A			
A	74.143	7	28
A			
A	74.000	7	23

F4) PolyT

g	Mean	N	fam
A	1.89541	7	13
A			
A	1.80884	7	1
A			
A	1.78806	7	29
A			
A	1.78246	7	25
A			
A	1.76733	7	27
A			
A	1.74668	7	24
A			
A	1.74668	7	5
A			
A	1.74668	7	21
A			
A	1.74668	7	12

A			
A	1.74668	7	10
A			
A	1.73548	7	26
A			
A	1.72598	6	4
A			
A	1.70529	7	15
A			
A	1.70529	7	30
A			
A	1.70529	7	8
A			
A	1.70529	7	3
A			
A	1.69016	7	11
A			
A	1.66391	7	28
A			
A	1.66391	7	2
A			
A	1.66391	7	18
A			
A	1.66391	7	14
A			
A	1.66391	7	6
A			
A	1.66391	7	19
A			
A	1.66391	7	9
A			
A	1.66391	7	16
A			
A	1.66391	7	20
A			
A	1.66391	7	22
A			
A	1.58114	7	23

F5) TB

g	Mean	N	fam
A	390.69	7	2
A			
A	382.44	7	29
A			
A	378.77	7	8
A			
A	370.59	7	10
A			
A	361.94	7	1
A			
A	360.20	7	26

A			
A	331.87	7	21
A			
A	317.50	7	5
A			
A	296.00	7	27
A			
A	287.57	7	20
A			
A	286.16	7	19
A			
A	283.24	7	14
A			
A	279.46	7	13
A			
A	276.33	7	22
A			
A	273.66	7	11
A			
A	268.26	7	30
A			
A	265.49	7	25
A			
A	264.81	7	28
A			
A	253.64	7	18
A			
A	241.26	7	6
A			
A	228.63	7	3
A			
A	222.19	7	15
A			
A	222.03	7	23
A			
A	216.95	6	4
A			
A	211.43	7	16
A			
A	210.39	7	9
A			
A	197.23	7	12
A			
A	180.01	7	24

F6) NB

g	Mean	N	fam
A	218.03	7	8
A			
A	212.66	7	10
A			
A	206.91	7	2

A			
A	202.01	7	29
A			
A	199.01	7	26
A			
A	190.90	7	1
A			
A	186.49	7	21
A			
A	182.90	7	5
A			
A	172.03	7	27
A			
A	168.30	7	20
A			
A	165.59	7	22
A			
A	164.46	7	14
A			
A	158.54	7	19
A			
A	158.46	7	13
A			
A	158.03	7	28
A			
A	155.53	7	11
A			
A	149.66	7	30
A			
A	141.80	7	18
A			
A	139.77	7	25
A			
A	139.51	7	3
A			
A	138.86	7	6
A			
A	132.53	7	15
A			
A	129.15	6	4
A			
A	127.61	7	23
A			
A	124.57	7	16
A			
A	123.37	7	9
A			
A	103.93	7	24
A			
A	97.91	7	12

F7) $\delta^{13}C$

g Mean N fam

A	-25.8982	7	3
A			
A	-26.1700	7	12
A			
A	-26.1856	7	25
A			
A	-26.1935	7	1
A			
A	-26.2640	7	11
A			
A	-26.3558	7	13
A			
A	-26.3693	7	22
A			
A	-26.3815	7	18
A			
A	-26.4198	7	14
A			
A	-26.4555	7	15
A			
A	-26.4654	7	5
A			
A	-26.5176	7	27
A			
A	-26.5309	6	4
A			
A	-26.5349	7	29
A			
A	-26.5414	7	6
A			
A	-26.5444	7	16
A			
A	-26.5446	7	30
A			
A	-26.5448	7	9
A			
A	-26.6177	7	20
A			
A	-26.6972	7	21
A			
A	-26.6987	7	2
A			
A	-26.7145	7	19
A			
A	-26.7285	7	28
A			
A	-26.7553	7	23
A			
A	-26.8478	7	26
A			
A	-26.9567	7	10
A			
A	-27.0020	7	24
A			
A	-27.0712	7	8

G) Population 43

G1) HT02

g	Mean	N	fam
A	39.429	7	6
A			
A	38.429	7	25
A			
A	38.429	7	16
A			
A	38.000	7	30
A			
A	36.833	6	9
A			
A	36.286	7	12
A			
A	36.167	6	27
A			
A	35.714	7	24
A			
A	35.667	6	1
A			
A	35.571	7	19
A			
A	35.000	7	2
A			
A	32.714	7	11
A			
A	32.571	7	13
A			
A	32.571	7	29
A			
A	32.429	7	4
A			
A	32.000	7	3
A			
A	32.000	7	28
A			
A	31.571	7	20
A			
A	31.000	7	17
A			
A	30.000	7	21
A			
A	29.714	7	10

A			
A	29.667	6	8
A			
A	29.429	7	14
A			
A	28.286	7	7
A			
A	27.000	7	22
A			
A	27.000	7	18
A			
A	26.714	7	23
A			
A	25.286	7	5
A			
A	24.600	5	15

G2) HT03

g	Mean	N	fam
A	99.429	7	25
A			
A	94.857	7	6
A			
A	93.286	7	30
A			
A	92.286	7	16
A			
A	90.833	6	9
A			
A	89.143	7	12
A			
A	89.000	7	4
A			
A	88.286	7	19
A			
A	87.143	7	13
A			
A	87.000	7	20
A			
A	85.571	7	2
A			
A	84.857	7	24
A			
A	84.833	6	27
A			
A	83.857	7	10
A			
A	83.833	6	8
A			
A	82.429	7	11
A			
A	82.000	7	23

A			
A	81.143	7	3
A			
A	80.571	7	18
A			
A	80.429	7	29
A			
A	80.143	7	28
A			
A	78.857	7	22
A			
A	78.833	6	1
A			
A	77.571	7	17
A			
A	76.429	7	14
A			
A	74.571	7	7
A			
A	73.857	7	21
A			
A	70.143	7	5
A			
A	66.800	5	15

G3) DC

g	Mean	N	fam
A	27.266	7	16
A			
A	25.487	7	12
A			
A	25.402	7	6
A			
A	24.150	6	9
A			
A	24.069	7	25
A			
A	23.792	7	13
A			
A	23.731	7	2
A			
A	23.619	7	30
A			
A	23.618	6	27
A			
A	23.478	7	4
A			
A	23.358	6	8
A			
A	23.239	7	19
A			
A	23.124	7	10

A			
A	23.009	7	20
A			
A	22.819	7	21
A			
A	22.768	7	18
A			
A	22.089	7	7
A			
A	22.066	7	11
A			
A	21.495	7	23
A			
A	21.470	7	22
A			
A	21.296	7	24
A			
A	21.008	7	14
A			
A	20.977	7	28
A			
A	20.871	7	29
A			
A	20.722	5	15
A			
A	20.452	6	1
A			
A	20.402	7	17
A			
A	19.528	7	3
A			
A	18.940	7	5

G4) PolyT

g	Mean	N	fam
A	1.81824	7	4
A			
A	1.78246	7	18
A			
A	1.78246	7	10
A			
A	1.74668	7	23
A			
A	1.74108	7	25
A			
A	1.74108	7	28
A			
A	1.74108	7	3
A			
A	1.74108	7	22
A			
A	1.72598	6	9

A			
A	1.70529	7	24
A			
A	1.70529	7	5
A			
A	1.70529	7	16
A			
A	1.70529	7	6
A			
A	1.69969	7	17
A			
A	1.69969	7	29
A			
A	1.66391	7	7
A			
A	1.66391	7	13
A			
A	1.66391	7	20
A			
A	1.66391	7	21
A			
A	1.66391	7	12
A			
A	1.66391	7	30
A			
A	1.65438	7	11
A			
A	1.62942	6	1
A			
A	1.61830	6	27
A			
A	1.58114	6	8
A			
A	1.58114	5	15
A			
A	1.57161	7	2
A			
A	1.57161	7	19
A			
A	1.53023	7	14

G5) TB

g	Mean	N	fam
A	406.83	7	16
A			
A	372.26	7	6
A			
A	364.24	7	12
A			
A	334.97	7	30
A			
A	326.67	7	25

A			
A	323.32	6	8
A			
A	314.09	7	13
A			
A	308.81	7	2
A			
A	305.93	7	20
A			
A	299.60	6	9
A			
A	295.73	6	27
A			
A	295.07	7	7
A			
A	294.07	7	19
A			
A	289.24	7	18
A			
A	284.91	7	22
A			
A	280.54	7	21
A			
A	279.23	7	11
A			
A	278.29	7	10
A			
A	274.31	7	4
A			
A	272.11	7	23
A			
A	268.56	7	17
A			
A	263.34	7	29
A			
A	252.94	7	24
A			
A	248.00	5	15
A			
A	232.62	6	1
A			
A	225.77	7	28
A			
A	225.49	7	14
A			
A	197.83	7	5
A			
A	196.29	7	3

G6) NB

g	Mean	N	fam
A	226.01	7	16

A	210.93	7	6
A	204.56	7	12
A	194.74	7	30
A	187.72	6	8
A	181.87	7	25
A	181.77	7	7
A	176.84	7	13
A	176.07	7	20
A	173.74	7	2
A	173.43	6	9
A	172.04	7	19
A	172.03	7	22
A	171.37	7	18
A	169.78	6	27
A	167.13	7	21
A	161.17	7	11
A	159.91	7	17
A	159.21	7	4
A	158.91	7	23
A	158.77	7	10
A	154.37	7	29
A	150.10	5	15
A	145.54	7	24
A	134.60	6	1
A	131.03	7	28
A	129.41	7	14
A	121.37	7	5
A	110.81	7	3

G7) $\delta^{13}C$

g	Mean	N	fam
A	-26.1245	7	3
A	-26.2707	7	25
A	-26.3230	7	18
A	-26.3356	6	9
A	-26.4119	7	22
A	-26.4994	7	10
A	-26.5034	7	5
A	-26.5039	7	13
A	-26.5096	7	30
A	-26.5307	7	14
A	-26.5394	7	23
A	-26.5510	5	15
A	-26.5583	7	11
A	-26.5590	7	28
A	-26.5700	6	27
A	-26.5798	7	24
A	-26.6129	7	19
A	-26.6238	7	6
A	-26.6288	7	20
A	-26.6476	7	12
A	-26.6711	7	29
A	-26.6727	7	16
A	-26.6894	7	4
A	-26.7033	6	8
A	-26.7388	7	21

A	-26.7752	7	2
A	-26.9689	7	7
A	-26.9729	6	1
A	-27.0885	7	17

H) Population 21

H1) HT02

SNK Grouping	Mean	N					
fam							
A	41.333	6	21	A	35.500	6	17
A				A			
A	39.500	6	22	A	35.429	7	9
A				A			
A	39.000	7	15	A	35.000	7	28
A				A			
A	38.000	7	4	A	34.000	7	23
A				A			
A	36.667	6	26	A	33.333	6	27
A				A			
A	36.571	7	8	A	33.200	5	30
A				A			
A	36.000	7	20	A	31.333	6	25
A				A			
A	36.000	5	24	A	31.143	7	29
A				A			
A	35.714	7	12	A	30.857	7	7
A				A			
A	35.667	6	18	A	30.667	6	10
A				A			
				A	30.286	7	11
				A			
				A	28.571	7	1
				A			
				A	28.429	7	3

A			
A	26.286	7	13
A			
A	25.571	7	19
A			
A	25.000	6	14

H2) HT03

SNK Grouping fam	Mean		N
A	93.714	7	15
A			
A	92.571	7	4
A			
A	91.429	7	28
A			
A	90.333	6	21
A			
A	86.167	6	26
A			
A	86.000	7	20
A			
A	85.429	7	23
A			
A	84.600	5	24
A			
A	83.800	5	30
A			
A	83.500	6	22

A			
A	82.000	7	9
A			
A	80.143	7	8
A			
A	79.857	7	7

A			
A	79.667	6	17
A			
A	79.286	7	29
A			
A	79.167	6	27
A			
A	78.000	6	25
A			
A	76.571	7	12
A			
A	76.000	7	11
A			
A	75.833	6	18
A			
A	74.714	7	2
A			
A	74.571	7	13
A			
A	72.667	6	10
A			
A	71.429	7	1

A			
A	69.000	7	19
A			
A	64.714	7	3
A			
A	62.833	6	14

H3) DC

SNK Grouping fam	Mean		N
A	29.917	6	22
A			
A	29.300	5	24
A			
A	29.083	6	21
A			
A	28.417	6	26
A			
A	28.200	5	30
A			
A	28.083	6	18
A			
A	27.769	7	15
A			
A	27.456	7	4
A			
A	27.429	7	28
A			
A	27.326	7	12

A			
A	26.989	7	20
A			
A	26.931	7	9
A			
A	25.932	6	17
A			
A	25.667	6	10
A			
A	25.429	7	23
A			
A	24.583	6	25
A			
A	24.567	7	2
A			
A	24.535	7	13
A			
A	24.516	7	7
A			
A	24.391	7	11
A			
A	24.118	7	1
A			
A	23.917	6	27
A			
A	23.734	7	8
A			
A	23.429	7	29

A			
A	22.320	7	19
A			
A	21.428	6	14
A			
A	20.761	7	3

H4) PolyT

SNK Grouping fam	Mean	N	
A	1.7743	6	27
A			
A	1.7471	5	24
A			
A	1.7411	7	28
A			
A	1.7411	7	4
A			
A	1.7053	7	7
A			
A	1.6997	7	23
A			
A	1.6808	7	29
A			
A	1.6639	7	1
A			
A	1.6294	6	26
A			

A	1.6294	6	10
A			
A	1.6294	6	14
A			
A	1.6257	5	30
A			
A	1.6225	7	13
A			
A	1.6225	7	12
A			
A	1.6225	7	20
A			
A	1.6130	7	15
A			
A	1.6130	7	3
A			
A	1.6035	7	9
A			
A	1.5811	6	17
A			
A	1.5716	7	19
A			
A	1.5716	7	11
A			
A	1.5700	6	25
A			
A	1.5700	6	18
A			

A	1.5302	7	2
A			
A	1.5302	7	8
A			
A	1.5217	6	21
A			
A	1.5217	6	22

H5) TB

SNK Grouping fam	Mean	N	
A	458.77	6	21
A			
A	453.97	6	22
A			
A	445.95	6	18
A			
A	432.28	5	30
A			
B A	416.76	5	24
B A			
B A	399.80	7	12
B A			
B A	398.23	7	15
B A			
B A	390.09	7	28
B A			
B A	386.46	7	20
B A			

B A	370.14	7	23
B A			
B A	369.87	7	4
B A			
B A	354.37	7	9
B A			
B A	343.65	6	17
B A			
B A	342.68	6	26
B A			
B A	342.15	6	27
B A			
B A	310.60	7	1
B A			
B A	306.85	6	25
B A			
B A	297.93	6	10
B A			
B A	291.44	7	13
B A			
B A	284.16	7	11
B A			
B A	280.10	7	8
B A			
B A	268.69	7	2
B A			
B A	262.49	7	29
B A			

B	A	260.96	7	7
B	A			
B	A	258.79	7	19
B	A			
B	A	244.53	6	14
B				
B		189.83	7	3

H6) NB

SNK Grouping fam	Mean	N
A	276.70	5 30
A		
B 22	261.45	6
B		
B 18	257.20	6
B		
B 21	254.25	6
B		
B 24	234.12	5
B		
B 12	232.51	7
B		
B 20	218.26	7
B		

B 28	A	C	211.57	7
B	A	C		
B 27	A	C	207.82	6
B	A	C		
B 17	A	C	202.43	6
B	A	C		
B 23	A	C	200.74	7
B	A	C		
B 15	A	C	198.57	7
B	A	C		
B 9	A	C	194.90	7
B	A	C		
B 4	A	C	193.89	7
B	A	C		
B 1	A	C	184.41	7
B	A	C		
B 10	A	C	181.38	6
B	A	C		
B 26	A	C	178.77	6
B	A	C		
B 25	A	C	176.68	6
B	A	C		

B 13	A	C	173.87	7
B	A	C		
B 11	A	C	164.37	7
B		C		
B 19		C	152.41	7
B		C		
B 7		C	151.13	7
B		C		
B 8		C	149.84	7
B		C		
B 29		C	145.31	7
B		C		
B 2		C	144.10	7
B		C		
B 14		C	144.02	6
C				
C	110.27	7 3		

H7) $\delta^{13}C$

SNK Grouping fam	Mean	N
A	-26.0641	7 28
A		
B A	-26.4250	6 18
B		
A		

B A	-26.4902	7 20
B A		
B A	-26.6105	7 15
B A		
B A	-26.6911	7 4
B A		
B A	-26.7291	7 9
B A		
B A	-26.7315	6 17
B A		
B A	-26.7607	7 2
B A		
B A	-26.7739	6 22
B A		
B A	-26.7766	7 3
B A		
B A	-26.8112	7 1
B A		
B A	-26.8134	6 25
B A		
B A	-26.8579	6 27
B A		
B A	-26.8641	5 24
B A		
B A	-26.9158	7 13
B A		
B A	-26.9257	7 7
B A		

B	A	-26.9482	6	10
B	A			
B	A	-26.9668	7	11
B	A			
B	A	-27.0415	5	30
B	A			
B	A	-27.0456	6	14
B	A			
B	A	-27.0668	6	21
B	A			
B	A	-27.1454	7	8
B	A			
B	A	-27.1494	7	23
B	A			
B	A	-27.1685	6	26
B	A			
B	A	-27.1886	7	19
B	A			
B	A	-27.3150	7	12
B				
B		-27.5668	7	29

I) Population 24

II) HT02

SNK Grouping fam	Mean	N
------------------	------	---

A	44.286	7	2
A			
A	43.143	7	3
A			
A	42.000	6	20
A			
A	41.571	7	9
A			
A	41.000	7	7
A			
A	37.143	7	16
A			
A	36.429	7	5
A			
A	36.286	7	8
A			
A	35.857	7	18
A			
A	35.143	7	17
A			
A	33.857	7	10
A			
A	33.714	7	4
A			
A	31.286	7	14
A			
A	30.667	6	19
A			

I2) HT03

SNK Grouping fam	Mean	N
A	30.429	7 15
A		
A	30.429	7 11
A		
A	30.429	7 6
A		
A	29.000	7 12
A		
A	24.571	7 13
A		
A	92.857	7 7
A		
A	91.857	7 2
A		
A	90.857	7 3
A		
A	90.500	6 20
A		
A	86.571	7 9
A		
A	85.857	7 4
A		
A	81.857	7 18
A		
A	81.857	7 8
A		

I3) DC

SNK Grouping fam	Mean	N
A	79.857	7 10
A		
A	79.714	7 16
A		
A	78.286	7 17
A		
A	75.333	6 19
A		
A	74.429	7 12
A		
A	73.857	7 5
A		
A	73.143	7 14
A		
A	72.714	7 15
A		
A	72.286	7 11
A		
A	71.429	7 6
A		
A	64.571	7 13
A		
A	30.216	7 2
A		
A	27.869	7 3
A		

A	27.697	7	9
A			
A	27.500	6	20
A			
A	27.214	7	17
A			
A	27.071	7	18
A			
A	26.535	7	7
A			
A	26.217	7	4
A			
A	26.049	7	11
A			
A	25.895	7	15
A			
A	25.418	7	16
A			
A	24.321	7	8
A			
A	24.259	7	10
A			
A	23.702	7	14
A			
A	23.604	7	12
A			
A	23.104	7	6
A			
A			

A	22.000	6	19
A			
A	21.627	7	5
A			
A	20.861	7	13
A			
I4) PolyT			
	SNK Grouping	Mean	N
	fam		
A	1.78806	7	6
A			
A	1.78246	7	10
A			
A	1.70529	7	3
A			
A	1.67117	6	20
A			
A	1.66391	7	7
A			
A	1.66391	7	8
A			
A	1.66391	7	12
A			
A	1.65831	7	5
A			
A	1.64878	7	4
A			
A	1.62942	6	19

A	1.62252	7	9
A			
A	1.62252	7	13
A			
A	1.62252	7	18
A			
A	1.62252	7	16
A			
A	1.61299	7	2
A			
A	1.58114	7	15
A			
A	1.57161	7	14
A			
A	1.47931	7	17
A			
A	1.47931	7	11
A			
I5) TB			
	SNK Grouping	Mean	N
	fam		
A	433.76	7	2
A			
B A	423.33	7	17
B A			
B A	382.66	7	18
B A			

B A	377.83	7	3
B A			
B A	373.27	6	20
B A			
B A	371.46	7	9
B A			
B A	357.27	7	7
B A			
B A	357.19	7	16
B A			
B A	345.99	7	15
B A			
B A	335.46	7	11
B A			
B A	333.41	7	8
B A			
B A	329.59	7	4
B A			
B A	305.70	7	10
B A			
B A	271.84	7	14
B A			
B A	271.20	7	6
B A			
B A	267.26	7	5
B A			
B A	265.45	6	19
B A			

B	A	261.31	7	12
B				
B		167.70	7	13

I6) NB

SNK Grouping	fam	Mean		N
A		228.84	7	17
A				
A		226.99	7	2
A				
A		223.19	7	18
A				
B	A	210.67	7	9
B	A			
B	A	201.55	6	20
B	A			
B	A	201.37	7	3
B	A			
B	A	198.01	7	15
B	A			
B	A	193.94	7	16
B	A			
B	A	192.79	7	7
B	A			
B	A	189.09	7	4
B	A			
B	A	186.33	7	11
B	A			

B	A	186.11	7	8
B	A			
B	A	170.80	7	10
B	A			
B	A	164.13	7	6
B	A			
B	A	160.59	7	5
B	A			
B	A	159.48	6	19
B	A			
B	A	156.40	7	14
B	A			
B	A	151.40	7	12
B				
B		89.43	7	13

I7) $\delta^{13}C$

SNK Grouping	fam	Mean		N
A		-26.3956	7	16
A				
A		-26.5623	7	3
A				
A		-26.6474	7	18
A				
A		-26.7340	7	9
A				
A		-26.7750	6	20
A				

A		-26.7812	7	12
A				
A		-26.7860	7	6
A				
A		-26.8010	7	8
A				
A		-26.8233	6	19
A				
A		-26.8285	7	7
A				
A		-26.8475	7	10
A				
A		-26.8542	7	5
A				
A		-26.9019	7	2
A				
A		-26.9666	7	15
A				
A		-27.0589	7	14
A				
A		-27.0906	7	13
A				
A		-27.1643	7	4
A				
A		-27.1872	7	17
A				
A		-27.5431	7	11

**Annexe 7. Résultats d'ANOVA par
population pour *HT02, HT03, DC, PolyT,*
TB, NB et $\delta^{13}C$**

Pour $\delta^{13}C$, l'ANOVA est également testée au sein de chaque type d'analyse.

Population	Caractère	Origine de l'analyse (a)	SSQ _{err} (b)	df _{err} (c)	MS _{err} (d)	SSQ _{fam} (b)	df _{fam} (c)	MS _{fam} (d)	p _{fam} (e)	SSQ _{bloc} (b)	df _{bloc} (c)	MS _{bloc} (d)	SSQ _{tot} (b)	R _{err} (f)	R _{fam} (f)	R _{bloc} (f)	CoefDet (g)
15	HTO2	total	10282.350	82	125.395	3970.500	14	283.607	0.01164	1908.310	6	318.052	16161.17	0.636	0.246	0.118	0.364
20	HTO2	total	6614.440	85	77.817	2769.900	17	162.935	0.01407	1926.100	6	321.017	11310.44	0.585	0.245	0.170	0.415
2124	HTO2	total	9783.900	142	68.901	3292.670	26	126.641	0.01323	4147.330	7	592.476	17223.9	0.568	0.191	0.241	0.432
22	HTO2	total	6572.540	109	60.299	1309.700	19	68.932	0.32021	2080.250	6	346.708	9962.49	0.660	0.131	0.209	0.340
23	HTO2	total	7810.940	160	48.818	4013.650	28	143.345	0.00001	3521.870	6	586.978	15346.46	0.509	0.262	0.229	0.491
25	HTO2	total	10076.420	158	63.775	1739.210	28	62.115	0.50902	1907.250	6	317.875	13722.87	0.734	0.127	0.139	0.266
26	HTO2	total	8055.970	155	51.974	2048.620	28	73.165	0.09906	2095.480	7	299.354	12200.07	0.660	0.168	0.172	0.340
27	HTO2	total	9833.340	97	101.375	5689.190	19	299.431	0.00026	2979.170	7	425.596	18501.7	0.531	0.308	0.161	0.469
29	HTO2	total	5060.260	154	32.859	2291.720	26	88.143	0.0001	2073.550	6	345.592	9425.53	0.537	0.243	0.220	0.463
30	HTO2	total	733.080	88	8.330	404.280	17	23.781	0.00072	190.910	6	31.818	1328.26	0.552	0.304	0.144	0.448
4041	HTO2	total	26678.420	322	82.852	7515.440	54	139.175	0.00354	7845.670	6	1307.612	42039.54	0.635	0.179	0.187	0.365
4243	HTO2	total	24819.130	329	75.438	7121.480	56	127.169	0.00292	5932.410	6	988.735	37873.01	0.655	0.188	0.157	0.345
44	HTO2	total	8787.710	123	71.445	1641.400	21	78.162	0.36328	3386.150	6	564.358	13815.26	0.636	0.119	0.245	0.364
46	HTO2	total	7269.000	82	88.646	1699.150	14	121.368	0.18729	2618.400	6	436.400	11586.55	0.627	0.147	0.226	0.373
50	HTO2	total	3688.000	54	68.296	810.800	9	90.089	0.24902	1337.140	6	222.857	5835.94	0.632	0.139	0.229	0.368
15	HTO3	total	33173.190	82	404.551	7850.610	14	560.758	0.17875	2349.040	6	391.507	43372.85	0.765	0.181	0.054	0.235
20	HTO3	total	18236.950	85	214.552	8991.310	17	528.901	0.00345	4692.030	6	782.005	31920.29	0.571	0.282	0.147	0.429
2124	HTO3	total	34798.030	142	245.057	11002.400	26	423.169	0.0235	8056.520	7	1150.931	53856.95	0.646	0.204	0.150	0.354
22	HTO3	total	17833.500	109	163.610	6529.860	19	343.677	0.00902	5555.010	6	925.835	29918.38	0.596	0.218	0.186	0.404
23	HTO3	total	35363.710	160	221.023	17339.760	28	619.277	0.00003	6365.050	6	1060.842	59068.52	0.599	0.294	0.108	0.401
25	HTO3	total	41047.850	158	259.797	10621.910	28	379.354	0.07696	4139.320	6	689.887	55809.08	0.736	0.190	0.074	0.265
26	HTO3	total	32232.910	155	207.954	9102.330	28	325.083	0.04655	5397.160	7	771.023	46732.4	0.690	0.195	0.115	0.310
27	HTO3	total	32758.580	97	337.717	14333.220	19	754.380	0.00569	4996.670	7	713.810	52088.48	0.629	0.275	0.096	0.371
29	HTO3	total	27473.340	154	178.398	13103.350	26	503.975	0.00004	6796.980	6	1132.830	47373.67	0.580	0.277	0.143	0.420
30	HTO3	total	9855.330	88	111.992	4734.980	17	278.528	0.00304	1365.080	6	227.513	15955.39	0.618	0.297	0.086	0.382
4041	HTO3	total	93493.120	322	290.351	32993.840	54	610.997	0.00004	13516.360	6	2252.727	140003.31	0.668	0.236	0.097	0.332
4243	HTO3	total	93393.890	329	283.872	24850.540	56	443.760	0.0095	11733.010	6	1955.502	129977.43	0.719	0.191	0.090	0.281
44	HTO3	total	27913.160	123	226.936	7565.310	21	360.253	0.06279	7032.120	6	1172.020	42510.6	0.657	0.178	0.165	0.343
46	HTO3	total	30505.010	82	372.012	6665.100	14	476.079	0.23788	4703.370	6	783.895	41873.48	0.729	0.159	0.112	0.272
50	HTO3	total	13766.910	54	254.943	2368.990	9	263.221	0.42686	2243.940	6	373.990	18379.84	0.749	0.129	0.122	0.251
15	DC	total	2137.240	82	26.064	758.630	14	54.188	0.02132	126.270	6	21.045	3022.14	0.707	0.251	0.042	0.293

20	DC	total	1862.180	85	21.908	538.860	17	31.698	0.13558	376.870	6	62.812	2777.91	0.670	0.194	0.136	0.330
2124	DC	total	2846.730	142	20.047	1017.760	26	39.145	0.00718	195.140	7	27.877	4059.63	0.701	0.251	0.048	0.299
22	DC	total	1898.700	109	17.419	544.110	19	28.637	0.05808	121.000	6	20.167	2563.81	0.741	0.212	0.047	0.259
23	DC	total	3018.610	160	18.866	1440.710	28	51.454	0.00004	89.002	6	14.834	4548.32	0.664	0.317	0.020	0.336
25	DC	total	3258.700	158	20.625	1170.350	28	41.798	0.00353	71.013	6	11.836	4500.07	0.724	0.260	0.016	0.276
26	DC	total	3083.410	155	19.893	832.690	28	29.739	0.0654	311.807	7	44.544	4227.9	0.729	0.197	0.074	0.271
27	DC	total	2150.420	97	22.169	978.830	19	51.517	0.00389	149.176	7	21.311	3278.43	0.656	0.299	0.046	0.344
29	DC	total	2706.410	154	17.574	1150.780	26	44.261	0.00026	224.225	6	37.371	4081.42	0.663	0.282	0.055	0.337
30	DC	total	1195.960	88	13.590	509.670	17	29.981	0.00897	49.223	6	8.204	1754.85	0.682	0.290	0.028	0.318
4041	DC	total	5784.580	322	17.965	2253.140	54	41.725	0	262.041	6	43.674	8299.75	0.697	0.271	0.032	0.303
4243	DC	total	5934.960	329	18.039	1405.980	56	25.107	0.04219	311.866	6	51.978	7652.81	0.776	0.184	0.041	0.224
44	DC	total	2079.780	123	16.909	480.860	21	22.898	0.15502	146.825	6	24.471	2707.46	0.768	0.178	0.054	0.232
46	DC	total	1447.290	82	17.650	460.220	14	32.873	0.04293	190.954	6	31.826	2098.47	0.690	0.219	0.091	0.310
50	DC	total	1034.570	54	19.159	167.300	9	18.589	0.4746	445.406	6	74.234	1647.28	0.628	0.102	0.270	0.372
15	PolyT	total	2.690	82	0.033	0.280	14	0.020	0.85077	0.105	6	0.018	3.08	0.875	0.091	0.034	0.125
20	PolyT	total	1.700	85	0.020	0.570	17	0.034	0.0637	0.292	6	0.049	2.56	0.664	0.222	0.114	0.336
2124	PolyT	total	4.960	142	0.035	0.910	26	0.035	0.47357	0.276	7	0.039	6.14	0.807	0.148	0.045	0.193
22	PolyT	total	2.510	109	0.023	0.650	19	0.034	0.10917	0.251	6	0.042	3.41	0.737	0.189	0.074	0.263
23	PolyT	total	4.810	160	0.030	1.010	28	0.036	0.23716	0.292	6	0.049	6.11	0.787	0.166	0.048	0.213
25	PolyT	total	4.250	158	0.027	1.350	28	0.048	0.01363	0.287	6	0.048	5.89	0.722	0.229	0.049	0.278
26	PolyT	total	4.410	155	0.028	0.520	28	0.019	0.90592	0.468	7	0.067	5.4	0.817	0.097	0.087	0.183
27	PolyT	total	4.040	97	0.042	1.200	19	0.063	0.09637	0.426	7	0.061	5.67	0.713	0.212	0.075	0.287
29	PolyT	total	2.780	154	0.018	0.900	26	0.035	0.00838	0.222	6	0.037	3.9	0.713	0.230	0.057	0.287
30	PolyT	total	4.080	88	0.046	0.780	17	0.046	0.48046	0.234	6	0.039	5.09	0.801	0.153	0.046	0.199
4041	PolyT	total	9.880	322	0.031	1.940	54	0.036	0.20799	0.730	6	0.122	12.56	0.787	0.154	0.058	0.213
4243	PolyT	total	10.070	329	0.031	1.780	56	0.032	0.40502	0.780	6	0.130	12.63	0.797	0.141	0.061	0.203
44	PolyT	total	2.740	123	0.022	0.620	21	0.030	0.16745	0.137	6	0.023	3.5	0.783	0.178	0.039	0.217
46	PolyT	total	2.200	82	0.027	0.740	14	0.053	0.0301	0.210	6	0.035	3.15	0.697	0.235	0.068	0.303
50	PolyT	total	1.690	54	0.031	0.600	9	0.067	0.04119	0.460	6	0.077	2.75	0.614	0.219	0.166	0.386
15	TB	total	2592506.870	82	31615.937	529072.530	14	37790.895	0.29489	68557.400	6	11426.233	3190136.8	0.813	0.166	0.021	0.187
20	TB	total	1428962.870	85	16811.328	328160.650	17	19303.568	0.32458	294807.420	6	49134.570	2051930.94	0.696	0.160	0.144	0.304
2124	TB	total	1921832.050	142	13534.029	894255.130	26	34394.428	0.00026	192663.990	7	27523.427	3008751.17	0.639	0.297	0.064	0.361
22	TB	total	1089129.320	109	9992.012	333316.980	19	17542.999	0.03753	94784.620	6	15797.437	1517230.92	0.718	0.220	0.062	0.282
23	TB	total	1823566.090	160	11397.288	1079818.410	28	38564.943	0	81173.980	6	13528.997	2984558.49	0.611	0.362	0.027	0.389
25	TB	total	2467100.960	158	15614.563	1138679.390	28	40667.121	0.0001	44289.020	6	7381.503	3650069.37	0.676	0.312	0.012	0.324

26	TB	total	2333544.460	155	15055.126	741656.840	28	26487.744	0.0165	298840.940	7	42691.563	3374042.23	0.692	0.220	0.089	0.308
27	TB	total	2188145.800	97	22558.204	1027410.890	19	54074.257	0.00285	132382.130	7	18911.733	3347938.83	0.654	0.307	0.040	0.346
29	TB	total	2127342.190	154	13813.910	747945.090	26	28767.119	0.00326	142090.360	6	23681.727	3017377.63	0.705	0.248	0.047	0.295
30	TB	total	827813.920	88	9406.976	310828.650	17	18284.038	0.02398	86444.110	6	14407.352	1225086.68	0.676	0.254	0.071	0.324
4041	TB	total	4766255.780	322	14802.037	1544535.560	54	28602.510	0.00027	412924.250	6	68820.708	6723715.59	0.709	0.230	0.061	0.291
4243	TB	total	4109803.460	329	12491.804	1150685.950	56	20547.963	0.00437	371794.110	6	61965.685	5632283.52	0.730	0.204	0.066	0.270
44	TB	total	1340127.020	123	10895.342	453804.250	21	21609.726	0.01115	213313.130	6	35552.188	2007244.4	0.668	0.226	0.106	0.332
46	TB	total	1254127.780	82	15294.241	596344.370	14	42596.026	0.00198	174840.020	6	29140.003	2025312.17	0.619	0.294	0.086	0.381
50	TB	total	1363118.400	54	25242.933	201698.880	9	22410.987	0.54209	358588.540	6	59764.757	1923405.83	0.709	0.105	0.186	0.291
15	NB	total	696968.560	82	8499.617	148259.690	14	10589.978	0.2596	17643.610	6	2940.602	862871.85	0.808	0.172	0.020	0.192
20	NB	total	352343.120	85	4145.213	101729.340	17	5984.079	0.13698	81785.850	6	13630.975	535858.31	0.658	0.190	0.153	0.342
2124	NB	total	511573.140	142	3602.628	289572.490	26	11137.403	0.00001	43113.040	7	6159.006	844258.66	0.606	0.343	0.051	0.394
22	NB	total	347046.710	109	3183.915	126741.420	19	6670.601	0.00924	25919.240	6	4319.873	499707.37	0.695	0.254	0.052	0.306
23	NB	total	579141.270	160	3619.633	346106.490	28	12360.946	0	15290.700	6	2548.450	940538.46	0.616	0.368	0.016	0.384
25	NB	total	796134.980	158	5038.829	373197.560	28	13328.484	0.00008	16893.170	6	2815.528	1186225.71	0.671	0.315	0.014	0.329
26	NB	total	766015.050	155	4942.033	257792.470	28	9206.874	0.00926	101352.040	7	14478.863	1125159.56	0.681	0.229	0.090	0.319
27	NB	total	599499.390	97	6180.406	251885.390	19	13257.126	0.00824	39160.430	7	5594.347	890545.21	0.673	0.283	0.044	0.327
29	NB	total	763548.790	154	4958.109	266690.960	26	10257.345	0.00352	38982.330	6	6497.055	1069222.09	0.714	0.249	0.036	0.286
30	NB	total	310077.370	88	3523.606	114197.190	17	6717.482	0.02749	32597.920	6	5432.987	456872.48	0.679	0.250	0.071	0.321
4041	NB	total	1327421.430	322	4122.427	394145.020	54	7298.982	0.00145	107918.450	6	17986.408	1829484.9	0.726	0.215	0.059	0.274
4243	NB	total	1172973.020	329	3565.268	329991.810	56	5892.711	0.00404	95811.450	6	15968.575	1598776.28	0.734	0.206	0.060	0.266
44	NB	total	355064.900	123	2886.707	143784.380	21	6846.875	0.00178	62386.830	6	10397.805	561236.11	0.633	0.256	0.111	0.367
46	NB	total	320719.120	82	3911.209	164024.980	14	11716.070	0.00097	45730.620	6	7621.770	530474.73	0.605	0.309	0.086	0.395
50	NB	total	410083.280	54	7594.135	51569.530	9	5729.948	0.6579	122223.640	6	20370.607	583876.46	0.702	0.088	0.209	0.298
15	$\delta^{13}C$	anFR	31.915	82	0.389	3.320	14	0.237	0.85013	6.756	6	1.126	41.991	0.760	0.079	0.161	0.240
20	$\delta^{13}C$	anUSA	33.045	85	0.389	57.856	17	3.403	0	6.158	6	1.026	97.059	0.340	0.596	0.063	0.660
2124	$\delta^{13}C$	total	58.910	142	0.415	15.803	26	0.608	0.08291	6.929	7	0.990	81.642	0.722	0.194	0.085	0.278
		anFR	25.429	78	0.326	4.599	14	0.328	0.4547	3.419	6	0.570	33.447	0.760	1.345	0.102	0.240
		anUSA	31.584	56	0.564	10.803	13	0.831	0.1568	5.177	7	0.740	47.564	0.664	2.087	0.109	0.336
22	$\delta^{13}C$	total	37.528	109	0.344	29.702	19	1.563	0	8.480	6	1.413	75.71	0.496	0.392	0.112	0.504
		anFR	22.377	77	0.291	6.210	14	0.444	0.1217	5.726	6	0.954	34.312	0.652	1.084	0.167	0.348
		anUSA	11.087	24	0.462	17.385	6	2.898	0.0005	5.310	6	0.885	33.782	0.328	3.274	0.157	0.672
23	$\delta^{13}C$	total	57.990	160	0.362	21.977	28	0.785	0.00151	8.138	6	1.356	88.105	0.658	0.249	0.092	0.342
		anFR	26.136	84	0.311	4.084	14	0.292	0.5232	3.247	6	0.541	33.467	0.781	1.258	0.097	0.219
		anUSA	30.032	70	0.429	17.963	13	1.382	0.0008	6.713	6	1.119	54.708	0.549	2.676	0.123	0.451

25	$\delta^{13}C$	total	51.596	158	0.327	29.370	28	1.049	0	4.606	6	0.768	85.572	0.603	0.343	0.054	0.397
		anFR	31.088	82	0.379	4.321	14	0.309	0.6518	5.112	6	0.852	40.520	0.767	0.845	0.126	0.233
		anUSA	18.756	70	0.268	25.042	13	1.926	<.0001	1.247	6	0.208	45.045	0.416	20.088	0.028	0.584
26	$\delta^{13}C$	total	55.682	155	0.359	17.057	28	0.609	0.02331	12.235	7	1.748	84.974	0.655	0.201	0.144	0.345
		anFR	30.901	82	0.377	8.356	14	0.597	0.1013	4.984	6	0.831	44.241	0.698	1.677	0.113	0.302
		anUSA	21.819	67	0.326	9.036	13	0.695	0.0228	10.212	7	1.459	41.068	0.531	0.885	0.249	0.469
27	$\delta^{13}C$	anUSA	33.278	97	0.343	15.989	19	0.842	0.00225	2.960	7	0.423	52.228	0.637	0.306	0.057	0.363
29	$\delta^{13}C$	total	49.501	154	0.321	17.114	26	0.658	0.00396	6.531	6	1.089	73.146	0.677	0.234	0.089	0.323
		anFR	28.935	83	0.349	8.401	14	0.600	0.0665	4.092	6	0.682	41.428	0.698	2.053	0.099	0.302
		anUSA	19.614	65	0.302	8.652	11	0.787	0.0082	3.391	6	0.565	31.657	0.620	2.552	0.107	0.380
30	$\delta^{13}C$	total	31.324	88	0.356	6.449	17	0.379	0.39942	1.441	6	0.240	39.214	0.799	0.164	0.037	0.201
		anFR	16.837	44	0.383	4.489	8	0.561	0.1971	0.496	6	0.083	21.822	0.772	9.050	0.023	0.228
		anUSA	12.909	38	0.340	1.884	8	0.235	0.695	2.522	6	0.420	17.316	0.746	0.747	0.146	0.254
4041	$\delta^{13}C$	anFR	97.490	322	0.303	25.297	54	0.468	0.0121	18.315	6	3.052	141.102	0.691	0.179	0.130	0.309
4243	$\delta^{13}C$	anFR	116.023	329	0.353	21.487	56	0.384	0.32116	18.196	6	3.033	155.706	0.745	0.138	0.117	0.255
44	$\delta^{13}C$	anFR	37.911	123	0.308	10.848	21	0.517	0.04342	5.261	6	0.877	54.02	0.702	0.201	0.097	0.298
46	$\delta^{13}C$	anFR	21.613	82	0.264	5.459	14	0.390	0.13746	5.882	6	0.980	32.954	0.656	0.166	0.178	0.344
50	$\delta^{13}C$	anFR	23.051	54	0.427	7.507	9	0.834	0.0634	5.579	6	0.930	36.137	0.638	0.208	0.154	0.362

(a) Total : jeu de données complet, *anFR* : jeu de données $\delta^{13}C$ analysées en France, *anUSA* : jeu de données $\delta^{13}C$ analysées aux USA.

(b) Somme des carrés des écarts de la résiduelle du modèle (SSQ_{err}), de l'effet famille hiérarchisé dans la population (SSQ_{fam}), de l'effet bloc (SSQ_{bloc}) et totale du modèle (SSQ_{tot}).

(c) Nombre de degré de liberté de l'erreur (df_{err}), des familles (df_{fam}) et des blocs (df_{bloc}). Le nombre maximum de blocs étant de 7, le degré de liberté df_{bloc} est de 6, à l'exception de 4 familles des populations 21, 26 et 27 où le 8^{ème} bloc a été utilisé pour compenser les nombreuses données manquantes dans les 7 premiers blocs.

(d) Carré Moyen de l'erreur (MS_{err}), de l'effet famille hiérarchisé dans la population (MS_{fam}) et de l'effet bloc (MS_{bloc}).

(e) Valeurs de probabilité de l'effet famille. Les nombres surlignés en jaune correspondent à des effets familles significatifs à 5%.

(f) Coefficient de détermination de l'erreur (R_{err}), de l'effet famille hiérarchisé dans la population (R_{fam}) et de l'effet bloc (R_{bloc}). Ces coefficients correspondent à la part de variation totale expliquée par chaque effet.

(g) Coefficient de détermination du modèle.

**Annexe 8. Moyenne, variance génétique
entre familles, variance génétique additive,
et héritabilité au sens strict (avec erreurs
standards de h^2) par population pour *HT03*,
TB et $\delta^{13}C$.**

Pour $\delta^{13}C$, ces indices sont également estimés au sein de chaque type d'analyse.

Population	Origine de l'analyse ^(a)	N fam ^(b)	HT03					TB					$\delta^{13}C$				
			Moyenne	Var(fam) ^(c)	V_A ^(d)	h^2 ^(e)	SE_{h^2} ^(f)	Moyenne	Var(fam) ^(c)	V_A ^(d)	h^2 ^(e)	SE_{h^2} ^(f)	Moyenne	Var(fam) ^(c)	V_A ^(d)	h^2 ^(e)	SE_{h^2} ^(f)
15	<i>an FR</i>	15	110.15	24.03	96.10	0.23	0.30	489.28	1147.90	4591.60	0.15	0.27	-26.49	-0.03	-0.11	-0.28	0.15
20	<i>an USA</i>	18	79.06	51.35	205.40	0.76	0.37	344.99	291.61	1166.44	0.07	0.23	-26.73	0.50	2.00	2.23	0.41
2124	<i>total</i>	49	79.79	25.63	102.50	0.34	0.17	332.58	3205.40	12821.60	0.51	0.194	-26.88	0.03	0.11	0.25	0.22
	<i>an FR</i>	15											-26.89	0.00	0.01	0.06	0.13
	<i>an USA</i>	14											-26.91	0.06	0.23	0.37	0.34
22	<i>total</i>	20	67.29	25.48	101.94	0.51	0.31	263.34	1133.10	4532.40	0.40	0.29	-26.73	0.18	0.72	1.37	0.40
	<i>an FR</i>	15											-26.88	0.02	0.10	0.30	0.31
	<i>an USA</i>	7											-26.31	0.48	1.94	2.12	0.70
23	<i>total</i>	29	69.52	60.56	242.24	0.86	0.30	289.71	4131.20	16524.80	1.07	0.32	-26.89	0.07	0.26	0.61	0.27
	<i>an FR</i>	15											-26.91	0.00	-0.01	-0.02	0.23
	<i>an USA</i>	14											-26.87	0.15	0.61	1.05	0.46
25	<i>total</i>	29	68.23	18.45	73.80	0.27	0.22	316.70	3781.00	15124.00	0.79	0.29	-26.84	0.11	0.44	1.03	0.31
	<i>an FR</i>	15											-26.81	-0.01	-0.04	-0.10	0.20
	<i>an USA</i>	14											-26.86	0.26	1.04	2.00	0.48
26	<i>total</i>	29	65.31	17.10	68.40	0.31	0.22	289.19	1702.80	6811.20	0.41	0.24	-26.69	0.04	0.16	0.40	0.24
	<i>an FR</i>	15											-26.69	0.03	0.14	0.34	0.32
	<i>an USA</i>	14											-26.69	0.06	0.23	0.55	0.38
27	<i>an USA</i>	20	92.42	69.36	277.45	0.70	0.34	390.91	5146.30	20585.20	0.76	0.35	-26.14	0.09	0.35	0.80	0.35
29	<i>total</i>	27	65.77	47.89	191.56	0.85	0.31	302.65	2224.70	8898.80	0.56	0.27	-26.72	0.05	0.20	0.55	0.27
	<i>an FR</i>	15	66.33	46.09	184.35	0.85	0.42	292.16	1065.50	4262.00	0.30	0.31	-26.72	0.04	0.15	0.41	0.34
	<i>an USA</i>	12	65.07	56.29	225.17	0.93	0.48	315.80	3718.20	14872.80	0.81	0.47	-26.73	0.07	0.29	0.82	0.47
30	<i>total</i>	18	48.31	26.71	106.83	0.77	0.37	215.00	1522.90	6091.60	0.59	0.34	-27.09	0.01	0.02	0.04	0.22
	<i>an FR</i>	9	50.03	17.44	69.74	0.46	0.46	206.89	1079.50	4318.00	0.66	0.51	-27.10	0.03	0.12	0.31	0.42
	<i>an USA</i>	9	46.40	34.94	139.77	1.19	0.60	224.02	2077.80	8311.20	0.57	0.49	-27.07	0.00	0.00	-0.22	0.22
4041	<i>an FR</i>	55	86.36	46.19	184.76	0.55	0.19	299.99	1998.90	7995.60	0.48	0.18	-26.68	0.02	0.10	0.29	0.16
4243	<i>an FR</i>	57	85.00	23.54	94.14	0.31	0.16	283.99	1185.50	4742.00	0.35	0.16	-26.56	0.01	0.02	0.06	0.12
44	<i>an FR</i>	22	82.85	18.63	74.53	0.30	0.25	280.44	1504.10	6016.40	0.48	0.29	-26.43	0.03	0.13	0.38	0.27
46	<i>an FR</i>	15	87.86	19.34	77.35	0.21	0.29	331.08	4066.10	16264.40	0.85	0.42	-26.53	0.02	0.08	0.27	0.31
50	<i>an FR</i>	10	92.13	0.65	2.58	0.01	0.29	501.07	-703.30	-2813.20	-0.11	0.25	-26.81	0.06	0.24	0.52	0.45

^(a) *anFR* : lot de données $\delta^{13}C$ analysées en France, *anUSA* : lot de données $\delta^{13}C$ analysées aux USA ; ^(b) Nombre de familles par population ; ^(c) Variance entre familles. Les valeurs surlignées en jaune correspondent à un effet famille significatif à 5% ; ^(d) Variance génétique additive ; ^(e) Héritabilité au sens strict. Les valeurs en gras sont significativement différentes de zéro ; ^(f) Erreur standard de l'héritabilité.

**Annexe 9. Résultats d'ANOVA global sur
l'ensemble des populations pour *HT02*,
HT03, *DC*, *PolyT*, *TB*, *NB* et $\delta^{13}C$**

Caractère	SSQ _{err} (a)	df _{err} (b)	MS _{err} (c)	SSQ _{pop} (a)	df _{pop} (b)	MS _{pop} (c)	P _{pop} (d)	SSQ _{fam} (a)	df _{fam} (b)	MS _{fam} (c)	P _{fam} (d)	SSQ _{bloc} (a)	df _{bloc} (b)	MS _{bloc} (c)	SSQ _{tot} (a)	R _{error} (e)	R _{pop} (e)	R _{fam} (e)	R _{bloc} (e)	CoefDet (f)
<i>HTO2</i>	165213.69	2337.00	70.69	85335.42	13.00	6564.26	0.000	50146.18	396.00	126.63	0.000	38654.37	7	6442.40	300695.29	0.55	0.28	0.17	0.13	0.58
<i>HTO3</i>	593513.90	2337.00	253.96	391460.03	13.00	30112.31	0.000	185844.80	396.00	469.31	0.000	69576.36	7	11596.06	1170818.73	0.51	0.33	0.16	0.06	0.55
<i>DC</i>	44703.09	2337.00	19.13	9944.13	13.00	764.93	0.000	14111.21	396.00	35.63	0.000	1721.73	7	286.96	68758.43	0.65	0.14	0.21	0.03	0.37
<i>PolyT</i>	68.08	2337.00	0.03	11.23	13.00	0.86	0.000	14.71	396.00	0.04	0.001	2.92	7	0.49	94.02	0.72	0.12	0.16	0.03	0.31
<i>TB</i>	34810197.24	2337.00	14895.25	9145234.62	13.00	703479.59	0.000	10474415.49	396.00	26450.54	0.000	1765886.52	7	294314.42	54429847.35	0.64	0.17	0.19	0.03	0.39
<i>NB</i>	10198774.12	2337.00	4364.05	2480357.98	13.00	190796.77	0.000	3058776.16	396.00	7724.18	0.000	464051.84	7	77341.97	15737908.26	0.65	0.16	0.19	0.03	0.38
$\delta^{13}C$	811.80	2337.00	0.35	103.56	13.00	7.97	0.000	293.00	396.00	0.74	0.000	87.94	7	14.66	1208.36	0.67	0.09	0.24	0.07	0.40

(a) Somme des carrés des écarts de la résiduelle du modèle (SSQ_{err}), de l'effet population (SSQ_{pop}), de l'effet famille hiérarchisé dans la population (SSQ_{fam}), de l'effet bloc (SSQ_{bloc}) et totale du modèle (SSQ_{tot}).

(b) Nombre de degré de liberté de l'erreur (df_{err}), des populations (df_{pop}), des familles (df_{fam}) et des blocs (df_{bloc}). Le nombre maximum de blocs étant de 7, le degré de liberté df_{bloc} est de 6, à l'exception de 4 familles des populations 21, 26 et 27 où le 8^{ème} bloc a été utilisé pour compenser les nombreuses données manquantes dans les 7 premiers blocs.

(c) Carré Moyen de l'erreur (MS_{err}), de l'effet population (MS_{pop}), de l'effet famille hiérarchisé dans la population (MS_{fam}) et de l'effet bloc (MS_{bloc}).

(d) Valeurs de probabilité de l'effet population (P_{pop}) et de l'effet famille hiérarchisé dans la population (P_{fam}).

(e) Coefficient de détermination de l'erreur (R_{err}), de l'effet population (R_{pop}), de l'effet famille hiérarchisé dans la population (R_{fam}) et de l'effet bloc (R_{bloc}). Ces coefficients correspondent à la part de variation totale expliquée par chaque effet.

(f) Coefficient de détermination du modèle.

**Annexe 10. Valeurs de Q_{ST} par paires ou
groupes de populations pour *HT03*, *TB* (B)
et $\delta^{13}C$**

Les valeurs surlignées en jaune sont inférieures à 0.05. Les valeurs de Q_{ST} qui ne sont pas significatives (c'est-à-dire lorsque la différence entre moyennes des deux populations n'est pas significative), sont remplacées par « NS ».

A) HT03

HT03	15	20	2124	22	23	25	26	27	29	30	4041	4243	44	46	50
15	*														
20	0.619	*													
2124	0.700	0.000	*												
22	0.825	0.181	0.269	*											
23	0.685	0.082	0.125	0.000	*										
25	0.847	0.183	0.267	0.000	0.000	*									
26	0.871	0.281	0.380	0.000	0.021	0.019	*								
27	0.277	0.161	0.193	0.466	0.345	0.499	0.563	*							
29	0.758	0.175	0.243	0.000	0.009	0.003	0.000	0.444	*						
30	0.907	0.603	0.703	0.461	0.367	0.535	0.462	0.717	0.324	*					
4041	0.457	0.061	0.059	0.357	0.257	0.354	0.433	0.041	0.360	0.686	*				
4243	0.629	0.064	0.059	0.449	0.295	0.445	0.535	0.094	0.423	0.778	0.000	*			
44	0.694	0.019	0.016	0.413	0.209	0.421	0.529	0.119	0.345	0.776	0.010	0.000	*		
46	0.598	0.109	0.136	0.537	0.306	0.564	0.650	0.019	0.441	0.813	0.000	0.006	0.053	*	
50	0.580	0.244	0.324	0.691	0.410	0.717	0.784	0.000	0.548	0.880	0.035	0.119	0.270	0.049	*
moy	0.676	0.185	0.232	0.322	0.210	0.335	0.390	0.271	0.284	0.649	0.202	0.251	0.244	0.284	0.391

B) TB

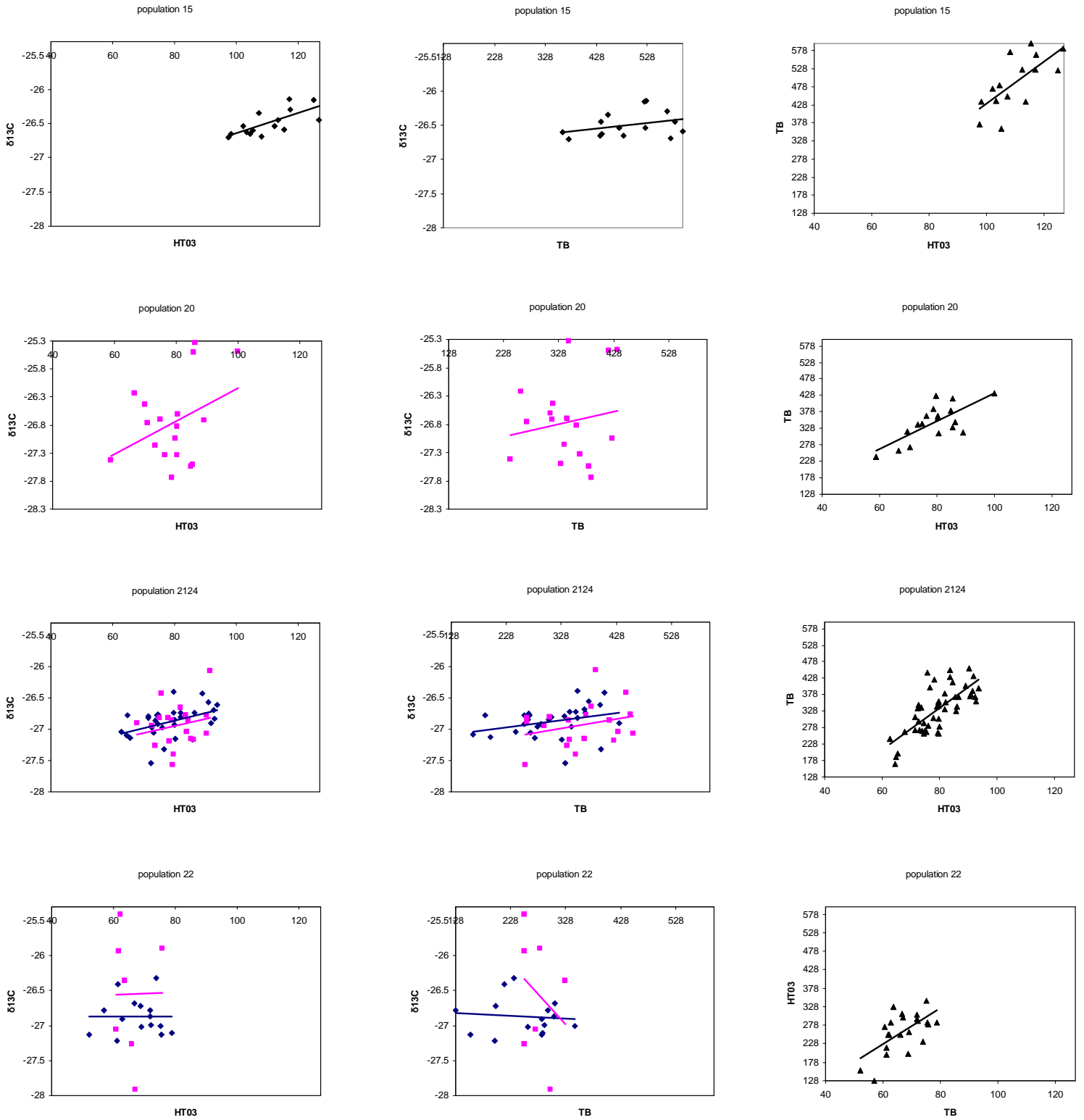
TB	15	20	2124	22	23	25	26	27	29	30	4041	4243	44	46	50
15	*														
20	0.630	*													
2124	0.369	0.000	*												
22	0.740	0.343	0.128	*											
23	0.442	0.058	0.032	0.008	*										
25	0.388	0.008	0.000	0.059	0.006	*									
26	0.624	0.127	0.050	0.019	0.000	0.010	*								
27	0.136	0.037	0.040	0.259	0.126	0.072	0.181	*							
29	0.538	0.057	0.020	0.046	0.000	0.000	0.000	0.128	*						
30	0.781	0.522	0.271	0.087	0.095	0.181	0.169	0.372	0.196	*					
4041	0.550	0.065	0.031	0.039	0.000	0.002	0.000	0.166	0.000	0.189	*				
4243	0.691	0.182	0.087	0.014	0.000	0.028	0.000	0.265	0.008	0.184	0.005	*			
44	0.667	0.191	0.070	0.004	0.000	0.021	0.000	0.197	0.006	0.147	0.004	0.000	*		
46	0.369	0.000	0.000	0.099	0.015	0.000	0.029	0.043	0.005	0.233	0.014	0.063	0.046	*	
50	0.000	1.000	0.441	0.882	0.488	0.442	0.724	0.179	0.629	0.889	0.611	0.764	0.794	0.449	*
moy	0.495	0.215	0.103	0.190	0.084	0.076	0.128	0.146	0.104	0.312	0.108	0.154	0.144	0.085	0.595

C) $\delta^{13}C$

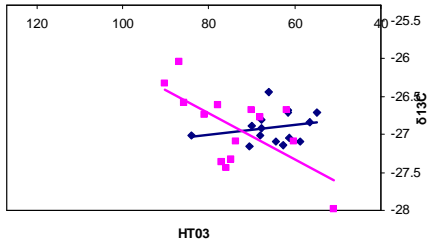
$\delta^{13}C$	15	20	2124	22	23	25	26	27	29	30	4041	4243	44	46	50
15	*														
20	0.007	*													
2124	0.472	0.000	*												
22	0.020	0.000	0.010	*											
23	0.260	0.000	0.000	0.011	*										
25	0.093	0.000	0.000	0.000	0.000	*									
26	0.098	0.000	0.055	0.000	0.043	0.011	*								
27	0.188	0.076	0.405	0.130	0.336	0.235	0.247	*							
29	0.101	0.000	0.033	0.000	0.027	0.004	0.000	0.248	*						
30	1.000	0.018	0.080	0.066	0.039	0.040	0.248	0.543	0.188	*					
4041	0.113	0.000	0.085	0.000	0.069	0.022	0.000	0.325	0.000	0.326	*				
4243	1.000	0.011	0.319	0.021	0.234	0.097	0.041	0.328	0.059	0.742	0.043	*			
44	0.000	0.021	0.298	0.041	0.221	0.118	0.099	0.079	0.108	0.565	0.129	0.076	*		
46	0.000	0.003	0.222	0.009	0.146	0.062	0.032	0.141	0.043	0.603	0.045	0.000	0.008	*	
50	0.401	0.000	0.000	0.000	0.000	0.000	0.003	0.269	0.000	0.122	0.019	0.201	0.179	0.107	*
moy	0.301	0.009	0.132	0.020	0.087	0.043	0.059	0.261	0.053	0.300	0.081	0.231	0.150	0.113	0.081

Annexe 11. Graphiques de régression des moyennes familiales de $\delta^{13}C$ en fonction de *HT03* et de *TB* pour chaque lot d'analyse, *anFR* et *anUSA*, et de *TB* en fonction de *HT03*

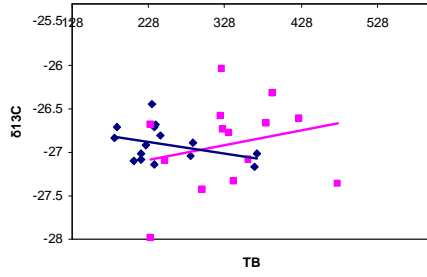
Les graphiques de régression sont présentés avec leurs courbes de régression associées pour les moyennes familiales de $\delta^{13}C$ en fonction de *HT03* et de *TB* pour chaque lot d'analyse, *anFR* (bleu) et *anUSA* (rose), et de *TB* en fonction de *HT03* (noir).



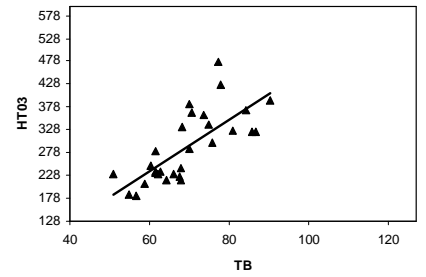
population 23



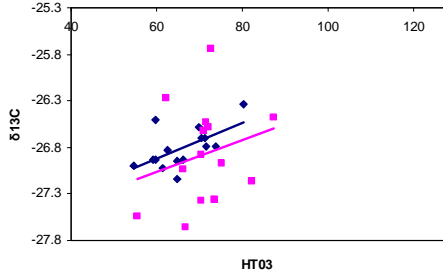
population 23



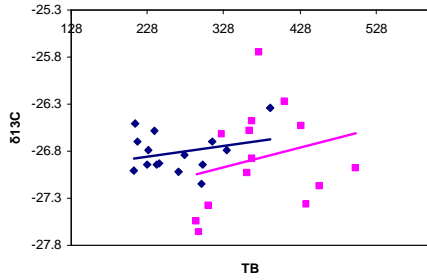
population 23



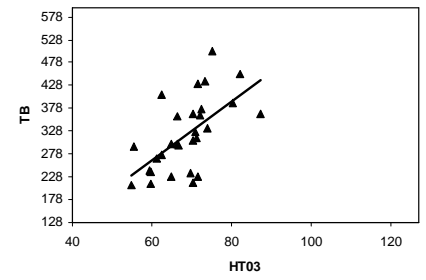
population 25



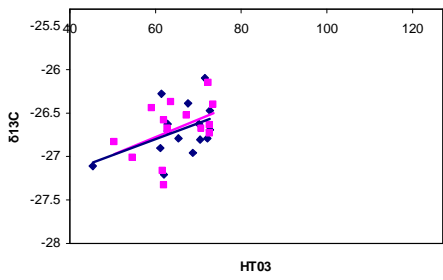
population 25



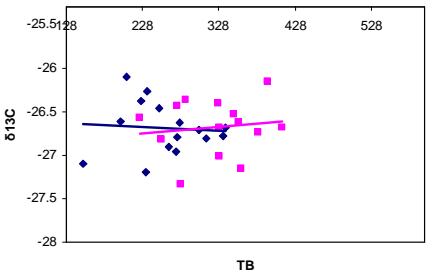
population 25



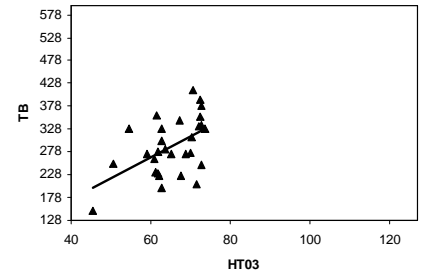
population 26



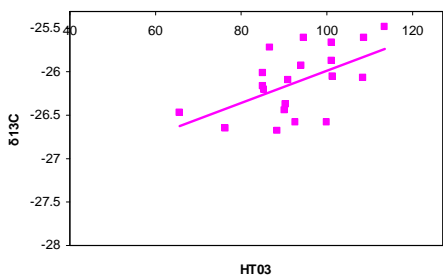
population 26



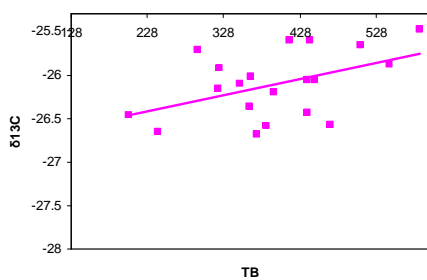
population 26



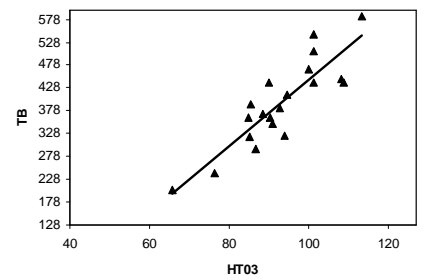
population 27



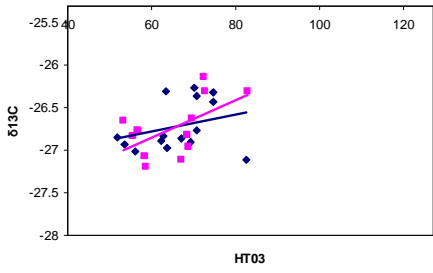
population 27



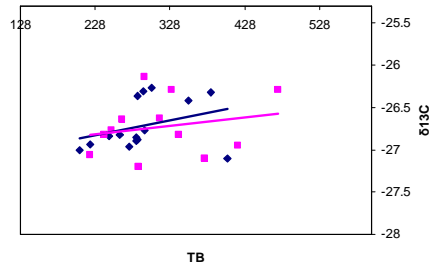
population 27



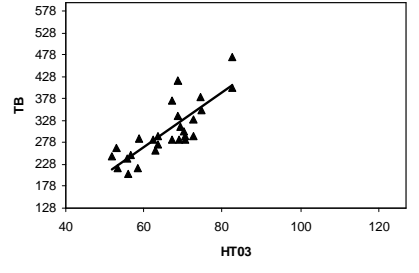
population 29



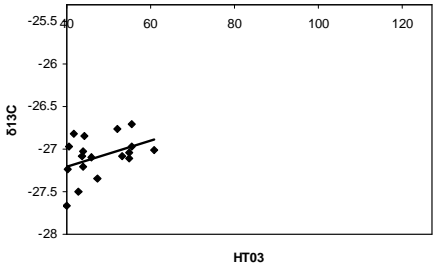
population 29



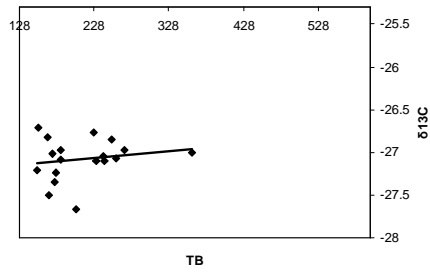
population 29



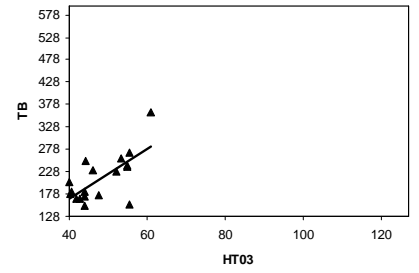
population 30



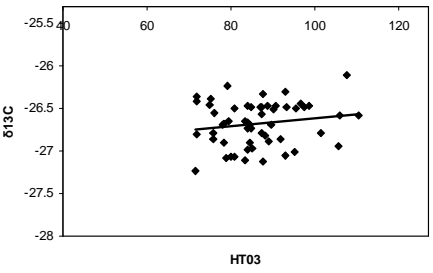
population 30



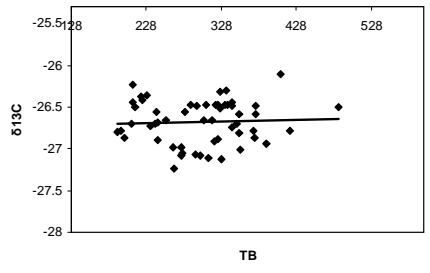
population 30



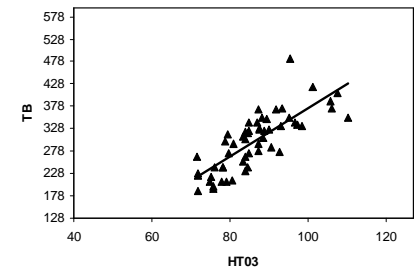
population 4041



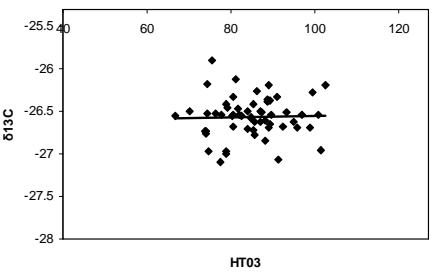
population 4041



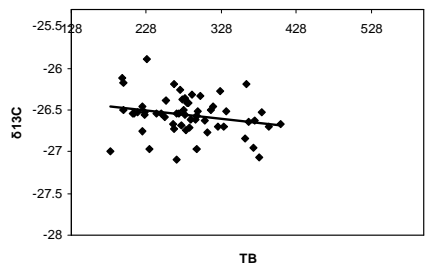
population 4041



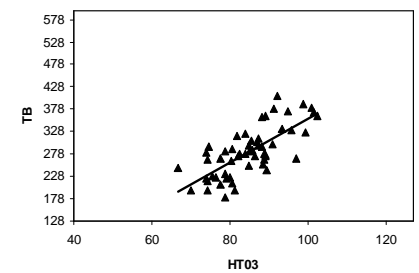
population 4243



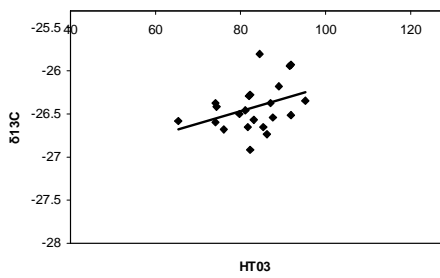
population 4243



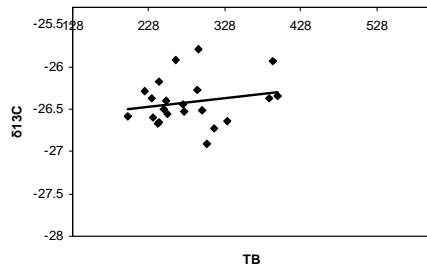
population 4243



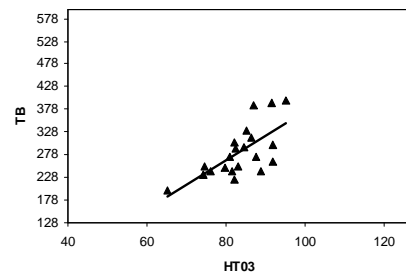
population 44



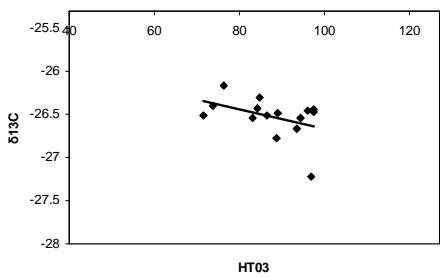
population 44



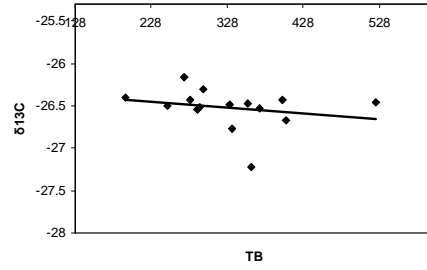
population 44



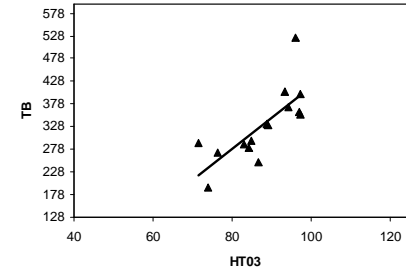
population 46



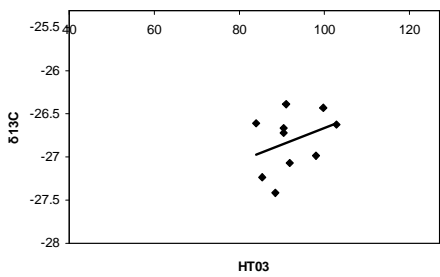
population 46



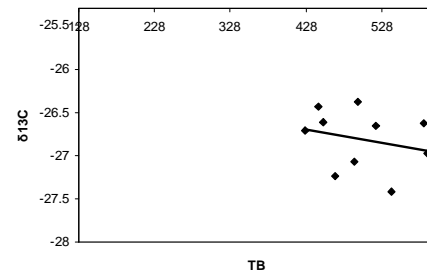
population 46



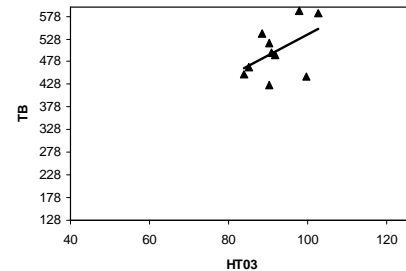
population 50



population 50



population 50



**Annexe 12. Matrices de F_{ST} par paires de
populations pour les gènes et SNPs
« outliers »**

Les matrices de F_{ST} par paires de populations sont présentées pour les gènes détectés comme “outliers” positifs ou négatifs avec “Fdist2” ou “BayesFst” (sans données manquantes), et pour les SNPs détectés comme “outliers” positifs avec “FstSNP” ou “BayesFst” (avec données manquantes). Les valeurs supérieures à 0.2 sont surlignées en bleu, et les valeurs supérieures à 0.5 sont surlignées en jaune.

A)

<i>PR-AGP4</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadec	Tabarka
Pineta										
Pinia	-0.144									
Arenas	0.301	0.197								
Coca	0.003	-0.017	0.228							
San Cipriano	0.377	0.231	-0.049	0.264						
Oria	-0.012	0.040	0.412	0.108	0.478					
Tamrabta	0.040	0.039	0.450	0.039	0.544	0.086				
Mimizan	0.218	0.125	-0.024	0.163	-0.019	0.365	0.359			
Pleucadec	0.172	0.077	-0.052	0.119	-0.024	0.324	0.323	-0.080		
Tabarka	0.309	0.195	-0.074	0.236	-0.059	0.409	0.464	-0.022	-0.056	
Mean	0.140	0.082	0.154	0.127	0.194	0.246	0.260	0.121	0.089	0.156

B)

<i>erd3</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadec	Tabarka
Pineta										
Pinia	0.000									
Arenas	0.000	0.000								
Coca	NA	NA	NA							
San Cipriano	0.000	0.000	0.000	NA						
Oria	0.262	0.228	0.336	NA	0.262					
Tamrabta	0.567	0.534	0.640	NA	0.567	0.038				
Mimizan	0.040	0.000	0.126	NA	0.040	0.071	0.352			
Pleucadec	0.040	0.000	0.126	NA	0.040	0.071	0.352	-0.111		
Tabarka	0.000	0.000	0.000	NA	0.000	0.228	0.534	0.000	0.000	
Mean	0.239	0.220	0.279	NA	0.239	0.182	0.413	0.113	0.124	0.196

C)

<i>GRP3</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadec	Tabarka
Pineta										
Pinia	0.12									
Arenas	0.103	0.032								
Coca	0.366	0.2538	0.0779							
San Cipriano	0.099	-0.056	-0.016	0.1621						
Oria	0.07	0.0334	0.0875	0.2513	0.0629					
Tamrabta	0.029	0.0651	0.1428	0.3392	0.1107	-0.11				
Mimizan	0.138	0.115	-0.037	0.0485	0.0126	0.0954	0.1711			
Pleucadec	0.49	0.3091	0.1366	-0.063	0.2351	0.3113	0.4154	0.091		
Tabarka	0.062	0.1159	0.2222	0.4135	0.1968	-0.041	-0.085	0.2642	0.515	
Mean	0.164	0.110	0.083	0.205	0.090	0.084	0.120	0.100	0.271	0.185

D)

<i>dhn-2</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadec	Tabarka
Pineta										
Pinia	-0.200									
Arenas	0.040	0.040								
Coca	-0.152	-0.152	-0.010							
San Cipriano	-0.091	-0.091	0.040	-0.127						
Oria	-0.015	-0.015	0.143	-0.070	-0.156					
Tamrabta	0.100	0.100	0.196	-0.072	0.000	-0.045				
Mimizan	-0.133	-0.133	-0.059	-0.145	-0.096	-0.022	0.057			
Pleucadec	0.033	0.033	0.318	-0.053	-0.018	-0.058	-0.019	0.029		
Tabarka	-0.005	-0.005	-0.132	-0.058	-0.012	0.075	0.122	-0.078	0.226	
Mean	-0.047	-0.047	0.064	-0.093	-0.061	-0.018	0.049	-0.064	0.055	0.015

E)

<i>lp3-1</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadec	Tabarka
Pineta										
Pinia	-0.186									
Arenas	-0.053	-0.107								
Coca	-0.094	-0.074	0.085							
San Cipriano	-0.132	-0.104	-0.017	-0.025						
Oria	0.171	0.073	0.026	0.171	0.201					
Tamrabta	0.091	-0.018	-0.125	0.091	0.062	-0.032				
Mimizan	-0.026	-0.111	-0.104	0.017	-0.029	-0.009	-0.104			
Pleucadec	0.022	-0.108	-0.127	0.022	0.062	0.045	-0.127	-0.160		
Tabarka	-0.094	-0.127	-0.177	-0.167	-0.025	0.126	0.000	-0.072	-0.140	
Mean	-0.033	-0.085	-0.088	-0.019	-0.001	0.086	-0.018	-0.066	-0.057	-0.075

F)

<i>dhn-1</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadec	Tabarka
Pineta										
Pinia	-0.108									
Arenas	-0.003	-0.038								
Coca	0.002	-0.062	-0.035							
San Cipriano	-0.003	-0.055	-0.102	-0.082						
Oria	-0.124	-0.130	0.020	-0.004	0.031					
Tamrabta	0.047	-0.030	0.032	-0.019	0.027	0.000				
Mimizan	0.072	0.003	0.008	-0.013	0.000	0.050	-0.074			
Pleucadec	-0.018	-0.068	-0.043	-0.058	-0.061	-0.029	-0.094	-0.080		
Tabarka	0.287	0.198	0.257	0.175	0.293	0.222	-0.002	0.090	0.098	
Mean	0.017	-0.032	0.011	-0.011	0.005	0.004	-0.013	0.006	-0.039	0.180

G)

<i>PR-AGP4 S182</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadec	Tabarka
Pineta										
Pinia	-0.067									
Arenas	0.440	0.330								
Coca	0.329	0.210	-0.042							
San Cipriano	0.358	0.236	-0.098	-0.090						
Oria	-0.049	-0.101	0.276	0.155	0.181					
Tamrabta	0.260	0.384	0.851	0.770	0.873	0.444				
Mimizan	0.339	0.221	-0.023	-0.047	-0.070	0.166	0.726			
Pleucadec	0.517	0.418	-0.047	-0.002	-0.075	0.370	0.918	0.015		
Tabarka	0.084	0.190	0.746	0.651	0.722	0.244	0.000	0.629	0.818	
Mean	0.246	0.202	0.270	0.215	0.226	0.188	0.581	0.217	0.326	0.454

H)

<i>PR-AGP4 S331</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadec	Tabarka
Pineta										
Pinia	0.010									
Arenas	0.467	0.219								
Coca	0.358	0.106	-0.042							
San Cipriano	0.389	0.129	-0.098	-0.090						
Oria	0.168	-0.051	0.040	-0.039	-0.031					
Tamrabta	0.231	0.485	0.851	0.770	0.873	0.667				
Mimizan	0.421	0.168	-0.037	-0.046	-0.081	0.000	0.768			
Pleucadec	0.540	0.309	-0.047	-0.002	-0.075	0.124	0.918	-0.008		
Tabarka	0.061	0.298	0.746	0.651	0.722	0.495	0.000	0.677	0.818	
Mean	0.294	0.186	0.233	0.185	0.193	0.152	0.618	0.207	0.286	0.496

I)

<i>PR-AGP4 S472</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadec	Tabarka
Pineta										
Pinia	0.056									
Arenas	0.675	0.452								
Coca	0.540	0.247	0.011							
San Cipriano	0.417	0.099	0.090	-0.078						
Oria	0.168	-0.087	0.334	0.124	-0.004					
Tamrabta	-0.088	0.074	0.755	0.604	0.468	0.195				
Mimizan	0.421	0.102	0.091	-0.008	-0.068	0.000	0.465			
Pleucadec	0.513	0.213	0.023	-0.071	-0.090	0.094	0.574	-0.021		
Tabarka	0.061	0.368	0.924	0.818	0.740	0.495	0.020	0.677	0.802	
Mean	0.307	0.169	0.373	0.243	0.175	0.147	0.341	0.184	0.226	0.545

J)

<i>PR-AGP4 S525</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadec	Tabarka
Pineta										
Pinia	-0.044									
Arenas	0.321	0.207								
Coca	0.156	0.008	0.011							
San Cipriano	0.231	0.111	0.000	-0.032						
Oria	0.231	0.111	0.000	-0.032	0.000					
Tamrabta	0.231	0.111	0.000	-0.032	0.000	0.000				
Mimizan	0.412	0.303	0.000	0.047	0.000	0.000	0.000			
Pleucadec	0.276	0.159	0.000	-0.009	0.000	0.000	0.000	0.000		
Tabarka	0.441	0.625	0.924	0.818	0.889	0.889	0.889	0.947	0.908	
Mean	0.251	0.177	0.163	0.104	0.133	0.133	0.133	0.190	0.148	0.814

K)

<i>PR-AGP4 S666</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadec	Tabarka
Pineta										
Pinia	-0.111									
Arenas	-0.018	-0.018								
Coca	-0.050	-0.050	-0.058							
San Cipriano	0.020	0.020	-0.067	-0.057						
Oria	0.120	0.120	0.359	0.300	0.411					
Tamrabta	0.495	0.495	0.677	0.636	0.756	0.088				
Mimizan	-0.051	-0.051	-0.032	-0.046	-0.022	0.278	0.582			
Pleucadec	0.020	0.020	-0.067	-0.057	-0.111	0.411	0.756	-0.022		
Tabarka	0.667	0.667	0.788	0.760	0.889	0.285	0.000	0.684	0.889	
Mean	0.121	0.121	0.174	0.153	0.204	0.263	0.498	0.147	0.204	0.625

L)

<i>PR-AGP4 S762</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadec	Tabarka
Pineta										
Pinia	0.074									
Arenas	0.439	0.078								
Coca	0.380	0.030	-0.058							
San Cipriano	0.337	-0.010	-0.076	-0.088						
Oria	-0.087	-0.026	0.315	0.254	0.206					
Tamrabta	0.020	0.368	0.677	0.636	0.625	0.127				
Mimizan	0.353	0.018	-0.032	-0.046	-0.069	0.233	0.582			
Pleucadec	0.495	0.127	-0.067	-0.057	-0.068	0.368	0.756	-0.022		
Tabarka	0.222	0.556	0.788	0.760	0.778	0.333	0.000	0.684	0.889	
Mean	0.248	0.135	0.229	0.201	0.182	0.192	0.421	0.189	0.269	0.557

M)

<i>PR-AGP4 S1181</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadec	Tabarka
Pineta										
Pinia	-0.062									
Arenas	0.612	0.393								
Coca	0.419	0.191	-0.017							
San Cipriano	0.295	0.074	-0.024	-0.109						
Oria	0.143	-0.044	0.133	-0.031	-0.114					
Tamrabta	-0.113	0.022	0.670	0.504	0.400	0.252				
Mimizan	0.412	0.184	0.008	-0.047	-0.097	-0.037	0.497			
Pleucadec	0.457	0.229	-0.055	-0.070	-0.106	-0.001	0.539	-0.048		
Tabarka	0.095	0.315	0.854	0.736	0.699	0.564	0.020	0.706	0.782	
Mean	0.251	0.145	0.286	0.175	0.113	0.096	0.310	0.175	0.192	0.530

N)

<i>PR-AGP4 S1409</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadec	Tabarka
Pineta										
Pinia	-0.062									
Arenas	0.249	0.047								
Coca	0.183	-0.004	-0.053							
San Cipriano	0.338	0.114	-0.062	-0.052						
Oria	-0.102	0.066	0.403	0.341	0.507					
Tamrabta	0.308	0.500	0.694	0.674	0.866	0.163				
Mimizan	0.246	0.045	-0.039	-0.048	-0.051	0.396	0.670			
Pleucadec	0.457	0.229	-0.002	0.019	-0.106	0.608	0.896	0.008		
Tabarka	0.328	0.518	0.702	0.684	0.873	0.181	0.000	0.676	0.901	
Mean	0.216	0.161	0.215	0.194	0.270	0.285	0.530	0.211	0.334	0.540

O)

<i>PR-AGP4 S1480</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadec	Tabarka
Pineta										
Pinia	0.013									
Arenas	0.718	0.405								
Coca	0.511	0.173	-0.008							
San Cipriano	0.458	0.114	-0.033	-0.103						
Oria	0.256	-0.044	0.144	-0.043	-0.086					
Tamrabta	-0.133	0.111	0.759	0.583	0.545	0.361				
Mimizan	0.539	0.206	0.005	-0.050	-0.085	-0.025	0.602			
Pleucadec	0.571	0.229	-0.052	-0.069	-0.106	-0.001	0.637	-0.051		
Tabarka	-0.010	0.315	0.858	0.725	0.722	0.564	-0.057	0.721	0.782	
Mean	0.325	0.169	0.311	0.191	0.159	0.125	0.379	0.207	0.216	0.513

P)

<i>PR-AGP4 S1488</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadec	Tabarka
Pineta										
Pinia	0.0125									
Arenas	0.7182	0.4052								
Coca	0.5109	0.1726	-0.0076							
San Cipriano	0.4583	0.1143	-0.0327	-0.1034						
Oria	0.4583	0.1143	-0.0327	-0.1034	-0.1429					
Tamrabta	-0.1332	0.1105	0.7586	0.5826	0.5448	0.5448				
Mimizan	0.4713	0.1422	0.0312	-0.0443	-0.0797	-0.0797	0.5422			
Pleucadec	0.5708	0.2293	-0.052	-0.0689	-0.1055	-0.1055	0.6371	-0.0346		
Tabarka	-0.0104	0.3148	0.8582	0.7252	0.7218	0.7218	-0.0569	0.6692	0.7816	
Mean	0.340	0.180	0.294	0.185	0.153	0.153	0.392	0.180	0.206	0.525

Q)

<i>GRP3 S164</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadec	Tabarka
Pineta										
Pinia	0.066									
Arenas	-0.092	0.152								
Coca	0.302	0.633	0.175							
San Cipriano	0.120	0.458	0.080	0.000						
Oria	0.066	-0.125	0.152	0.633	0.458					
Tamrabta	0.203	-0.096	0.285	0.737	0.588	-0.096				
Mimizan	0.312	0.642	0.180	0.000	0.000	0.642	0.745			
Pleucadec	0.198	0.535	0.122	0.000	0.000	0.535	0.655	0.000		
Tabarka	0.171	-0.097	0.252	0.683	0.536	-0.097	-0.110	0.691	0.600	
Mean	0.149	0.241	0.145	0.395	0.280	0.241	0.323	0.357	0.331	0.292

R)

<i>erd3-S42</i>	Pineta	Pinia	Arenas	Coca	San Cipriano	Oria	Tamrabta	Mimizan	Pleucadec	Tabarka
Pineta										
Pinia	0.000									
Arenas	0.000	0.000								
Coca	NA	NA	NA							
San Cipriano	0.000	0.000	0.000	NA						
Oria	0.233	0.233	0.308	NA	0.233					
Tamrabta	1.000	1.000	1.000	NA	1.000	0.525				
Mimizan	0.000	0.000	0.072	NA	0.000	-0.052	0.800			
Pleucadec	0.040	0.040	0.126	NA	0.040	-0.111	0.774	-0.220		
Tabarka	0.000	0.000	0.000	NA	0.000	0.200	1.000	-0.035	0.000	
Mean	0.253	0.253	0.278	NA	0.253	0.190	0.789	0.130	0.132	0.241

Résumé

Les changements climatiques annoncés risquent de constituer des pressions de sélection importantes sur des caractères tels que la résistance à la sécheresse, notamment pour des espèces forestières comme le pin maritime. Dans ce contexte, une question essentielle est de savoir si ces espèces pourront s'adapter suffisamment rapidement aux nouvelles conditions climatiques. Les objectifs de cette thèse étaient d'évaluer le potentiel adaptatif de populations naturelles de pin maritime pour des caractères liés à la résistance au déficit hydrique, et d'identifier des gènes potentiellement impliqués dans la variation adaptative intra- et inter-populations. Pour y répondre, la variabilité phénotypique pour des caractères de croissance, de biomasse, et d'efficacité d'utilisation en eau (*EUE*) a été évaluée dans un essai de populations naturelles provenant de milieux contrastés sur les plans climatiques et écologiques. Une grande variabilité phénotypique a été mise en évidence entre performances moyennes des populations, et pour les valeurs relatives d'héritabilité et de corrélations entre biomasse et *EUE* estimées dans chaque population. Ces résultats ont confirmé le rôle de la sélection dans l'adaptation locale du pin maritime, suggérant des stratégies de réponses à un déficit hydrique différentes selon les populations. Les patrons de diversité et différenciation nucléotidiques de gènes candidats à la résistance au stress hydrique ont également été étudiés. Certains gènes s'écartent significativement de patrons d'évolution neutre, avec des signatures cohérentes et complémentaires entre plusieurs méthodes, qui ont été interprétées en termes de scénarios de sélection. L'hétérogénéité de ces signatures entre gènes candidats et entre populations pour certains gènes, serait aussi cohérente avec un rôle de la sélection naturelle plutôt qu'avec celui d'évènements démographiques qui auraient affecté l'ensemble du génome de façon plus homogène. Enfin, quelques associations remarquables ont été mises en évidence entre la variation de l'*EUE* entre populations et au sein des populations, et celle de marqueurs SNPs de gènes de déhydrines, de lignification ou de formation de la paroi cellulaire. Ce travail ouvre ainsi des perspectives encourageantes pour la compréhension de l'architecture génétique moléculaire de l'adaptation à la sécheresse chez cette espèce.

Mots clés : adaptation, sécheresse, pin maritime, populations naturelles, gènes candidats, sélection naturelle.

Summary

The predicted changes of climate may constitute important selective pressures on traits such as drought resistance for forest species as maritime pine. In this context, one of the essential questions is to know if these species could adapt sufficiently rapidly to the new climatic conditions. The targets of this work were to evaluate adaptive potential of natural populations of maritime pine for traits related to drought resistance, and to identify genes potentially involved in adaptive variation within and between populations. Phenotypic variability for traits linked to growth, biomass, and water use efficiency (*WUE*) was evaluated in a field trial including natural populations coming from contrasted environments at both climatic and ecological levels. Large phenotypic variability has been shown between average performances of populations, and for relative values of heritability and correlations between biomass and *WUE* estimated within populations. These results would confirm the role of natural selection in local adaptation of maritime pine, suggesting different strategies of response to drought resistance among populations. Patterns of nucleotide diversity and differentiation of candidate genes to drought resistance have also been studied. Some of these genes showed significant departures from neutral evolution patterns, with complementary and consistent signatures across methods, and interpreted in terms of selective scenarios. Heterogeneity of signatures between candidate genes and between populations for some genes, would be more consistent with a role of natural selection rather than to the role of demographic events that would affect homogeneously the whole genome. Statistical associations have also been shown between the variation of *WUE* between and within populations, and the variation of SNPs markers from genes of dehydrines, lignification and cellular wall formation. This work presents encouraging perspectives for understanding the molecular genetic architecture of drought adaptation in this species.

Key words: adaptation, drought, maritime pine, natural populations, candidate genes, natural selection.