N d'ordre :

THÈSE

PRÉSENTÉE A

L'UNIVERSITÉ BORDEAUX 1

ÉCOLE DOCTORALE DES SCIENCES ET ENVIRONNEMENT

Par ALBERTO, Florian

POUR OBTENIR LE GRADE DE

DOCTEUR

SPECIALITE : ECOLOGIE EVOLUTIVE, FONCTIONNELLE ET DES COMMUNAUTES

DIFFERENCIATION PHENOLOGIQUE ET MOLECULAIRE DU CHENE SESSILE LE LONG DE GRADIENTS ENVIRONNEMENTAUX

Thèse dirigée par KREMER, Antoine
DELZON, Sylvain

Soutenue le :

Devant la commission d'examen formée de :

Mme. SAVOLAINEN, Outi	PR, Université d'Oulu	Rapporteur
M. FADY, Bruno	DR, INRA Avignon	Rapporteur
Mme. LE CORRE Valérie	CR, INRA Dijon	Examinateur
Mme. JORGE Véronique	CR, INRA Orléans	Examinateur
M. RENAUDIN Jean-Pierre	PR, Université Bordeaux 1	Président de Jury
M. KREMER, Antoine	DR, INRA Bordeaux	Directeur de thèse

“Bonjour !
Comme un diable au fond de sa boîte,
Le bourgeon s’est tenu caché…
Mais dans sa prison trop étroite
Il baille et voudrait respirer.
Il entend des chants, des bruits d’ailes,
Il a soif de grand jour et d’air…
Il voudrait savoir les nouvelles,
Il fait craquer son corset vert.
Puis, d’un geste brusque, il déchire
Son habit étroit et trop court
« Enfin, se dit-il, je respire,
Je vis, je suis libre…bonjour ! ”

Paul Géraldy *(1885-1983)*

TABLE DES MATIERES

INTRODUCTION GENERALE

Contexte général

Dans le cadre des changements actuels du climat à l'échelle planétaire, de nombreuses questions se posent quant aux capacités d'adaptation des écosystèmes et des êtres vivants qui les composent. La température du globe a augmentée d'environ 0,74°C en moyenne au cours du dernier siècle et les premières années du XXI^{ème} siècle ont été parmi les plus chaudes jamais enregistrées, notamment 2005 et 2009 (IPCC 2007). Ce réchauffement, qui est en grande partie imputable aux activités humaines, devrait se poursuivre voire s'accentuer d'ici la fin du siècle. Selon les projections basées sur des modèles de circulation atmosphérique la température moyenne pourrait augmenter entre 1.8 et 4.0 °C d'ici 2100 suivant les scénarios socio-économiques envisagés. Dans ce contexte les préoccupations sont grandissantes quant à la capacité de réponse des êtres vivants, en particulier pour les espèces végétales dont le mode de vie fixé limite les capacités de migration en comparaison des espèces animales.

Face aux changements climatiques les réponses possibles des espèces végétales peuvent être de trois types (Aitken *et al.* 2008) :

(i) la migration en latitude ou en altitude afin de suivre les conditions environnementales qui leur sont favorables.
(ii) l'adaptation (*sensu lato*) des populations existantes aux nouvelles conditions du milieu.
(iii) l'extinction locale des populations.

Par le passé, des changements climatiques de grande ampleur se sont déjà produits à l'échelle planétaire. En particulier, les épisodes de glaciation et de réchauffement successifs qui ont eu lieu durant le Quaternaire ont eu comme principales conséquences de fortes modifications des aires de répartition des espèces végétales. Sur le continent Européen les épisodes de glaciation ont contraint les plantes à restreindre leurs aires de répartition aux limites méridionales dans des zones appelées « refuges glaciaires » (Petit *et al.* 2002a). Cependant les cycles de refroidissements/réchauffements successifs se sont produits sur des échelles de temps de l'ordre de plusieurs milliers d'années qui permettaient aux espèces de suivre l'évolution du climat. La distinction des changements climatiques actuels par rapport à ceux du passé repose donc dans leur rapidité qui devrait limiter les capacités de migration des espèces (Davis & Shaw 2001). Par conséquent les recherches visant à estimer le potentiel d'adaptation des espèces végétales se sont fortement développées ces dernières années.

Adaptation et diversité génétique

L'adaptation *sensu lato* peut être décomposée en deux mécanismes évolutifs distincts : la plasticité phénotypique et l'adaptation génétique. Le premier correspond à la capacité d'un organisme à modifier sa physiologie, sa morphologie ou son développement en réponse à des changements de son environnement (Callahan *et al.* 1997). La plasticité phénotypique peut donc permettre une réponse rapide d'un même génotype mais est souvent considérée comme étant de faible amplitude. Il s'agit donc essentiellement d'un mécanisme au niveau individuel et qui s'exerce sur une durée relativement courte qui est celle de la vie de l'organisme. L'adaptation génétique est le résultat de l'action de la sélection sur la diversité génétique qui favorise les individus en fonction de leurs génotypes. Il faut également que le trait soumis à sélection présente soit héritable et lié à la valeur adaptative des individus, *i.e.* que la valeur du trait soit contrôlée par des gènes transmissibles d'une génération à l'autre, pour que l'adaptation puisse s'opérer. Par comparaison à la plasticité, l'adaptation génétique est donc un mécanisme populationnel dont l'effet s'accumule au cours des générations, si les pressions de sélection sont continues au cours du temps. L'ampleur des changements évolutifs induits par l'adaptation dépend en premier lieu du niveau de diversité génétique du caractère.

La diversité génétique d'un trait liée à a valeur adaptative peut être appréhendée au niveau phénotypique (on parle alors de variabilité génétique) ou au niveau génomique en étudiant le polymorphisme moléculaire de l'ADN. Il existe donc deux manières de mesurer la diversité génétique des populations. Par abus de langage un trait lié à la valeur adaptative d'un arbre est appelé trait adaptatif, et la mesure de sa variabilité peut être réalisée dans des tests de provenances/descendances. Ces expérimentations consistent à planter dans un même environnement le plus homogène possible des individus de plusieurs populations d'origines différentes (les provenances) et de familles différentes (les descendances). La connaissance des relations de parenté à l'intérieur des populations (familles de plein-frères ou de demi-frères) permet de mesurer à la fois la variabilité génétique à l'intérieur de chaque provenance et la différenciation génétique entre les provenances, c'est-à-dire la part de la variabilité génétique totale attribuable aux différences entre provenances. L'estimation de la diversité génétique basée sur les marqueurs moléculaires est obtenue par amplification à partir de l'ADN des individus puis détection et identification du polymorphisme (les allèles) à ces marqueurs. Une des grandes questions de la génétique des populations est de savoir si la diversité génétique ainsi estimée a été soumise à la sélection naturelle (diversité adaptative) ou si elle ne résulte que des autres forces évolutives ayant eu cours dans la population

(diversité neutre), dont la démographie est souvent considérée comme la plus influente. Par conséquent un grand nombre d'approches ont été développées afin de tester la neutralité des marqueurs utilisés dans les populations étudiées. Ces approches reposent essentiellement sur le principe que les événements démographiques affectent le génome dans son ensemble tandis la sélection n'agit que sur les niveaux de diversité des régions génomiques impliquées dans le contrôle de traits adaptatifs. La diversité peut donc être estimée à partir d'un nombre important de marqueurs. La diversité neutre peut permettre d'étudier l'empreinte de la démographie sur la diversité génétique totale des populations en estimant des paramètres tels que les tailles efficaces des populations ou les flux de gènes et la différenciation entre les populations. Ces paramètres peuvent ensuite servir à déterminer des événements démographiques tels que des réductions de tailles de populations (« bottle-necks ») ou des phénomènes d'effet de fondation (établissement d'une population à partir d'un faible nombre d'individus) ayant amené des différences de niveaux de diversité génétique neutre mais aussi adaptative. La diversité adaptative peut ensuite être appréhendée en « déduisant » de la diversité totale la part liée aux événements démographiques par la recherche de signatures de sélection. Pour cela on recherche généralement les marqueurs qui présentent une diversité génétique s'écartant du patron de diversité génétique neutre. Une autre approche consiste à directement démontrer l'implication des marqueurs dans la variabilité d'un trait adaptatif. Ce type d'approche est cependant souvent limité aux espèces modèles (l'Humain, *Arabidopsis*, *Drosophila*) dont la diversité génétique est décrite de manière exhaustive. Chez les espèces non-modèles tels que les arbres forestiers une approche basée sur la sélection de gènes ayant un rôle putatif sur la variabilité du trait adaptatif (les gènes candidats) semble être la plus appropriée pour rechercher les bases moléculaires de l'adaptation (Gonzalez-Martinez *et al.* 2006).

Dans cette thèse, nous nous attacherons à estimer la diversité neutre et adaptative du chêne sessile dans des contextes environnementaux bien définis correspondant à des gradients climatiques pour un trait exposé aux pressions sélectives générées par les changements climatiques : le débourrement du bourgeon apical.

La phénologie du débourrement

Parmi les traits qui déterminent l'adaptation des espèces végétales à leur milieu, la phénologie, qui correspond à l'étude de l'apparition d'événements biologiques cycliques en relation avec les variations saisonnières du climat (Schwartz 2003) est un de ceux qui devrait être le plus sensible à l'élévation des températures. Des modifications de la phénologie des

plantes en réponses au réchauffement climatique ont déjà été observées durant les dernières décennies (Menzel & Fabian 1999, Beaubien & Freeland 2000, Menzel *et al.* 2006). Le débourrement des arbres, qui correspond à l'ouverture des bourgeons végétatifs au printemps (Figure 1) est un trait adaptatif majeur puisqu'il détermine le début de la saison de végétation et donc en partie la croissance de l'arbre (Kramer *et al.* 2000). Des études de modélisation de la phénologie du débourrement ont montré que celle-ci est en grande partie déterminée par le régime de température chez les espèces de zones tempérées (Chuine & Cour 1999, Chuine 2000).

Figure 1 : Stades phénologiques du débourrement (adapté de Derory *et al.* 2006)

Le bourgeon végétatif représente un organe de protection des méristèmes foliaires, qui leur permet de passer les conditions défavorables de l'hiver sans être endommagés. Sa formation a lieu généralement à la fin de l'été et s'accompagne d'un arrêt de la croissance de l'arbre. Le bourgeon rentre alors dans un stade de dormance, c'est-à-dire qu'il cesse presque complètement toute activité physiologique. Selon Lang *et al.* (1987) trois types de dormance ayant lieu successivement peuvent être distingués :

(i) La paradormance, qui a couramment lieu à la fin de l'été avec la diminution de la photopériode est une inhibition de la croissance des méristèmes foliaires qui est initiée par un organe distinct et éloigné du tissu dormant.

(ii) L'endodormance, qui est un état de dormance profond où des facteurs endogènes du bourgeon bloquent la croissance même si les conditions redeviennent favorables. Une accumulation suffisante de températures froides (appelées températures de « chilling ») est nécessaire pour que l'endodormance soit levée.

(iii) L'écodormance, qui correspond à un état dit de quiescence où le blocage de l'activité physiologique est lié aux conditions environnementales. La reprise d'activité peut avoir lieu lorsqu'une accumulation suffisante de températures chaudes (appelées températures de « forcing ») est atteinte.

Biologie et écologie du Chêne sessile

Le chêne sessile *Quercus petraea* (Matt.) Liebl. appartient au genre Quercus qui est rattaché à la famille des Fagacées. Son aire de répartition s'étend à toute l'Europe, du nord de l'Espagne au sud de la côte Scandinave, et de l'Irlande à l'Oural (Figure 2).

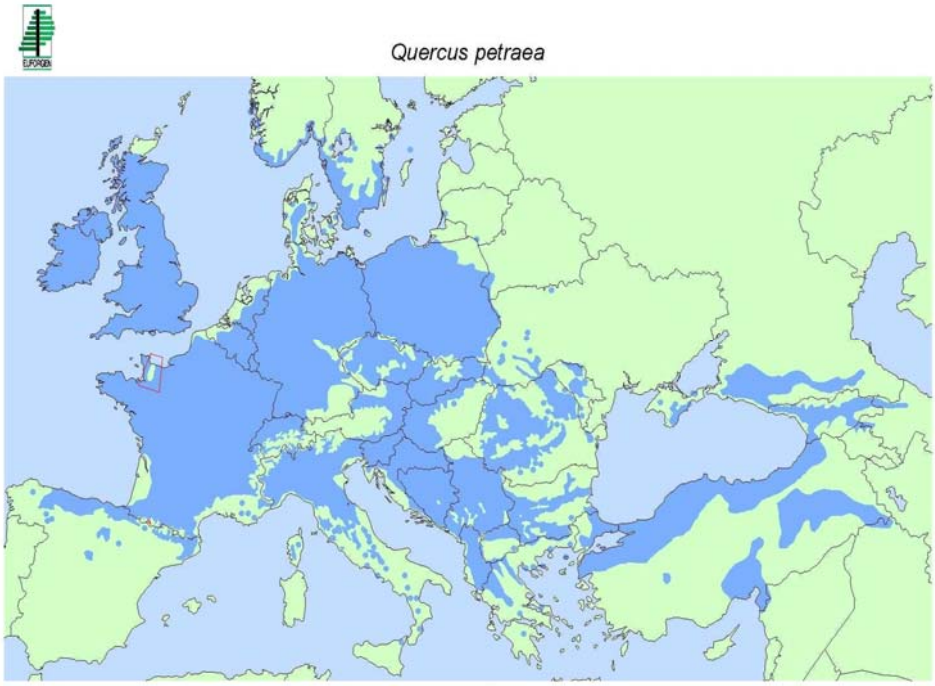


Figure 2 : Aire de répartition du chêne sessile (tiré de Ducousso *et al.* 2005).

De par leurs étendues, les chênaies représentent une composante majeure des forêts européennes sur le plan de la superficie, le chêne sessile (*Quercus petraea*) et le chêne pédonculé (*Quercus robur*) constituant à eux seuls la grande majorité des chênaies de la zone tempérée. Sur le continent européen, les chênaies sont considérées parmi les écosystèmes forestiers tempérés qui renferment la plus grande biodiversité en nombre d'espèces associées (*cf.* Figure 3).

Le chêne sessile forme avec les chênes pédonculé et pubescent un complexe d'espèces qui se caractérisent par des possibilités d'hybridation importantes (Lepais *et al.* 2009). Ces trois espèces se rencontrent fréquemment en peuplements mixtes où comprenant également de nombreux arbres aux phénotypes intermédiaires issus d'hybridation et dont le statut taxonomique est difficilement identifiable (Bacilieri *et al.* 1996). Alors que le chêne pubescent est limité au pourtour du bassin méditerranéen (espèce plutôt xérophile), le chêne sessile et le chêne pédonculé possèdent tous deux une très large amplitude géographique.

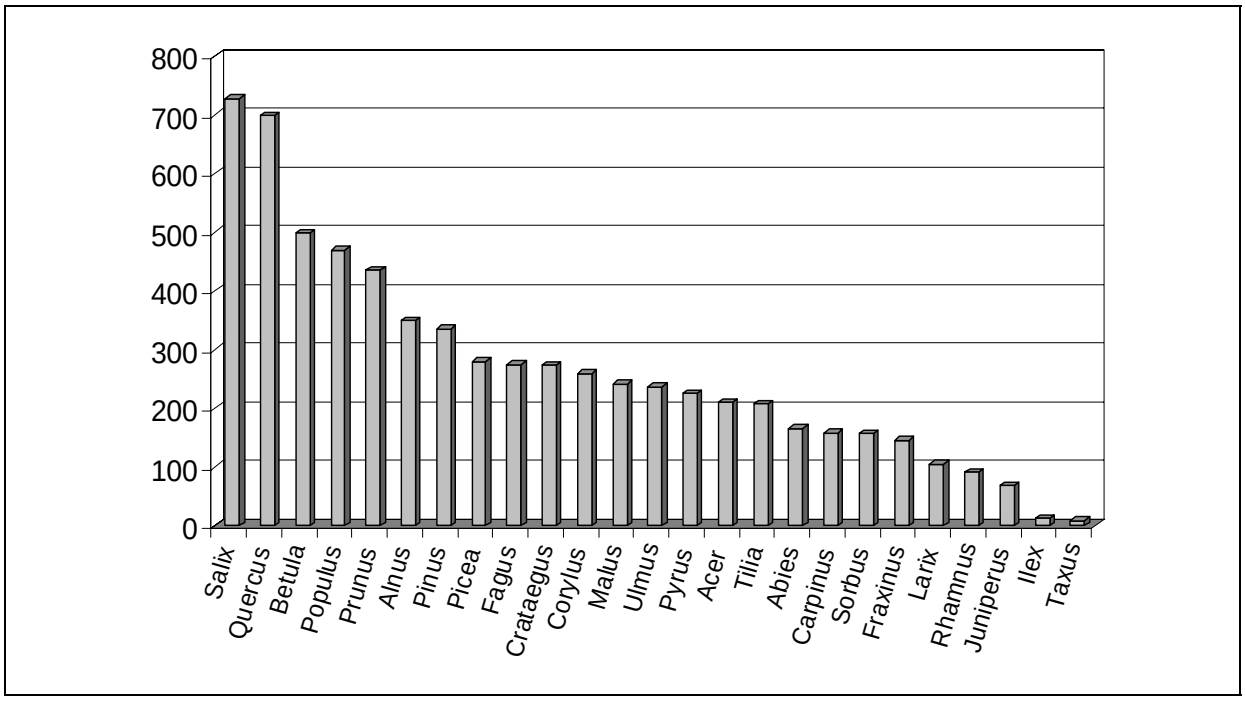


Figure 3 : Nombre d'espèces d'insectes associés à des espèces ligneuses (adapté de Brändle *&* Brandl 2001)

Ces deux espèces se distinguent par des affinités écologiques différentes, principalement au niveau de l'alimentation en eau et de la richesse du sol. Le chêne pédonculé occupe généralement des zones relativement humides à sol fertile assez profond, comme les fonds de vallées. Il est plus fréquent également dans les milieux ouverts et les lisières de peuplements. Le chêne sessile est une espèce méso-xérophile qui tolère relativement bien la sécheresse et peut se retrouver sur des sols bien drainés comme par exemple sur les collines et les plateaux. Il peut également se rencontrer à des altitudes élevées, par exemple jusqu'à 1800 m dans les Pyrénées.

Historique des études menées à l'UMR BIOGECO

Ces dernières années, la diversité génétique des chênes sessile et pédonculé a été évaluée à la fois pour des marqueurs nucléaires et chloroplastiques ainsi que pour des traits adaptatifs. L'analyse de la diversité spatiale du génome chloroplastique a permis de reconstituer les principales voies de colonisation empruntées par les chênes suite au dernier épisode de glaciation (Petit *et al.* 2002a, b). Chez les chênes, les chloroplastes ont la particularité de n'être transmis que par les graines. La structuration géographique des variants d'ADN chloroplastique a permis de regrouper des lignées qui montraient à la fois une similarité génétique et une proximité géographique. La distribution spatiale de ces lignées témoigne

ainsi de la recolonisation de l'Europe par les chênes à partir de zones appelées refuges glaciaires (Figure 4). La diversité génétique de marqueurs izozymes a également été étudiée à l'échelle de l'aire de répartition chez le chêne sessile (Zanetto & Kremer 1995).

Cette étude a montré une très faible différenciation génétique entre les forêts pour ces marqueurs. Le génome nucléaire étant transmis à la fois par le pollen et par les graines, le faible niveau de différenciation génétique observé pour les marqueurs nucléaires indique que des flux de pollens intenses se sont produits entre forêts depuis la recolonisation. Ces flux de pollens ont donc permis d'effacer les effets de structuration hérités de la recolonisation et d'homogénéiser ainsi la composition génétique du génome nucléaire neutre à l'échelle du continent. L'étude réalisée sur des traits adaptatifs en tests de provenances a montré que les populations avaient été soumises à des pressions environnementales locales depuis leur établissement (Kremer *et al.* 2002). Ces pressions sélectives ont entraîné une différenciation génétique pour les traits adaptatifs indépendante de l'histoire évolutive des populations (et donc indépendante des lignées chloroplastiques). La sélection naturelle au cours de l'holocène semble donc avoir totalement redistribué la composition génétique pour ces traits en fonction des facteurs sélectifs qui les gouvernent. L'étude réalisée par Ducousso *et al.* (1996) montre que chez le chêne sessile, le débourrement est contrôlé génétiquement et qu'il présente une différenciation génétique le long d'un gradient latitudinal. Ainsi, dans les tests de provenances, les populations méridionales débourrent plus tôt que celles du nord de l'aire de répartition. Dernièrement les efforts de recherche se sont accentués sur le déterminisme génétique du débourrement. Une étude de cartographie génétique a ainsi permis de déterminer le nombre et la position de QTL du débourrement (« Quantitative Trait Loci ») et leurs effets sur la variabilité du trait en population de ségrégation (Scotti-Saintagne et al. 2004 ; Derory *et al.* 2010). Plus récemment, une étude transciptomique a identifié les gènes dont l'expression est régulée au cours des différents stades du débourrement (Derory et al. 2006). Ces données ont pu servir de base pour des études de diversité génétique sur un sous-ensemble de gènes « candidats » présentant des critères de choix intéressant (Derory *et al.* 2010). Enfin les travaux de Vitasse (2009a, b) ont permis de comparer la réponse phénologique de six espèces ligneuses dont le chêne sessile le long d'un gradient altitudinal. Les patrons de variabilité du débourrement et de la sénescence ont ainsi pu être mesurés en populations naturelles le long de deux gradients altitudinaux s'étalant entre 130 et 1630 m dans deux vallées du versant nord occidental des Pyrénées (Vitasse *et al.* 2009a).

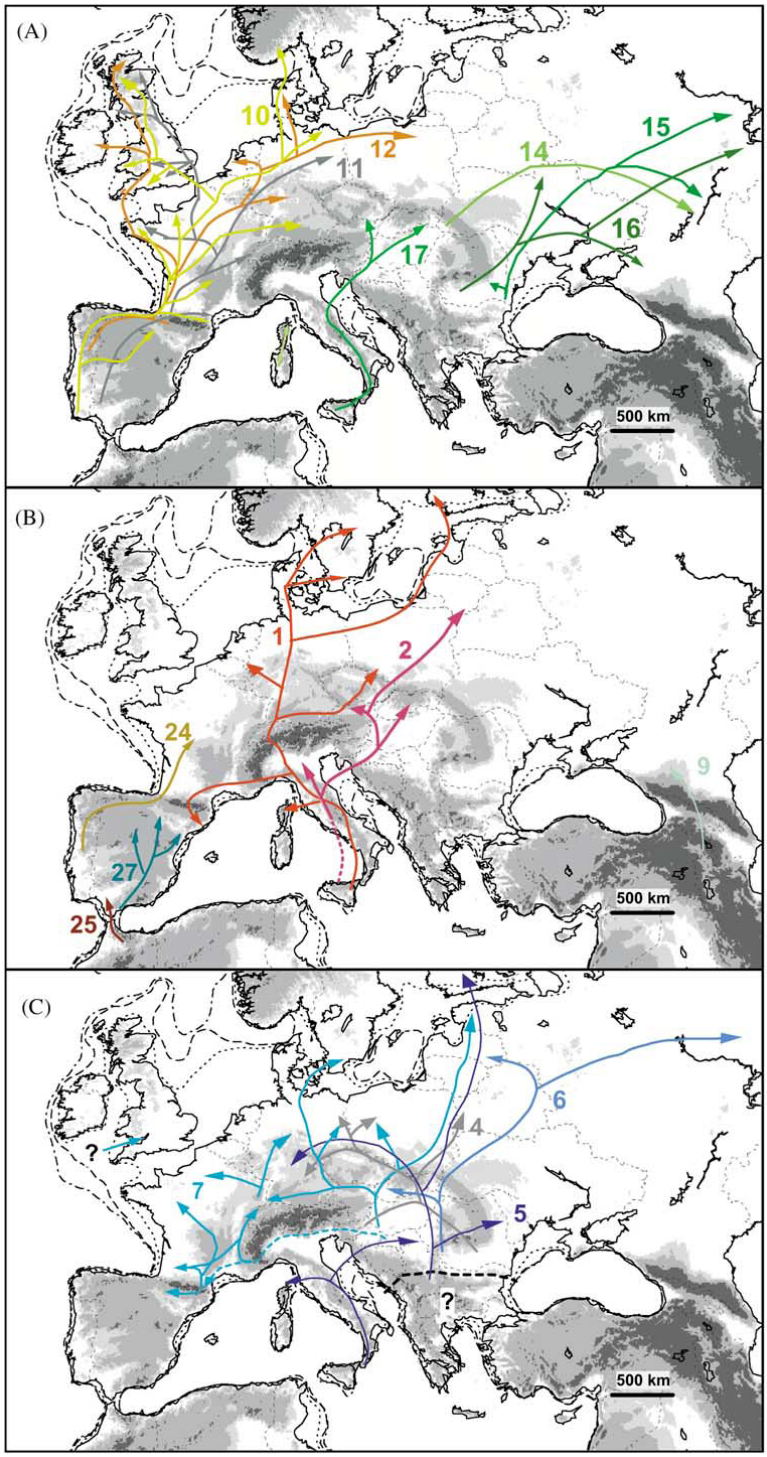


Figure 4 : Routes de colonisation post glaciaire des principaux haplotypes d'ADN chloroplastique des chênes blancs en Europe (tiré de Petit *et al.* 2002a).

L'altitude est indiquée par les ombres grisées (250–500, 500–1000, >1000 m) et les niveaux d'élévation de la mer à 21 ka BP (18 14C ka BP), 15 ka BP (13 14C ka BP) and 12 ka BP (10 14C ka BP) sont indiqués par les lignes pointillées. (A) Lignées B (haplotypes 10–12) and E (haplotypes 14–17). (B) Lignées B (24–25), C (1–2), D (27) and F (9). (C) Lignée A (haplotypes 4–7).

Un test de provenances a été installé afin de mesurer la différenciation génétique entre populations pour la phénologie foliaire et a révélé une faible adaptation locale du débourrement chez le chêne sessile suivant un cline, les populations de hautes altitudes étant plus tardives que celles de basses altitudes (Vitasse *et al.* 2009b).

Plan de la thèse

Les travaux effectués dans le cadre de cette thèse se placent dans la continuité des précédentes études menées au sein de l'UMR BIOGECO. Les résultats accumulés pour le débourrement chez le chêne sessile font de ce trait phénologique un objet d'étude idéal pour rechercher les bases moléculaires qui sous-tendent son adaptation. Le débourrement présente un fort déterminisme génétique et une différenciation modérée à forte en populations naturelles (Ducousso *et al.* 1996 ; Vitasse *et al.* 2009b). L'architecture génétique du trait montre qu'un nombre important de locus avec des effets relativement faibles contrôlent sa variabilité (Scotti-Saintagne *et al.* 2004, Derory *et al.* 2010). Enfin la séquence de nombreux gènes issus de l'étude transcriptomique (Derory *et al.* 2006) ou des bases de données de séquences publiques permet d'adopter une approche de gènes candidats pour rechercher les bases moléculaires de l'adaptation pour le débourrement.

Pour étudier l'adaptation pour le débourrement, l'échantillonnage de populations naturelles localisées le long d'un fort gradient environnemental qui permet d'exacerber la différenciation entre populations semble être la stratégie la plus appropriée. Nous avons donc choisi de nous focaliser sur deux gradients de température générés par un gradient altitudinal et un gradient latitudinal. Le gradient altitudinal est composé de 12 populations situées dans les Pyrénées et représente le cadre d'étude principal de cette thèse (Figure 5). Le gradient latitudinal est lui composé de 21 populations échantillonnées sur une grande partie de l'aire de répartition du chêne sessile (Figure 6) et sert de point de comparaison avec le gradient altitudinal pour la recherche des bases moléculaires de l'adaptation pour le débourrement. Ces populations ont été précédemment étudiées pour la diversité génétique de marqueurs supposés neutres (Mariette *et al.* 2002) et pour l'adaptation locale pour le débourrement (Ducousso *et al.* 2005).

L'objectif de cette thèse est donc de rechercher les signatures phénotypiques et moléculaires de l'adaptation qui a eu cours depuis l'établissement de ces populations. Le contexte temporel de cette adaptation est bien connu : en effet les données palynologiques et phylogéographiques ont montré que les populations de chênes situées sur le gradient latitudinal sont en place depuis 6000 ans (Brewer *et al.* 2002, Petit *et al.* 2002a). De la même

manière les premières apparitions de populations de chênes situées à plus de 1600 mètres remontent à 10 000 ans dans les Pyrénées. Des recherches récentes ont par ailleurs montré que la différenciation génétique actuelle de la phénologie n'est pas attribuable à des événements évolutifs qui se seraient produits antérieurement à leur établissement (Kremer *et al.* 2010). Cette thèse vise donc à faire une étude rétrospective des événements et processus microévolutifs qui ont accompagné la différenciation de ces populations de chêne depuis leur installation au travers du décryptage des empreintes que ces événements ont laissé dans le génome (gènes candidats ou marqueurs « neutres ») ou dans le phénotype (date de débourrement du bourgeon apical). Cette analyse rétrospective a pour objectif également d'élaborer une vision prospective sur l'évolution future des forêts, en considérant que les mêmes mécanismes qui ont agi dans le passé seront à l'œuvre dans l'avenir

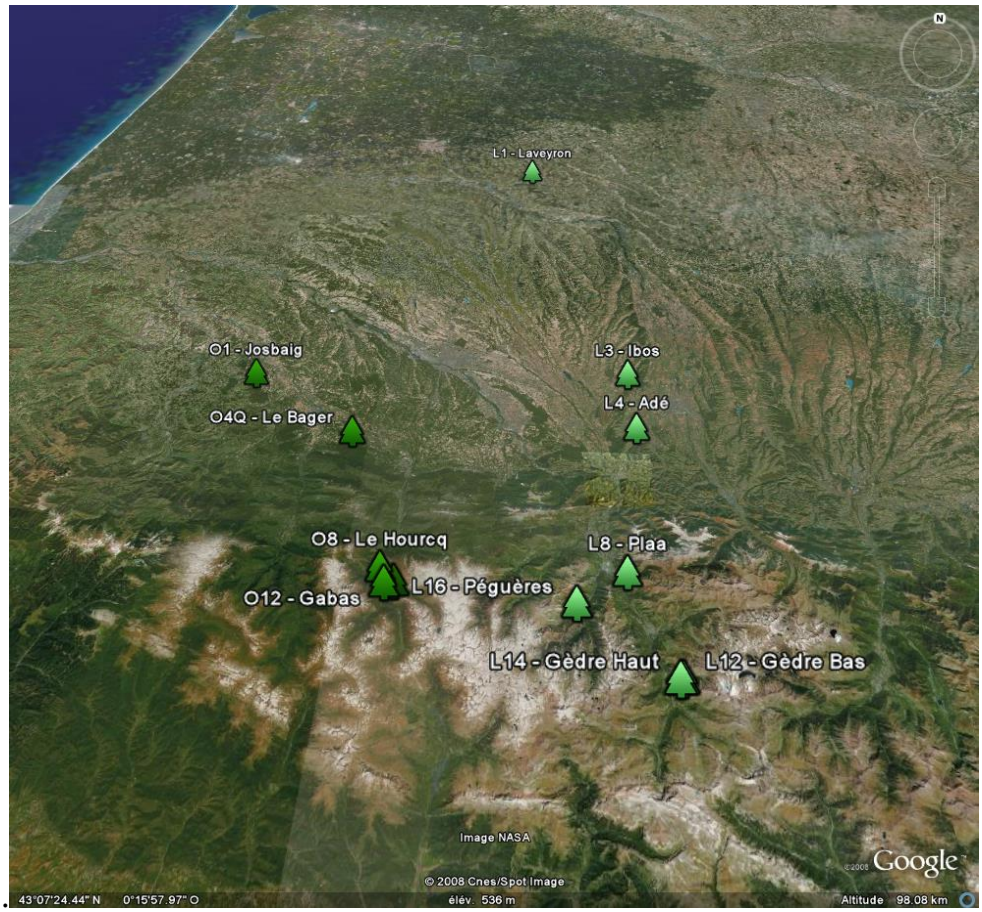


Figure 5 : Localisation des 12 populations du gradient altitudinal au sein des deux vallées pyrénéennes.

Figure 6 : Localisation des populations du gradient altitudinal (tiré de Ducousso *et al.* 2005)

Les différentes approches mises en œuvre dans le cadre de cette thèse peuvent donc se résumer ainsi :

(iv) Mesurer l'empreinte de la démographie sur la structure de la diversité génétique neutre des populations du gradient altitudinal.

(v) Mettre en évidence l'adaptation locale pour le débourrement le long du gradient altitudinal et la variabilité génétique à l'intérieur des ces populations.

(vi) Etudier les niveaux de diversité génétique des gènes candidats du débourrement sur les populations des gradients altitudinal et latitudinal et rechercher des signatures moléculaires de la sélection naturelle sur ces gènes.

(vii) Réaliser une étude d'association entre la diversité génétique du débourrement et la diversité génétique des gènes candidats. Cette étape constitue l'aboutissement des études précédentes puisqu'elle permet de tester statistiquement l'effet de la diversité génétique observée au niveau moléculaire sur la diversité génétique du trait.

Dans le Chapitre I des marqueurs microsatellites nucléaires et des marqueurs chloroplastiques sont utilisés pour analyser l'impact de l'altitude sur les niveaux de diversité génétique neutre et de différenciation, d'un point de vue à la fois rétrospectif par rapport à la colonisation du gradient suite au dernier épisode glaciaire, mais également prospectif dans le cadre de la réponse de ces populations aux changements climatiques en cours. Cette étude a également pour but de déterminer si l'empreinte démographique de ces populations ne risquerait pas de perturber les signatures de sélection détectées au sein des gènes candidats du débourrement.

Le Chapitre II traite de la mise en place de plusieurs tests de provenances à partir de descendances récoltées dans les populations naturelles du gradient altitudinal dans le but de mesurer la variabilité génétique du débourrement à l'intérieur des populations et l'adaptation locale le long du gradient. Par ailleurs nous avons profité de ce dispositif pour tester si le déterminisme génétique de la germination des graines était corrélé à celui du débourrement, ces deux traits présentant tous deux une phase de dormance précédant la croissance et l'expression de gènes identiques qui laissaient supposer une régulation commune.

Le Chapitre III porte sur la diversité nucléotidique des gènes candidats du débourrement dans les populations du gradient altitudinal et latitudinal et se divise en deux sous-parties. La première vise à évaluer les niveaux de diversité nucléotidique au sein de 9 fragments de gènes candidats ainsi que le déséquilibre de liaison global entre les sites polymorphes détectés au sein de ces gènes. La deuxième partie consiste en la recherche de signatures de sélection dans chacun des deux gradients pour un sous-ensemble de sites polymorphes représentatifs de 73 gènes candidats. La détection de ces signatures de sélection est basée à la fois sur les niveaux de différenciation génétique des sites polymorphes et sur la recherche de variations clinales de leurs fréquences allèliques.

Enfin le Chapitre IV présente la première étude d'association réalisée chez le chêne entre la variabilité du débourrement et le polymorphisme des gènes candidats. Cette étude est conduite indépendamment pour les populations du gradient altitudinal et celles du gradient latitudinal et permet de comparer les bases moléculaires de l'adaptation génétique pour le débourrement entre ces populations.

REFERENCES

Aitken SN, Yeaman S, Holliday JA, Wang TL, Curtis-McLane S (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* **1**, 95-111.

Bacilieri R, Ducousso A, Kremer A (1996) Comparison of morphological characters and molecular markers for the analysis of hybridization in sessile and pedunculate oak. *Annals of Forest Science* **53**, 79-91.

Beaubien EG, Freeland HJ (2000) Spring phenology trends in Alberta, Canada: links to ocean temperature. *International Journal of Biometeorology* **44**, 53-59.

Brändle M, Brandl R (2001) Species richness of insects and mites on trees: expanding Southwood. *Journal of Animal Ecology*. **70**, 491-504

Brewer S, Cheddadi R, de Beaulieu JL, Reille M (2002) The spread of deciduous Quercus throughout Europe since the last glacial period. *Forest Ecology and Management* **156**, 27-48.

Callahan HS, Pigliucci M, Schlichting CD (1997) Developmental phenotypic plasticity: Where ecology and evolution meet molecular biology. Bioessays **19**, 519-525.

Davis MB, Shaw RG (2001) Range shifts and adaptive responses to Quaternary climate change. *Science* **292,** 673-679.

Derory J, Leger P, Garcia V, *et al.* (2006) Transcriptome analysis of bud burst in sessile oak (Quercus petraea). *New Phytologist* **170**, 723-738.

Derory J, Scotti-Saintagne C, Bertocchi E *et al.* (2010) Contrasting relationships between the diversity of candidate genes and variation of bud burst in natural and segregating populations of European oaks. *Heredity* (in press).

Ducousso A, Guyon JP, Kremer A (1996) Latitudinal and altitudinal variation of bud burst in western populations of sessile oak (*Quercus petraea* (Matt.) Liebl). *Annales Des Sciences Forestieres* **53**, 775-782.

Ducousso A, Louvet JM, Jarret P, Kremer A (2005). Geographic variations of sessile oaks in French provenance tests. In *Proceedings of the Joint Meeting of IUFRO working groups Genetic of oaks and Improvement and Silviculture of oaks*, ed. R. Rogers, A. Ducousso and A. Kanazashi, FFPRI (Forestry and Forest Products Research Institute) Scientific Meeting Report 3, Tsukuba, Japan, pp 128-138

Chuine I, Cour P (1999) Climatic determinants of budburst seasonality in four temperate-zone tree species. *New Phytologist* **143**,339-349.

Chuine I (2000) A unified model for budburst of trees. *Journal of Theoretical Biology* **207**, 337-347.

Gonzalez-Martinez SC, Krutovsky KV, Neale DB (2006) Forest-tree population genomics and adaptive evolution. *New Phytologist* **170**, 227-238.

Kramer K, Leinonen I, Loustau D (2000) The importance of phenology for the evaluation of impact of climate change on growth of boreal, temperate and Mediterranean forests ecosystems: an overview. *International Journal of Biometeorology* **44**, 67-75.

Kremer A, Kleinschmit J, Cottrell J, *et al.* (2002) Is there a correlation between chloroplastic and nuclear divergence, or what are the roles of history and selection on genetic diversity in European oaks? *Forest Ecology and Management* **156**, 75-87.

Kremer A, Le Corre V, Petit RJ, Ducousso A (2010) Historical and contemporary dynamics of adaptive differentiation in European oaks. In : DeWoody A, Bickham J, Michler C, Nichols K, Rhodes G, Woeste K (eds) *Molecular approaches in natural resource Conservation* . Cambridge University Press (in press)

Lang GA, Early JD, Martin GC, Darnell RL (1987) Endo-, para-, and ecodormancy: physiological terminology and classification for dormancy research. *Hortscience* **22**,371-377.

Lepais O, Petit RJ, Guichoux E, *et al.* (2009) Species relative abundance and direction of introgression in oaks. *Molecular Ecology* **18**, 2228-2242.

Mariette S, Cottrell J, Csaikl UM, *et al.* (2002) Comparison of levels of genetic diversity detected with AFLP and microsatellite markers within and among mixed *Q. petraea* (Matt.) Liebl. and *Q. robur* L. stands. *Silvae Genetica* **51**, 72-79.

Menzel A, Fabian. P (1999) Growing season extended in Europe. Nature **397**, 659.

Menzel A, Sparks TH, Estrella N *et al.* (2006) European phenological response to climate change matches the warming pattern. *Global Change Biology* **12**, 1969-1976.

Petit RJ, Brewer S, Bordacs S, *et al.* (2002) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management* **156**, 49-74.

Petit RJ, Csaikl UM, Bordacs S, *et al.* (2002) Chloroplast DNA variation in European white oaks - Phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management* **156**, 5-26.

Schwartz MD (2003) *Phenology: An Integrative Environmental Science*. Kluwer Academic Publishers, Netherlands.

Scotti-Saintagne C, Bodenes C, Barreneche T, *et al.* (2004) Detection of quantitative trait loci controlling bud burst and height growth in Quercus robur L. *Theoretical and Applied Genetics* **109**, 1648-1659.

Vitasse Y, Delzon S, Dufrene E, *et al.* (2009a) Leaf phenology sensitivity to temperature in European trees: Do within-species populations exhibit similar responses? *Agricultural and Forest Meteorology* **149**, 735-744.

Vitasse Y, Delzon S, Bresson CC, Michalet R, Kremer A (2009b) Altitudinal differentiation in growth and phenology among populations of temperate-zone tree species growing in a common garden. *Canadian Journal of Forest Research* **39**, 1259-1269.

Zanetto A, Kremer A (1995) Geographical Structure of gene diversity in *Quercus petraea* (Matt) Liebl .1. Monolocus patterns of variation. *Heredity* **75**, 506-517.

Population differentiation of sessile oak at the altitudinal front of migration in the French Pyrenees

Alberto F. [1,2], Niort J. [1,2], Derory J. [1,2], Lepais O. [1,2], Vitalis R. [3] , Galop D. [4] and Kremer A. [1,2*]

[1] INRA, UMR 1202 Biodiversité Gènes et Communautés, Cestas, F-33610, France

[2] Université de Bordeaux, UMR1202 Biodiversité Gènes et Communautés, Talence, F-33400, France

[3] CNRS, UMR CBGP (INRA – IRD – CIRAD – Montpellier SupAgro), Campus International de Baillarguet, CS 30016, Montferrier-sur-Lez, F-34988, France

[4] CNRS, UMR 5602 Géographie de l'Environnement, Maison de la Recherche de l'Université du Mirail, 5 Allée A. Machado, Toulouse, F-31058, France

Accepted in *Molecular Ecology*

INTRODUCTION

Altitudinal gradients encompass contrasted environments and impose strong adaptive challenges for plants. At high altitude, environmental constraints as cold temperatures, frost and shorter growing season induce strong selection pressures that can affect reproduction and survival of populations (Premoli 2003). Because located at the ecological margins of the species, these peripheral populations could exhibit peculiar demographic and biological features as reduced population sizes or ability for clonal reproduction (Ettl & Peterson 2001; Premoli 2003). Isolation of populations can also be strongly increased by the topographical barriers and sharp variations of phenology between lower and higher altitudes. These physical and temporal constraints on gene flow can limit the recruitment and may consequently increase genetic drift. High altitude populations are not only exposed to extreme environmental conditions, but they also face genetic and demographic stochasticity, which may further challenge their survival (Quiroga & Premoli 2007). As a result strong genetic differentiation between high and low altitude populations may be expected not only for adaptive traits but also for neutral markers (Premoli 2003; Quiroga & Premoli 2007).

Interest for high altitude populations has risen recently, as they may be expanding at even higher altitudes due to climate change (Jump *et al.* 2006; Aitken *et al.* 2008). Surveys of plant inventories have indicated that during the last century their distributions have shifted about 29 meters per decade in altitude (Lenoir *et al.* 2008). Consequently, the populations that are currently located at the altitudinal tree species line constitute source populations for their altitudinal expansion. But the upslope shift of these populations as a response to changing climatic conditions will not only depend on their evolutionary potential (*i.e.* the genetic variability for adaptive traits) but will also depend on gene flow (Byars *et al.* 2007). Gene flow may be particularly important as alleles required for adaptation may actually be present at lower altitudes. In the context of climate change, evolutionary responses of populations at their range limits may therefore be dependent on alleles coming from warmer climates (Bridle & Vines 2009) contradicting the traditional view that gene swamping from central populations will actually limit species expansion (Kirkpatrick & Barton 1997). It is therefore of upmost importance to draw attention to the processes that shape the genetic diversity of populations located along altitudinal gradients (Engler *et al.* 2009). In Europe, as mountains are mostly present under southern latitudes, altitudinal migration is expected for a large spectrum of species whose southern distribution will be challenged by climatic changes (see

for example Peñuelas *et al.* 2007). European white oaks are designated case studies in this respect. Temperate white oaks (*Quercus petraea, Quercus robur)* are widely distributed across Europe from Spain to the Ural Mountains and their southern extant limits extend right at the southern slopes of European Mountains (Pyrenees, Alps, Carpathians), where they co-occur with some of the Mediterranean white oaks (mainly *Q. pubescens* and *Q. pyrenaica*). Along their post glacial migration from the glacial refugial zones (Petit *et al.* 2002b), the two temperate oak species have also expanded within valleys of European mountains (Csaikl *et al.* 2002). Among the four temperate white oak species present along the northern side of the Pyrenees Mountains *(Q. petraea, Q. robur, Q. pubescens* and *Q. pyrenaica*), forest inventory data indicate that only *Q. petraea* and *Q. robur* extend at higher altitudes and can reach up to 1800 meters. Relative abundance of *Q. petraea* increases above 1600 meters, where *Q. robur* is only occasionally present. *Q. petraea* and *Q. robur* had already reached these limits between 11,000 and 10,000 BP (Aubert 2001; Belet 2001). A recent study on hybridization dynamics among these four oak species showed that introgression is possible between all species but depends on species abundance (Lepais *et al.* 2009). A comparative study of various broadleaved species distributed in the Pyrenees Mountains has suggested that oaks may benefit more from climatic changes in terms of growing season length and are prone to faster altitudinal expansion than other species during the next decades (Vitasse *et al.* 2009a; Vitasse *et al.* 2009b).

To understand the processes that may stimulate future altitudinal expansion of oaks we have concentrated our attention to a retrospective analysis of the processes that have shaped the current genetic variation along altitudinal gradients in the Pyrenees. We chose to focus on *Q. petraea* which reaches higher altitudes than the other white oaks and monitored genetic diversity along the altitudinal distribution of this species. The investigations were replicated within two altitudinal transects corresponding to two valleys. We used chloroplast DNA markers and nuclear microsatellites to retrace the historical migration patterns along the valleys and to characterize the extant genetic differentiation among populations. We wanted first to identify historical and contemporary processes that may have proceeded along the whole altitudinal gradient and contributed to the current genetic structure of populations. Second, we attempted to check whether high altitude populations located at the margin of the species' range exhibit peculiar traits that would facilitate their maintenance or future expansion as climate change is proceeding.

MATERIAL AND METHODS

Study area and sampling

Twelve populations of *Q. petraea* were sampled along two parallel valleys (Ossau and Luz) on the Northern side of the Pyrenees Mountains (Figure 1a and b). These two valleys are approximately 30 km apart and extend South-North. Oaks are spread along the northern side of the Pyrenees and occur in continuous stands at lower altitudes and in scattered isolated stands at their higher limits. Four species are present with relative abundances decreasing from *Q. robur*, *Q. petraea*, *Q pubescens*, to *Q. pyrenaica*. *Q. pyrenaica* extends only in lowlands, while *Q. pubescens* spreads throughout the foothills below 800 meters (http://www.ifn.fr). The altitudinal range of pedunculate oak (*Q. robur*) extends within the Pyrenees valleys up to 1600 m, with occasional stands present up to 1800 m. Similarly the current altitudinal distribution of sessile oak (*Q. petraea*) ranges from quite continuous forests in the plains with tall trees of almost 30 m to stands of smaller trees distributed in scattered populations at high altitudes (up to 1800 m), and is more frequent at higher altitudes than *Q. robur*. We explored the two valleys and used the data of the French National Inventory (http://www.ifn.fr) to sample the study populations. In each valley, we sampled natural populations distributed along an altitudinal gradient varying between 131 and 1630 m (Table 1 and Figure 1) and over a distance of 80 km. At least 20 individuals were sampled in each population except in populations O8 and O16 where only 10 and 14 trees respectively were available. While the oak study populations at lower altitudes were sampled from widespread stands (a few hundreds of hectares), study populations at higher altitudes were sampled within scattered patches of a few hectares and in two cases the sampling was exhaustive (O8 and O16). The sampling was therefore representative of the overall distribution of oaks in the Pyrenees valleys. We collected either leaves in mid-summer 2004 or buds in winter 2005 on a total of 295 adult trees distributed among the 12 sampled populations. The taxonomic status of each tree was checked based on the leaf morphology (Kremer *et al.* 2002) assessed during the growing season on fully elongated leaves. DNA was extracted from buds and leaves following a protocol adapted from Doyle & Doyle (1990).

a)

b)

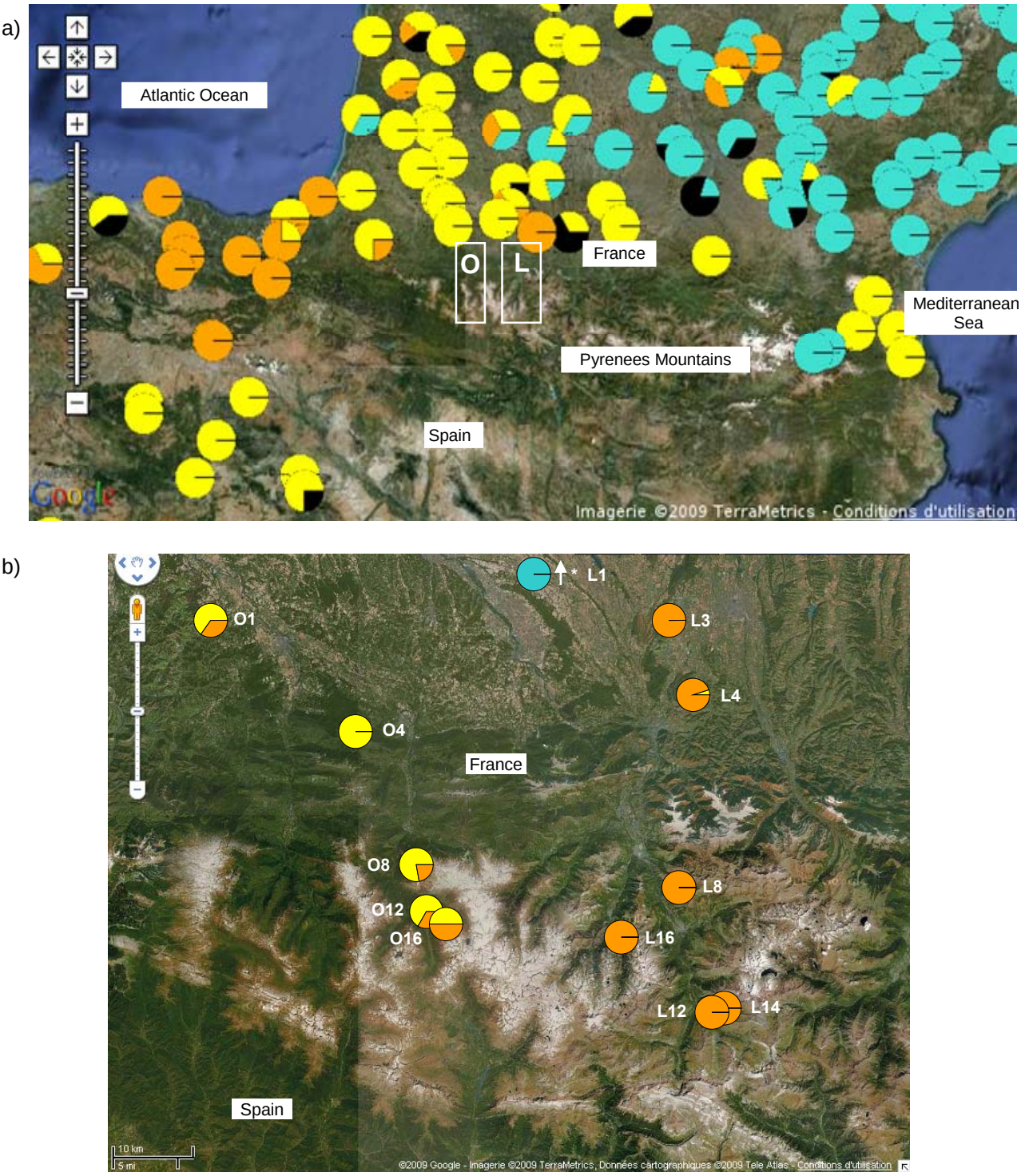


Figure 1: Geographic maps of chloroplast haplotype frequencies a) localisation of the two studied Pyrenean valleys and distribution of chloroplast haplotypes along the Pyrenees foothills (adapted from Petit *et al.* (2002a) and b) localisation and chloroplast haplotype frequencies of the studied populations.

Pie charts represent the haplotype frequencies in the populations sampled in Petit *et al.* (2002a).
Blue: haplotype 7. Yellow: haplotype 10. Orange: haplotype 12. Black: any other haplotypes bulked together. The two white rectangles indicates the location of the two valleys in which populations were sampled (O: Ossau valley; L: Luz valley).
* L1 is located approximately 50 km north of the position indicated on the map.

Table 1: Description of the 12 natural populations sampled.

Valley	Site	Code	Sample size	Altitude (m)	Latitude	Longitude	Group
Luz valley	**Laveyron**	**L1**	20	131	43°45' N	00°13' W	LL
	Ibos	**L3**	25	387	43°15' N	00°00' W	
	Adé	**L4**	40	427	43°08' N	00°00' W	
	Chèze	**L8**	22	803	42°55' N	00°02' W	
	Gèdre	**L12**	32	1235	42°47' N	00°01' E	LH
	Gèdre haut	**L14**	23	1349	42°47' N	00°02' E	
	Péguère	**L16**	25	1630	42°52' N	00°07' W	
Ossau valley	**Josbaig**	**O1**	28	259	43°13' N	00°44' W	OL
	Bager	**O4**	22	422	43°07' N	00°32' W	
	Le Hourcq	**O8**	10	841	42°54' N	00°26' W	
	Gabas	**O12**	34	1194	42°53' N	00°25' W	OH
	Artouste	**O16**	14	1614	42°53' N	00°24' W	

Groups meanings:
LL: Luz valley low altitudes. LH: Luz valley high altitudes. OL: Ossau valley low altitudes. OH: Ossau valley high altitudes.

Chloroplast haplotypes genotyping

In this study we used a set of five chloroplast markers (µdt1, tf42, dt73b, dt13, dt74b) described by Deguilloux *et al.* (2003). These markers were initially developed in oaks to reconstruct the colonization pathway of *Quercus sp.* after the last glacial episode (Petit *et al.* 2002a) and allow identifying the 8 major haplotypes present in France (Petit *et al.* 2002c). We used the same scoring as in Petit *et al.* (2002c) and Deguilloux *et al.* (2003).

Nuclear microsatellites genotyping

A total of 16 nuclear microsatellite markers previously developed on oaks species (Dow *et al.* 1995; Steinkellner *et al.* 1997; Kampfer *et al.* 1998) were used in this study (QpZAG110, QrZAG11, QrZAG112, QrZAG39, QrZAG96, QrZAG20, QrZAG5, QrZAG65, QrZAG7, QrZAG87, QpZAG15, QpZAG46, QpZAG9, QpZAG1/5, QpZAG36, MSQ13). These dinucleotide microsatellite markers are located on 9 linkage groups of the *Quercus* map (Barreneche *et al.* 2004), and were considered as unlinked. They are anonymous markers, *i.e.* we do not know whether they are located in coding or non-coding regions of the genome. Among these 16 microsatellites, 10 markers were amplified in two multiplex kits following

Lepais *et al.* (2006) and the other 6 markers were grouped in two distinct multiplex kits (QpZAG15 with QpZAG46, QpZAG9, and QpZAG1/5, and QpZAG36 with MSQ13). Polymerase Chain Reaction (PCR) amplifications were all performed in the conditions described by Lepais *et al.* (2006). PCR product electrophoresis were performed with a Megabace1000 multicapillary sequencer (Amersham Biosciences Molecular Dynamic, Uppsala, Sweden) and the genotypes were scored and visually controlled using the Fragment Profiler software version 1.2 provided by the manufacturer.

Genetic diversity and population differentiation

The frequencies of each chloroplast haplotype were calculated in each population.

Gene diversity statistics (gene diversity H_e (Nei, 1987) and allelic richness A) were estimated for nuclear microsatellites using the program Fstat version 2.9.3.2 (Goudet 2001). Allelic richness (A) was calculated using the rarefaction method developed by El Mousadik *&* Petit (1996a). Rarefaction was done based on a minimum sample size of 9 diploid individuals, which corresponded to the smallest number of individuals successfully genotyped for a given locus in a population. Pairwise differences of genetic diversity and allelic richness were tested by means of a paired t-test, following the method of Zhang *&* Allard (1986) with the statistical program R (R Development Core Team 2005). We performed Bonferroni corrections to account for multiple testing.

For both the chloroplast haplotypes and the nuclear microsatellites we measured among population differentiation as Wright's F_{ST} according to Weir *&* Cockerham (1984) by performing a locus by locus analysis of molecular variance (AMOVA) using Arlequin version 3.01 (Excoffier *et al.* 2005). We calculated the overall F_{ST} and the F_{ST} between populations within each valley. A hierarchical AMOVA was used to partition the genetic variation at three levels, among all populations (F_{ST}), among populations within valleys (F_{SV}) and among valleys (F_{VT}).

Selection imprints

We used the software package BayeScan version 1.0 (Foll *&* Gaggiotti 2008) to test for the neutrality of the microsatellite markers. BayeScan extends Beaumont *&* Balding's (2004) Bayesian method that implements the multinomial-Dirichlet likelihood, which arises in a wide range of neutral population genetics models. The rationale is to discriminate between the effects on F_{ST} values that are specific to each population and to each locus. A locus is suspected to be under selection if the locus-specific effect is significantly different from zero.

The method developed by Foll & Gaggiotti (2008) uses a hierarchical Bayesian approach to estimate the posterior probabilities of two alternative models, one including the effects of selection and one excluding it. The results are expressed as a Bayes factor which indicates for each locus the ratio of posterior probability of the selection model against the neutral model and can be translated as different levels of evidence of selection according to Jeffreys' interpretation (1998). We ran BayeScan with 10 pilot runs with a burn-in of 500,000 followed by 50,000 iterations each, a sample size of 50,000 and a thinning interval of 20.

For each outlier locus, we examined the overall F_{ST} and the F_{ST} within each valley and we inspected the altitudinal variations of allelic frequencies along the two altitudinal gradients in order to check for potential clinal patterns of variation in both valleys. The loci departing significantly from neutral expectation were removed from further analyses of genetic structure and gene flow (see below).

Genetic structure of populations

We used the Bayesian clustering method implemented in Structure version 2.3.1 (Pritchard *et al.* 2000) to determine the genetic structure of the sampled populations. We used the new clustering method implemented in this version which is not only based on the individual multilocus genotypes but also takes into account the sampling locations (Hubisz *et al.* 2009). The LocPrior model considers that the prior distribution of cluster assignments can vary among populations. This approach is recommended by the authors when the genetic data are not very informative to help the detection of population structure. A parameter r indicates the extent to which the sampling locations are informative (small values < 1 indicate that locations are informative). We conducted five independent runs for each value of K (the number of putative clusters) ranging from 1 to 12 (the number of populations sampled). We used the admixture model with sampling locations as prior information (Hubisz *et al.* 2009) and assumed correlated allele frequencies among populations (Falush *et al.* 2003). Each run consisted in 500,000 burn-in steps followed by 10^6 iterations. To determinate the most likely number of clusters we used the method developed by Evanno *et al.* (2005), based on an *ad hoc* statistic, ΔK, which depends on the rate of change in the log probability of the data between successive values of K and we also followed the recommendations of the software documentation (Pritchard *et al.* 2009). Once the most likely K value was determined, we chose the run with the higher posterior probability and lower variance for interpreting results.

Gene flow and population sizes

We used the Bayesian approach implemented in Migrate version 3.0.3 (Beerli & Felsenstein 2001; Beerli 2006) to assess the direction and amount of gene flow among populations. Based on the genetic data this software estimates the scaled effective population size ($\theta = 4N\mu$) for each population and the scaled immigration rates ($M = m/\mu$) between each pair of populations. The program Migrate calculates the posterior probability distribution of these parameters, using a Metropolis-Hastings algorithm to explore all possible genealogies. A coalescent simulation explores the likelihood space for θ and M. Within each valley we grouped populations in two groups according to altitude (Table 1) which limited the number of parameters to estimate. We performed Migrate using the Brownian motion model with starting values of θ and M generated from the F_{ST} estimates, exponential window prior for both parameters (min = 0, mean = 5, max = 50 and $\Delta = 5$ for θ; min = 0, mean = 5, max = 50 and $\Delta = 5$ for M), slice sampling (Neal 2003) for the proposal distribution, and mutation rates calculated from data for each locus. We replicated 2 long chains of 10^6 genealogies recorded every 20 steps after a burn-in period of 20,000. Then we used the resulting θ and M estimates as starting values for another run in order to test the reliability of the analysis. We repeated this procedure until the starting parameters and the resulting estimates were congruent. If necessary we adjusted the Δ values of the exponential window priors to get a better sample of the parameter landscapes.

Introgression

We further used the software Structure for species assignment and introgression analysis since morphological variation is multimodal but overlapping across trees of different species (Kremer *et al.* 2002) and introgressed trees could not be excluded. To check for introgression, the sampled trees were included in a larger sample of 2107 oak trees representing the 4 oak species present in the study area (*Q. petraea*, *Q. robur*, *Q. pubescens*, *Q. pyrenaica*) using a subset of 10 common microsatellites as described in Lepais *et al.* (2009). Bayesian clustering of the trees was achieved by using Structure version 2.1 (Pritchard *et al.* 2000; Falush *et al.* 2003) and results were analyzed for the most likely number of clusters ($K = 4$). Each species was represented by one cluster and the admixture coefficient Q, corresponding to the assignment probability of each tree to each cluster (Pritchard *et al.* 2000), was used to infer its taxonomic status. Following Lepais *et al.* (2009), we classified individuals as pure species for $Q > 0.90$ (or $Q < 0.90$ for one cluster but $Q < 0.10$ for the three others) and as hybrids for $Q < 0.90$.

RESULTS

In the two higher altitude populations (L16 and O16), a few trees shared the same multilocus genotype for the 16 loci and were assumed to be clonal copies. Two multilocus genotypes were found in two exact copies in the L16 population and 4 copies of a single genotype were found in the O16 population. In the L16 population trees sharing the same genotype were separated by one meter, but in O16 population they were located in a sloping avalanche corridor and separated by five meters each. In addition one tree exhibited a tri-banded electrophoregram at 10 loci in L3 population, which suggested a natural triploid genotype (Dzialuk *et al.* 2007). These 7 trees were excluded for the genetic analyses, leading to a final dataset of 288 genotypes.

Chloroplast DNA diversity and differentiation

A total of 280 chloroplast haplotypes could be reconstructed, out of 288, due to 8 incomplete genotypes. Three chloroplast haplotypes were identified over the 12 populations (haplotype 7, 10 and 12 according to Petit *et al.* 2002a, see also Figure 1b). According to these authors who carried out a continental wide survey of cpDNA variation, haplotypes 10 and 12 belong to the western lineage (lineage B), and haplotype 7 to an eastern lineage (lineage A). These haplotypes were unevenly distributed among the two valleys and populations (Table 2). Five out of seven populations of the Luz valley were fixed for haplotype 12 and one (L1) for haplotype 7. The remaining population of the Luz valley (L4) was almost fixed for haplotype 12 (38 out of 40 trees carried haplotype 12). In the Ossau valley, one population (O4) was fixed for haplotype 10 and the remaining populations were all of mixed composition with haplotypes 10 and 12 distributed roughly as 70 to 30% in each stand. Hence while the populations of the Luz valley were almost all fixed for two haplotypes and thus highly differentiated the level of differentiation between populations in the Ossau valley was extremely low (Table 3).

Nuclear genetic diversity and allelic richness

Oak populations exhibited high levels of gene diversity across all loci (Table 4). Fifteen of the 16 microsatellites presented high levels of genetic diversity (from 0.757 to 0.937) and allelic richness (from 6.62 to 12.43). Only locus QrZAG112 showed significantly lower levels of genetic diversity (0.372) and allelic richness (3.388). After Bonferroni corrections, no pairwise comparisons of the average level of genetic diversity or the allelic richness were significant. However, both parameters were of similar magnitude across populations except

for the O4 population, in which we found the lowest values as compared to the other populations (Table 2). No difference of genetic diversity was noticeable between the two valleys.

Nuclear genetic differentiation among populations

The partitioning of genetic variance into different components (between and within valleys) using AMOVA, indicated that differentiation resides rather between populations within valleys than between the two valleys. While the differentiation among all 12 populations amounted to 0.023, F_{VT} value between the two valleys amounted only to 0.003 ($p < 0.01$; Table 3). Differentiation was however higher within Ossau valley than within Luz valley (F_{ST} = 0.028 and 0.018 respectively). Larger differentiation within Ossau valley was mainly caused by the divergence of O4 from the other populations as indicated by pairwise F_{ST} values (data not shown). The BayesScan analysis revealed that two loci, QrZAG39 and QpZAG15, presented patterns of genetic differentiation among populations that were not compatible with neutral expectations (Table 4). For QrZAG39, the model including selection effects was clearly favoured (BF = 270.7) corresponding to a decisive evidence of selection according to Jeffrey's interpretation (Jeffreys 1998), whereas QpZAG15 presented a substantial evidence of selection (BF = 7.7). The Bayes factors of the remaining 14 markers were lower than 1.3 which did not witness any selection signature.

Table 2: Frequencies of the chloroplast haplotypes and genetic diversity (H_e) and allelic richness (A) of the nuclear microsatellites.

	Chloroplast haplotypes frequencies			Nuclear microsatellites diversity	
	7	**10**	**12**	H_e	A
L1	1.00	0.00	0.00	0.832	7.847
L3	0.00	0.00	1.00	0.820	7.823
L4	0.00	0.05	0.95	0.819	7.917
L8	0.00	0.00	1.00	0.837	8.231
L12	0.00	0.00	1.00	0.836	7.929
L14	0.00	0.00	1.00	0.829	7.653
L16	0.00	0.00	1.00	0.836 (0.833)	8.300 (8.133)
O1	0.00	0.65	0.35	0.819	7.896
O4	0.00	1.00	0.00	0.765	6.766
O8	0.00	0.78	0.22	0.823	7.136
O12	0.00	0.67	0.33	0.812	7.379
O16	0.00	0.50	0.50	0.844 (0.821)	8.089 (7.211)
Overall	0.06	0.26	0.68	0.822	8.400

Allelic richness (A) was computed using the rarefaction method (El Mousadik *&* Petit, 1996a) based on 9 individuals. Values between parentheses correspond to the genetic diversity (H_e) and allelic richness (A) of populations L16 and O16 calculated when clonal copies are included in the sample size.

Table 3: Genetic differentiation statistics.

	F_{SV}	F_{VT}	F_{ST} Luz	F_{ST} Ossau
Nuclear microsatellites	0.022 ***	0.003 **	0.018 ***	0.028 ***
Chloroplast haplotypes	0.509 ***	0.592 **	0.901 ***	0.106 *

F_{SV}: differentiation among populations within valleys. F_{VT}: differentiation between the Luz and Ossau valley. F_{ST} Ossau: differentiation among populations within the Ossau valley. F_{ST} Luz: differentiation among populations within the Luz valley.Differentiation for nuclear microsatellites and chloroplast haplotypes was estimated by AMOVA using Arlequin. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

Table 4: Genetic statistics and neutrality test performed for the 16 nuclear microsatellites.

Locus	***A***	***H_e***	**Overall *F_{ST}***	***F_{ST}* Luz**	***F_{ST}* Ossau**	**BF**
QpZAG110	8.005	0.820	0.035	0.029	0.053	0.2
QrZAG11	6.617	0.757	0.018	0.021	0.013	0.5
QrZAG112	3.388	0.372	0.015	0.002	0.029	0.5
QrZAG39	9.357	0.844	0.047	0.039	0.038	270.7
QrZAG96	8.990	0.858	0.019	0.020	0.019	0.2
QrZAG20	9.181	0.878	0.016	0.008	0.026	0.3
QrZAG5	10.186	0.900	0.024	0.016	0.040	0.2
QrZAG65	12.428	0.937	0.022	0.022	0.022	0.2
QrZAG7	10.535	0.901	0.028	0.024	0.030	0.2
QrZAG87	7.862	0.843	0.016	0.004	0.011	0.3
QpZAG15	7.578	0.793	0.039	0.042	0.043	7.7
QpZAG46	8.094	0.859	0.015	0.012	0.010	0.3
QpZAG9	7.667	0.846	0.025	0.017	0.046	0.3
QpZAG1/5	7.622	0.842	0.020	0.019	0.022	0.3
QpZAG36	9.553	0.888	0.015	0.009	0.023	0.9
MSQ13	7.330	0.821	0.013	0.004	0.024	1.3
All	**8.400**	**0.823**	**0.023**	**0.018**	**0.028**	

Genetic diversity (H_e) and allelic richness (A) were calculated with Fstat, genetic differentiation (F_{ST}) was calculated with Arlequin and the Bayes factor BF was obtained with BayeScan.

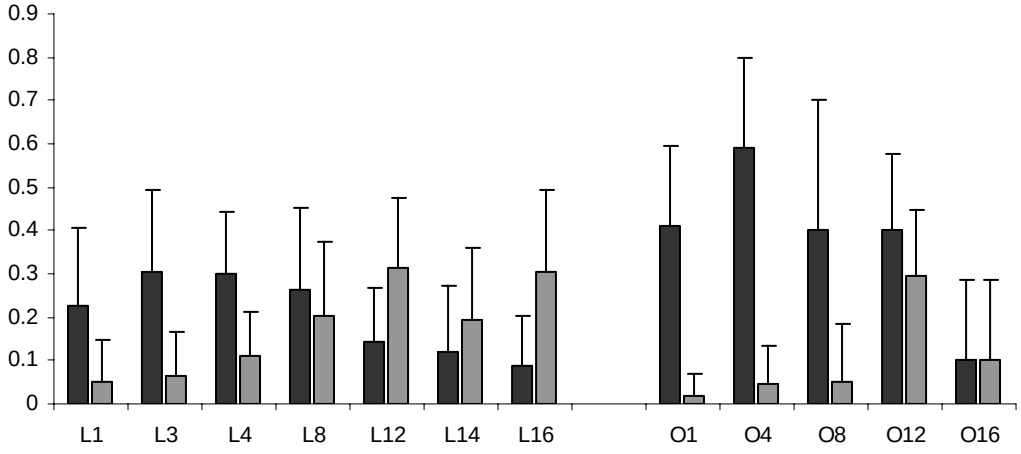


Figure 2: Allelic frequencies of allele 109 (locus QrZAG39) and allele 113 (locus QpZAG15) along the two altitudinal gradients.

Population codes according to Table 1 are given in the horizontal axis. Plots in black: allele 109 (microsatellite QrZAG39). Plots in grey: allele 113 (microsatellite QpZAG15). Bars represent the 95 % confidence intervals.

Genetic differentiation within valleys varied across loci (Table 4) with three loci, QrZAG110, QrZAG39 and QpZAG15, contributing to most of the genetic differentiation in both valleys. For loci QrZAG39 and QpZAG15 we detected congruent variation of allelic frequencies with altitude. Allele 109 at locus QrZAG39 decreased with increasing altitude and allele 113 at locus QpZAG15 decreased with decreasing altitude (Figure 2). Although the trend of variation was the same in both valleys, the linear regressions were significant in the Luz valley (slope = -1.35×10^{-4}, r = -0.87and $p < 0.01$ for QrZAG39, and slope = 1.70×10^{-4}, r = 0.91 and $p < 0.01$ for QpZAG15), but not in the Ossau valley.

Genetic structure

The logarithm of the probability of the data [LnP(D)] as a function of *K* reached a peak for *K* = 4 (mean values: LnP(D) = -18169.4, Var[LnP(D)] = 934.3) and then decreased, but we found a higher ΔK value for *K* = 2 than for *K* = 4 using Evanno's criterion (Evanno *et al.* 2005). According to the software documentation (Pritchard *et al.* 2009), we choose to interpret the results for *K* = 4 because the two additional clusters exhibited strong assignments with two existing populations (L12 for the blue cluster and O4 for the green cluster; Figure 3). For the most likely run the *r* value was 0.67, indicating that the sample locations were informative and helped greatly to find the population structure. The genetic structure detected followed the altitudinal gradient with many populations presenting high levels of admixture and few populations almost completely assigned to one cluster (Figure 3). Low altitude populations were mainly admixed between the yellow and the green cluster. The green cluster was considerably represented in populations located until 400 m with an average assignment probability amounting to 78.5 % in population O4 and lower values in the other populations (between 29.9 % in O1 and 40.0 % in L4). High assignment probabilities to the yellow cluster were found in populations L1, L3, L4, L8, O1 and O8 (between 42.1 % in L4 and 74.7 % in L8), whereas trees of population L16 were almost totally assigned to this cluster (87.0 % on average). The red cluster characterized high altitude populations in Ossau valley where assignment probabilities increased along the altitudinal gradient from 32.6 % in population O8 to 77.5 % and 68.5 % in populations O12 and O16 respectively. Finally the blue cluster was mainly represented in two high altitude populations of Luz valley (67.2 % and 35.2 % for L12 and L14 populations respectively).

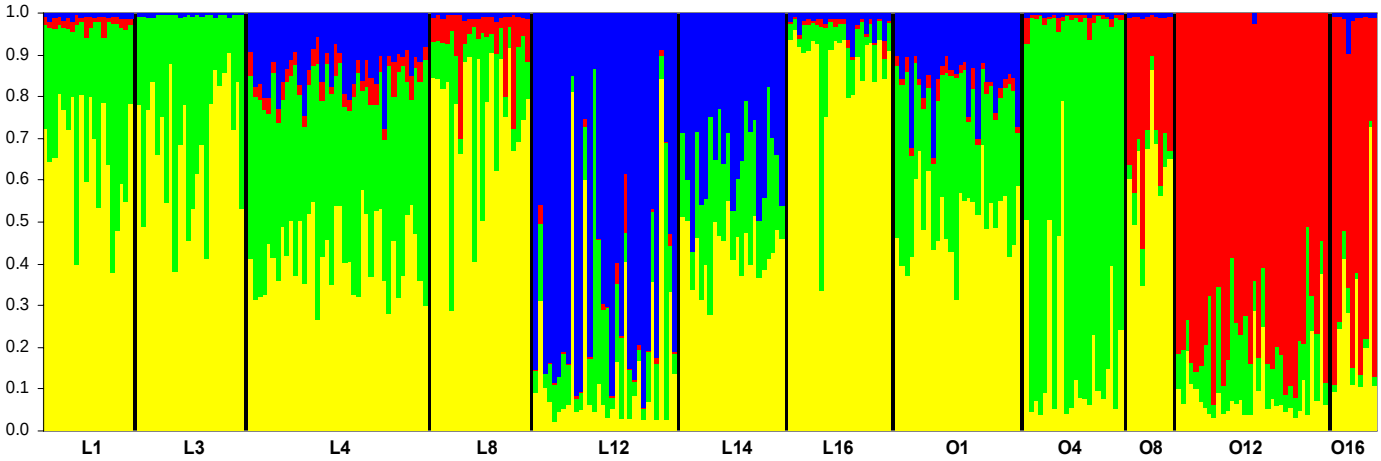


Figure 3: Structure clustering results obtained for *K* = 4 with the LocPrior model.
Each individual is represented by a thin bar corresponding to the sum of the assignment probabilities to the *K* clusters. Populations are separated by black bars and identified at the bottom.

Table 5: Results from the Migrate analysis on groups of populations a) scaled effective population sizes (θ) and b) scaled immigration rates (*M*).

a)

θ_{LL}	θ_{LH}	θ_{OL}	θ_{OH}
14.7 (11.6 – 17.9)	**9.7** (7.7 – 11.8)	**7.0** (5.1 – 9.1)	**2.9** (1.4 – 4.5)

b)

	LL	LH	OL	OH
LL	-	**1.6** (0.4 – 2.7)	**1.0** (0.1 – 1.9)	**0.5** (0.0 – 1.4)
LH	**2.3** (1.0 – 3.5)	-	**1.2** (0.2 – 2.2)	**0.0** (0.0 – 0.7)
OL	**2.6** (1.4 – 3.9)	**0.3** (0.0 – 0.7)	-	**0.0** (0.0 – 0.7)
OH	**2.4** (1.0 – 3.7)	**0.0** (0.0 – 0.7)	**0.0** (0.0 – 0.7)	-

$\theta = 4N_e\mu$ with N_e: effective population size and μ: mutation rate per site per generation.
$M = m/\mu$ with m: immigration rate.
Source groups of populations for migration rates are in columns and sink groups of populations are in rows.
Values between parentheses correspond to the 95 % confidence interval.

Gene flow and population size

After three runs of Migrate the results were congruent and we choose the latter run for interpretations. We observed a clear trend between groups of populations located in the two valleys (Table 5). At a same altitudinal level, populations located in Luz valley presented higher θ values compared to the Ossau valley. In both valleys the scaled effective population sizes decreased noticeably with increasing altitude. The LL group showed thus the highest θ value. Moreover, almost all significant *M* values were found between this group and the other groups of populations. The scaled immigration rates *M* were higher from the LL group to the OL group than to the two high altitudes groups ($M_{LL->OL}$ = 2.6 against $M_{LL->LH}$ = 2.3 and $M_{LL->OH}$ = 2.4) whereas the reciprocal migration rates were lower and preferentially within valleys ($M_{LH->LL}$ = 1.6 against $M_{OL->LL}$ = 1.0). Finally a low migration rate was also found from low altitudes in Ossau valley to high altitudes in Luz valley ($M_{OL->LH}$ = 1.2).

Introgression

The assignment of the sampled trees to the 4 oak species showed that the sampled populations are mainly composed by pure *Quercus petraea* trees at 77.4 % (Table 6). However 4 trees were assigned to the *Q. pubescens* cluster and one tree to the *Q. robur* cluster and hybrid trees were present in all populations. Among the 60 trees considered as hybrids, more than half (31) were *Q. petraea* x *Q. robur* hybrids, 14 were *Q. petraea* x *Q. pyrenaica* hybrids and 10 were *Q. petraea* x *Q. pubescens* hybrids. Five trees were considered as hybrids between other species than *Q. petraea* but showing also sizeable assignment probabilities to the *Q. petraea* cluster (> 0.20). Along the altitudinal gradient we observed a significant increase in the proportion of hybrids with increasing altitude (regression coefficient significant at $p < 0.05$), ranging 17.9 % to 20 % at low altitude to 47.8 % to 40 % at high altitude in the Luz and Ossau valley respectively (Table 6).

Table 6: Number (and percentage) of pure species and hybrid oaks in the sampled populations.

	Q. petraea	*Q. robur*	*Q. pubescens*	*Q. pyrenaica*	H_{PetRob}	H_{PetPub}	H_{PetPyr}	H_{RobPub}	H_{RobPyr}	H_{PubPyr}	% Hyb
L1	16				2	1		1			20.0%
L3	16				4		3		1		33.3%
L4	38				1		1				5.0%
L8	17		2		1		1	1			13.6%
L12	24		2		3	2	1				18.8%
L14	16				2	2	3				30.4%
L16	12				8	1	1			1	47.8%
O1	23				4	1					17.9%
O4	21						1				4.5%
O8	7	1			1		1				20.0%
O12	27				3	2	2				20.6%
O16	6				2	1		1			40.0%
Total	223	1	4	0	31	10	14	3	1	1	20.8%
%	*77.4 %*	*0.3 %*	*1.4 %*	*0.0 %*	*10.8 %*	*3.5 %*	*4.9 %*	*1.0 %*	*0.3 %*	*0.3 %*	

Assignment to pure species or hybrid status is based on the admixture coefficient *Q* (Lepais *et al.* 2009).
$H_{PetRob:}$ hybrid between *Q. petraea* and *Q. robur*.

DISCUSSION

Oak colonization dynamics of Pyrenean valleys

Palinological historical reconstruction indicates that deciduous oaks have colonized the northern side of the Pyrenees between 13000 and 12000 BP stemming from the Spanish glacial refugial areas (Reille & Adrieu 1995). Migration pathways from Spain skirted the Pyrenees on the western edge in the Basque region and oaks spread on the northern foothills from west to east (Petit *et al.* 2002b). An eastern colonization route coming from the Mediterranean basin is also suggested from phylogeographical investigations (Petit *et al.* 2002c). There is indeed a sharp geographical split in chloroplast lineages (between the Eastern lineage A and the Western lineage B) indicating that both movements met at mid distance from the Atlantic ocean and the Mediterranean sea at the Northern side of the Pyrenees (Figure 1a). We found signatures of these records in our data set since all sampled populations comprise two chloroplast haplotypes that belong to the Western lineage (haplotypes 10 and 12), and only one plain population is fixed for a haplotype belonging to the Eastern lineage (haplotype 7).

Pollen records at higher altitude indicate that deciduous oaks reached their upper limit (between 1,700 to 2,000 meters) in less than 2,000 years (Aubert 2001, Belet 2001) since the earliest records mentioned oak presence between 11,000 and 10,000 BP. We suspect that both valleys were colonized by source populations located in the plain and our data suggest that the source populations were composed of different haplotypes (haplotype 10 and 12 for the Ossau valley, and haplotype 12 for the Luz valley). Interestingly haplotype 10 is the most frequent haplotype occurring in extant oak populations located in the plain along the Pyrenees Mountains (Petit *et al.* 2002b), and the persistence of mixed populations (haplotype 10 and 12) in the Ossau valley suggests past or recent contributions from the Luz valley that is entirely fixed for haplotype 12. Persistence of extant mixed populations can be obtained in only two different scenarios. Either the founder populations were already mixed and large enough to prevent the fixation of one haplotype by drift over the 200 to 500 generations since foundation, or recurrent seed flow was maintained from the source population of the neighbouring Luz valley that is fixed for one haplotype. The latter scenario is less probable as seed dispersion witnessed by cpDNA variation, was reported to be limited to less than 1 km in oak species (Valbuena-Carabaña *et al.* 2005).

Preferential unidirectional pollen flow

Colonization along altitudinal gradients has generated population differentiation within each valley (Table 3) that has been tempered by pollen flow since establishment of the populations. Because of the strong asymmetry between pollen and seed flow in oak (El Mousadik *&* Petit 1996b), we suspect that most of the gene flow occurring since stand establishment is due to pollen movement. Even if the F_{ST} estimates are low, they are close to those observed over the entire natural range (Mariette *et al.* 2002). The population clustering analysis using Structure further showed that the overall differentiation was mainly due to distinct altitudinal trends of variation occurring in each valley. Whereas populations located until 400 m are mostly admixed (with each individual being assigned to several putative clusters), populations located at higher altitudes are not (with each individual being predominantly assigned to a single cluster). In the Ossau valley, populations from the highest elevations (O12, O16 and to a lesser extent O8) appear as a distinct genetic cluster differentiated from lower altitude populations. Similarly, in the Luz valley populations located from 800 m upward exhibited differences from lower populations. Populations L12 on one hand and L16 and L8 on the other hand are each assigned to a single cluster, whereas population L14 is admixed between these two clusters. Divergence of high altitude populations is observed even if the markers presumably under altitudinal selection are removed from the analysis which indicates that the genetic differentiation is caused by demographic events. Genetic differentiation of high altitude populations is found in several studies focusing on the genetic diversity of tree species along an altitudinal gradient (reviewed by Ohsawa *&* Ide, 2008). Our survey suggests that disruption of gene flow along altitude or introgression (see below) may cause the observed divergence. Limitations to pollen flow can be created by physical barriers such as the mountain topography disrupting wind streams (Fournier *et al.* 2006, Herrera *&* Bazaga 2008). Differences in flowering phenology between populations located at extreme altitudes can also create temporal restrictions to gene flow (Schuster *et al.* 1989, Premoli *et al.* 2003). Temperature is the main environmental factor controlling the vegetative and reproductive phenology of most tree species (Chuine 2000, Vitasse *et al.* 2009b) and populations along altitudinal gradients are exposed to highly contrasted temperature regimes. Leaf unfolding phenology of the same oak populations that we sampled was monitored by Vitasse *et al.* (2009b). Flushing dates which are correlated to flowering dates, varied by 50 days between the earliest populations located at lower altitudes to the latest at the highest altitudes. Even if a

large variation of flushing dates persisted also among trees within populations (Vitasse *et al.* 2009b), such delays between populations separated by only a few tens of kilometres can limit pollen flow. Furthermore, the Migrate analysis suggested that gene flow occurred preferentially from lower to higher altitudes. This directional gene flow may be the direct consequences of altitudinal variation in phenological dates and demographic differences in population sizes. While populations at higher altitudes flush later, they also exhibit lower variations of flushing dates among trees (F. Alberto, INRA – Université de Bordeaux, Bordeaux; unpublished data). Hence, because of the partial overlap of flushing dates between altitudes and the larger between tree variation at lower than higher altitudes, pollen migration rates are expected to be higher from low to high altitude than the reverse.

Imprints of natural selection

Among the 16 nuclear microsatellites, two loci (QrZAG39, QpZAG15) exhibited divergence patterns which deviated from neutral expectations. The congruent allelic variations displayed by these loci in the two valleys illustrate directional selection acting along an environmental gradient. Previous studies have already reported clinal variations of allozyme frequencies with respect to altitude (Mitton *et al.* 1980; Belleti *&* Lanteri 1996; Kara *et al.* 1997). More recently, Jump *et al.* (2006) found a sharp differentiation of AFLP markers in European beech populations located at different altitudes, which departed significantly from neutral expectation. Interestingly, QpZAG15 is located on linkage group 9 of the Quercus genetic map, where the strongest QTL for bud burst is located (Derory *et al.* 2010). As genetic differentiation for date of bud burst has also been observed on the same populations in common garden experiments (Vitasse *et al.* 2009c), one may conclude that allelic frequencies have been shifted by hitchhiking effects between the QTL of budburst and QpZAG15, although strong linkage is needed between the microsatellite and QTL to prevent recombination in natural populations. QrZAG39 is located close to one extremity of linkage group 5, with no known QTL position in its vicinity, but hitchhiking effects with another adaptive trait shaped by altitudinal variation cannot be excluded. An alternative interpretation to the clinal variation of allelic frequencies is selective introgression. As hybridization rates increase with altitude, allele frequencies at loci involved in species divergence may vary along an altitudinal gradient. Although located at less than 20 cM from genomic hotspots of species divergence, the reported interspecific F_{ST} values of the two loci were in the range of neutral expectations, suggesting that the loci are not affected by introgression (Scotti-

Saintagne *et al.* 2004). We therefore suspect that the congruent variations of allelic frequencies are caused by diversifying natural selection occurring along the altitudinal gradients in the two valleys.

Maintenance of genetic diversity associated to a decrease of population size

Gene diversity and allelic richness reached similar levels across the altitudinal gradient and between the two valleys except for one population located in low altitude (O4). The small population size and the isolation of this stand on a north-exposed mountainside may have exposed O4 to drift. In their review on genetic diversity of plants along altitudinal gradients Ohsawa & Ide (2008) reported maintenance of diversity in 29 % of the studies. As an example, Truong *et al.* (2007) found similar levels of heterozygosity in *Betula pubescens* populations located between 450 and 1200 m, which was explained by the maintenance of extensive pollen and seed flow of this species along the altitudinal gradient. On the opposite, in studies reporting steadily decrease of genetic diversity, limitations of gene flow, small population size or founder effects at higher altitudes were advocated (Premoli *et al.* 2003; Quiroga & Premoli 2007). In our study, despite the disruption of gene flow suggested by genetic differentiation, genetic diversity was maintained along the altitudinal gradients. However, we found a trend towards decreasing population size with increasing altitude in both valleys and population size was higher in Luz than in Ossau valley. In both valleys, the scaled population size parameter ($\theta = 4N_e\mu$) was always lower at higher altitude. The apparent paradox between maintenance of diversity and decrease of population size can be explained by the relationship linking diversity and population size ($H_e = \theta/(\theta+1)$), at equilibrium between drift and mutation and under the Wright Fisher model (Hartl & Clark 1989). In the upper ranges of N_e ($N_e > \sim 5000$ for $\mu = 5.10^{-4}$) population size may further increase while diversity saturates (*i.e.* variations of N_e will have little impact on H_e). Assuming an average mutation rate of 5.10^{-4} for microsatellites (Selkoe & Toonen 2006), our estimates of N_e vary between 1,450 and 7,350, corresponding to H_e varying between 0.744 and 0.936 which lie in the range where H_e is not linearly related to N_e (Table 2).

Peculiar reproductive characteristics at higher altitudes

The survey of genetic diversity across a wide range of altitudinal variation showed that populations at the higher margin of distribution share peculiar reproductive characteristics. First we found evidence of vegetative propagation. Clonal copies of trees sharing the same multilocus genotypes were separated by more than five meters in O16 population, and at

smaller distances in the L16 population. Oaks are known to produce stumps sprouts when they are logged or hedged but clonal copies may be separated at most by a few meters if repeated over generations. Re-rooting of shrubby trees (Copini *et al.* 2006), root suckers (Mayes *et al.* 1998; Valbuena-Carabaña *et al.* 2008) or repeated stump sprouting caused by snow coverage and avalanches may cause vegetative reproduction of *Q. petraea* at high altitudes.

Finally this study showed that the high altitudinal *Q. petraea* populations in both valleys comprised a significant proportion of hybrids between *Q. petraea* and *Q. robur* or *Q. pubescens,* although species status was checked from leaf morphology (Kremer *et al.* 2002). The more frequent occurrence of hybrids at the margins of the distribution fits to the scenario of species dispersion facilitated by unidirectional gene flow between a recipient local species (*Q. petraea*) and the migrant species (*Q. robur* or *Q. pubescens*) (Petit *et al.* 2004). Our study indicated that hybridization occurred between one local parental species (*Q petraea* acting as female) and one distant species (acting as male parent *Q. robur* or *Q. pubescens*). As *Q. robur* and *Q. pubescens* are present at lower altitudes, our observations suggested that pollen flow occurs from lower to higher altitude. Interestingly, this is the reverse situation to the one observed at the latitudinal margin of the species distribution, where *Q. robur* is the local species and *Q. petraea* is the migrant species. Our results therefore suggest that the mechanism of dispersion facilitated by hybridization is bidirectional depending on the spatial and demographic distribution of the parental species. Additionally, as suggested by Premoli (2003) for *Nothofagus pumilio* populations, hybridization may also contribute to the maintenance of diversity in high altitude populations. By providing new genetic variation and creating new combinations of alleles, hybridization increases genetic diversity. Furthermore, this gain of genetic diversity can improve the ecological tolerance of a species and facilitate its adaptation to extreme environments (Lewontin & Birch 1966; Rieseberg *et al.* 2003). Under climate warming, introgression from species inhabiting lower altitudes may rather enhance then constrain adaptation. Hybridization may therefore play a major role in the adaptation of *Q. petraea* along the altitudinal gradients in the Pyrenees Mountains and these oak populations represent an interesting case for the study of natural hybrid zones (Lexer *et al.* 2004).

In conclusion our results show that *Q. petraea* populations maintain high level of genetic diversity at higher altitudes despite a reduction in population size and weak differentiation along altitudinal transects. Due to the high overall level of population sizes, genetic diversity is not affected by the altitudinal variation. Pollen flow is preferential from lower to higher altitudes and contributes to the maintenance of diversity. Hybrids are more frequent at higher altitudes and the direction of introgression is the opposite of the latitudinal limit of the distribution suggesting that this mechanism depends of the context of the local and migrant species. Hybridization may play an important role in the adaptation of *Q. petraea* in these peripheral populations. Consequently, by bringing new genetic variability, pollen flow and hybridization may both facilitate the migration of oak at higher altitudes as a response to climate change.

REFERENCES

Aitken SN, Yeaman S, Holliday JA, Wang TL, Curtis-McLane S (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* **1**, 95-111.

Aubert S (2001) Limite supérieure de la forêt, climat et anthropisation: dynamique tardiglaciaire et holocène de la végétation dans la vallée du Marcadau (Hautes-Pyrénées, France). Thèse de doctorat, Université Toulouse III, 365 p.

Barreneche T, Casasoli M, Russell K*, et al.* (2004) Comparative mapping between *Quercus* and *Castanea* using simple-sequence repeats (SSRs). *Theoretical and Applied Genetics* **108**, 558-566.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**, 969-980.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 4563-4568.

Beerli P (2006) Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* **22**, 341-345.

Belet JM, (2001) Variations tardiglaciaires et Holocène de la limite supérieure de la forêt dans les Pyrénées centrales : exemple du vallon d'Estibère (Hautes Pyrénées, France). Thèse de doctorat. Université Toulouse III, 241 p.

Belletti P, Lanteri S (1996) Allozyme variation among European beech (*Fagus sylvatica* L) stands in Piedmont, north-western Italy. *Silvae Genetica* **45**, 33-37.

Bridle JR, Vines TH (2009) Limits to evolution at range margins. When and why does adaptation fail? *Trends in Ecology & Evolution* **22**, 140-147.

Byars SG, Papst W, Hoffmann AA (2007) Local adaptation and cogradient selection in the alpine plant, *Poa hiemata*, along a narrow altitudinal gradient. *Evolution* **61**, 2925-2941.

Chuine I (2000) A unified model for budburst of trees. *Journal of Theoretical Biology* **207**, 337-347.

Copini P, Buiteveld J, Sass-Klaassen U, Ouden Jd (2006) Oak clusters on the Veluwe. *Vakblad Natuur Bos Landschap* **3**, 24-27.

Csaikl UM, Burg K, Fineschi S, König AO, Matyas G, Petit RJ (2002) Chloroplast DNA variation of white oaks in the alpine region. *Forest Ecology and Management* **156**, 131-145.

Deguilloux MF, Pemonge MH, Bertel L, Kremer A, Petit RJ (2003) Checking the geographical origin of oak wood: molecular and statistical tools. *Molecular Ecology* **12**, 1629-1636.

Derory J, Scotti-Saintagne C, Bertocchi E *et al.* (2010) Contrasting relationships between the diversity of candidate genes and variation of bud burst in natural and segregating populations of European oaks. *Heredity* (in press).

Dow BD, Ashley MV, Howe HF (1995) Characterization of highly variable $(GA/CT)_n$ microsatellites in the bur oak, *Quercus macrocarpa*. *Theoretical and Applied Genetics* **91**,137-141.

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* **12**, 13-15.

Dzialuk A, Chybicki I, Welc M, Śliwińska E, Burczyk J (2007) Presence of triploids among oak species. *Annals of Botany* **99**, 959-964.

El Mousadik A, Petit RJ (1996a) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L) Skeels] endemic to Morocco. *Theoretical and Applied Genetics* **92**, 832-839.

El Mousadik A, Petit RJ (1996b) Chloroplast DNA phylogeography of the argan tree of Morocco. *Molecular Ecology* **5,** 547-555.

Engler R, Randin CF, Vittoz P, *et al.* (2009) Predicting future distributions of mountain plants under climate change: does dispersal capacity matter? *Ecography* **32**, 34-45.

Ettl GJ, Peterson DL (2001) Genetic variation of subalpine fir (*Abies lasiocarpa* (Hook.) Nutt.) in the Olympic Mountains, WA, USA. *Silvae Genetica* **50**, 145-153.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.

Excoffier L, Laval G, and Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47-50.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.

Foll M, Gaggiotti O (2008) A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* **180**, 977-993.

Fournier N, Rigling A, Dobbertin M, Gugerli F (2006) Random amplified polymorphic DNA (RAPD) patterns show weak genetic differentiation between low- and high-elevation types of Scots pine (*Pinus sylvestris* L.) in dry continental valleys in the Alps. *Annals of Forest Science* **63**, 431-439.

Goudet J (2001). FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html. Updated from Goudet (1995).

Hartl DL, Clark AG (1989) *Principles of population genetics*. Sinauer associates, Sunderland, Massachusetts.

Herrera CM, Bazaga P (2008) Adding a third dimension to the edge of a species' range: altitude and genetic structuring in mountainous landscapes. *Heredity* **100**, 275-285.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* **9**, 1322-1332.

Jeffreys H (1998) *Theory of Probability*. Oxford University Press, Oxford.

Jump AS, Hunt JM, Martinez-Izquierdo JA, Peñuelas J (2006) Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*. *Molecular Ecology* **15**, 3469-3480.

Kampfer S, Lexer C, GloÈssl J, Steinkellner H (1998) Characterization of $(GA)_n$ microsatellite loci from *Quercus robur*. *Hereditas* **129**, 183-186.

Kara N, Korol L, Isik K, Schiller G (1997) Genetic diversity in *Pinus brutia* Ten.: Altitudinal variation. *Silvae Genetica* **46**, 155-161.

Kirkpatrick M, Barton NH (1997) Evolution of a species' range. *American Naturalist* **150**, 1-23.

Kremer A, Dupouey JL, Deans JD*, et al.* (2002) Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak stands. *Annals of Forest Science* **59**, 777-787.

Lenoir J, Gégout JC, Marquet PA, de Ruffray P, Brisses H (2008). A significant upward shift in plant species optimum elevation during the 20th century. *Science* **320**, 1768-1771.

Lepais O, Leger V, Gerber S (2006) Short note: High throughput microsatellite genotyping in oak species. *Silvae Genetica* **55**, 238-240.

Lepais O, Petit RJ, Guichoux E*, et al.* (2009) Species relative abundance and direction of introgression in oaks. *Molecular Ecology* **18**, 2228-2242.

Lewontin RC, Birch LC (1966) Hybridization as a Source of Variation for Adaptation to New Environments. *Evolution* **20**, 315-336.

Lexer C, Heinze B, Alia R, Rieseberg LH (2004) Hybrid zones as a tool for identifying adaptive genetic variation in outbreeding forest trees: lessons from wild annual sunflowers (Helianthus spp.). *Forest Ecology and Management* **197**, 49-64.

Mariette S, Cottrell J, Csaikl UM*, et al.* (2002) Comparison of levels of genetic diversity detected with AFLP and microsatellite markers within and among mixed *Q. petraea* (Matt.) Liebl. and *Q. robur* L. stands. *Silvae Genetica* **51**, 72-79.

Mayes SG, McGinley MA, Werth CR (1998) Clonal population structure and genetic variation in sand-shinnery oak, *Quercus havardii* (Fagaceae). *American Journal of Botany* **85**, 1609-1617.

Mitton JB, Sturgeon KB, Davis ML (1980) Genetic differentiation in ponderosa pine along a steep elevational transect. *Silvae Genetica* **29**, 100-103.

Neal RM (2003) Slice sampling. *Annals of Statistics* **31**, 705-741.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Ohsawa T, Ide Y (2008) Global patterns of genetic variation in plant species along vertical and horizontal gradients on mountains. *Global Ecology and Biogeography* **17**, 152-163.

Peñuelas J, Ogaya R, Boada M, Jump AS (2007) Migration, invasion and decline: changes in recruitment and forest structure in a warming-linked shift of European beech forest in Catalonia (NE Spain). *Ecography* **30**, 829-837.

Petit RJ, Csaikl UM, Bordacs S*, et al.* (2002a) Chloroplast DNA variation in European white oaks - Phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management* **156**, 5-26.

Petit RJ, Brewer S, Bordacs S*, et al.* (2002b) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management* **156**, 49-74.

Petit RJ, Latouche-Halle C, Pemonge MH, Kremer A (2002c) Chloroplast DNA variation of oaks in France and the influence of forest fragmentation on genetic diversity. *Forest Ecology and Management* **156**, 115-129.

Petit RJ, Bodenes C, Ducousso A, Roussel G, Kremer A (2004) Hybridization as a mechanism of invasion in oaks. *New Phytologist* **161**, 151-164.

Premoli AC (2003) Isozyme polymorphisms provide evidence of clinal variation with elevation in *Nothofagus pumilio*. *Journal of Heredity* **94**, 218-226.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

Pritchard JK, Wen W, Falush D (2009) *Documentation for structure software: Version 2.3*. Available from http://pritch.bsd.uchicago.edu/structure.html.

Quiroga MP, Premoli AC (2007) Genetic patterns in *Podocarpus parlatorei* reveal the long-term persistence of cold-tolerant elements in the southern Yungas. *Journal of Biogeography* **34**, 447-455.

Reille M, Andrieu V (1995) The late Pleistocene and Holocene in the Lourdes Basin, Western Pyrenees, France: new pollen analytical and chronological data. *Vegetation History and Archaeobotany* **4**, 1-21.

Rieseberg LH, Raymond O, Rosenthal DM*, et al.* (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**, 1211-1216.

Schuster WS, Alles DL, Mitton JB (1989) Gene flow in limber pine - Evidence from pollination phenology and genetic differentiation along an elevational transect. *American Journal of Botany* **76**, 1395-1403.

Scotti-Saintagne C, Mariette S, Porth I, *et al.* (2004) Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. *Genetics* **168**, 1615-1626.

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* **9**, 615-629.

Steinkellner H, Fluch S, Turetschek E*, et al.* (1997) Identification and characterization of (GA/CT)(n)-microsatellite loci from *Quercus petraea*. *Plant Molecular Biology* **33**, 1093-1096.

Truong C, Palme AE, Felber F (2007) Recent invasion of the mountain birch *Betula pubescens* ssp tortuosa above the treeline due to climate change: genetic and ecological study in northern Sweden. *Journal of Evolutionary Biology* **20**, 369-380.

Valbuena-Carabaña M, Gonzalez-Martinez SC, Sork VL*, et al.* (2005) Gene flow and hybridisation in a mixed oak forest (*Quercus pyrenaica* Willd. and *Quercus petraea* (Matt.) Liebl.) in central Spain. *Heredity* **95**, 457-465.

Valbuena-Carabaña M, Gonzalez-Martinez SC, Gil L (2008) Coppice forests and genetic diversity: A case study in *Quercus pyrenaica* Willd. from Central Spain. *Forest Ecology and Management* **254**, 225-232.

Vitasse Y, Porté AJ, Kremer A, Michalet R, Delzon S (2009a) Response of canopy duration to temperature changes in four temperate tree species: relative contributions of spring and autumn leaf phenology. *Oecologia* **161**, 187-198.

Vitasse Y, Delzon S, Dufrene E*, et al.* (2009b) Leaf phenology sensitivity to temperature in European trees: Do within-species populations exhibit similar responses? *Agricultural and Forest Meteorology* **149**, 735-744.

Vitasse Y, Delzon S, Bresson CC, Michalet R, Kremer A (2009c) Altitudinal differentiation in growth and phenology among populations of temperate-zone tree species growing in a common garden. *Canadian Journal of Forest Research* **39**, 1259-1269.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.

Zhang Q, Allard RW (1986) Sampling variance of the genetic diversity index. *Journal of Heredity* **77**, 54-55.

Adaptive responses of seed and leaf phenology in natural populations of sessile oak along altitudinal gradients

Alberto F. * †, Bouffier L. * ‡, Louvet J.M. * †, Delzon S. * †, and Kremer A. * † [1]

* INRA, UMR1202 Biodiversité Gènes et Communautés, Cestas, F-33610, France

† Université de Bordeaux, UMR1202 Biodiversité Gènes et Communautés, Talence, F-33400, France

‡ FCBA, Pôle Biotechnologie Sylviculture, Domaine de Sivaillan-Les-Lamberts, Moulis-en-Médoc, F-33480, France

Submitted to *Journal of Evolutionary Biology*

INTRODUCTION

During the last decade, adaptive responses of animal and plant species to climate change have been increasingly investigated (reviews by Parmesan, 2006; Bradshaw & Holzapfel, 2008). One major concern in this context arose about the ability of tree species to respond to climate change (Saxe *et al.,* 2001; Lindner *et al.,* 2009). While palinological and phylogeographical studies on temperate tree clearly showed that they experienced several environmental changes in the past (Brewer *et al.*, 2002, Petit *et al.*, 2003, Magri *et al.*, 2006), their responses to future changes remain speculative (Kremer, 2007; Aitken *et al.,* 2008). The succession of cold and warm periods during the quaternary era came along with large changes in the geographic range of tree species, and subsequent adaptation to these changes is supported by extant population differentiation (Kremer *et al.,* 2010). But nowadays, climate change caused by human activities is occurring at an unprecedented rapid rate (IPCC 2007). According to Aitken *et al.,* (2008), expected responses of tree populations to ongoing environmental changes have been sketched in three scenarios: migration, adaptation or extirpation. In mountainous landscapes, an upslope shift of plant species has already been observed (Lenoir *et al.*, 2008), although the fragmentation of habitats may limit range shifts (Davis & Shaw, 2001). Extirpation due to climate change have been predicted by niche modelling (Thomas *et al.*, 2004), but not yet observed. Local adaptation as a response to ongoing climate change is less tractable, although micro-evolutionary studies suggested rapid adaptations (Daubree *&* Kremer, 1993; Skroppa & Kohlmann, 1997). Here we explore the capacity of oak adaptation by a retrospective analysis of population differentiation and diversity for seed and bud phenological traits along steep temperature gradients mimicking future environmental change. Along with phenotypic plasticity of genotypes which permit a fast but limited response to environmental changes the level of genetic diversity of natural populations will be critical for future adaptation of trees species (Jump *et al.*, 2009). Local adaptation witnessed by genetic differentiation between populations feeds on genetic diversity but also on external gene flow (Aitken *et al.,* 2008). Indeed high latitude and altitude populations will benefit of gene flow from central populations which will introduce alleles preadapted to warmer climates (Davis & Shaw, 2001; Aitken *et al.,* 2008). It is therefore of crucial importance to estimate the extent of within and between population genetic variation in an ecological context, *e.g.* when populations are linked by gene flow. The distribution of genetic variability of tree populations for adaptive traits can be assessed in common garden experiments. Common garden

experiments were initially used by foresters as provenance tests, to monitor genetic variation for traits of economical and ecological relevance within a set of natural populations originating from different parts of the natural range (see for example Ducousso *et al.,* 1996; Viveros-Viveros *et al.*, 2005; Saenz-Romero *et al.*, 2006). These studies were generally conducted to establish seed transfer zones and manage the genetic resources of tree species with a forestry perspective (Morgenstern, 1996), but they also provided estimates of the level of genetic variation in natural populations. Among the most important adaptive traits, phenology is probably one of the most affected by global change (Bertin, 2008). Indeed, modifications of the phenology of flowering, bud burst or bud set in relation to global warming have been observed for many tree species (Menzel *et al.,* 2006; Nordli *et al.,* 2008) and predicted by mechanistic models (Bennie *et al.,* 2009). Later bud set and earlier bud burst increases the length of the growing season and the amount of net assimilation and the competitive ability of trees (Vitasse *et al.,* 2009a; Bennie *et al.,* 2009), while earlier bud set and later bud burst will improve cold resistance (Howe *et al*., 2000; Howe *et al.,* 2003). Variation of the timing of bud set and bud burst will also result in different synchronization with herbivorous insects and pathogens fungi (Visser & Holleman, 2001; Van Asch *et al.,* 2007; Ghelardini & Santini, 2009). Modelling approaches have shown that temperature was the main driver of the timing of bud burst for most of temperate tree species, along with photoperiod for few species (Chuine & Cour, 1999). Overall, bud burst is thought to be strongly related to the fitness of trees. Local adaptation of populations for bud burst in relation to climatic conditions has been recently described in several tree species and few studies have reported clinal variations according to latitude or altitude (Howe *et al.,* 2003, Premoli *et al.,* 2007, Vitasse *et al.,* 2009b, Jensen & Hansen, 2008). Generally tree populations from the colder climates tend to flush earlier than populations from the warmer climates when grown in common gardens. This is the case for most of conifers (Wright, 1976; Morgenstern, 1976), *Quercus rubra* (Kriebel *et al.,* 1976; Kremer, 1994) and *Quercus robur* (Jensen & Hansen, 2008) along latitudinal gradients. The same adaptation to temperature regimes was reported along altitudinal gradients for two *Abies* species (Worrall, 1983), *Pseudotsuga sp.* (Acevedo-Rodriguez *et al.,* 2006), *Fagus sylvatica* (Chmura & Rozkowski, 2002; Von Wuehlisch *et al.,* 1995; Vitasse *et al.,* 2009b) and *Quercus petraea* (Deans & Harvey, 1995; Ducousso *et al.,* 1996). However opposed results according to temperature were reported for the same species depending on the environmental gradient. Southern provenances of sessile oak (*Quercus petraea*) showed earlier bud burst compared to northern provenances (Liepe, 1993; Deans & Harvey, 1995; Ducousso *et al.,* 1996). Hence the clinal patterns of differentiation for bud

burst phenology in tree species requires further investigations to elucidate the sources of the differences among species and among populations within species. The timing of germination is also an important ecological trait intensively studied in the frame of plant adaptation to their environment (Donohue *et al.,* 2005; Finch-Savage & Leubner-Metzger 2006). Young seedlings are indeed exposed to the same selection pressures than flushing buds but they are critically more vulnerable to injuries than adult trees. The timing of seed germination has been the focus of intensive study, mostly in annual or biannual plant species which displayed a high variance both within and among seasons despite the strong selection expected on this trait (Donohue *et al.,* 2005; Simons & Johnston 2006; Evans *et al.,* 2007; Venable, 2007; Simons, 2009). This diversification was considered as the archetypal illustration of the bet hedging theory which hypothesises that a trait submitted to unpredictable environment is selected for its variance rather than for its value (Cohen, 1966). Bet hedging strategy thus corresponds to the production of a high phenotypic variance by a given genotype to maximize its long-term fitness under fluctuating selection (Simons, 2009). In tree species however the pattern of variation for germination timing has been poorly described (but see Campbell, 1979) and thus the evolutionary potential for responding to climate change are yet totally unknown.

The main objective of this study was to estimate the level of genetic variability of seed and bud phenology in sessile oak populations located in the Pyrenees along altitudinal gradients. Sessile oak (*Quercus petraea* Matt. Liebl.) is a widespread European white oak species extending from Great Britain to the Iberic peninsula, and from the eastern Poland to the Atlantic Ocean. We used common garden experiments to quantify the genetic variability within and among populations sampled within natural populations connected by gene flow. Our objective was to estimate the impact of past divergent selection induced by the altitudinal gradient, and the capacity of populations to maintain genetic diversity despite selection pressures. In other words, can local adaptation and diversity coexist, and what are the evolutionary mechanisms that may maintain both processes? Our second objective was to compare the distribution of genetic variation between two phenological traits (timing of seed germination and bud burst) that follow dormancy release and may therefore be the consequence of similar physiological pathways (Rohde *et al.*, 2000; Derory *et al.,* 2006; Ruttink *et al.,* 2007; Rohde & Bhalerao, 2007). Our investigations may lead to practical applications for future research regarding the discovery of candidate genes for these traits.

MATERIAL AND METHODS

Sampling populations

Quercus petraea belongs to the complex of the white European oaks and is rather mesoxerophyle, preferentially located in south aspects and on acid soils. In the Pyrenees Mountains, sessile oak range from almost pure populations in the plain to more scattered stands at higher altitudes, reaching up to 1800 m.

Natural populations of sessile oak (*Q. petraea*) located in two valleys (Ossau and Luz) in the Pyrenees Mountains (south-west of France) were sampled for this study. In each valley, five populations were sampled along an altitudinal gradient ranging from 131 m to 1630 m (see Table 1 and Alberto *et al.,* 2010 for details)**.** These populations have been described in previous studies. Earlier *in situ* investigations were conducted on phenological traits (Vitasse *et al.,* 2009c; Vitasse *et al.*, 2009b) and genetic diversity was monitored (Alberto *et al.,* 2010) on the same set of populations. In each population, air temperature was monitored every hour during year 2006, 2007 and 2008 using data loggers (HOBO Pro RH/Temp; Onset Computer, Bourne, Massachusetts) as described in Vitasse *et al.,* (2009c).

Harvesting of acorns

Acorns were harvested in September 2006 on 152 adult trees within the 10 natural populations, resulting in a total collection of 8833 acorns (Table 1). Depending on the level of fruiting within each population, open pollinated progenies were collected from 7 to 33 trees per population. Monitoring of leaf unfolding in the spring and leaf coloration in the fall was conducted in 2005, 2006 and 2007 on the mother trees of the collected progenies (Vitasse *et al.,* 2009c).The number of acorns greatly varied according to population and mother tree due to the heterogeneity of the seed crop. Acorns were weighted for each open pollinated progeny. The average weight of acorns was homogeneous between families of the same population but decreased between populations with respect to altitude (Table 1). Acorns were stored in a cold room at + 4°C during the following winter to ensure chilling requirements prior to the germination.

Table 1: Description of the sample size of provenances in each common garden.

Prov.	Altitude *in situ* (m)	Altitudinal level (m)	Altitudinal group †	Mean acorn weight	PIE (sowing and plantation)					TOUL		
					fam.	acorns	GERM ‡	plants	LU ‡	fam.	plants	LU ‡
L1	**131**	**100**		**2.81**	*11*	*324*	***150***	166	***163***	*11*	114	***110***
L3	**387**	**400**	**Lalti**	**2.80**	*18*	*471*	***336***	347	***344***	*18*	235	***228***
L8	**803**	**800**		**1.73**	*16*	*1094*	***345***	490	***479***	*16*	332	***323***
L12	**1235**	**1200**	**Halti**	**2.20**	*14*	*468*	***248***	283	***282***	*14*	198	***191***
L16	**1630**	**1600**		**0.97**	*14*	*829*	***291***	450	***445***	*14*	353	***349***
O1	**259**	**100**		**2.51**	*24*	*1605*	***456***	594	***592***	*23*	436	***428***
O4	**422**	**400**	**Lalti**	**3.38**	*7*	*126*	***34***	34	***33***	*7*	34	***29***
O8	**841**	**800**		**2.41**	*8*	*721*	***188***	350	***348***	*7*	266	***262***
O12	**1194**	**1200**	**Halti**	**2.05**	*33*	*2611*	***1090***	1750	***1744***	*33*	1374	***1358***
O16	**1614**	**1600**		**0.88**	*7*	*584*	***131***	168	***166***	*7*	106	***103***
Total				**2.17**	*152*	8833	**3269**	4632	***4596***	*150*	3448	***3381***

* Code: provenances beginning by the letter L come from Luz valley and provenances beginning by the letter O come from Ossau valley.

† Lalti: provenances coming from altitudes below 1000 m, Halti: provenances coming from altitudes above 1000 m

‡ For GERM and LU the number of individuals measured per provenance in each experiment is given.

Sowing and plantations establishment

Acorns from the 152 open pollinated families were sown in a greenhouse at the INRA Pierroton station (PIE; 44°44'N, 00°46'W) in March 2007 within a complete block design (152 entries in 5 blocks), but the number of acorns per entry per block varied due to the uneven seed crop (on average 12 acorns per entry per block). Sowing was done on individual clods of loam homogeneously recovered by levelled sand in small containers. The overall germination rate amounted to 52.4 % but differed between provenances from 73.7 % (L3) to 28.8 % and 27 % (O16 and O4 respectively). Low germination rates were essentially caused by a fungus pathogen of the *Ciboria* genus (O1, 37.0 %) or by the low viability of acorns stemming from high altitude (mainly O16). Trees from population O4 produced few but heavy acorns which had the lowest germination rate. Date of germination was recorded on four blocks corresponding to a subset of 6475 acorns during spring 2007.

In winter 2007 the containers with germinated acorns were transferred in a common garden experiment located in the nursery of the INRA research station at Pierroton (PIE). The experimental plantation comprised complete blocks with random spatial allocation of entries within blocks. Seedlings were planted in rows with 1 m * 0.5 m spacing. Monitoring of leaf unfolding was conducted on each seedling in the spring 2008. Individual identity of seedlings was maintained during the transfer from the greenhouse to the nursery. Survival after the first year of plantation was high since only 36 plants did not survived.

During autumn 2008, seedlings were uprooted and transplanted in a field site located on the fruit tree research field of INRA at Toulenne (TOUL; 44°34'N, 00°16'W, 23 m), very close to the nursery where the plants were raised. At TOUL the recorded mean annual temperature between 1984 and 2006 was 13.2°C and the mean annual rainfall was 836 mm. The common garden was established on a sandy loam type soil. The plantation comprised 3448 plants representing 150 families (Table 1). Five randomized complete blocks were set up in a 3.5 m x 1.5 m density plantation.

Throughout the paper, the term population refers to the trees in the natural stands on which acorns were harvested, and the term provenance refers to the offspring of these mother trees which were sown and planted in the common garden experiments.

Phenological measurements

Seed germination was recorded in the greenhouse at PIE during spring 2007. The date of shoot emergence out of the sand surface was recorded, and considered as an indirect assessment of the time of germination. Previous observations have indeed shown that despite

a short lag, there was a high correlation between the time of root and shoot emergence (F. Alberto, personal observations). Date of emergence of each acorn was scored as the number of days since January 1st when the emergence of the shoot was observed (GERM). Monitoring was conducted twice per week on the whole set of germinated acorns.

The development stage of the apical bud from fully dormant bud to internodes elongation was monitored on each seedling according to Derory *et al.,* (2006). Scoring was realized twice a week during spring 2008 at PIE and during spring 2009 at TOUL. Leaf unfolding stage, corresponding to at least one leaf completely unfolded out of the bud, was here considered as the target trait and the number of days (LU) since January 1st when leaf unfolding was observed was recorded. Linear interpolation was used to estimate LU when leaf unfolding occurred earlier or later then the day of observation.

For all mother trees sampled within the ten populations, leaf unfolding has also been previously monitored in situ on the mother trees during springs 2005, 2006 and 2007 following the same procedure (Vitasse *et al.,* 2009a). Yearly *in situ* observations of LU were transformed as standardized measurements to account for population differences recorded *in situ*.

Statistical Analyses

Quantitative genetic analyses were conducted on GERM and LU to divide the phenotypic variance into its genetic and environmental components. Several mixed models were performed using a restricted maximum likelihood (REML) method with the ASReml software (Gilmour *et al.,* 2002).

Model 1

Genetic parameters were first estimated in each site separately. The following mixed model was used for each phenological variable, GERM, LU at PIE, and LU at TOUL:

$$Y_{ijk} = \mu + b_k + Prov_i + F_{j(i)} + (bProv)_{ki} + (bF)_{kj(i)} + \varepsilon_{ijk} \qquad \text{(Equ.1)}$$

where μ is the trait mean, b_k is the fixed block effect, $Prov_i$ and $F_{j(i)}$ are random effects for provenance and family within provenance, respectively, $(bp)_{ki}$ and $(bF)_{kj(i))}$ are the interactions between block and random effects and ε_{ijk} is the residual effect. The significance of each effect was estimated by a likelihood ratio test (threshold at 0.05).

Model 2

As LU was measured on common families in both sites (PIE and TOUL), these two data sets were gathered in a general model to increase the genetic parameter accuracy. This multi-site analysis was conducted with the following model:

$$Y_{ijk} = \mu + e_l + b_{k(l)} + Prov_i + F_{j(i)} + (eProv)_{li} + (eF)_{lj(i)} + (ebProv)_{kli} + (ebF)_{klj(i)} + \varepsilon_{ijk}$$

(Equ.2)

with the previous effects from Model 1, e_l the fixed experimental site effect and $(be)_{kl}$, $(eProv)_{li}$, $(eF)_{lj(i)}$, $(beF)_{klj(i)}$, $(beProv)_{kli}$ the interaction effects between fixed and random effects. Model 1 and Model 2 provided an overall estimate of the within population genetic variance. While we were also interested in comparing the within population genetic variance along the altitudinal gradient, models 1 (for GERM) and 2 (for LU) were also run within altitudinal classes: either by bulking populations of the same altitude between the two valleys, or by bulking populations of low (Lalti) and high (Halti) altitudes (see Table 1).

Genetic parameters and differentiation estimates

The phenotypic variance within provenance (σ^2_P) was estimated as:

$\sigma^2_P = \sigma^2_f + \sigma^2_{(bf)} + \sigma^2_\varepsilon$ for Model 1

$\sigma^2_P = \sigma^2_f + \sigma^2_{(ef)} + \sigma^2_{(ebf)} + \sigma^2_\varepsilon$ for Model 2

In both models, the additive genetic variance (σ^2_G) was estimated as:

$\sigma^2_G = 4 \times \sigma^2_f$ (σ^2_f is the variance of family within provenances) assuming that open pollinated offsprings were half sibs.

We tested the significance of the differences between the genetic variances of the altitudinal classes using a test Z.

The genetic differentiation between populations was estimated by the Q_{ST} which expresses the amount of provenance variation (σ^2_{Prov}) related to the overall genetic variation (Spitze 1993) such that:

$$Q_{ST} = \sigma^2_{Prov} / (\sigma^2_{Prov} + 2 \times \sigma^2_G)$$

Narrow sense heritability h^2 was estimated as:

$$h^2 = \sigma^2_G / \sigma^2_p$$

Finally breeding values, *e.g.* additive genetic values, of each mother tree and provenance values were estimated for GERM and LU using BLUPs (Best Linear Unbiased Predictors) in ASReml.

The standard errors of genetic parameters were calculated with ASReml using a standard Taylor series approximation (Gilmour *et al.,* 2002).

To make a comparison with the estimation by mixed linear model the heritability of LU was also estimated through the parent-offspring relationship. According to Lynch & Walsh (1998), in the case of half-sibs the heritability can be directly estimated from the regression coefficient b_{op} of the parent-offspring regression, where $h^2_{op} = 2 \times b_{op}$. Linear regressions were conducted between the standardized values of LU measured at PIE and the standardized values of LU measured *in situ* in 2006 using the statistical package R (R Development Core Team 2005). We performed the linear regressions and calculated heritability for the overall sample and for each altitudinal class.

Linear regressions between traits and environmental variables

Correlations between environmental variables and traits were tested by linear regressions conducted using R software. For all analyses, the provenance values and the family breeding values of GERM and LU estimated in the mixed models were considered. First, the potential effect of the amount of resources contained in acorns (acorn weight) on the phenology of germination was tested. Linear regressions between GERM and acorn weight were performed using the average acorn weight at the provenance level and the standardized values of acorn weight at the family level. Second, in order to link the among provenances differentiation with environmental variables we performed linear regressions between provenance values and either altitude or spring and annual temperatures of the provenances' sites of origin. Finally, we calculated the genetic correlation between GERM and LU by linear regression between family and provenance values of both traits and the phenotypic correlation by computing at the single tree level.

For each linear regression, we used the Pearson's product-moment method to test the significance of the correlation coefficient *r*.

RESULTS

Provenance differentiation

For both models, main effects were all significant while interactions effects did not exhibit significant differences (Table 2). Provenance differentiation was important as shown by the high Q_{ST} values estimated for GERM (0.29) and LU (0.23 in PIE and 0.21 in TOUL). Both phenological traits, GERM and LU, followed a consistent clinal trend of variation according to altitude. Provenances from high altitudes germinated and flushed later than populations from low altitudes (Fig. 1).

However, altitudinal clines exhibited slight differences between GERM and LU. Altitudinal variation for GERM was rather discontinuous: whereas provenances coming from up to 1200 m germinated between day 116 and 124, provenances from high altitude (1600 m) germinated much later (between day 134 and 144 for L16 and O16 respectively). Timing of leaf unfolding displayed a continuous clinal pattern of variation that was very stable among sites and years. At PIE, the earlier provenance flushed on average at day 101 and the latest at day 115 (respectively provenances L1 and L16)**.** Delays of LU according to altitude amounted to 0.6 and 0.7 days per 100 m increase of elevation at TOUL and PIE, respectively.

Table 2: Genetic parameters estimated for phenology traits in the common garden experiments.

	GERM	LU		
Site	PIE	PIE	TOUL	Multi-site *
Sample	3269	4596	3381	7977
Mean DOY	119.5	107.8	108.9	108.2
Q_{ST}	0.28 (0.11)	0.23 (0.09)	0.21 (0.09)	0.22 (0.09)
h^2	0.51 (0.08)	0.87 (0.11)	1.07 (0.13)	0.87 (0.11)
σ^2_G	117.7 (20.2)	31.7 (4.9)	30.4 (4.9)	29.1 (4.4)
σ^2_P	232.2 (7)	36.6 (1.3)	28.5 (1.3)	33.5 (1.2)
σ^2_{Prov}	90.2	18.6	15.7	16.8
σ^2_F	29.4	7.9	7.6	7.3
σ^2_{bF}	-	1.7	2.0	-
σ^2_{eProv}	-	-	-	0.4
σ^2_{eF}	-	-	-	0.9
σ^2_{beProv}	-	-	-	0.2
σ^2_{beF}	-	-	-	1.8
σ^2_{bProv}	-	0.3	-	-
σ^2_{ε}	202.8	26.9	18.9	23.6

For GERM and LU the site of measurement, the sample of individuals measured and the mean DOY of the trait (day from 1st January) are given. For each analysis the among provenances differentiation (Q_{ST}), the heritability (h^2) were calculated from the genetic variance σ^2_G and the phenotypic variance σ^2_P.
* The Multi-site analysis was conducted by grouping measured of LU at PIE and TOUL.

Within population genetic variation

Family effects were highly significant in both models used to estimate components of the phenotypic variance. According to Model 1, we found high estimates of heritability for both phenological traits (Table 2). The heritability of GERM amounted to 0.52. This trait displayed high phenotypic and additive genetic variances within provenances (Table 2). Although the additive genetic variances of LU were lower in the two sites, LU exhibited higher heritability values in PIE and in TOUL (h^2 = 0.87 and 1.07, respectively). Standardized values of LU measured on the offsprings in PIE were highly correlated with those measured on mother trees *in situ* in 2006 (Fig. 2, $p < 0.0001$).

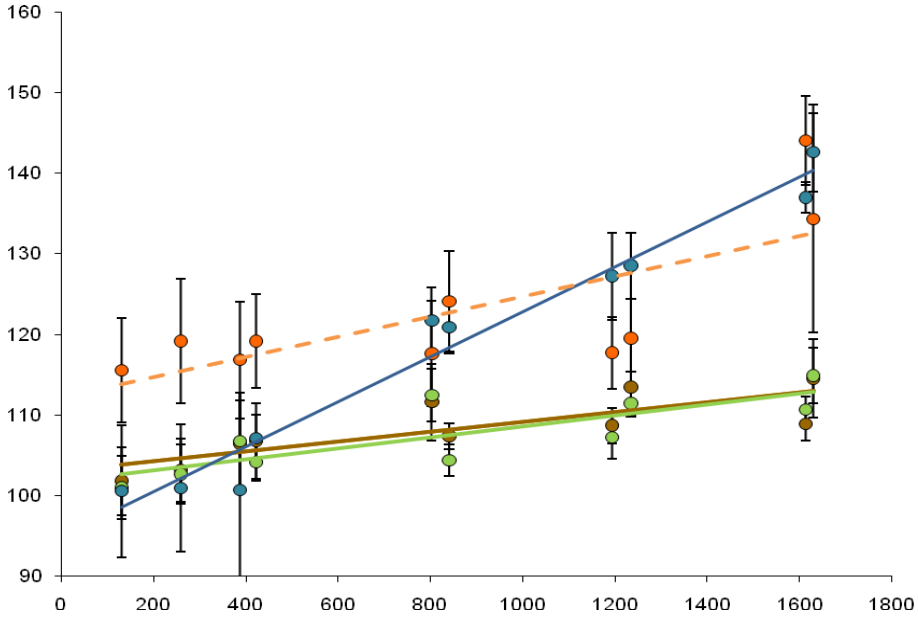


Figure 1: Altitudinal trends of variation for germination and leaf unfolding.

Each point represents the provenance mean of GERM or LU according to the altitude of the provenance's site of origin.
GERM is in orange, LU is in green at PIE and in brown at TOUL and LU *in situ* measured in 2007 is in blue.

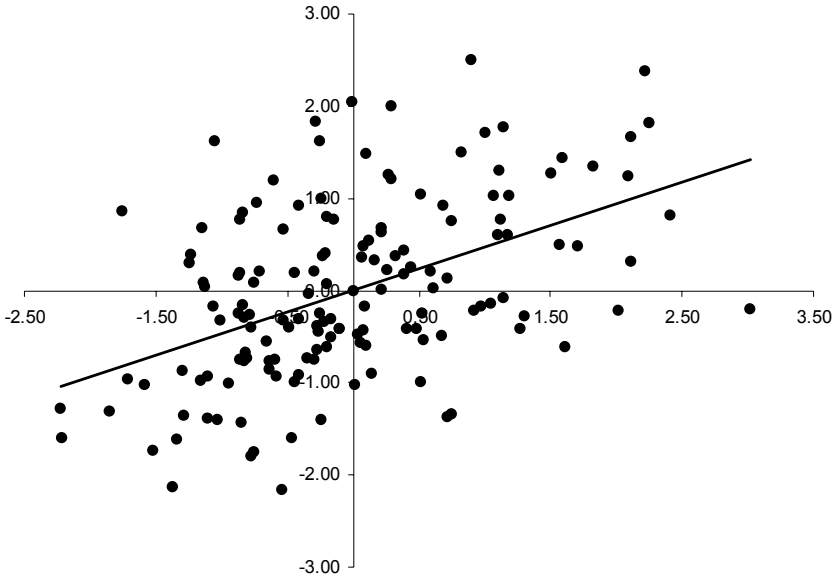


Figure 2: Parent-offspring relationship for leaf unfolding

Linear regression between standardized values of LU measured on offspring at PIE and LU measured on mother trees in natural populations in 2006.

The correlation was also highly significant with LU measured at TOUL ($p < 0.0001$). Linear regressions performed with *in situ* measurements in 2005 and 2007 were also highly significant but correlation coefficients were lower (data not shown). The heritability values estimated by parent-offspring regression ($h^2_{op} = 0.94$ at PIE, Fig. 3; and $h^2_{op} = 1.01$ at TOUL) was of the same order of magnitude than the estimated value obtained by the intra class correlation coefficient within the open pollinated progenies. However when calculated by altitudinal classes we found discrepancies between both estimates (Table 3). Discrepancies between both estimates may either be due to the assumption of half sib relationships within the open pollinated progenies or to the varying environmental variances occurring between the *in situ* forest environment and the common garden experiment. Nevertheless the estimates of heritability by the two methods were always higher than 0.50 (except for the 1200 m populations with $h^2 = 0.42$).

Table 3: Main genetic parameters estimated by grouping provenances according to altitude.

		Altitudinal levels					Altitudinal groups	
		100 m	**400 m**	**800 m**	**1200 m**	**1600 m**	**Lalti**	**Halti**
	Sample	606	370	533	1338	422	1509	1760
	Mean	115.4	115.8	118.2	117.4	137.2	116.5	122.1
GERM	**Q_{ST}**	**0.01 (0.03)**	**0.03 (0.11)**	**0.1 (0.15)**	**0.00 (0.00)**	**0.04 (0.09)**	**0.02 (0.02)**	**0.43 (0.21)**
	h^2	**0.87 (0.22)**	**0.3 (0.18)**	**0.47 (0.18)**	**0.26 (0.08)**	**0.67 (0.25)**	**0.63 (0.13)**	**0.44 (0.1)**
	σ^2_G	**184.7 (59)**	**46.9 (29.4)**	**99.2 (42.2)**	**52.9 (17.1)**	**292.9 (130.7)**	**125.8 (29)**	**114.8 (28.4)**
	Sample	1293	634	1412	3575	1063	3339	4638
	Mean	102.8	106.3	109.9	108.5	113	106.4	109.5
LU	**Q_{ST}**	**0.00 (0.0)**	**0.00 (0.0)**	**0.29 (0.32)**	**0.27 (0.29)**	**0.19 (0.24)**	**0.13 (0.08)**	**0.21 (0.16)**
	h^2	**0.89 (0.22)**	**1.55 (0.35)**	**0.67 (0.21)**	**0.42 (0.12)**	**0.96 (0.3)**	**1.04 (0.16)**	**0.54 (0.12)**
	h^2_{op} *	**1.01 *****	**0.92 ****	**1.51 *****	**0.59 ***	**1.23 ****	**1.05 *****	**0.78 ****
	σ^2_G	**37.1 (11.3)**	**78.7 (28)**	**22.9 (8.7)**	**11.3 (3.6)**	**25.8 (10.4)**	**42.9 (8.6)**	**14.4 (3.6)**

* The significance of h^2_{op} corresponds to the p-value of the linear regressions conducted between the standardized values of LU measured at PIE and LU measured *in situ* in 2006. Symbols meanings: * p <0.05; ** p < 0.01; *** p < 0.001.

When variance components were separately estimated within altitudinal classes, there was a clear reduction in the level of genetic variation of LU as altitude increased (Table 3). Indeed the genetic variances of LU varied from 78.7 at 400 m to 11.3 at 1200 m of elevation, and were dramatically lower for the high elevation group (Table 3). Differences between genetic variances of altitudinal levels were significant when comparing the provenance from 1200 m with provenances of 100 and 400 m ($p < 0.05$) and when provenances were grouped by

altitudinal groups ($p < 0.01$). The combined provenance and genetic variation is illustrated by the distribution of the mean values of leaf unfolding within each population (Fig. 3). The overall picture was the maintenance of a large within population genetic variation associated to a clinal population differentiation along the altitudinal gradient. The latest flushing tree from the lowest altitude unfolds its leaves at about the same period as the earliest flushing tree from the highest altitude.

While timing of germination also exhibited a large genetic variation (Table 2), the genetic variance σ^2_G of GERM did not show any altitudinal trend (Table 3), as genetic variances were not significantly different between altitudinal groups. The greatest σ^2_G values were found for provenances from the lowest and highest altitudes (184 .7 and 292.9 for 100 and 1600 m, respectively) whereas the weakest value was found at 400 m (46.9).

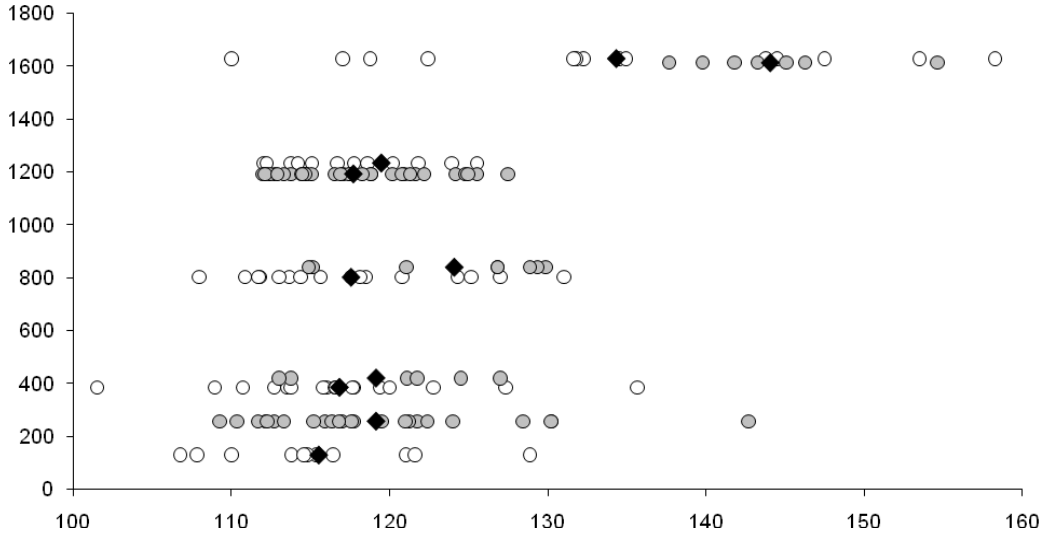
Correlations between traits and environmental variables

We found no genetic correlation between families and provenance values of GERM and LU ($r = 0.07$, $p = 0.40$ and $r = 0.44$, $p = 0.20$ respectively). However the correlation was highly significant between breeding values ($r = 0.67$; $p < 0.0001$) and provenance values ($r = 0.94$; $p < 0.0001$) of LU at PIE and TOUL.

Correlation between acorn weight and GERM was highly significant ($r = -0.77$, $p < 0.01$) at the provenance level. This correlation is mainly driven by the late germination of provenances from 1600 m and which acorns weighted largely less than those of others provenances (Table 1). The correlation was also significant between acorn weight and LU at PIE ($r = - 0.78$; $p < 0.01$) but not at TOUL ($r = -0.62$; $p = 0.05$). The slopes of the linear regressions decreased from -8.92 for GERM to -4.14 and -3.01 for LU at PIE and TOUL, respectively. However acorn weight was not correlated with the variability observed between families within provenances for any trait ($p > 0.05$).

The correlation between provenance values and environmental variables of the sites of origin were highly significant (Table 4). The provenance values of GERM were best correlated to spring and annual temperatures, whereas LU exhibited stronger correlation with altitude.

a)



b)

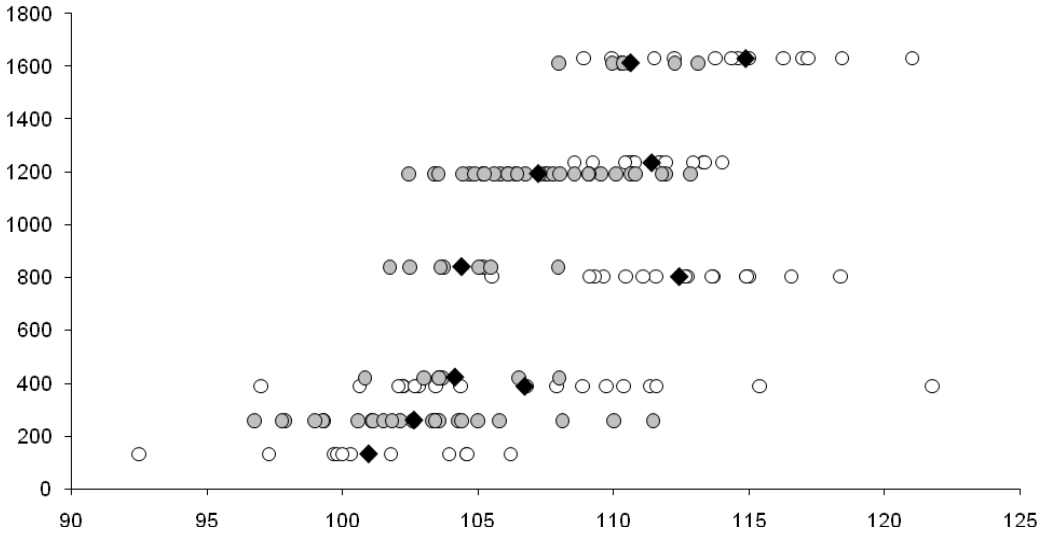


Figure 3: Variability of a) germination and b) leaf unfolding along the altitudinal gradients.

The black diamond-shaped points represent the mean of the provenances of a) GERM and b) LU measured at PIE, and the circle points represent the mean values of the families in white for Luz valley are and in light grey for Ossau valley.
The horizontal axis corresponds to the DOY of GERM and LU and the vertical axis corresponds to the altitude of provenances' sites.

Table 4: Correlations between provenance values (PV) of germination and leaf unfolding and environmental factors of provenances' sites of origin.

	Altitude *			Tspr06 †			Tyear06 ‡		
	slope	***r***	**df**	**slope**	***r***	**df**	**slope**	***r***	**df**
PV GERM	0.013	**0.70 ***	8	-3.086	**-0.80 ****	8	-3.455	**-0.79 ****	8
PV LU (PIE)	0.006	**0.82 ****	8	-1.365	**-0.78 ****	8	-1.374	**-0.69 ***	8
PV LU (TOUL)	0.006	**0.79 ****	8	-1.152	**-0.72 ***	8	-1.198	**-0.66 ***	8
PV LU (Multi-site §)	0.006	**0.82 ****	8	-1.265	**-0.76 ***	8	-1.289	**-0.68 ***	8

* Altitude: Altitude of the provenance's sites of origin; † Tspr06: mean temperatures from 1st March to 31th May 2006 and ‡ Tyear06: mean annual temperatures of year 2006 at the provenance's sites of origin.
§ The Multi-site analysis was conducted by grouping measured of LU at PIE and TOUL.
The slope, the correlation coefficient *r* and the degree of freedom df are given for each linear regression.
The significance of each regression corresponds to the following symbols: ns: non significant; * p <0.05; ** p < 0.01; *** p < 0.001.

DISCUSSION

Seed and leaf phenology displayed high levels of differentiation among provenances in common garden experiments. Differentiation of both traits followed a clear and strong altitudinal trend, with provenances from high elevation displaying later phenology than provenances from low elevation. In addition differentiation coexisted with very large within population genetic diversity. These results suggest peculiar interplays between gene flow and natural selection ensuring local extant adaptation. We suggest that these mechanisms may also allow populations to cope with future environmental changes.

Clinal variation and local adaptation

The altitudinal clinal variation of leaf unfolding in *Quercus petraea* was consistent across the two experimental plantations, and was also observed within each valley. These results contradict earlier findings that showed an opposite altitudinal cline of bud burst phenology in the same species (Deans *&* Harvey 1995; Ducousso *et al.,* 1996). However in these two studies, all provenances originated from below 500 m a.s.l., where temperature inversions often occur. Moreover, because of the large sampling over the natural range of the species, altitudinal effects were confounded with longitudinal and latitudinal effects. Indeed populations from southern latitudes flush earlier (Ducousso *et al.,* 1996) and mountainous areas are mainly located under southern latitudes in Europe. In our case, sampling of populations within the same valleys allows to disentangle altitudinal from latitudinal trends. We confirmed the results of Vitasse *et al.,* (2009b) who assessed altitudinal genetic

differentiation of six woody species (of which *Q. petraea*) coming from the same Pyrenean valleys. Despite a higher overall differentiation among provenances, the altitudinal trend of variation was almost identical (0.5 days.100 m^{-1}; Y. Vitasse, personal communication). Vitasse *et al.,* (2009b) suggested that the late flushing of high altitude provenances resulted from an adaptation to avoid leaf damages caused by late spring frosts which occur more frequently at high elevation. Population differentiation in our study amounted to the same level than differentiation observed among Scandinavian provenances of *Q. robur* and *Q. petraea* (Jensen & Hansen, 2008). Furthermore, altitudinal clines can be compared to latitudinal clines that have repeatedly been observed in *Quercus petraea* and showing that northern populations displayed later flushing than south populations (Liepe, 1993; Deans & Harvey, 1995; Ducousso *et al.,* 1996). Parallel clines for altitude and latitude suggest that the avoidance of late spring frosts either at higher altitudes or latitudes may be the selecting factor shaping population differentiation. We suspect therefore selection to be stronger at higher than lower altitudes, which is supported by the lower genetic variation residing within populations from higher altitudes (Fig. 2). Alternatives hypothesis to natural selection were recently proposed to account for high population differentiation for bud phenology in other tree species, and particularly in spruce. Indeed it was shown that "after effects", *e.g.* effect of the environmental conditions where seed development occurred, may modify bud set of *Picea abies* (Skroppa, 1994). Whether the response is due to maternal or epigenetic effects (Johnsen *et al.,* 2005), the consequences lead to rapid population differentiation within one single generation (Skroppa & Kohlmann, 1997). Epigenetic effects have been rarely described in deciduous trees and are poorly understood (Rohde & Junttila, 2008), and their contribution to population differentiation remains speculative.

Altitudinal variation was also observed for the date of germination. Later germination of high altitude populations was also found by Campbell (1979) on Douglas fir seeds. However in our study altitudinal variation was due to the very late germination of seed coming from the high altitude populations. As for leaf unfolding, altitudinal trends may be the result of natural selection favouring late germinating seeds to avoid late frost that would be fatal for the young seedlings. It would certainly account for the very late germination of populations from the highest altitudes. However population differentiation due to natural selection may be inflated or blurred by maternal effects. As seed mature on standing trees, their attributes are dependent on the maternal genotype and on the environmental conditions of the mother tree. Hence so called "maternal effects" may combine with genetic and environmental effects in our case. Maternal environment effects have been reported in the seed germination of *Arabidopsis*

thaliana. Donohue *et al.,* (2005) showed that temperature, photoperiod and light quality during seed maturation influenced significantly the germination timing. For example cold temperatures during seed development reinforce dormancy and thus delay germination. Along the altitudinal gradient, temperature and light quality vary considerably between populations, generating differences in seed maturation which may subsequently induce variation of the timing of acorn germination. However effects of seed size on the timing of germination have also been reported in several plant species without a common trend emerging among species (Kalisz, 1989; Zammit & Zedler, 1990; Platenkamp & Shaw, 1993; Simons & Johnston, 2000). Maternal provision is known to usually result in better-nurtured seeds that possess an obvious advantage for seedling establishment (Rohde & Junttila, 2008), as shown by positive correlation of seed weight and seedling development (Oleksyn *et al.,* 1998; Vitasse *et al.,* 2009b). When seeds contain few resources, a late germination thus appears as an advantageous strategy to ensure favourable conditions during seedlings emergence. In conclusion earlier records of the effects of environmental conditions prevailing during seed maturation on seedlings traits moderate our conclusions about genetic differentiation resulting from natural selection on the timing of acorn germination.

Maintenance of genetic diversity within population

Leaf and seed phenology exhibit very large heritability values in our experiments. Strong genetic control of bud phenology in forest tree species was reported in earlier studies (review in Howe *et al.,* 2003). Germination timing has been less investigated, although large variation was observed within populations of Douglas fir (Campbell 1979). In white oaks, our results confirm the high heritability previously reported. In a mapping study of bud burst QTLs in *Q. robur*, Scotti-Saintagne *et al.,* (2004) found a moderate to high heritability (0.15 – 0.52). In two provenance tests of 20/25 *Q. robur* and *Q. petraea* populations originating from northern Europe, heritability of bud burst amounted to 0.87 (Jensen, 1993). The same value was observed by Baliuckas *&* Pliura (2003), who studied open pollinated families coming from six Lithuanian populations of *Q. robur*. High heritability values may have a twofold source: either a low environmental and/or large genetic variances. Our results suggest that large genetic variance is the predominant cause of heritability, because large values were obtained by two different methods (parent-offspring correlations and intra class correlation) featured by obvious differences in environmental variances. Indeed the range of within population variation extends the range of the between population variation (Fig. 2). Interestingly this large genetic variation was maintained while local adaptation occurred at the same time.

There are at least three different non exclusive scenarios that may be advocated: (1) directional selection with low intensity and/or high gene flow, (2) directional selection changing directions over time and (3) disruptive selection favouring extreme flushing times. Scenario (1) has already been advocated in regards to the clinal pattern of leaf unfolding suggesting that late bud burst provides higher fitness at high altitudes. In this context large genetic variation can be maintained if selection intensity is low and/or new genes are introduced in populations via gene flow. Scenario (2) has been suggested earlier in response to environmental heterogeneity occurring through space (Campbell, 1979) but also over years. Fluctuating environmental changes may modify the selective values of traits and represent a powerful force for maintaining genetic diversity within natural populations (Jump *et al.*, 2009). While late frosts during spring can critically damage leaves of early flushing trees, the lengthening of growing season is advantageous if frost is avoided (Vitasse *et al.,* 2009a). The cumulative effect over time may result in the same outcome than balancing selection, *e.g.* maintaining large within population variation. Finally scenario (3) stipulates that different selective pressures act on leaf phenology in opposing directions. Besides the temperature mediated effect on bud burst, the synchrony between bud flushing of oaks and the emergence of pests populations such as herbivore insects or pathogens fungi may also be a determinant factor of selection in natural populations. Whereas late frosts can damage early flushing trees and select late flushing ones, the herbivore insects and pathogens fungi can be synchronized with late flushing trees, favouring early flushing trees. Indeed Desprez-Loustau *et al.,* (2010) have recently studied the synchrony between *Erisyphe alphitoides*, the pathogen responsible of the oak-powdery mildew, and oaks along the same altitudinal gradient in the Pyrenees and found that the pathogen was preferentially synchronized with late flushing oaks at low altitudes. However, synchronization changed with altitude and early flushing oaks were more infected at high altitudes. This result could thus also explain the decreasing trend of genetic variation with altitude for bud phenology. If the balance between biotic and abiotic selective pressures is modified along the altitudinal gradient, changes of selection types (disruptive at lower altitude and directional at higher altitude) can actually account for the reduction in genetic variation along altitude. Additionally directional selection intensity generated by late spring frosts may be stronger at high altitudes (Agrawal *et al.,* 2004) and contribute to the reduction of genetic variation. Finally the preferential unidirectional gene flow from lower to higher altitudes (Alberto *et al.,* 2010) may also be responsible for this trend.

For germination timing, the level of within population variability is larger than for leaf unfolding but does not follow an altitudinal trend. We suspect that biotic and abiotic selection

pressures shaped the distribution of genetic variation within populations just as for leaf unfolding. However, since germination timing is a highly critical stage for seedling survivorship, stronger diversification is expected for this trait. In this study the diversification of germination timing is largely caused by the high genetic variance although microenvironmental effects may also contribute to the variation existing among sibs of the same mother tree. In contrast to species where bet hedging was proposed (which are mainly selfing; Simons & Johnston, 2006), sessile oak is largely outcrossing resulting in large genetic variation between sibs. Our results therefore suggest that bet hedging may also present a strong genetic component in this species, and hence undergo its evolution.

Finally, the lack of genetic correlation between germination and leaf unfolding indicates different genetic controls between these phenology traits. It is likely that different genes are involved in the dormancy release of bud and seed. Consequently, research of candidate genes should not focus systematically on similar genes for the two traits.

Microevolution and adaptation to climate change

According to pollen records, the altitudinal colonization of *Q. petraea* in the Pyrenees lasted less than 2,000 years and was achieved between 11,000 and 10,000 BP (Aubert, 2001; Belet, 2001). In the two studied valleys, phylogeographical interpretations using chloroplast markers indicated that common source populations located in the low-land plain colonized the altitudinal gradients (Alberto *et al.,* 2010). Hence extant populations originated from the same gene pool and became locally adapted due to diversifying selection during the Holocene, while climate was progressively warming. Our results indicated that local adaptation along the altitudinal gradients occurred despite high gene flow connecting natural populations (Alberto *et al.,* 2010), while high levels of genetic variation was maintained.

However, several constraints could limit the potential response of tree species to global warming (Jump & Peñuelas, 2005). Assessing the level of genetic variation in date of bud burst in *Betula pubescens* and *Betula pendula*, Billington *&* Pelham, (1991) found high heritable variation within populations. The authors stated that the predicted rate of warming will be too rapid for the evolution of native populations without new variability brought by gene flow from earlier flushing populations. More recently Savolainen *et al.,* (2004) simulated the evolution of bud set and frost hardiness in *Pinus sylvestris* under progressive warming during 100 years but also in presence of gene flow between southern and northern populations. The rate of evolution was also slower than the rate of climate change but this result was mainly caused by the long lifespan of trees which reduced the opportunities for the

establishment of new genotypes. Moreover the delayed reproductive maturity of trees compared to annual plants may cause changes in the environment between establishment and reproduction of an individual (Jump & Peñuelas, 2005). Finally, because of negative genetic correlations between adaptive traits, evolutionary responses may be slower than predicted (Etterson & Shaw, 2001). Adaptive constraints to environmental changes have been considered in a theoretical frame highlighting the genetic and demographic factors driving the responses of populations (Bürger & Krall, 2004; Aitken *et al.*, 2008). Under changing environmental conditions the optimum value of a population (*e.g.* trait value for which fitness is highest) changes and natural selection will drive population to meet the optimum. Continuous environmental change maintains a lag between the population and its optimum inducing a reduction of the mean fitness of the population. There is a critical rate of environmental changes beyond which the lag may increase to the point that the population will go extinct. This critical level is dependent on demographic and genetic properties of the populations: standing genetic variation, demographic rate of population growth (fecundity), population size, and strength of selection. We have addressed in this contribution genetic variation and to a lesser extend demographic rate of growth by considering germination timing which is critical to population recruitment. In a previous study (Alberto *et al.*, 2010) we provided indirect estimates of effective population sizes and showed that, although decreasing with altitudes, they exceeded a few thousands. While assessments of population reproduction are still missing, our results indicate that altitudinal oak populations have important resources facilitating future adaptation and allowing upslope shift induce by global warming.

REFERENCES

Acevedo-Rodriguez, R., Vargas-Hernandez, J.J., Lopez-Upton, J. & Mendoza, J.V. 2006. Effect of geographic origin and nutrition on shoot phenology of Mexican Douglas-Fir (Pseudotsuga sp.) seedlings. *Agrociencia* **40:** 125-137.

Agrawal, A.A., Conner, J.K. & Stinchcombe, J.R. 2004. Evolution of plant resistance and tolerance to frost damage. *Ecol. Lett.* **7:** 1199-1208.

Aitken, S.N., Yeaman, S., Holliday, J.A., Wang, T.L. & Curtis-McLane, S. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evol. Appl.* **1:** 95-111.

Alberto, F., Niort, J., Derory, J., Lepais, O., Vitalis, R., Galop, D., & Kremer, A. 2010. Population differentiation of sessile oak at the altitudinal front of migration in the French Pyrenees. *Mol. Ecol.* (accepted)

Aubert S. 2001. Limite supérieure de la forêt, climat et anthropisation: dynamique tardiglaciaire et holocène de la végétation dans la vallée du Marcadau (Hautes-Pyrénées, France*). Thèse de doctorat*, Université Toulouse III, 365 p

Baliuckas, V. & Pliura, A. 2003. Genetic variation and phenotypic plasticity of Quercus robur populations and open-pollinated families in Lithuania. *Scand. J. Forest Res.* **18:** 305-319.

Belet J.M. 2001. Variations tardiglaciaires et Holocène de la limite supérieure de la forêt dans les Pyrénées centrales : exemple du vallon d'Estibère (Hautes Pyrénées, France). *Thèse de doctorat*. Université Toulouse III, 241 p.

Bennie, J., Kubin, E., Wiltshire, A., Huntley, B. & Baxter, R. 2009. Predicting spatial and temporal patterns of bud-burst and spring frost risk in north-west Europe: the implications of local adaptation to climate. *Global Change Biol.* Published Online. (DOI: 10.1111/j.1365-2486.2009.02095.x)

Bertin, R.I. 2008. Plant phenology and distribution in relation to recent climate change. *J. Torrey Bot. Soc.* **135:** 126-146.

Billington, H.L. & Pelham, J. 1991. Genetic Variation in the Date of Budburst in Scottish Birch Populations: Implications for Climate Change. *Funct. Ecol.* **5:** 403-409.

Bürger, R. & Krall, C., 2004. Quantitative-genetic models and changing environments. In *Evolutionary Conservation Biology*. Ferrière, R., Dieckmann, U., Couvet, D. (eds.). Cambridge University Press, pp. 171–187.

Bradshaw, W.E. & Holzapfel, C.M. 2008. Genetic response to rapid climate change: it's seasonal timing that matters. *Mol. Ecol.* **17:** 157-166.

Brewer, S., Cheddadi, R., de Beaulieu, J.L. & Reille, M. 2002. The spread of deciduous Quercus throughout Europe since the last glacial period. *For. Ecol. Manag.* **156:** 27-48.

Campbell, R.K. 1979. Genecology of Douglas-Fir in a Watershed in the Oregon Cascades. *Ecology* **60:** 1036-1050.

Chmura, D.J. & Rozkowski, R. 2002. Variability of beech provenances in spring and autumn phenology. *Silvae Genet.* **51:** 123-127.

Chuine, I. & Cour, P. 1999. Climatic determinants of budburst seasonality in four temperate-zone tree species. *New Phytol.* **143:** 339-349.

Cohen, D. 1966. Optimizing reproduction in a randomly varying environment. *J. Theor. Biol.* **12:** 119-129.

Davis, M.B. & Shaw, R.G. 2001. Range shifts and adaptive responses to Quaternary climate change. *Science* **292:** 673-679.

Daubree, J.B. & Kremer, A. 1993. Genetic and phenological differentiation between introduced and natural populations of Quercus rubra L. *Ann. For. Sci.* **50:** 271s-280s.

Deans, J.D. & Harvey, F.J. 1995. Phenologies of 16 European provenances of sessile oak growing in Scotland. *Forestry* **68:** 265-273.

Derory, J., Leger, P., Garcia, V., Schaeffer, J., Hauser, M.T., Salin, F., Luschnig, C., Plomion, C., Glossl, J. & Kremer, A. 2006. Transcriptome analysis of bud burst in sessile oak (Quercus petraea). *New Phytol.* **170:** 723-738.

Desprez-Loustau, M.L., Vitasse, Y., Delzon, S., Capdevielle, X., Marcais, B. & Kremer, A. 2010. Are plant pathogen populations adapted for encounter with their host? A case study of phenological synchrony between oak and an obligate fungal parasite along an altitudinal gradient. *J. Evol. Biol.* **23:** 87-97.

Donohue, K., Dorn, L., Griffith, C., Kim, E., Aguilera, A., Polisetty, C.R. & Schmitt, J. 2005. The evolutionary ecology of seed germination of Arabidopsis thaliana: Variable natural selection on germination timing. *Evolution* **59:** 758-770.

Ducousso, A., Guyon, J.P. & Kremer, A. 1996. Latitudinal and altitudinal variation of bud burst in western populations of sessile oak (Quercus petraea (Matt) Liebl). *Ann. For. Sci.* **53:** 775-782.

Etterson, J.R. & Shaw, R.G. 2001. Constraint to adaptive evolution in response to global warming. *Science* **294:** 151-154.

Evans, M.E.K., Ferriere, R., Kane, M.J. & Venable, D.L. 2007. Bet hedging via seed banking in desert evening primroses (Oenothera, Onagraceae): Demographic evidence from natural populations. *Am. Nat.* **169:** 184-194.

Finch-Savage, W.E. & Leubner-Metzger, G. 2006. Seed dormancy and the control of germination. *New Phytol.* **171:** 501-523.

Ghelardini, L. & Santini, A. 2009. Avoidance by early flushing: a new perspective on Dutch elm disease research. *iForest* **2:** 143-153.

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J. & Thompson, R. 2002. ASReml User Guide Release 1.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Howe, G.T., Saruul, P., Davis, J. & Chen, T.H.H. 2000. Quantitative genetics of bud phenology, frost damage, and winter survival in an F-2 family of hybrid poplars. *Theor. Appl. Genet.* **101:** 632-642.

Howe, G.T., Aitken, S.N., Neale, D.B., Jermstad, K.D., Wheeler, N.C. & Chen, T.H.H. 2003. From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. *Can. J. Bot.* **81:** 1247-1266.

IPCC, 2007. Climate Change 2007: Impacts, Adaptation and Vulnerability. *Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. University Press, Cambridge, United Kingdom and New York, NY, USA., Cambridge.

Jensen, J.S. 1993. Variation of growth in Danish provenance trials with oaks (*Quercus robur* L. and *Quercus petraea* Mattuschka Liebl.). *Ann. For. Sci.* **50:** 203s-207s

Jensen, J.S. & Hansen, J.K. 2008. Geographical variation in phenology of Quercus petraea (Matt.) Liebl and Quercus robur L. oak grown in a greenhouse. *Scand. J. Forest Res.* **23:** 179-188.

Johnsen, O., Fossdal, C.G., Nagy, N., Molmann, J., Daehlen, O.G. & Skroppa, T. 2005. Climatic adaptation in Picea abies progenies is affected by the temperature during zygotic embryogenesis and seed maturation. *Plant Cell Environ.* **28:** 1090-1102.

Jump, A.S. & Penuelas, J. 2005. Running to stand still: adaptation and the response of plants to rapid climate change. *Ecol. Lett.* **8:** 1010-1020.

Jump, A.S., Marchant, R. & Penuelas, J. 2009. Environmental change and the option value of genetic diversity. *Trends Plant Sci.* **14:** 51-58.

Kalisz, S. 1989. Fitness consequences of mating system, seed weight, and emergence date in a winter annual, Collinsia-verna. *Evolution* **43:** 1263-1272.

Kremer, A. 1994. Genetic diversity and phenotypic variability of forest trees. *Genet. Sel. Evol.* **26:** S105-S123.

Kremer, A. 2007. How well can existing forests withstand climate change? In *Climate change and forest genetic diversity. Implications for sustainable forest management in Europe*, Koskela, J., A. Buck, Teissier du Cros E. (eds.). Bioversity International, Rome. pp. 3-17.

Kremer, A., Le Corre, V., Petit, R.J. & Ducousso, A. 2010. Historical and contemporary dynamics of adaptive differentiation in European oaks. In *Molecular approaches in natural resource Conservation*. DeWoody A., Bickham J., Michler C., Nichols K., Rhodes G., Woeste K. (eds.). Cambridge University Press (in press).

Kriebel, H.B., Bagley, W.T., Deneke, F.J., Funsch, R.W., Roth, P., Jokela, J.J., Merritt, C., Wright, J.W. & Williams, R.D. 1976. Geographic variation in Quercus rubra in north central United-states plantations. *Silvae Genet.* **25:** 118-122.

Lenoir, J., Gegout, J.C., Marquet, P.A., de Ruffray, P. & Brisse, H. 2008. A significant upward shift in plant species optimum elevation during the 20th century. *Science* **320:** 1768-1771.

Liepe, K. 1993. Growth-chamber trial on frost hardiness and field trial on flushing of sessile oak (Quercus petraea Liebl). *Ann. For. Sci.* **50:** 208s-214s.

Lindner, M., Maroschek, M., Netherer, S., Kremer, A., Barbati, A., Garcia-Gonzalo, J., Seidl, R., Delzon, S., Corona, P., Kolström, M., Lexer, M.J. & Marchetti, M. 2009. Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems *For. Ecol. Manag.* (in press)

Lynch, M., & Walsh, B. 1998. *Genetics and Analysis of Quantitative Traits.* Sinauer Associates, Inc. 980 pp.

Magri, D., Vendramin, G.G., Comps, B., Dupanloup, I., Geburek, T., Gomory, D., Latalowa, M., Litt, T., Paule, L., Roure, J.M., Tantau, I., van der Knaap, W.O., Petit, R.J. & de Beaulieu, J.L. 2006. A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytol.* **171:** 199-221.

Menzel, A., Sparks, T.H., Estrella, N., Koch, E., Aasa, A., Ahas, R., Alm-Kubler, K., Bissolli, P., Braslavska, O., Briede, A., Chmielewski, F.M., Crepinsek, Z., Curnel, Y., Dahl, A., Defila, C., Donnelly, A., Filella, Y., Jatcza, K., Mage, F., Mestre, A., Nordli, O., Penuelas, J., Pirinen, P., Remisova, V., Scheifinger, H., Striz, M., Susnik, A., Van Vliet, A.J.H., Wielgolaski, F.E., Zach, S. & Zust, A. 2006. European phenological response to climate change matches the warming pattern. *Global Change Biol.* **12:** 1969-1976.

Morgenstern, E.K. 1996. *Geographic variation in forest trees*. University of British Columbia Press, Vancouver, 209 pp

Nordli, O., Wielgolaski, F.E., Bakken, A.K., Hjeltnes, S.H., Mage, F., Sivle, A. & Skre, O. 2008. Regional trends for bud burst and flowering of woody plants in Norway as related to climate change. *Int. J. Biometeorol.* **52:** 625-639.

Oleksyn, J., Modrzynski, J., Tjoelker, M.G., Zytkowiak, R., Reich, P.B. & Karolewski, P. 1998. Growth and physiology of Picea abies populations from elevational transects: common garden evidence for altitudinal ecotypes and cold adaptation. *Funct. Ecol.* **12:** 573-590.

Parmesan, C. 2006. Ecological and evolutionary responses to recent climate change. *Annu. Rev. Ecol. Evol. Syst.* **37:** 637-669.

Petit, R.J., Csaikl, U.M., Bordacs, S., Burg, K., Coart, E., Cottrell, J., van Dam, B., Deans, J.D., Dumolin-Lapegue, S., Fineschi, S., Finkeldey, R., Gillies, A., Glaz, I., Goicoechea, P.G., Jensen, J.S., Konig, A.O., Lowe, A.J., Madsen, S.F., Matyas, G., Munro, R.C., Olalde, M., Pemonge, M.H., Popescu, F., Slade, D., Tabbener, H., Taurchini, D., de Vries, S.G.M., Ziegenhagen, B. & Kremer, A. 2003. Chloroplast DNA variation in European white oaks phylogeography and patterns of diversity based on data from over 2600 populations (vol 156, pg 5, 2002). *For. Ecol. Manag.* **176:** 595-599.

Platenkamp, G.A.J. & Shaw, R.G. 1993. Environmental and genetic maternal effects on seed characters in Nemophila-menziesii. *Evolution* **47:** 540-555.

Premoli, A.C., Raffaele, E. & Mathiasen, P. 2007. Morphological and phenological differences in Nothofagus pumilio from contrasting elevations: Evidence from a common garden. *Austral. Ecol.* **32:** 515-523.

Rohde, A., Howe, G.T., Olsen, J.E., Moritz, T., van Montagu, M., Junttila, O. & Boerjan, W. 2000. Molecular Aspects of bud dormancy in trees. In: *Molecular biology of woody plants*. Jain, S.M., Minocha, S.C., (eds.). Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 89-134.

Rohde, A. & Bhalerao, R.P. 2007. Plant dormancy in the perennial context. *Trends Plant Sci.* **12:** 217-223.

Rohde, A. & Junttila, O. 2008. Remembrances of an embryo: long-term effects on phenology traits in spruce. *New Phytol.* **177:** 2-5.

Ruttink, T., Arend, M., Morreel, K., Storme, V., Rombauts, S., Fromm, J., Bhalerao, R.P., Boerjan, W. & Rohde, A. 2007. A molecular timetable for apical bud formation and dormancy induction in poplar. *Plant Cell* **19:** 2370-2390.

Saenz-Romero, C., Guzman-Reyna, R.R. & Rehfeldt, G.E. 2006. Altitudinal genetic variation among Pinus oocarpa populations in Michoacan, Mexico - Implications for seed

zoning, conservation, tree breeding and global warming. *For. Ecol. Manag.* **229:** 340-350.

Savolainen, O., Bokma, F., Garcia-Gil, R., Komulainen, P. & Repo, T. 2004. Genetic variation in cessation of growth and frost hardiness and consequences for adaptation of Pinus sylvestris to climatic changes. *For. Ecol. Manag.* **197:** 79-89.

Saxe, H., Cannell, M.G.R., Johnsen, B., Ryan, M.G. & Vourlitis, G. 2001. Tree and forest functioning in response to global warming. *New Phytol.* **149:** 369-399.

Scotti-Saintagne, C., Bodenes, C., Barreneche, T., Bertocchi, E., Plomion, C. & Kremer, A. 2004. Detection of quantitative trait loci controlling bud burst and height growth in Quercus robur L. *Theor. Appl. Genet.* **109:** 1648-1659.

Simons, A.M. & Johnston, M.O. 2000. Plasticity and the genetics of reproductive behaviour in the monocarpic perennial, Lobelia inflata (Indian tobacco). *Heredity* **85:** 356-365.

Simons, A.M. & Johnston, M.O. 2006. Environmental and genetic sources of diversification in the timing of seed germination: Implications for the evolution of bet hedging. *Evolution* **60:** 2280-2292.

Simons, A.M. 2009. Fluctuating natural selection accounts for the evolution of diversification bet hedging. *Proc. R. Soc. Biol. Sci. Ser. B* **276:** 1987-1992.

Skroppa, T. 1994. Growth rythm and hardiness of Picea abies progenies of high altitude parents from seed produced at low elevations. *Silvae Genet.* **43:** 95-100.

Skroppa, T. and K. Kohlmann, 1997. Adaptation to local conditions after one generation in Norway spruce. *Forest genetics* 4 (3): 171-177.

Spitze, K. 1993. Population structure in Daphnia obusta - Quantitative genetic and allozymic variation. *Genetics* **135:** 367-374.

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C. , Erasmus, B.F.N., Ferreira de Siqueira, M., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Townsend Peterson, A., Phillips, O.L. & Williams, S. E. 2004. Extinction risk from climate change. *Nature* **427:**145-148.

Van Asch, M., Tienderen, P.H., Holleman, L.J.M. & Visser, M.E. 2007. Predicting adaptation of phenology in response to climate change, an insect herbivore example. *Global Change Biol.* **13:** 1596-1604.

Venable, D.L. 2007. Bet hedging in a guild of desert annuals. *Ecology* **88:** 1086-1090.

Visser, M.E. & Holleman, L.J.M. 2001. Warmer springs disrupt the synchrony of oak and winter moth phenology. *Proc. R. Soc. Lond. B Biol. Sci.* **268:** 289-294.

Vitasse, Y., Porte, A.J., Kremer, A., Michalet, R. & Delzon, S. 2009a. Responses of canopy duration to temperature changes in four temperate tree species: relative contributions of spring and autumn leaf phenology. *Oecologia* **161:** 187-198.

Vitasse, Y., Delzon, S., Bresson, C.C., Michalet, R. & Kremer, A. 2009b. Altitudinal differentiation in growth and phenology among populations of temperate-zone tree species growing in a common garden. *Can. J. For. Res.* **39:** 1259-1269.

Vitasse, Y., Delzon, S., Dufrene, E., Pontailler, J.Y., Louvet, J.M., Kremer, A. & Michalet, R. 2009c. Leaf phenology sensitivity to temperature in European trees: Do within-species populations exhibit similar responses? *Agric. For. Meteorol.* **149:** 735-744.

Viveros-Viveros, H., Saenz-Romero, C., Lopez-Upton, J. & Vargas-Hernandez, J.J. 2005. Altitudinal genetic variation in plant growth of Pinus pseudostrobus Lindl. in field testing. *Agrociencia* **39:** 575-587.

Von Wuehlisch, G., Krusche, D. & Muhs, H.J. 1995. Variation in temperature sum requirement for flushing of beech provenances. *Silvae Genet.* **44:** 343-346.

Worrall, J. 1983. Temperature - bud-burst relationships in amabilis and subalpine fir provenance tests replicated at different elevations. *Silvae Genet.* **32:** 203-209.

Wright, J.W. 1976. *Introduction to forest genetics*. Academic Press, New York, NY, USA.

Zammit, C. & Zedler, P.H. 1990. Seed yield, seed size and germination behavior in the annual Pogogyne-abramsii. *Oecologia* **84:** 24-28.

Population genomics of clines: Nucleotide diversity of bud burst candidate genes in sessile oak populations from latitudinal and altitudinal gradients

Florian Alberto[1,2], Jérémy Derory[1,2], Antoine Kremer[1,2]

(1) INRA, UMR1202 BIOGECO, F-33610 Cestas, France

(2) Université de Bordeaux, UMR1202 BIOGECO, F-33610 Cestas, France

To be submitted to *Genetics*

INTRODUCTION

Local adaptation of temperate forest trees in relation to environmental variation has been widely investigated using common gardens experiments or provenance tests (Langlet 1971; Matyas 1996; König 2005 for reviews). When range wide collections of populations were compared continuous patterns of differentiation were often observed following environmental gradients (Wright 1976; Morgenstern 1996). These patterns called clines were very frequently reported for phenological traits (Howe *et al.* 2003; Savolainen *et al.* 2007) along geographic gradients (latitude, altitude, longitude). Congruent clines were identified across species and experiments within species suggesting that they witnessed systematic directional selection in response to identical environmental gradients. For example, in north temperate conifers (either spruces or pines), populations from northern latitudes flush earlier than southern populations (Wright, 1976; Morgenstern, 1996) and steep clines were observed in species extending over large latitudinal ranges (for examples *Pinus sylvestris* in Scandinavia, Giertych, 1991; *Picea Sitchensis* in North America, Mimura *&* Aitken, 2007). In north temperate broadleaves, clines for bud burst were also observed, but follow opposite trends. In beech (von Wuehlish, 1995), birch (Worrell *et al.,* 2000) and sessile oak (Liepe, 1993; Deans *&* Harvey, 1995; Ducousso *et al.* 1996) southern populations flush earlier than northern. Similar clinal variations were also reported along altitude for oaks, where populations from higher altitude flushed later than populations from lower altitude (Vitasse *et al.* 2009a), suggesting that the time of bud burst in sessile oak is mainly driven by temperature signals (Vitasse *et al.* 2009b). Although clinal variation can be generated by isolation by distance or admixture (Vasemägi, 2006), we suspect that congruent clinal variation across different environmental gradients would be the genetic imprint of systematic directional selection. Populations showing high differentiation for adaptive traits along environmental gradients thus constitute ideal case studies to detect the molecular basis of adaptation. In this study, we explore the genetic basis of the clinal altitudinal and latitudinal variation of bud burst in sessile oak, by monitoring the nucleotide diversity in candidate genes underlying bud burst.

The extant distribution of nucleotide diversity results from systematic and stochastic evolutionary forces, and disentangling demographic from selection signals is a major concern in this respect (Thornton *et al.* 2007). Demographic events such as bottlenecks or population expansion leave genome wide signatures whereas selection imprints are only expected in genes involved in fitness related traits. Hence one way to unravel the molecular signals left by both evolutionary forces is the use of large genome-wide datasets which allow to catch the

overall signal of demography and to detect loci under selection departing from this background. One approach to track signals of selection in the genome is based on the comparison between levels of genetic diversity and differentiation within and between populations. A multi-locus genome scan is performed to identify the loci submitted to selection displaying either particularly high (directional selection) or low (balancing selection) differentiation compared to the majority of loci representing the differentiation caused by demographic events. Such approach has been increasingly used in non-model species with the development of several statistical models to test the departure of loci from a 'neutral' model inferred from the data (Beaumont *&* Nichols 1996, Beamont *&* Balding 2004, Foll *&* Gaggiotti 2008). In tree species genome wide scans have not yet been implemented and genome explorations of population differentiation were limited today to large sets of candidate genes. In *Picea glauca* Namroud *et al.* (2008) used 534 SNPs from candidate genes involved in several fitness-related traits. Searching adaptive signatures for drought tolerance Eveno *et al.* (2008) studied the differentiation of 11 candidate genes in *Pinus pinaster* populations. In these two studies comparisons between different methods were made to evaluate the advantages and limitations of the demographic models to detect selection. The first method relies on the estimation of the overall differentiation from a symmetrical island model (Beaumont *&* Nichols 1996). The second method developed by Beaumont *&* Balding (2004) distinguishes population and locus effects thus accounting for heterogeneous migration rates and population sizes.

Candidate genes in forest trees are selected among existing EST banks that have accumulated in recent years, and are either expressional, functional, or positional candidates. Genetic surveys were conducted in natural populations of European aspen (Ingvarsson *et al.* 2008), Douglas fir (Eckert *et al.* 2009a), maritime pine (Eveno *et al.* 2008), scots pine (Wachowiak *et al.* 2009), loblolly pine (Gonzalez-Martìnez *et al.* 2006) and Norway spruce (Heuertz *et al.* 2006) to monitor nucleotide diversity. In recent years the discovery of candidate genes in phenological traits of trees increased with the development of gene expression and mapping studies (Hall *et al.* 2007, Eckert *et al.* 2009a). In oaks, monitoring of gene expression during bud burst by microarray and quantitative PCR (Derory *et al.* 2006) and mapping of the differentially expressed genes (Derory *et al.* 2010) resulted in the selection of more than 100 candidate genes that are investigated here. Our study species, sessile oak, is distributed over a wide latitudinal range in Europe from Spain to Southern Scandinavia. It extends also within mountainous areas in the Alps, Carpathians and Pyrenees (Arbez *&* Lacaze, 1998). We conducted a genetic survey in populations sampled from the overall range distribution, and

from two valleys in the Pyrenees to assess polymorphism variation along latitudinal and altitudinal gradients. Our main objective is to check whether the clinal variation of bud burst observed in common garden experiments translates into a clinal variation of allelic frequencies at the underlying genes. We first dissect nucleotide polymorphism at the gene level and assess linkage disequilibrium within genes. We further estimate population differentiation with the aim to identify SNPs with allelic profiles deviating from neutral expectation using the two genome-scan methods described above. Finally we compare the variation of allelic frequencies along latitude and altitude to detect clinal patterns.

MATERIAL AND METHODS

Sampled populations

Selected populations for this study were sampled according to their clinal variation of apical bud phenology recorded in provenance tests. A range wide provenance test of *Q. petraea* comprising 112 populations was installed in four different plantations located in the central and eastern part of France between 1990 and 1995 and used to investigate the latitudinal clinal variation of apical bud burst (Ducousso *et al.* 1996, Ducousso *et al.* 2005, Kremer *et al.* 2010). Assessments of time of bud phenology conducted at 3 years after plantations were used to select for this study a subset of 21 provenances distributed along the latitudinal cline. Similarly the altitudinal clinal variation was investigated in a provenance test comprising populations of *Q. petraea* from two valleys between 100 m and 1600 m of elevation on the northern side of the Pyrenees (see for details Alberto *et al.* 2010b). These populations were previously studied for leaf unfolding and leaf coloration phenology (Vitasse *et al.* 2009). We selected 12 populations for this study. One population, Ad, was included in both gradients as it was located at the same time in the southern range of the latitudinal gradient and at low altitude in one valley. Geographic data of the 32 sampled populations are shown on Table 1. For both sets of populations (latitudinal and altitudinal cline), population structure was investigated with isozymes (for the latitudinal cline; Zanetto & Kremer, 1995) and microsatellites (for the altitudinal cline; Alberto *et al.* 2010a) and indicated very low population differentiation (F_{ST} amounted respectively to 0.025 and 0.023).

Bud or leaf tissues were collected on 30 trees for each population of the latitudinal cline within the provenance test. Similarly buds or leaves were harvested on 9 to 40 trees per population belonging to the altitudinal cline (Table 1). The latter collections were made *in situ* on the trees where earlier acorn collections were made to install the provenance test. The overall sample consisted of 32 natural populations and 758 trees (Table 1).

Candidate genes diversity

Candidate genes for bud burst were selected according to three criteria (i) their differential expression before and after bud flush (Derory *et al.* 2006), (ii) their position on the genetic map, *e.g.* their collocation with QTL of bud burst (Derory *et al.* 2010), and (iii) their functional role as assessed in model plants. Following these criteria, a total of 105 candidate genes were selected. Nucleotide polymorphism within candidate genes was assessed by two different methods according to the criteria used for their selection.

(i) For 9 genes showing strong contrasting expression profiles during bud burst and/or exhibiting collocation with QTLs of bud burst, haplotype sequencing was conducted on a sub sample trees of the altitudinal and latitudinal cline (see next paragraph and Table 2). Sequencing was done on partial fragments of the genes corresponding to the EST fragments as obtained in the transcriptomic analysis (Derory *et al.* 2006).

(ii) Nucleotide diversity of 105 candidate genes (including the previous genes) for bud burst was assessed for 384 single nucleotide polymorphisms (SNPs) on the whole sample representing 758 individuals from both environmental gradients.

Haplotype sequencing of gene fragments

Sample

A subsample of each environmental gradient was used as a diversity panel for the description of the nucleotide diversity of 9 candidate genes. This diversity panel was also used to estimate the haplotypic diversity and the extent of linkage disequilibrium within and between the candidate genes and for the discovery of single nucleotide polymorphism (SNPs). For the latitudinal gradient, 6 individuals were sequenced in 9 out of the 21 populations presenting contrasted bud phenology and between 8 and 16 individuals were sequenced in five populations from the same valley for the altitudinal gradient (Table 1).

Table 1: Localization and sample size of the 32 sampled populations from the latitudinal (purple color) and altitudinal (blue color) gradients in the sequencing and genotyping of candidate genes.

		Geographic coordinates			Sequencing									Genotyping	
Code	**Population**	**Longitude**	**Latitude**	**Altitude**	***ASI***	***AUX_REP***	***DAG2***	***GALA***	***H3***	***YSL1***	***GA3***	***GA20***	***PM23***	**Illumina**	***GALA***
Ad	**Ade**	-0.01	43.15	450										43	36
Bl	**Blakeney**	-2.50	51.78	76										21	19
Co	**Cochem**	7.05	50.08	400	4	5	5	5	5	5	6	2	4	21	23
Fo	**Fontainebleau**	2.70	48.42	80										25	20
Go	**Gohrde**	10.85	53.10	90										25	22
Ha	**Haslach**	7.38	48.55	265	6	6	4	5	6	6	4	2	6	18	22
Jo	**Johanneskreuz**	7.83	49.40	460	4	6	6	5	6	6	6	2	6	27	23
Kl	**Klostermarienberg**	16.57	47.41	310	4	5	5	3	6	3	6	2	6	26	22
La	**Lapwald**	10.95	52.28	180										22	20
Le	**Lembach**	7.83	48.98	260										26	21
Lo	**Longchamps**	5.30	47.27	235										23	24
Lu	**Luss**	10.30	52.84	110										24	20
Mo	**Mölln**	10.75	53.62	36	5	5	3	6	6	6	5	2	5	24	18
Ne	**La Neuville en Hez**	2.33	49.40	70	5	6	5	5	6	6	5	3	6	27	21
Ob	**Obora**	19.08	48.58	400										26	19
Ro	**Romersberg**	6.73	48.82	220										21	22
Sa	**Saint Jean**	6.72	48.82	227	5	6	5	5	6	6	5	3	6	25	17
Sp	**Sprakensehl**	10.60	52.80	115	6	6	4	5	6	6	6	2	5	25	24
St	**Still**	7.25	48.58	688										26	22
We	**Westhoffen**	7.45	48.60	400	5	4	5	5	5	4	3	3	6	25	21
Wo	**Wolfganz**	9.05	50.15	160										24	23
L1	**Laveyron**	-0.22	43.76	130	6	8	6	*0*	6	7	*0*	*0*	*0*	19	20
L3	**Ibos**	-0.01	43.26	350										25	24
L4	**Ade**	-0.01	43.15	450	7	8	16	7	5	8	*0*	*0*	*0*	43	36
L8	**Papillon**	-0.03	42.92	800	7	8	8	7	7	7	*0*	*0*	*0*	21	20
L12	**Gèdres**	0.03	42.79	1220										32	*0*
L14	**Gèdres**	0.04	42.79	1350	7	8	6	6	6	6	*0*	*0*	*0*	23	21
L16	**Péguères**	-0.13	42.87	1610	9	9	7	2	8	8	*0*	*0*	*0*	23	22
O1	**Josbaig**	-0.74	43.23	200										22	26
O4	**Bager**	-0.54	43.07	400										22	23
O8	**Le Hourcq**	-0.43	42.91	897										10	10
O12	**Gabas**	-0.42	42.89	1250										28	29
O16	**Artouste**	-0.40	42.89	1660										9	9
Total					***80***	***90***	***85***	***66***	***84***	***84***	***46***	***21***	***50***	***758***	***663***

Candidate genes

Galactinol synthase (*GALA*), Dof Affecting Germination 2 (*DAG2*), Alpha-amylase/subtilisin inhibitor (ASI), metal-nicotianamine transporter (*YSL*1) an auxin-repressed protein (*AUX-REP*) and a seed maturation protein (*PM23*) were all down regulated from quiescent to developing buds (Derory *et al.* 2006) in microarray and quantitative PCR experiments and thus constitute relevant expressional candidate genes for molecular signals regulating bud burst transition. Details on these candidate genes and their biological function are given in Derory *et al.* (2010). Two genes involved in the biosynthesis gibberellins were also selected as gibberellins are known to be implied in bud flushing (Falusi and Calamassi 2003, Or *et al.* 2000). We included Gibberellin 20-oxydase (*GA20*) and Gibberellin 3-β-hydroxylase (*GA3*), which catalyse the synthesis of the active form of gibberellins in plants GA1 (Calvo *et al.* 2004, Israelsson *et al.* 2004, Perez-Flores *et al.* 2003). Moreover, 5 out of these 9 genes were located within QTL regions displaying high contributions to bud burst variation (Derory *et al.* 2010). The 9 fragments of genes were sequenced in the populations of the latitudinal gradient whereas only 6 (*ASI*, *AUX-REP*, *DAG2*, *GALA*, *H3* and *YSL1*) were sequenced in the populations of the altitudinal gradient (Table 1). DNA extraction, primer design and PCR conditions are described in previous studies (Derory *et al.* 2006, Casasoli *et al.* 2006, Derory *et al.* 2010).

Sequencing of CG fragments

PCR conditions and cloning strategy are described for the populations of the latitudinal gradient in Derory *et al.* (2010). The same protocol was used for the populations of the altitudinal gradient with one haplotype obtained for each individual.

High throughput genotyping

Sample

The genotyping was conducted on the overall sample comprising 21 populations for the latitudinal gradient and 12 populations for the altitudinal gradient and a total of 798 individuals (Table 1).

Candidate genes and SNP discovery

We selected candidate genes among the 190 unigenes originating from the expressional study of Derory *et al.* (2006) and displaying interesting patterns of expression regulation before, during and after bud burst.

Among the 190 unigenes, 21 were already sequenced as fragments of genes in the populations of the diversity panel (unpublished data). These genes were selected according to the three criterions mentioned above and comprised the nine fragments of genes sequenced for the diversity panel and genes involved in the flowering pathway, the oxidative stress response, the hormone biosynthesis, light perception or also transcription factors. We used the sequences of the fragments to design SNPs from alignments containing at least 20 sequences (referred as divSNPs). The chromatograms were previously visually checked and nucleotides with phred scores below 30 were considered as missing data. The bud burst EST library (Derory *et al.* 2006) was successively implemented with ESTs from other libraries which led to the clustering of the 190 initial unigenes in 124 contigs of ESTs with on average more than 800 bp length. The SNPs discovered in these contigs of ESTs will be referred as electronic SNPs (elecSNPs). The identification of SNPs was done separately for the divSNPs and for the elecSNPs using the automatic pipeline *snp2illumina*. This Perl script automatically extracts SNPs from multifasta sequence files and generates and output SequenceList file compatible with the Illumina Assay Design Tool software (ADT). This file contains the SNP names and surrounding sequences with polymorphic loci indicated by IUPAC codes for degenerated bases. SNPs were selected on functionality scores provided by ADT and minor allele frequency. We discarded all the SNPs presenting a functionality score below 0.5 as recommended by the manufacturer. A total of 78 divSNPs and 306 elecSNPs representing 105 candidate genes for bud burst phenology were included in the 384-plex assay (Table S1). Among these SNPs we included 26 insertions-deletions (indels) with sizes ranging from one to three base pairs. The genotyping of indels was possible by coding either for the first base of the insertion or for the first base after the insertion.

SNP genotyping

Genotyping was carried out using the Illumina GoldenGate SNP genotyping platform (Illumina, San Diego, CA, USA) and conducted at the Genotyping and Sequencing facility of Bordeaux (INRA and University of Bordeaux II). The fluorescent signal intensities of the specific alleles were determined by the BeadArray Reader (Illumina) and then quantified and matched using GenomStudio ver. 3.1.14 (Illumina). The genotypic clusters were manually adjusted when necessary. To ensure the good reliability of the genotyping data we used the GenCall50 (GC50) which indicates the reliability of the genotypic clusters of a given SNP and the call rate (CR) which represents the fraction of 96 samples that has been successfully genotyped for a given SNP. We choose to keep SNPs with GC50 higher than 0.35 and CR

higher than 0.85, which are the thresholds commonly used for Illumina genotyping data (Pavy *et al.* 2008, Eckert *et al.* 2009b).

In addition, 14 SNPs from the candidate gene *GALA* were obtained from diploid sequencing on the overall sample (J. Derory, unpublished data). These SNPs were analyzed as genotypic data jointly with the illumina SNP genotyping dataset.

Statistical analyses

The sequences coming either from the 9 fragments of genes were analysed in order to describe the structure and level of diversity within and among the candidate genes, prior the analyses of population genomics.

Nucleotide diversity and differentiation

Two measures of DNA polymorphism were computed for each of the 9 fragments of candidate genes: π, the average number of pairwise nucleotide differences per site in the sample (Nei 1987), and *S* the number of segregating sites. We calculated the number of single nucleotide polymorphisms (SNPs) and insertion-deletions (indels), using the software Arlequin version 3.11 (Excoffier *et al.* 2005). These parameters were also computed with DnaSP software version 5.0 (Rozas *et al.* 2003) for noncoding regions (including introns, 3´ and 5´ untranslated regions), and coding regions subdivided in two components (synonymous and replacement). However DnaSP excludes sites with indels and missing data which potentially contain SNPs, and thus estimations of DNA polymorphism on the whole sequences were slightly lower with this software than when calculated with Arlequin. Consequently we also calculated the number of haplotypes *H* and the haplotypic diversity *Hd* (Nei 1987) using Arlequin.

For the three candidate genes sequenced in full-length we calculated the number of SNPs and indels in the sequences and we assessed the total nucleotide diversity per site π for each full-length gene using Arlequin. We also calculated π separately for the coding and non coding regions because the automatic analysis of synonymous and non synonymous mutations according to the ORF (open reading frame) of the genes was not available for this type of data.

Linkage disequilibrium

We measured the gametic linkage disequilibrium between each pair of polymorphic site within the 9 fragments of genes using Tassel version 2.1 (Bradbury *et al.* 2007). Because the number of sequences per gene was generally lower than one hundred we discarded all the sites with a MAF (minor allele frequency) lower than 0.15. Linkage disequilibrium (LD) was

calculated as the squared correlation of allelic frequencies r² (Hill & Robertson, 1968) which is affected by both recombination and allelic frequencies between sites. The significance of r^2 was tested for each pair of loci using a two-tailed Fisher's exact test with Bonferroni corrections for multiple tests. The decay of LD with the distance between the polymorphic sites was estimated using a non linear regression of r^2 with the function *nls* implemented in the package implemented in the statistical software R (R Development Core Team 2005). We obtained an estimation of the recombination rate C corresponding to $4N_ec$ (with N_e: effective population size and c: recombination fraction between sites) and we calculated the expected value of r^2 for a low level of mutation, adjusted for a sample of size n according to Hill and Weir's (1988) formula:

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

The significance of each r^2 value was tested by a one-tailed Fisher's exact test using the software Genepop with Bonferroni corrections. We grouped together the polymorphic sites of the 9 fragments to obtain an averaged estimation of the decay of r^2.

We used the LDheatmap package of R to plot r² values between the polymorphic sites. Only sites with MAF > 0.05 were considered in the calculations. We calculated the decay of LD with the distance between polymorphic sites using the stat R-package. We also measured the intergenic linkage disequilibrium between all the polymorphic loci coming either from the diploid sequenced genes or from the SNP genotyping.

Neutrality tests

We tested departures of nucleotide diversity patterns from expectations under the standard neutral model (SNM). For each of the nine genes we performed four neutrality tests on the whole sample using DnaSP. First we applied Tajima's *D* statistic (Tajima 1989), which measures a standardized difference between two estimators of the expected diversity under the SNM ($\theta = 4N_e\mu$ with N_e: effective population size and μ: mutation rate) θ_π and θ_S. We also performed the Fu's *Fs* statistic (Fu 1997) which compares the observed and expected numbers of haplotypes given the observed nucleotide diversity. Because the phase between SNPs of the three full-length genes was unknown we only performed Tajima's *D* tests using Arlequin.

Genetic differentiation

Differentiation analyses of the SNPs were conducted separately on the 12 populations from the altitudinal gradient and on the 21 populations from the latitudinal gradient. We calculated the genetic differentiation as Wright's F_{ST} (Weir and Cockerham 1984) for each polymorphic site and between all populations of each environmental gradient using Genepop version 4.3 (Raymond & Rousset 1995). The genetic differentiation was also measured between each pair of populations using the illumina genotyping dataset.

We also examined the distribution of F_{ST} values for each environmental gradient based on the common set of markers comprising the polymorphic sites from the illumina genotyping and from the diploid sequencing of *GALA*. We compared the shape of the distributions using a Wilcoxon test implemented in R commander. The F_{ST} distributions were also calculated for the two valleys of the altitudinal gradient in order to detect loci displaying repeatable patterns of differentiation across the two valleys

Outlier SNPs

The detection of outlier SNPs displaying differentiation deviating from neutral expectations was conducted separately for the populations of each gradient using two different approaches. Prior to the analyses we removed indels from the dataset and used only SNPs in the analyses to compare polymorphic sites with similar mutational processes.

First we conducted analyzes using the software Fdist2 (Beaumont & Nichols 1996) implemented in the package Lositan (Antao *et al.* 2008). By plotting the neutral distribution of F_{ST} against H_e under an island model (Wright 1951), the method allows to detect outlier loci that exhibit higher or lower genetic differentiation than expected under neutrality. Using Lositan we performed a global analysis for each environmental gradient separately. We run simulations with the infinite allele mutation model, and the expected total populations depending on the environmental gradient. The mean F_{ST} value was automatically estimated from the data and we used the neutral F_{ST} option which performs a second run with the mean F_{ST} recalculated by excluding the loci detected as outliers in the first run. We simulated 95,000 loci for each run and calculated the 90 % confidence interval of the neutral envelope. The results are expressed as the probability for each locus that it exhibits a higher F_{ST} than the F_{ST} simulated. Hence p-values lower than 5 % indicate balancing selection whereas p-values higher than 95 % indicate directional selection.

Second we used the software BayeScan version 1.0 (Foll & Gaggiotti 2008) which presents an extension of Beaumont & Balding's (2004) method to detect outlier loci from the

distribution of F_{ST} under a neutral model. In comparison to the frequentist method, BayeScan takes into account heterogeneous migration rates between populations which are more likely in natural populations. The method developed by Foll & Gaggiotti (2008) uses a Bayesian approach to estimate the posterior probabilities of two alternative models, one including the effects of selection and another that excludes it. The two models are tested to estimate the locus-population-specific F_{ST} according to Balding *et al.* (1996): the "neutral model" includes only population effects β_j and thus accounts only for demography and the "selection model" includes both population effects β_j and locus effects α_i which indicate selection: a negative α_i value indicates balancing selection whereas a positive α_i value indicates directional selection. The results are expressed for each locus as a Bayes Factor (BF) which is the ratio of the posterior probabilities of the "selection model" over "the neutral model". According to Jeffreys (1998), the Bayes Factors can be converted in a scale of evidence of selection acting on the locus. We run 10 pilot runs of 5,000 iterations followed by a burn-in of 500,000 and 5,000 iterations sampled every 20. The identification of outlier loci was based on considering loci with BF > 3.

One advantage of the Bayesian method against the frequentist method is that it deals with the problem of multiple testing through the definition of prior distributions. Corrections for multiple testing available for the frequentist method, such as Bonferroni corrections or the false discovery rate (Storey & Tibshirani 2003), represent however too conservative criterions (Namroud *et al.* 2008, Eveno *et al.* 2008, Oetjen *et al.* 2010). We thus preferred to compare the results between both approaches and we retained finally the loci displaying congruent departures from neutral expectations.

Detection of clines

In addition to the detection of outlier SNPs we tested if polymorphic sites presented clinal patterns of allelic variation according to altitude or latitude. In this respect we conducted linear regressions between allelic frequencies at each polymorphic locus and the geographic data of populations. For the altitudinal gradient we also performed the regression analyses separately for each valley to validate clinal patterns by biological repetitions.

Since the frequency of one of the two alleles arbitrary chosen was tested, sign of the regression slopes were not interpreted. We were mainly interested by the existence of a clinal response regardless it was positive or negative. All the linear regressions were performed using the Hmisc package implemented in R and the significance of the correlation coefficient *r* was tested by the Pearson's product-moment method. We performed corrections for

multiple tests using Q-value (Storey & Tibshirani 2003). We also compared the distributions of the coefficient of determination r^2 of the linear regressions between both gradients by performing a Wilcoxon test.

RESULTS

Diversity and differentiation of candidate genes

For the 6 gene fragments sequenced in the populations of the latitudinal and altitudinal gradient, between 80 and 94 haplotypes were obtained (Table 2), while for GA20, GA3 and PM23 we sequenced respectively 21, 46 and 50 haplotypes in the nine populations of the latitudinal gradient. A total of 4.7 kb were sequenced among which 2.8 kb were in coding regions and 1.9 kb were in non coding regions (introns or untranslated regions). The fragment sizes ranged from 315 bp for YSL1 to 1104 bp for PM23. DAG2 was sequenced for a slightly longer fragment for the altitudinal gradient than for the latitudinal gradient (440 bp and 360 bp respectively), but no additional polymorphism was detected. A total of 161 SNPs were found with on average one SNP every 24.8 bp,(GA3; Table 2). We also detected 31 indels distributed among 7 genes which varied in size between 1 and 31 bp (Table 2a) and were mostly located in non coding regions.

Among the nine gene fragments the total nucleotide diversity per site π_{tot} varied between 17.4 x 10^{-3} for GALA and 1.3 x 10^{-3} for DAG2 and GA3 (Table 3). A large variation among genes was also found for the number of polymorphic sites S. Whereas ASI, GALA and PM23 displayed respectively one SNP every 12.6, 14.9 and 18.7 base pairs, DAG2, GA20 and GA3 exhibited only one SNP every 110, 138.5, 217.3 base pairs respectively. The level of diversity in coding sequence was lower than in non coding sequences. In coding sequences synonymous sites were more variable than replacement sites (13.4 x 10^{-3} versus 2.3 x 10^{-3}), except for GA20 for which the only SNP in coding region was non synonymous. Interestingly two genes, *ASI* and *PM23*, presented high levels of replacement sites compared to silent sites (0.38 and 0.50 respectively) which notably differed from the remaining 7 genes. The number of haplotypes per gene varied between 5 and 50 (for *DAG2* and *GALA* respectively). The average haplotype diversity was 0.792, corresponding to more than 23 haplotypes per gene for the sample sizes analyzed.

Table 2: Summary data of the fragments of 9 candidate genes

Gene	Annotation	Length (bp)	*Nb*	*S*	Indels	Indel sizes	SNPs	bp / SNP
ASI	Alpha-amylase/subtilisin inhibitor	327 (327)	*80*	29	3	6pb / 3 bp / 1 bp	26	12.6
AUX-REP	Auxine response factor	446 (258)	*90*	15	2	2 bp / 1 bp	13	34.3
DAG2	DOF zinc finger protein affecting germination	440 (440)	*85*	4	0	-	4	110.0
GALA	Galactinol Synthase	507 (213)	*82*	40	7	31 bp / 30bp / 18 bp / 2 bp (3) / 1 bp	34	14.9
H3	Histone 3	375 (126)	*88*	15	5	2 bp (2) / 1 bp (3)	10	37.5
YSL1	Metal-nicotianamine transporter	315 (315)	*94*	8	0	-	8	39.4
GA3	Gibberellin 3β-hydroxylase	652 (342)	*46*	4	1	6 bp	3	217.3
GA20	Gibberellin 20-oxydase	554 (425)	*21*	5	1	27 bp	4	138.5
PM23	Seed maturation protein	1104 (390)	*50*	70	12	12 bp / 5 bp / 4 bp / 3bp (3) / 2 bp (2) / 1bp (4)	59	18.7
Total		*4720 (2836)*	*660*	*190*	*31*	-	*161*	*24.8*

Nb: number of sequences.

S: number of polymorphic sites (SNPs and indels which can be found at the same position for *GALA* and *PM23*).

For Length values between parentheses correspond to the length of the coding regions. For Indel sizes values between parentheses indicate the number of indels of each size class.

Table 3: Diversity statistics of the 9 fragments of candidate genes.

Gene	π_{tot}	π_{syn}	π_{repl}	π_{sil}	π_{repl} / π_{sil}	*H*	*Hd*	Tajima's *D*	Fu's *Fs*
ASI	12.3	23.1	8.7	23.1	0.38	48	0.945	-0.825 ns	-32.397***
AUX-REP	4.0	0.7	0.0	5.3	0.00	18	0.720	-1.328 ns	-8.396 ***
DAG2	1.3	10.9	0.5	10.9	0.05	5	0.526	-0.572 ns	-1.196 ns
GALA	17.4	40.6	3.5	20.9	0.17	50	0.970	0.141 ns	-20.399 ***
H3	6.4	12.8	0.3	5.3	0.06	29	0.912	-0.673 ns	-7.569 ***
YSL1	6.3	19.8	1.9	19.8	0.10	9	0.812	0.268 ns	-0.404 ns
GA3	1.3	1.7	0.0	1.9	0.00	6	0.663	0.070 ns	-1.134 ns
GA20	2.7	0.0	0.6	5.1	0.11	5	0.600	0.373 ns	-0.466 ns
PM23	10.7	10.9	5.1	10.2	0.50	41	0.984	-1.051 ns	-18.224 ***
Total	*6.9*	*13.4*	*2.3*	*11.4*	*0.15*	*23*	*0.792*		

π_{tot} : Nucleotide diversity calculated on the whole sequences.

π_{syn} : Nucleotide diversity calculated for the synonymous sites.

π_{repl} : Nucleotide diversity calculated for the non synonymous sites.

π_{sil} : Nucleotide diversity calculated for the silent sites (non synonymous and non coding).

H : Number of haplotypes.

Hd : Haplotype diversity.

Linkage disequilibrium

For the 9 gene fragments the analysis of gametic LD resulted in estimates of r^2 for 90 site pairs, of which 79 were significant at the 5 % level, and 64 were still significant after Bonferroni corrections (Figure 1). Overall there was weak linkage disequilibrium within the candidate genes as witnessed by the absence of complete LD in all genes except in *GALA* between two adjacent sites separated by 6 bp. The average squared allelic frequency correlation was 0.37 over the 9 gene fragments and 0.31 over the three full-length genes.

LD decreased rapidly with distance within the candidate genes studied (Figure 1). The estimated r^2 values dropped from 0.46 to 0.20 within 700 bp for the 9 fragments and within 1200 bp for the three full-length genes.

Neutrality tests

The 9 gene fragments exhibited diversity profiles close to neutral expectation as suggested by the values of the Tajima's *D* test (Table 3). The Fu's *Fs* values were negatively significant for five genes ($p < 0.001$ for *ASI*, *AUX-REP*, *GALA*, *H3* and *PM23*) indicating a greater number of haplotypes than expected.

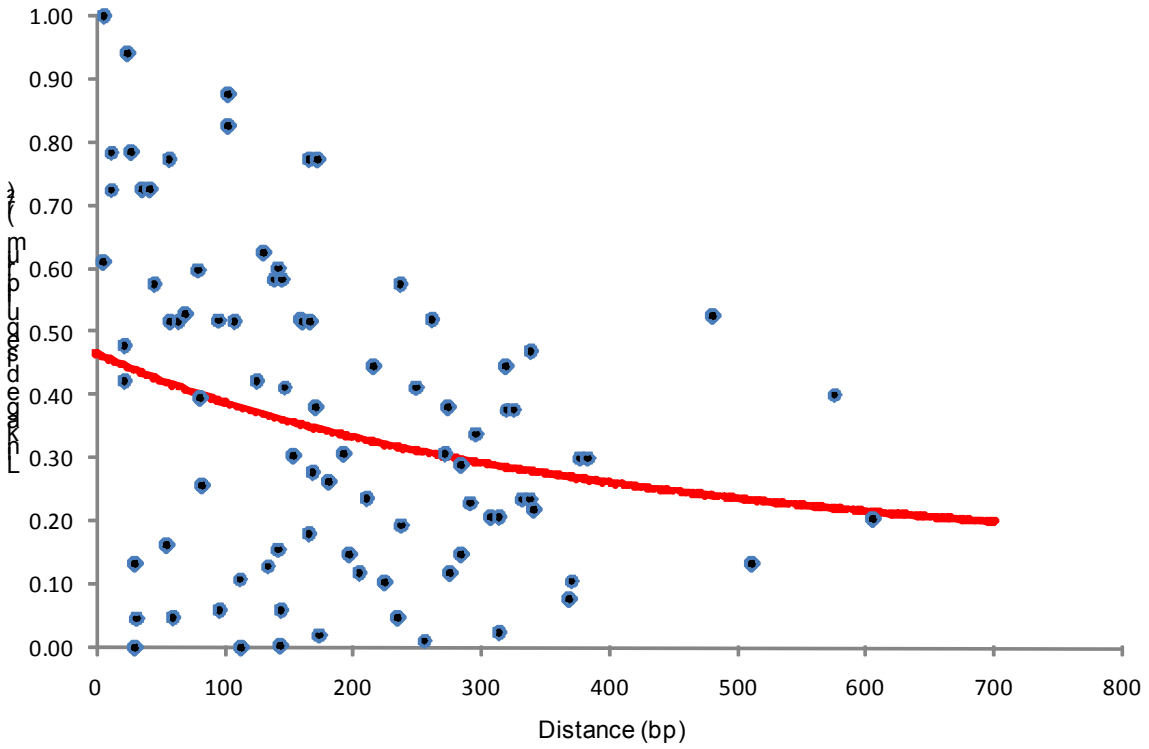


Figure 1: Gametic linkage disequilibrium (r^2) estimated for the 9 fragments of candidate genes plotted against the nucleotide distance between polymorphic sites.

SNP genotyping

SNP scoring was not successful at all SNP sites. The number of genotyping failures reached similar amounts for the divSNPs and the elecSNPs and varied between 25 and 30 % depending on the array. In the overall sample 26.1 % of the elecSNPs were monomorphic and 15.4 % for the divSNPs, which led to a total of 165 polymorphic sites, among which 6 indels. Among the 165 loci, 41 divSNPs and 4 indels were scored in 18 candidate genes and 118 elecSNPs and 2 indels were scored in 55 genes. Hence among the 105 genes, 73 could be assessed and were polymorphic in at least one SNP. The mean number of polymorphisms (SNP or indels) per gene amounted to 2.3. The 165 loci were coded by serial numbers from 1 to 45 for the divSNPs and 46 to 165 for the elecSNPs, preceded either by 'SNP_' or 'IND_' depending on the type of polymorphism (SNP or indel). In the populations from the altitudinal gradient 5 SNPs (2 divSNPs and 3 elecSNPs) could not be scored and 3 other elecSNPs were monomorphic, leading to a dataset of 157 polymorphic sites comprising 6 indels. Similarly 4 elecSNPs could not be scored in the populations of the latitudinal gradient leading to a dataset of 161 polymorphic sites including the 6 indels that were shared with the altitudinal gradient. We also included the 14 SNPs of *GALA* to the genotyping dataset leading to respectively 171 and 175 polymorphic loci for the altitudinal and latitudinal gradients.

Pairwise population differentiation

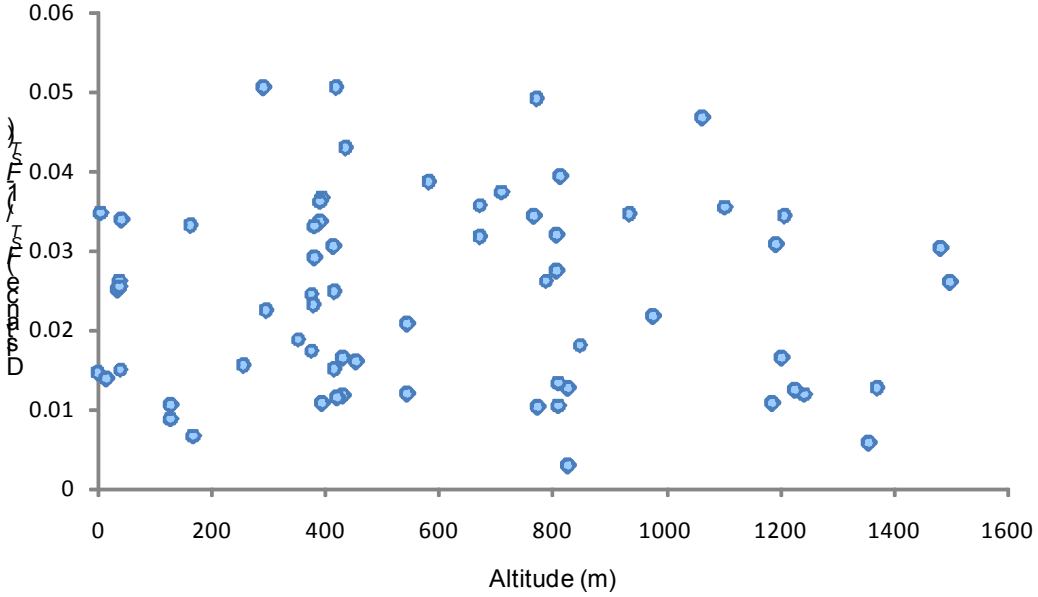
Altitudinal gradient

Pairwise population differentiation was lower in the Luz valley than in the Ossau valley. In the Luz valley differentiation between populations located at consecutive altitudes ranged from 0.003 to 0.017. The highest differentiations were found between populations L1 and L4 and populations L12 and L14 (pairwise F_{ST} between 0.024 and 0.034). In the Ossau valley populations O1 and O4 displayed the highest pairwise differentiations with populations O8 and O12 (between 0.033 and 0.048). Finally, when pairwise differentiation was calculated among populations of the two valleys, largest F_{ST} values were observed between populations O4, O8 and O12 and population L1 (amounting to 0.048).

Latitudinal gradient

Along the latitudinal gradient, the genetic differentiation between pairs of populations based on all 161 loci indicated that population Ob differentiated sharply from the remaining populations. The averaged pairwise F_{ST} value involving this population was 0.050 (ranging from 0.037 with La to 0.070 with Mo).

a)



b)

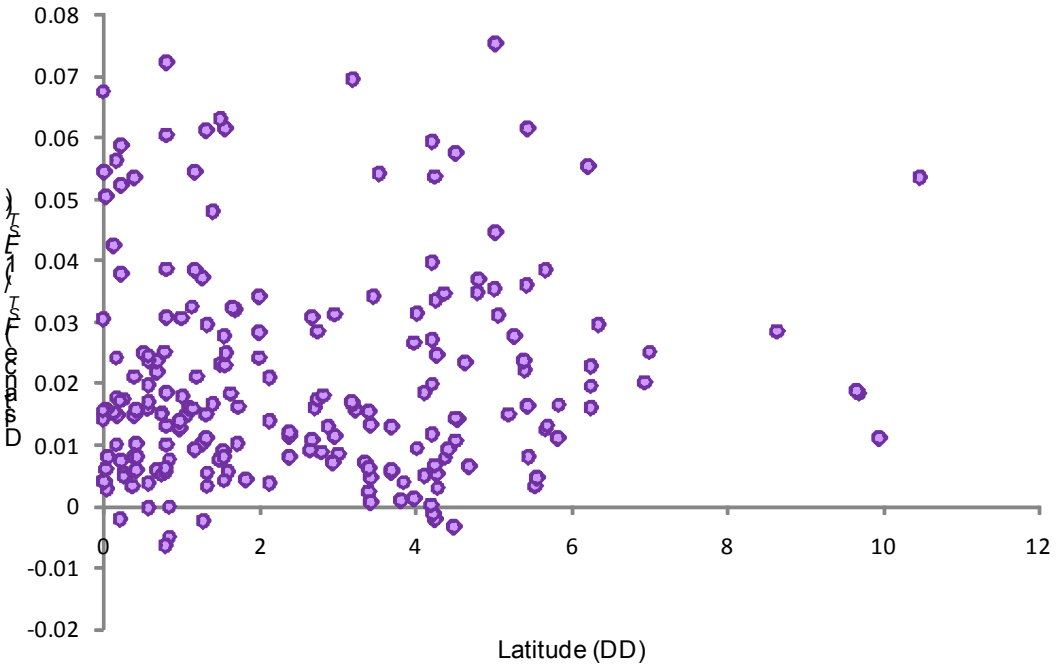


Figure 2: Genetic pairwise differentiation plotted against a) altitudinal or b) latitudinal distances between populations.

In order to evaluate whether differentiation is correlated with environmental factors we plotted pairwise population differentiation (measured as F_{ST} / (1-F_{ST})) against altitudinal or latitudinal distances and performed linear regressions. Figure 2 shows that no correlation based on the differentiation assessed with all SNPs, regardless if they are neutral or selected, is created by altitudinal or latitudinal distance. Differentiation of single SNPs can therefore be compared with the overall picture and inference about selection can be drawn in case of clinal deviating patterns.

Overall differentiation

Altitudinal gradient

Mean F_{ST} of the 157 polymorphic sites amounted to 0.023. The average F_{ST} for *GALA* was 0.011 with single-locus F_{ST} values ranging from -0.006 to 0.050. F_{ST} values were on average larger in the Ossau than in the Luz valley (data not shown). The highest F_{ST} value within Luz valley was 0.075 while 18 loci presented higher differentiation within Ossau valley, and the maximum (SNP_19) amounted to 0.133. Interestingly three loci (SNP_142, SNP_154 and SNP_165) displayed high F_{ST} values in both valleys and two (SNP_154 and SNP_165) exhibited clinal variation of their allelic frequencies along altitude.

Latitudinal gradient

The mean genetic differentiation between the 21 populations computed over the 161 loci was 0. 020. *GALA* was also weakly differentiated between populations of the latitudinal gradient (F_{ST} = 0.016) with single-locus F_{ST} values not exceeding 0.057.

Comparison between the two gradients

The distributions of F_{ST} along the two environmental gradients differed significantly (p < 0.05) because of the longer tail of the latitudinal F_{ST} distribution (Figure 3), although on average differentiation was higher in the altitudinal than latitudinal gradient. Highest F_{ST} values between populations of the altitudinal gradient varied between 0.040 and 0.080 depending on the SNPs. Loci were generally less differentiated between populations of the latitudinal gradient but two loci, IND_70 and SNP_164, displayed the highest values (F_{ST} = 0.108 and 0.098). Finally 24 loci presented F_{ST} > 0.025 in both gradients among which the locus IND_70.

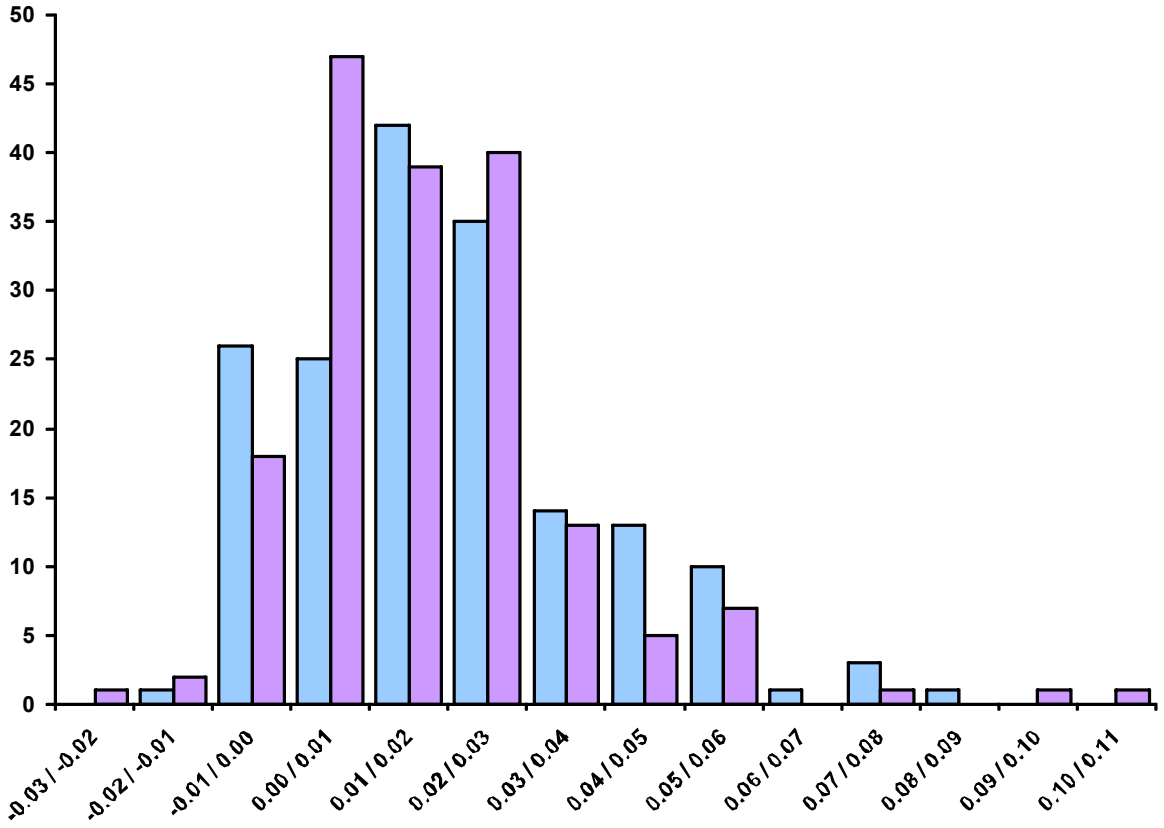


Figure 3: Distributions of single-locus F_{ST} values in the two gradients.

Blue bars correspond to the F_{ST} values calculated between populations of the altitudinal gradient. Purple bars correspond to the F_{ST} values calculated between populations of the latitudinal gradient.

Outlier detection for F_{ST}

Altitudinal gradient

Among the 165 SNPs tested we found a total of 36 (21.8 %) loci displaying non-neutral patterns of differentiation with at least one method. Among these loci respectively 25 and 2 loci were detected by only one method, either Fdist2 or BayeScan method, and 9 SNPs (5.5 %) deviated from neutral expectations in both methods: they fell outside the 95 % confidence interval envelope in Fdist2 and their Bayes Factor was larger than 3 in the BayeScan method (Table 4). Three loci were detected under balancing selection of which two from *GALA* showing a moderate linkage disequilibrium (r^2 = 0.64) and SNP_12 located in Gigantea (*GI*). The 6 outlier loci under directional selection showed higher levels of differentiation (F_{ST} varying between 0.054 and 0.085) and three were found in the cell rescue-defense category: a catalase (*CAT*), an endochitinase (*ECP*) and a seed maturation protein (*PM3*).

Table 4: Loci detected as outliers submitted to selection in the populations of the altitudinal and latitudinal gradients.

Altitudinal gradient							
Locus	**Gene (annotation)**	**Functional category**	**r**	**F_{ST}**	**Fdist2**	**BayeScan**	
GALA_143	Galactinol synthase (*GALA*)	Metabolism / Cell rescue-defense	0.12	-0.004	**0.000**	**-1.158**	(4.0)
GALA_288	Galactinol synthase (*GALA*)	Metabolism / Cell rescue-defense	-0.25	0.000	**0.001**	**-1.106**	(3.2)
SNP_12	Gigantea (*GI*)	Metabolism / Cell rescue-defense	0.47	-0.010	**0.013**	**-1.140**	(3.7)
SNP_152	Ribosomal protein (*L18a*)	Protein synthesis	-0.54	0.078	**0.994**	**0.955**	(3.5)
SNP_77	Endochitinase class I (*ECP*)	Cell rescue-defense	**0.59**	0.057	**0.975**	**1.019**	(5.2)
SNP_142	Seed maturation protein (*PM3*)	Cell rescue-defense	0.47	0.079	**1.000**	**1.013**	(14.7)
SNP_136	Catalase (*CAT*)	Cell rescue-defense	-0.22	0.056	**1.000**	**1.288**	(4.3)
SNP_130	Photosystem II polypeptide (*PSII*)	Energy	-0.29	0.085	**0.999**	**0.946**	(6.1)
SNP_156	Unknown protein 3	Unclassified / Hypothetical	-0.25	0.054	**0.991**	**1.112**	(15.6)
Latitudinal gradient							
Locus	**Gene (annotation)**	**Functional category**	**r**	**F_{ST}**	**Fdist2**	**BayeScan**	
GALA_262	Galactinol synthase (*GALA*)	Metabolism / Cell rescue-defense	0.12	0.007	**0.009**	**-1.122**	(4.3)
SNP_78	Endochitinase class I (*ECP*)	Cell rescue/defense	-0.27	-0.004	**0.018**	**-1.229**	(6.5)
SNP_129	Oxygen-evolving enhancer protein (*OEE*)	Energy	0.43	-0.002	**0.024**	**-1.106**	(4.4)
SNP_37	Gigantea (*GI*)	Metabolism / Cell rescue-defense	-0.38	0.034	**0.959**	**0.764**	(4.4)
SNP_35	Gigantea (*GI*)	Metabolism / Cell rescue-defense	0.32	0.044	**0.986**	**0.840**	(4.1)
SNP_152	Ribosomal protein (*L18a*)	Protein synthesis	**-0.52**	0.019	**0.967**	**0.836**	(6.0)
SNP_123	Early light-induced protein 1 (*ELIP1*)	Cell rescue-defense	0.14	0.058	**0.999**	**1.011**	(73.6)
SNP_66	Pseudo-response regulator (*PRR*)	Cell rescue-defense	-0.17	0.053	**1.000**	**0.955**	(6.1)
SNP_57	Transcription factor (*TF*)	Transciption	0.20	0.052	**0.987**	**0.863**	(4.7)
SNP_34	Histone 3 (*H3*)	Transciption	0.17	0.071	**0.996**	**1.207**	(21.2)
SNP_120	Plastocyanin A (*PC*)	Energy	**-0.55**	0.043	**0.988**	**0.948**	(11)
SNP_164	Unknown protein 2	Unclassified / Hypothetical	-0.29	0.111	**1.000**	**1.267**	(332.3)

r: Correlation coefficient between frequencies at the SNP and the environmental factor (altitude or latitude). Values in bold characters are significant at p=0.05 (see Table 5).

Fdist2: p-value of the Fdist2 analysis.

BayeScan: α value of the BayeScan analysis and Bayes Factor between parentheses.

F_{ST}: Differentiation between the populations of the gradient.

We also found in a photosystem II polypeptide (*PSII*), a ribosomal protein (*L18a*), and one gene not yet annotated. Interestingly one SNP of *ASI* (SNP_16) was found as outlier under directional selection in BayeScan and fall just under the detection threshold in Fdist2 (p = 0.93).

Latitudinal gradient

A total of 66 outlier loci out of 169 (39.0 %) were detected by either one method. 50 loci were detected by Fdist2 only and 5 by BayeScan only. Three SNPs of GALA were detected as negative outliers by Fdist2 of which one was also detected by BayeScan. Among the 12 outliers (7.1 %) detected by both methods, four loci displayed low levels of differentiation indicating balancing selection and 7 loci highly differentiated were found under directional selection (Table 4). The three negative outliers were located in *GALA* and *ECP* belonging to the cell rescue-defense category and in a gene coding for an Oxygen-evolving enhancer protein (*OEE*) belonging to the energy category. Two SNPs located in *GI* were detected as submitted to directional selection although displaying no LD (r^2 = 0.01). We also found one positive outlier in genes coding for an early light-induced protein (*ELIP1*), a plastocyanin (*PC*), a putative pseudo-response regulator (*PRR*), a ribosomal protein (*L18a*), a transcription factor (*TF*), Histone *H3* and an unknown protein.

Clines of allelic frequencies

We calculated the Pearson correlation (r) between allelic frequencies at each polymorphic locus and altitude or latitude, and the corresponding determination coefficient (r^2). The distribution of r^2 followed an L-shaped curve, with a maximum number of values near 0. However there was a long tail in the distribution suggesting that some SNPs exhibited clinal variations (Figure 4). The distributions of r^2 values were slightly different between the two gradients although not significantly (p = 0.52). This difference may be generated by the different numbers of populations sampled in each gradient (Figure 4).

We found respectively 13 and 12 loci within the altitudinal and latitudinal gradient (Table 5) displaying significant clinal variation of allelic frequencies when tested by the Pearson's product-moment method ($p < 0.05$). However none of these correlations remain significant at a q-value threshold of 5 %. However straightforward corrections for multiple testing leads most likely to very conservative thresholds, as some of the SNPs are not independent due to LD. We therefore used p-values for detecting SNPs showing clinal variations. The loci showing significant clinal variations belonged to 10 genes for the altitudinal gradient and to

10 genes for the latitudinal gradients, two being common to the two clines belong to a Chloroplast chaperonin 21 (*CPN21*) and a Dihydroflavonol-4-reductase (*DFR*).

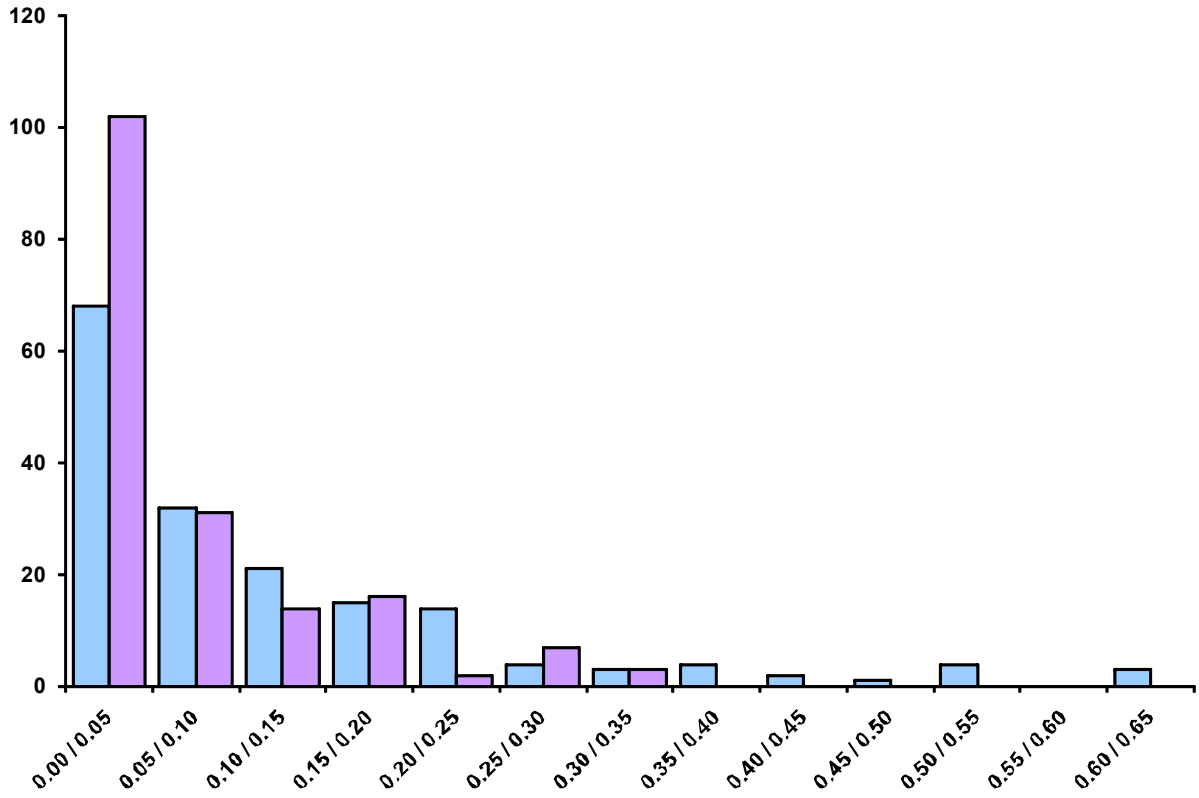


Figure 4: Distribution of the coefficient of determination values (r^2) between the allelic frequencies and environmental factors.

Blue bars correspond to the r^2 values calculated for the altitudinal gradient. Purple bars correspond to the r^2 values calculated for the latitudinal gradient.

Table 5: Loci exhibiting significant clinal variations in the two environmental gradients.

Altitudinal gradient					
Locus	**Gene (annotation)**	**Functional category**	**r**	*p*	F_{ST}
SNP_57	Cys-3-HIS zinc finger protein (CHZFP)	Transcription	**0.80**	0.017	0.021
SNP_59	Squamosa-promoter binding protein-like (*SQP*)	Transcription	**0.73**	0.007	0.017
SNP_74	Chloroplast chaperonin 21 (*CPN21*)	Cell rescue-defense	**-0.63**	0.028	0.012
SNP_77	Endochitinase class I (*ECP*)	Cell rescue-defense	**0.59**	0.043	0.057
SNP_90	Dihydroflavonol-4-reductase (DFR)	Metabolism / Cell rescue-defense	**0.62**	0.030	0.004
SNP_106	Chlorophyll A-B binding protein (*ChlBP*)	Energy / Protein with binding function	**0.78**	0.021	-0.010
SNP_107	Chlorophyll A-B binding protein (*ChlBP*)	Energy / Protein with binding function	**0.63**	0.029	0.019
SNP_131	Photosystem II polypeptide (*PSII*)	Energy	**0.72**	0.044	0.003
SNP_154	5'-adenylylsulfate reductase (*APS*)	Metabolism / Cell rescue-defense	**0.73**	0.007	0.046
SNP_161	Unknown protein 1	Unclassified / Hypothetical	**0.79**	0.019	0.019
SNP_163	Unknown protein 1	Unclassified / Hypothetical	**-0.65**	0.022	0.010
SNP_164	Unknown protein 2	Unclassified / Hypothetical	**-0.71**	0.009	0.022
SNP_165	Unknown protein 2	Unclassified / Hypothetical	**0.68**	0.015	0.038
Latitudinal gradient					
Locus	**Gene (annotation)**	**Functional category**	**r**	*p*	F_{ST}
GALA_310	Galactinol synthase (*GALA*)	Metabolism / Cell rescue-defense	**-0.58**	0.006	0.020
GALA_391	Galactinol synthase (*GALA*)	Metabolism / Cell rescue-defense	**0.52**	0.016	0.030
GALA_413	Galactinol synthase (*GALA*)	Metabolism / Cell rescue-defense	**-0.53**	0.014	0.008
SNP_15	Alpha-amylase/subtilisin inhibitor (*ASI*)	Metabolism / Cell rescue-defense	**0.56**	0.036	0.021
SNP_47	NADP-dependent malic enzyme (*NADP-ME*)	Metabolism / Protein with binding function	**-0.43**	0.050	0.030
SNP_73	Chloroplast chaperonin 21 (CPN21)	Cell rescue-defense	**0.44**	0.044	0.001
SNP_89	Dihydroflavonol-4-reductase (*DFR*)	Metabolism	**-0.53**	0.013	0.021
SNP_111	Early light-induced protein 2 (*ELIP2*)	Cell rescue-defense	**0.55**	0.010	0.003
SNP_120	Plastocyanin A (*PC*)	Energy	**-0.55**	0.010	0.052
SNP_134	Ribosomal protein (*S11*)	Protein synthesis	**-0.51**	0.017	0.047
SNP_143	Seed maturation protein (*PM3*)	Cell rescue/defense	**0.48**	0.028	0.050
SNP_152	Ribosomal protein (*L18a*)	Protein synthesis	**-0.52**	0.016	0.046

r: Correlation coefficient between allelic frequencies at the SNP and the environmental factor (altitude or latitude). *p*: p-value of the Pearson's test.

F_{ST}: Differentiation between the populations of the gradient

DISCUSSION

Nucleotide diversity pattern in oaks

The overall level of nucleotide diversity in the candidate genes for bud burst in sessile oak (6.9×10^{-3}) is in the range of earlier assessments in other tree species. In conifer species reported values range between, 1.9×10^{-3} in *Pinus radiata* (Pot *et al.* 2005) and 8.6×10^{-3} in *Pinus densata* (Ma *et al.* 2006). Recently, Quang *et al.* (2008) studied 11 genes originated from the same EST library than ours in an Asian oak species *Q. crispula* and found $6.0 – 6.9 \times 10^{-3}$. Larger diversity in broadleaves than in conifers has been suggested in regards to the first results obtained (Derory *et al.* 2010). However large differences exist among genes within species (Brown *et al.* 2004). As an example, we found a 15-fold variation of nucleotide diversity between *GA3* and *GALA*. In *Populus tremula* the high diversity earlier reported (11.1×10^{-3}, Ingvarsson 2005) was based on 5 genes and was not confirmed in a more recent contribution based on 77 fragments of genes (4.2×10^{-3}, Ingvarsson 2008). As a result, future comparisons among species should consider a much larger sample of genes.

Our genetic diversity survey also showed low linkage disequilibrium in oaks, which is in line with previous reports on forest trees (Savolainen *&* Pyhäjärvi 2007, Ingvarsson *et al.* 2008, Eckert *et al.* 2009a). We found a rapid decay of LD (r^2 dropping under 0.20 within at most 600bp), which is expected for large outcrossing populations as in oaks. Our results contrast with the large haploblocks (> 50 kb) found in humans or self-crossing species such as *Arabidopsis* (Rafalski *&* Morgante 2004, Hinds *et al.* 2005).

As earlier gene diversity studies conducted with neutral markers indicated low population differentiation along the two clines (Zanetto *&* Kremer 1995, Alberto *et al.* 2010a), we performed neutrality test based on allelic frequency profiles over the whole sample of trees. We found on average negative values for Tajima's *D* and Fu's *Fs* tests, indicating a greater number of haplotypes and an excess of low frequency polymorphisms compared to neutral expectations. Such results could be generated by either population expansions with maintenance of recent mutations or directional selection or a combination of both processes (Tajima 1989, Fu 1997). Negative Tajima's *D* values can also result from pooling differentiated populations but population differentiation was globally low in our sample. We have no means to estimate the contribution of demographic and selection effects to the diversity patterns and we suspect that both shaped the genomic signature observed today. Simulations of population expansion of forest trees after the last glaciations have indeed shown how nucleotide diversity profiles can be imprinted by demographic events (Heuertz *et*

al. 2006, Pyhäjärvi *et al.* 2007, Ingvarsson *et al.* 2008, Eckert *et al.* 2009a). Population expansion has also paved the postglacial recolonisation of European oaks (Brewer *et al.* 2008, Petit *et al.* 2002), and we would expect similar consequences in these species. However it was also shown that directional selection has strongly differentiated extant populations subsequently to their establishment (Kremer *et al.* 2009). Hence confounding results are expected from directional selection and population expansion at the genomic level, which cannot be disentangled by Tajima's or Fu's neutrality tests. However tests based on population differentiation may be more appropriate to identify selective factors, particularly in the case of species known to be poorly differentiated for neutral markers.

Detection of outlier loci along gradients

The proportion of outlier loci detected jointly by the two methods (on average 6.3 %) is higher than previously reported for candidate genes in other tree species. In *Picea glauca*, Namroud *et al.* (2008) found 20 positive outliers out of 534 polymorphic SNPs (3.7 %) using Fdist2 at the 5 % significance level but only two of them were also detected using the Beaumont *&* Balding's (2004) method. However the authors also considered 49 loci (5.5 %) displaying high γ_{ij} values, which indicate locus-population interaction, as candidate for local adaptation. Eveno *et al.* (2008) analysed patterns of genetic differentiation for 11 candidate genes for drought stress response in 10 populations of *Pinus pinaster* using 3 methods of multi-locus scan. When analyzed at the SNP level 15 loci (16 %) were detected by at least one method at the 10 % significance level but only two were detected by two methods. In our case most of the outlier loci (76 %) detected using BayeScan were confirmed by the Fdist2 analysis applying the same 10 % significance level. On the opposite a large number of outliers (72 %) detected by Fdist2 did not deviate from neutral expectations in BayeScan. Discrepancies between both methods can be attributed to different causes. First we may advocate that the significance level is not comparable between the two methods. Although, no theoretical background is available to compare the significance levels between both approaches, empirical comparisons based on simulations suggest that a cut off value of 0.7 of the posterior probability (corresponding to a Bayes Factor of 3) in BayeScan corresponds to a 1 % p-value of Fdist2 (Foll *&* Gaggiotti, 2008). In contrast to previous studies (Eveno *et al.* 2008, Namroud *et al.* 2008) the significance level of 10 % for the Fdist2 method appears too liberal in our case and may result in a higher rate of false positives. Second, outlier detection is Fdist2 is dependent on the presence of selected SNP in the original set of SNPs. Indeed, the method proceeds iteratively starting from the full set of SNPs by drawing the "neutral

envelope", then removing outliers based on the results obtained. Hence the "neutral envelope" is dependent on the presence or not of selected loci. As shown by Foll & Gaggiotti (2008) by simulations, the rate of false negative or false positive SNP detected by BayeScan is not varying according to the number of a selected SNP in the data set. Third, the basic assumptions differ between both models. The demographic models implemented in Fdist2 lie on a symmetric island model where population sizes and migration rates between populations are considered equal (Beaumont & Nichols 1996). In contrast to Fdist2, BayeScan considers implicitly the heterogeneity of migration rates and population sizes by estimating locus and population-specific effects on differentiation. Hence BayesScan accounts for population specific effects that are likely to occur in our case where unidirectional gene flow and unequal population sizes were found between altitudes and valleys (Alberto *et al.* 2010a) and are also likely occur along the latitudinal gradient. Finally Eveno *et al.* (2008) showed that the infinite mutation model used for the simulations in Fdist2 is not well suited for biallelic markers such as SNPs. The higher variance of F_{ST} values of SNPs compared to multi-allelic markers leads to an underestimation of the upper bound of the confidence interval and the detection of false-positive outliers. These differences may explain the contrasting number of positive outlier loci found between the two methods. Finally, concerning balancing selection, we found 10 outlier loci with BayeScan among which 6 were also detected by Fdist2. These results are surprising as Foll & Gaggiotti (2008) highlighted the difficulty in detecting negative outlier loci with BayeScan when populations are poorly differentiated. We suspect that the detection has improved in our case due to the larger sample sizes. As indicated by Foll & Gaggiotti (2008) in their simulations, detection of balancing selection is more dependent on the sample size than directional selection.

Selective response along the environmental gradients

This study is the first to report SNP differentiation for a large number of genes in a large number of populations in oaks, and in a comparative way along two geographic gradients. The overall genomic population differentiation was low within each gradient, and was of the same magnitude than F_{ST} values assessed with microsatellites or isozymes along the same gradients (Zanetto & Kremer 1995, Mariette *et al.* 2002, Alberto *et al.* 2010a). These results contrast with the high differentiation of phenological traits observed in common garden experiments (Ducousso *et al.* 1996, Alberto *et al.* 2010b) along the same gradients. The 73 genes polymorphic in the SNP scan were part of 190 genes that were down or up regulated during bud burst of sessile oaks (Derory *et al.* 2006). Despite the overall low differentiation of

the 73 candidate genes, 15 exhibited outlier genomic signatures across methods in at least one cline: 4 deviated significantly from neutral expectations (*GALA*, *GI*, *ECP* and *L18a*) in both clines, and 11 in one cline only. According to their annotation in data bases and literature (see below), they belong to 5 functional categories (cell rescue-defense, general metabolism, protein synthesis, cell cycle, and transcription factors). Seven out of the 15 genes detected as outliers by Fdist2 and BayeScan belong to the cell rescue-defense category which comprises proteins involved in responses to biotic and abiotic stresses. Other major functional gene categories represented were energy (3 genes), and transcription (2 genes). Directional selection was the most frequent form of selection detected (10 genes out of 15). The overall response across clines can be pictured in three cases:

Same outlier genes, same form of selection but different SNPs in different clines

For *GALA* and *L18a*, which showed the same deviating profiles in each cline, the signal was however not visible at the same SNP in the two gradients. Given the very low linkage disequilibrium that was depicted in the sequenced genes (Figure 1), this is not a surprise. Loose linkage within genes may generate slightly different signatures depending on populations. Our results therefore suggest that nucleotide diversity should now be exhaustively explored in these genes to locate the causal polymorphism.

Same outlier genes but different form of selection in different clines

GI and *ECP*, while exhibiting outlier profiles in both clines, showed however contrasting forms of selection in the two clines; the deviating profiles were also observed at different SNPs in the two clines. We may therefore suspect that different alleles at the same genes may induce different forms of selection. This may occur if an epistatic effect is associated to one allele changing its selective value, while other alleles may be purely additive and maintain their selective value. We cannot exclude that the same gene may respond differently in different genetic backgrounds. Finally in a companion paper investigating genetic differentiation of phenological traits, we suggested that the variation of the strength of selection within population can generate different diversity profiles of the trait, which may in turn modify selection responses at the gene level (Alberto *et al.* 2010b).

Different outlier genes in different clines

Eleven genes among the 15 exhibiting outlier SNPs, were detected in one or the another cline. Such response is not unexpected as the different environmental gradients may induce adaptive responses for different traits. While temperature is following the same clinal gradient along

altitude and latitude, other environmental changes (rainfall, light, photoperiod) are differently associated to these gradients, which will in turn stimulate different responses. Therefore the overall set of genes detected as outliers may comprise genes involved in different traits in the two gradients.

Clinal response along the environmental gradients

Allelic frequencies of 18 candidate genes exhibited clinal variations, *e.g.* continuous linear variations as a function of latitude or altitude, among the 73 candidate genes for which nucleotide diversity was explored. Clinal variations can be generated by various causes ranging from isolation by distance (Barbujani 1988), selection along continuous ecological gradients (Endler 1977) to admixture (Griebeler *et al.* 2006). Le Corre *et al.* (1998) have dealt in detail with the spatial variation expected across the range distribution of *Q. petreae* and showed that the preferential direction of frequency gradients is East-West in Europe, and was mostly generated by the longitudinal distribution of glacial refugial populations (form Spain to the Balkans). Using geostatistical analysis they have also shown that geographical trends remain low and cannot be observed beyond 500 to 1000 km. We therefore exclude that "isolation by distance" would have created a significant latitudinal genomic imprint. Thereafter, we showed in this contribution that pairwise differentiation calculated on all loci did not increase with latitudinal or altitudinal distance (Figure 2). Admixture as a result of hybridization between *Q. petraea* and *Q. robur* may also have created clinal variation especially if the rate of hybridization follows a latitudinal or altitudinal trend (Alberto *et al.* 2010a). However genomic signatures should then be located in hotspots of genomic divergence between the two species. These hotspots were located in previous studies (Scotti-Saintagne *et al.* 2004), and mapping positions of outlier loci in our study can be compared to these hotspots in the future. We therefore suspect that the 18 loci showing significant clinal variation shifted their allelic frequencies in response to divergent selection along the altitudinal or latitudinal gradients. Two of them showed clinal variations along the two gradients (*DFR* and *CPN21*) but not at the same SNPs (SNP_73 and SNP_74 for *CPN21*; SNP_89 and SNP_90 for *DFR*). That different genes show clinal variations along the two gradients can be attributed to the different adaptive responses of oak populations stimulated in the two clines.

Functional role of genes displaying responses to selection

Among the 18 genes showing clinal variations and the 15 exhibiting outlier F_{ST} values, five were common (*GALA*, *ECP*, *L18a*, *PSII* and *PC*). In addition *ELIP1* (F_{ST} outlier) and *ELIP2* (clinal variation) belong to the same gene family. The rather low number of genes detected by both methods (F_{ST} and clinal variation) can be explained in two ways. Genes showing high differentiation but not exhibiting clinal variation may be involved in local adaptation not related to the temperature gradient generated by altitude or latitude. On the other hand, clinal variation of allelic frequencies can be generated without significant differentiation, in the presence of divergent selection among populations, but weak selection intensity within population. Such patterns were observed in simulation studies comparing differentiation of traits and their underlying genes (Kremer & Le Corre, pers com). Weak selection intensity within populations will maintain high level of diversity and therefore constraint F_{ST} values, even when allelic frequencies differ among populations.

Finally the overall screening of genes responding to selection is summarized in Table 6 where the methods of detection are crossed with the two gradients. We conclude that genes responding either in the two clines with a given method, or within a given cline with the two methods, could deserve further attention for the search of functional genes related to phenological traits. This list comprises 9 genes of which six belong to the cell rescue-defense class (*GALA*, *GI*, *ECP*, *CPN21*, *DFR* and *ELIP*), two to the energy class (*PSII* and *PC*) and one to the protein synthesis class (*L18a*). *GALA* catalyses the first step in the biosynthesis of raffinose family oligosaccharides (RFOs) which are thought to protect cellular integrity during desiccation and imbibition, to extend longevity in the dehydrated state, and to provide substrates for energy generation during germination (Downie *et al.* 2003). It was shown to be under balanced selection (Table 6) in the altitudinal and latitudinal cline.

Table 6: Number of genes responding to selection detected in the two gradients with the 2 methods.

Detection method	Altitudinal gradient	Latitudinal gradient	Genes common to both gradients
F_{ST} outliers	8	11	*GALA*, *ECP* *GI*, *L18a*
Clinal patterns	10	10	*DFR*, *CPN21*
Genes detected by both methods	*ECP, PSII*	*GALA, ELIP2, PC*	

F_{ST} outliers: Genes detected as positive or negative outliers by both Fdist2 and BayeScan

Clinal patterns: Genes presenting a significant correlation ($p < 0.05$) with either altitude or latitude.

The multifunctional role of this gene therefore seems to have contributed to the maintenance of a larger diversity owing to balancing selection in natural populations of both gradients. *GI* is a circadian clock-controlled gene involved in photoperiodic flowering and seed germination in *Arabidopsis* by regulating the response to abcissic acid and gibberellins (Fowler *et al.* 1999; Penfield *&* Hall 2009) and it has also been shown to control the plant oxidative and cold stress responses (Cao *et al.* 2005; Cao *et al.* 2006). By degrading the chitin contained in the cellular wall of fungi, *ECP* plays a role in the defense against pathogens (Zamani *et al.* 2003). Recently three chitinase genes were studied in *Arabidopsis* plants exposed to several environmental stresses (cold, freezing, heat and strong light) and their expressions were correlated with stress tolerance (Takenaka *et al.* 2009). Consequently the contrasting results obtained for *GI* and *ECP* (balancing selection in one cline and directional selection in the other) may result from the multifunctional role of these genes. The *ELIP*s are involved in the photoprotection of plants exposed to photo-inhibitory conditions (Hutin *et al.* 2003), which can occur particularly in southern populations of the latitudinal cline. *CPN21* is a chaperonin protein which is involved in protein assembly in the chloroplast and may play also an important role in seed development (Forkmann *&* Martens 2001). *DFR* which is the first enzyme in the flavonoid pathway committed to biosynthesis of anthocyanins, which are mainly involved in flower coloration, defense against pathogens and in photoprotection against UV light (Harborne *&* Williams 2000). *PSII* and *PC* are involved in the photosystem apparatus which can be stimulated differently along wide altitudinal or latitudinal ranges (Katoh 2003). Finally *L18a* is a component of the large ribosomal subunit and is thus involved in the synthesis of proteins. Despite that this function seems not specifically related to bud burst, recent studies have shown that ribosomal proteins may play a role in the activation of leaf growth in *Arabidopsis* and *Populus* (Matsubara *et al.* 2006, Tatematsu *et al.* 2008).

In conclusion we have shown that candidate genes for bud burst display high nucleotide diversity and low linkage disequilibrium within genes. The detection along the two environmental gradients of either outlier patterns regarding differentiation, or clinal variations of allelic frequencies revealed few genes responding to natural selection. These genes deserve further investigations to assess their contribution to adaptive traits in natural populations.

REFERENCES

Alberto F, Niort J, Derory J, Lepais O, *et al.* (2010a) Population differentiation of sessile oak at the altitudinal front of migration in the French Pyrenees. *Molecular Ecology* (accepted)

Alberto F, Bouffier L, Louvet JM, Delzon S, Kremer A (2010b) Adaptive responses for seed and leaf phenology in natural populations of sessile oak along altitudinal gradient. *Journal of Evolutionary Biology* (submitted)

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a F-st-outlier method. *Bmc Bioinformatics* **9**.

Arbez M, Lacaze JF, (1998). Les ressources génétiques forestières en France. *Tome 2. Les feuillus*, Editions de l'INRA,

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London Series B-Biological Sciences* **263**, 1619-1626.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**, 969-980.

Brabujani G, (1988). Detecting and comparing the direction of gene-frequency gradients. *Journal of Genetics* **67**: 129-140

Bradbury PJ, Zhang Z, Kroon DE*, et al.* (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633-2635.

Brewer S, Cheddadi R, de Beaulieu JL, Reille M (2002) The spread of deciduous Quercus throughout Europe since the last glacial period. *Forest Ecology and Management* **156**, 27-48.

Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15255-15260.

Calvo AP, Nicolas C, Nicolas G, Rodriguez D (2004) Evidence of a cross-talk regulation of a GA 20-oxidase (FsGA20ox1) by gibberellins and ethylene during the breaking of dormancy in Fagus sylvatica seeds. *Physiologia Plantarum* **120**, 623-630.

Cao SQ, Jiang ST, Zhang RX (2006) The role of GIGANTEA gene in mediating the oxidative stress response and in Arabidopsis. *Plant Growth Regulation* **48**, 261-270.

Cao SQ, Ye M, Jiang ST (2005) Involvement of GIGANTEA gene in the regulation of the cold stress response in Arabidopsis. *Plant Cell Reports* **24**, 683-690.

Casasoli M, Derory J, Morera-Dutrey C*, et al.* (2006) Comparison of quantitative trait loci for adaptive traits between oak and chestnut based on an expressed sequence tag consensus map. *Genetics* **172**, 533-546.

Deans, J.D, Harvey, FJ, (1995). Phenologies of sixteen European provenances of sessile oak growing in Scotland. *Forestry* **68**, 265-273.

Derory J, Leger P, Garcia V*, et al.* (2006) Transcriptome analysis of bud burst in sessile oak (*Quercus petraea*). *New Phytologist* **170**, 723-738.

Derory J, Scotti-Saintagne C, Bertocchi E *et al.* (2010) Contrasting relationships between the diversity of candidate genes and variation of bud burst in natural and segregating populations of European oaks. *Heredity* (in press).

Downie B, Gurusinghe S, Dahal P*, et al.* (2003) Expression of a GALACTINOL SYNTHASE gene in tomato seeds is up-regulated before maturation desiccation and again after imbibition whenever radicle protrusion is prevented. *Plant Physiology* **131**, 1347-1359.

Ducousso A, Guyon, JP, Kremer A (1996). Latitudinal and altitudinal variation of bud burst in western populations of sessile oak (*Quercus petraea* (Matt.) Liebl.). *Annales des Sciences Forestières*, **53**, 775-782.

Ducousso A., Louvet JM, Jarret P, Kremer A (2005). Geographic variations of sessile oaks in French provenance tests. In *Proceedings of the Joint Meeting of IUFRO working groups Genetic of oaks and Improvement and Silviculture of oaks*, ed. R. Rogers, A. Ducousso and A. Kanazashi, FFPRI (Forestry and Forest Products Research Institute) Scientific Meeting Report 3, Tsukuba, Japan, pp 128-138

Eckert AJ, Wegrzyn JL, Pande B*, et al.* (2009a) Multilocus Patterns of Nucleotide Diversity and Divergence Reveal Positive Selection at Candidate Genes Related to Cold Hardiness in Coastal Douglas Fir (Pseudotsuga menziesii var. menziesii). *Genetics* **183**, 289-298.

Eckert AJ, Pande B, Ersoz ES*, et al.* (2009b) High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (*Pinus taeda* L.). *Tree Genetics & Genomes* **5**, 225-234.

Endler JA (1977) *Geographic variation, speciation and clines*. Princeton University Press.

Eveno E, Collada C, Guevara MA*, et al.* (2008) Contrasting patterns of selection at Pinus pinaster Ait. drought stress candidate genes as revealed by genetic differentiation analyses. *Molecular Biology and Evolution* **25**, 417-437.

Excoffier L, Laval G, and Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47-50.

Falusi R, Calamassi C (2003). Dormancy of *Fagus sylvatica* L. buds. III. Temperature and hormones in the evolution of dormancy in one-node cuttings. *Plant Biosystems* **137**:185

Foll M, Gaggiotti O (2008) A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* **180**, 977-993.

Forkmann G, Martens S (2001) Metabolic engineering and applications of flavonoids. *Current Opinion in Biotechnology* **12**, 155-160.

Fowler S, Lee K, Onouchi H*, et al.* (1999) GIGANTEA: a circadian clock-controlled gene that regulates photoperiodic flowering in Arabidopsis and encodes a protein with several possible membrane-spanning domains. *Embo Journal* **18**, 4679-4688.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915-925.

Giertrych M, (1991) Provenance variation in growth and phenology. *In "genetics of Scots Pine", Giertych M., Matyas Cs eds, Elsevier*, pp 87-98

Gonzalez-Martinez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB (2006) DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics* **172**, 1915-1926.

Griebeler EM, Muller JC, Seitz A (2006) Spatial genetic patterns generated by two admixing genetic lineages: a simulation study. *Conservation Genetics* **7**:753-766

Hall D, Luquez V, Garcia VM*, et al.* (2007) Adaptive population differentiation in phenology across a latitudinal gradient in European Aspen (Populus tremula, L.): A comparison of neutral markers, candidate genes and phenotypic traits. *Evolution* **61**, 2849-2860.

Harborne and Williams (2000) Advances in flavonoid research since 1992. *Phytochemistry*. **55** (6), 481-504.

Heuertz M, De Paoli E, Kallman T*, et al.* (2006) Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [Picea abies (L.) Karst]. *Genetics* **174**, 2095-2105.

Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical Population Biology* **33**, 54-78.

Hill WG, Robertson A. (1968) Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* **38**, 226-231.

Hinds DA, Stuve LL, Nilsen GB*, et al.* (2005) Whole-genome patterns of common DNA variation in three human populations. *Science* **307**, 1072-1079.

Howe GT, Aitken SN, Neale NB, Jermstad KD, Wheeler NC, Chen THH, (2003) From genotype to phenotype: unraveling the complexities offcold adaptation in forest trees. *Canadian Journal of Botany*. **81**, 1247-1266

Hutin C, Nussaume L, Moise N*, et al.* (2003) Early light-induced proteins protect arabidopsis from photooxidative stress. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 4921-4926.

Ingvarsson PK, (2005) Nucleotide polymorphism and linkage disequilbrium within and among natural populations of European Aspen (Populus tremula L., Salicaceae). *Genetics* **169**, 945-953.

Ingvarsson PK, (2008) Multilocus patterns of nucleotide polymorphism and the demographic history of Populus tremula. *Genetics* **180**, 329-340.

Israelsson M, Mellerowicz E, Chono M, Gullberg J, Moritz T (2004) Cloning and overproduction of gibberellin 3-oxidase in hybrid aspen trees. Effects of gibberellin homeostasis and development. *Plant Physiology* **135**, 221-230.

Katoh S (2003) Early research on the role of plastocyanin in photosynthesis. *Photosynthesis Research* **76**, 255-261.

König AO, (2005). Provenance research: evaluating the spatial pattern of genetic variation. In: Geburek, T., Turok, J. (Eds.), *Conservation and Management of Forest genetic Resources in Europe. Arbora Publishers, Zvolen, Slovakia*, pp. 275-238.

Kremer A, Le Corre V, Petit R J, Ducousso A (2010) Historical and contemporary dynamics of adaptive differentiation in European oaks. In : DeWoody A, Bickham J, Michler C, Nichols K, Rhodes G, Woeste K (eds) *Molecular approaches in natural resource Conservation* . Cambridge University Press (in press), pp

Langlet, O, (1971). Two hundred years of genecology. *Taxon* **20**, 653–722.

Le Corre, V., Roussel G, Zanetto A, and. Kremer A, (1998). Geographical structure of gene diversity in *Quercus petraea* (Matt.) Liebl. III. Patterns of variation identified by geostatistical analyses. *Heredity* **80**:464-473.

Liepe K, (1993) Growth chamber trial on frost hardiness and field trial on bud burst of sessile oak (*Quercus petraea* Liebl and *Q. robur* L.). *Annals of Forest Sciences* **50,** 208-214.

Ma XF, Szmidt AE, Wang XR (2006) Genetic structure and evolutionary history of a diploid hybrid pine Pinus densata inferred from the nucleotide variation at seven gene loci. *Molecular Biology and Evolution* **23**, 807-816.

Mariette S, Cottrell J, Csaikl UM*, et al.* (2002) Comparison of levels of genetic diversity detected with AFLP and microsatellite markers within and among mixed *Q. petraea* (Matt.) Liebl. and *Q. robur* L. stands. *Silvae Genetica* **51**, 72-79.

Matsubara S, Hurry V, Druart N*, et al.* (2006) Nocturnal changes in leaf growth of Populus deltoides are controlled by cytoplasmic growth. *Planta* **223**, 1315-1328.

Matyas C (1996) Climatic adaptation of trees: Rediscovering provenance tests. *Euphytica* **92,** 45-54.

Mimura M, Aiken SN (2007) Adaptive gradients and isolation-by-distance with postglacial migration in Picea sitchensis. *Heredity* **99**: 224–232

Morgenstern EK (1996) *Geographic variation in forest trees*. University of British Columbia Press, Vancouver, Canada.

Namroud MC, Beaulieu J, Juge N, Laroche J, Bousquet J (2008) Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology* **17**, 3599-3613

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Or E, Belausov E, Popilevsky I, Ben Tal Y (2000). Changes in endogenous ABA level in relation to the dormancy cycle in grapevines grown in a hot climate. *Journal of Horticultural Science and Biotechnology* **75**: 190-194

Pavy N, Pelgas B, Beauseigle S*, et al.* (2008) Enhancing genetic mapping of complex genomes through the design of highly-multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce. *Bmc Genomics* **9**.

Penfield S, Hall A (2009) A Role for Multiple Circadian Clock Genes in the Response to Signals That Break Seed Dormancy in Arabidopsis. *Plant Cell* **21**, 1722-1732.

Perez-Flores L, Carrari F, Osuna-Fernandez R*, et al.* (2003) Expression analysis of a GA 20-oxidase in embryos from two sorghum lines with contrasting dormancy: possible participation of this gene in the hormonal control of germination. *Journal of Experimental Botany* **54**, 2071-2079.

Petit RJ, Brewer S, Bordacs S*, et al.* (2002) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management* **156**, 49-74.

Pot D, McMillan L, Echt C*, et al.* (2005) Nucleotide variation in genes involved in wood formation in two pine species. *New Phytologist* **167**, 101-112.

Pyhajarvi T, Garcia-Gil MR, Knurr T*, et al.* (2007) Demographic history has influenced nucleotide diversity in European Pinus sylvestris populations. *Genetics* **177**, 1713-1724.

Quang ND, Ikeda S, Harada K (2008) Nucleotide variation in Quercus crispula Blume. *Heredity* **101**, 166-174.

Rafalski A, Morgante M (2004) Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends in Genetics* **20**, 103-111.

Raymond M, Rousset F (1995) Genepop (Version 1.2) - Population-genetics software for exact tests and Ecumenicism. *Journal of Heredity* **86**, 248-249.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496-2497.

Savolainen O, Pyhajarvi T (2007) Genomic diversity in forest trees. *Current Opinion in Plant Biology* **10**, 162-167.

Savolainen O, Pyhajarvi T, Knurr T (2007) Gene flow and local adaptation in trees. *Annual Review of Ecology Evolution and Systematics* **38**, 595-619.

Scotti-Saintagne C, Mariette S, Porth I, *et al.* (2004) Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. *Genetics* **168**, 1615-1626.

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 9440-9445.

Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.

Takenaka Y, Nakano S, Tamoi M, Sakuda S, Fukamizo T (2009) Chitinase Gene Expression in Response to Environmental Stresses in Arabidopsis thaliana: Chitinase Inhibitor Allosamidin Enhances Stress Tolerance. *Bioscience Biotechnology and Biochemistry* **73**, 1066-1071.

Tatematsu K, Kamiya Y, and Nambara E (2008) Co-regulation of ribosomal protein genes as an indicator of growth status: comparative transcriptome analysis on axillary shoots and seeds in Arabidopsis *Plant Signaling and Behavior* **3**, 450-452;

Thornton KR, Jensen JD, Becquet C, Andolfatto P (2007) Progress and prospects in mapping recent selection in the genome. *Heredity* **98**, 340-348.

Vasemagi A (2006) The adaptive hypothesis of clinal variation revisited: Single-locus clines as a result of spatially restricted gene flow. *Genetics* **173**, 2411-2414.

Vitasse Y, Delzon S, Bresson CC, Michalet R, Kremer A (2009a) Altitudinal differentiation in growth and phenology among populations of temperate-zone tree species growing in a common garden. *Canadian Journal of Forest Research* **39**, 1259-1269.

Vitasse Y, Delzon S, Dufrene E*, et al.* (2009b) Leaf phenology sensitivity to temperature in European trees: Do within-species populations exhibit similar responses? *Agricultural and Forest Meteorology* **149**, 735-744.

von Wuehlisch G, Krusche D, Muhs HJ (1995) Variation in temperature sum requirement for flushing of beech provenances. *Silvae Genetica* **44**, 343-346

Wachowiak W, Balk PA, Savolainen O (2009) Search for nucleotide diversity patterns of local adaptation in dehydrins and other cold-related candidate genes in Scots pine (Pinus sylvestris L.). *Tree Genetics & Genomes* **5**, 117-132.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.

Worrell R, Cundall, E.P., Malcolm, D.C., Ennos, R.A., (2000) Variation among seed sources of silver birch in Scotland. *Forestry* **73,** 419–435.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics* **15**, 323-354.

Wright JW (1976) *Introduction to forest genetics*. Academic Press, New York, NY, USA

Zamani A, Sturrock R, Ekramoddoullah AKM, Wiseman SB, Griffith M (2003) Endochitinase activity in the apoplastic fluid of Phellinus weirii-infected Douglas-fir and its association with over wintering and antifreeze activity. *Forest Pathology* **33**, 299-316.

Zanetto A, Kremer A (1995) Geographical structure of gene diversity in *Quercus petraea* (Matt.) Liebl. I. Monolocus patterns of variation. *Heredity* **75**, 506-517.

Association study for bud burst in sessile oak populations along altitudinal and latitudinal gradients

Florian Alberto [1,2], Antoine Kremer [1,2]

(1) INRA, UMR1202 BIOGECO, F-33610 Cestas, France

(2) Université de Bordeaux, UMR1202 BIOGECO, F-33610 Cestas, France

To be submitted to Heredity

INTRODUCTION

Deciphering the molecular basis of genetic variation in adaptive traits is a major objective of ecological and evolutionary genetics. During the last decades numerous approaches have attempted to link the variability of fitness-related traits to the underlying genes. Historically investigations in this field were essentially conducted using QTL mapping approaches (Lander & Botstein, 1989). QTL mapping provided detailed insights of the genetic architecture of a trait, *e.g.* the number of loci controlling a trait, their approximate position in the genome, the fraction of variation explained. However QTL mapping requires the genotyping of a large number of segregating markers and the phenotyping of the trait of interest on a large number of progenies. Furthermore QTL mapping are usually deployed on a very few number of mapping pedigrees, thus underestimating the number of QTLs existing in natural populations. In tree species, long generation time is a biological constraint for implementing projects aiming at detecting QTLs, and such experiments have been limited to economically important species for which breeding programmes have been developed (White *et al.*, 2007). Finally identification of causal genes through QTLs mapping remains limited, as confidence intervals of QTL positions extend usually over several centimorgans, thus comprising several hundred of genes (Sewell & Neale 2000).

In the last decade, association mapping has emerged has a promising approach to unravel the molecular basis of adaptation, taking advantage of the high recombination rates and large genetic variation usually occurring in natural populations. In this respect, tree species offer many characteristics making them appropriate for association genetic studies: they are predominantly outcrossing, compose large natural populations, resulting in substantial levels of nucleotide diversity (Neale & Savolainen 2004). However to date the few studies reporting associations in tree species have only found a low number of markers associated to traits and explaining a rather low proportion of their variation (Thumma *et al.* 2005, Gonzalez-Martinez *et al.* 2007, Gonzalez-Martinez *et al.* 2008, Ingvarsson *et al.* 2008, Eckert *et al.* 2009). The efficiency of association mapping depends on the extent of linkage disequilibrium but also on the ability to distinguish physical linkage disequilibrium from linkage caused by other evolutionary forces (Flint-Garcia *et al.* 2003). In this respect population structure is one of the major sources of linkage disequilibrium producing false-positive associations (Marchini *et al.* 2004). Assessing population structure is therefore a necessary preliminary step to association studies (Pritchard *et al.* 2000a, Price *et al.* 2006, Zhao *et al.* 2007). The aim of this step is to

"remove" genetic correlations caused by population structure by incorporating structure as covariate into the model of association.

Here we report the first association mapping study conducted on sessile oak (*Quercus petraea* Matt. Liebl.) using genotypic data from 73 candidate genes for bud burst recorded in populations sampled along an altitudinal and a latitudinal gradient. In two companion papers we reported on the distribution of the phenotypic variation of phenological traits (Alberto *et al.* 2010 b) and on the diversity of SNPs at the candidate genes (Alberto *et al.* 2010c). In this contribution, we compare the two data sets in order to detect associations between SNP diversity and variation of the timing of bud burst and acorn germination. Using populations sampled across environmental gradients offers an appealing approach to detect causal genomic mutations of adaptive variation, as reproducible clinal variation of the traits and genes has been reported in the previous papers. However they offer also drawbacks to the approach, as they increase the risk that population structure generates non causal associations. Indeed populations sampled across large gradients may originate from different genetic backgrounds or histories leading which can induce spurious associations. We therefore proceeded in two steps. Despite the low population differentiation reported by traditional F_{ST} estimations in oaks (Zanetto & Kremer 1995, Mariette *et al.* 2002) we first assessed population structure among populations within each gradient using either microsatellite or SNP markers. In a second step we conducted association tests between markers and phenological traits taking into account population structure. Finally we compared associations detected between the two clines and provide hints for future research on genetic variation of adaptive traits in natural populations of trees.

MATERIAL AND METHODS

Association populations

Thirty two natural populations of *Q. petraea* were sampled to investigate statistical associations between the phenology of the apical bud and nucleotide diversity. These populations were sampled along an altitudinal and a latitudinal gradient.

The altitudinal gradient comprises 10 populations sampled in two Pyrenean valleys between 131 m and 1630 m and the latitudinal gradient 21 populations located from 43'N and 51'N in western Europe. The sampled populations were raised in provenance tests and assessed for phenological traits (Ducousso *et al.* 1996, Alberto *et al.* 2010b) and nucleotide diversity was further monitored on the same set of populations (Alberto *et al.* 2010c). Details about the

geographic distribution of the populations in both gradients are available in a preceding companion paper (Alberto *et al.* 2010c).The overall sample corresponds to 758 individuals distributed in 32 populations (see for details Alberto *et al.* 2010c).

Table 1: Phenotypic data used for the association studies.

Trait	Meaning	Sample
BV_GERM	Family breeding values of	135
BV_LU	germination and leaf unfolding	135
GV_GERM	Genotypic values of germination	135
GV_LU	and leaf unfolding	135
SV_LU06	Standardized value of leaf unfolding	260
SV_LU07	*in situ*	258
BBS	Developmental score of bud burst	453
SV_BBS	Standardized value of BBS	450

Phenotypic data

Altitudinal gradient

8833 acorns were harvested in September 2006 on 152 adult trees within 10 natural populations distributed in two valleys on the Northern side of the Pyrenees mountains. The seed were raised in a greenhouse in spring 2007 and time of seed germination was recorded. In winter 2007 the containers with germinated acorns were transferred in a common garden experiment located in the nursery of the INRA research station at Pierroton. Finally the two year seedlings were transplanted during the fall 2008 in a field provenance test located at Toulenne in the Southwest of France. Time of germination (GERM) was recorded on each acorn in spring 2007 and time of leaf unfolding (LU) on first year and third year seedlings in the nursery and in the provenance tests. Details of the experimental design, the assessments of the traits, and data analysis are described in detail in a companion paper (Alberto *et al.* 2010b). In addition to the assessments on offspring, bud burst was also recorded *in situ* during the spring of 2006 and 2007 on mature trees in the natural stands, on a total number of 260 trees, among which 152 produced the seed collected in September 2006.

The phenotypic data available for the association studies were the phenotypic values of the mature trees recorded *in situ* during 2006 and 2007 and expressed as standardized values of leaf unfolding (SV_06 and SV_07) and the breeding values of the mother trees as estimated from the offspring data for timing of germination (BV_GERM) and for timing of leaf unfolding (BV_LU) recorded in the nursery and field test (Table 1).

Latitudinal gradient

A range wide provenance test of *Q. petraea* comprising 112 populations was installed in four different plantations located in the central and eastern part of France between 1990 and 1995 and used to investigate the latitudinal clinal variation of apical bud burst (Ducousso *et al.* 1996, Ducousso *et al.* 2005, Kremer *et al.* 2010). Each population is represented in each plantation by 10 to 15 replications (24 trees per replication), *e.g.* by 240 to 360 trees. Twenty one populations were selected among the total collection of 112 for their contrasting time of bud burst. Thirty trees of each sampled population, located in two replications of one plantation (Petite Charnie, located in the northwestern part of France; Ducousso *et al.* 2005) were sampled for this association study.

The phenotypic data used consisted of the bud burst score (BBS) recorded at 3 years of age in the Petite Charnie plantation. A grading system ranging from 0 (fully dormant) to 5 (fully elongated) was used to assess the stage of development of the apical bud at one observation (Ducousso *et al.* 2005) and trees were assessed once during the flushing period

Genotypic data

High-throughput genotyping was conducted with the Illumina GoldenGate technology (Illumina Inc., San Diego, CA, USA) using a multiplex assay including 384 SNPs from 105 candidate genes for bud burst phenology as described in Alberto *et al.* (2010c). A total of 165 polymorphic loci (159 SNPs and 6 indels) representing 73 genes were successfully genotyped in the populations of both gradients (Alberto *et al.* 2010c). We also included 14 SNPs from the gene *GALA* coming from direct diploid sequencing on the populations of both gradients. The total genotypic dataset corresponding to 179 polymorphic loci was used to test for associations with phenology traits. We also assessed the linkage disequilibrium between loci, measured as the squared allelic correlation coefficient (r^2) using the Genetics package implemented in R (R Development Core Team 2005).

Population structure

Genetic structure was assessed for the populations of both gradients separately using the software Structure version 2.3 (Pritchard *et al.* 2000b; Falush *et al.* 2003). Because cluster definition in Structure supposes independence between markers we discarded loci displaying high linkage disequilibrium ($r^2 > 0.80$). We used a set of 90 SNPs previously tested for neutral patterns regarding to selection (Alberto *et al.* 2010c) and presenting less than 10 % missing data. For the populations of the altitudinal gradient we also re-analyzed a dataset comprising 14 nuclear microsatellites (referred as SSRs) previously used to assess gene flow and

demography in the same populations of the Pyrenean valleys (Alberto *et al.* 2010a). This dataset was used in order to compare the structure obtained with SSRs and SNPs that are partly also used for the association tests. SSRs are known to be more informative owing to their high variability compared to biallelic markers such as SNPs. The number of alleles among the 14 SSRs varied between 10 and 29 (Alberto *et al.* 2010a). We intended to compare whether the low information of SNPs could be compensated by a higher number of loci by comparing the structure detected with both type of markers. For each analyse, we used the admixture model with allele frequencies correlated and performed 5 runs for each K value (from 1 to the number of populations sampled) with a burn-in period of 50,000 followed by 100,000 iterations. The more likely number of cluster (*K*) was defined using both the mean likelihood of the data LnP(D) given *K* over the 5 runs and the Δ*K* criterion defined by Evanno *et al.* (2005).

Statistical analysis for association tests

Association tests were performed separately for each environmental gradient since phenotype measurements were assessed in two different common garden experiments and recorded either as the date of leaf unfolding or the score of bud burst.

We used a general linear model (GLM) implemented in the software Tassel version 2.1 (Bradbury *et al.* 2007). This function fits a least squares fixed effects linear model to test for association between genotypes at a segregating site and phenotypes, while also taking in account population structure. A global analysis was performed considering each trait independently and testing the fixed effect of each SNP. Population structure was incorporated into the model as the percentage of assignment *Q* of each individual to the number of cluster K defined by Structure. This Q-matrix was identified as a covariate in the model and we removed the percentage of assignment to one cluster in order to obtain valid F-tests of the population structure (sum of *Q* not equal to 100 %). We used the permutation test (Churchill & Doerge 1994) incorporated in the GLM analysis to calculate the significance level corresponding to the experiment-wise error accounting for multiple testing. The great advantage of the permutation test comparing to other multiple-test correction procedures is that it considers dependence between hypotheses *i.e.* linkage disequilibrium between loci. We thus applied 10,000 permutations for each trait analysis and calculated the p-value adjusted for each marker (p-adj).

For the altitudinal gradient association tests were performed between genotypes and the BLUP values of the phenological traits from the phenotypic analyses described in Alberto *et*

al. (2010b). BLUP values were calculated for germination timing (GERM) and leaf unfolding (LU), for which we analyzed measurements from two consecutive years in different environments using restricted maximum-likelihood (REML) using the ASReml software (Gilmour *et al.* 2002).

Given the hierarchical subdivision of the data (provenance, mother trees, offspring) BLUP estimates were calculated for provenances (PV: provenance value) and mother trees (BV: breeding value).

Hence two different values were computed for each mother tree for the association analysis:

- Mother tree values (excluding provenance values) that actually amount to ½ of the breeding value of the mother tree, but are referred as BV (Table 1).
- Mother tree values including provenance effects: PV+BV, which are referred as genotypic value of the mother tree, GV (Table 1).

As leaf unfolding phenology was also followed on 260 adult trees in the natural populations during springs 2006 and 2007 we used the standardized values (SV_LU) to test for associations on a larger sample. For the latitudinal gradient we tested associations directly between SNP genotypes and the individual phenotypic values corresponding to the score of bud burst (BBS) or its standardized value (SV_BBS). The sample sizes of each phenotype used for the different analyses are given in Table 1.

For each significant trait-marker association detected we calculated the ratio of dominance (d) to additive effects (a). Additivity was defined as $|d/a| \leq 0.50$, dominance as $0.50 < |d/a| < 1.25$ and under and over dominance corresponded to $|d/a| > 1.25$.

In order to test whether the associations detected were mainly caused by clines of allelic frequencies following the pattern of phenotypic differentiation we performed linear regressions with either the PV of GERM and LU or the mean BBS using the Hmisc package implemented in the statistical software R.

RESULTS

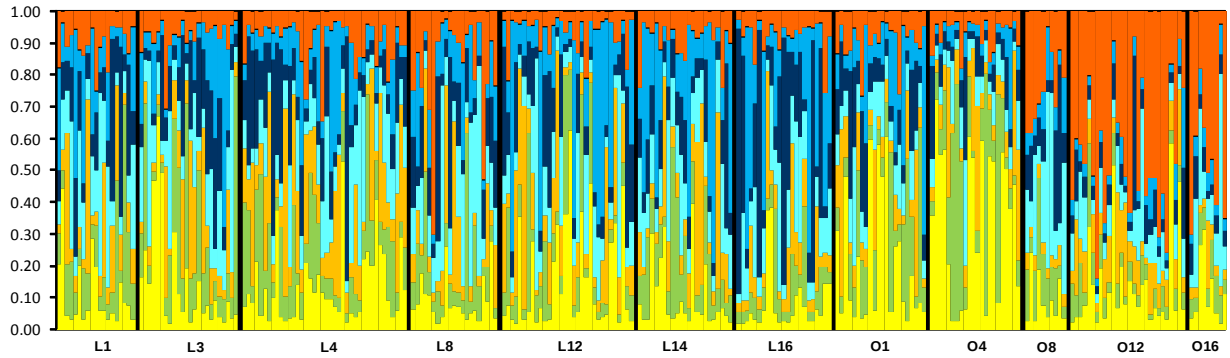
Population structure

Altitudinal gradient

Genetic structure between the populations of the altitudinal gradient was slightly more pronounced using SSRs than SNPs. For SSRs the probability of the data LnP(D) value according to *K* increased until *K* = 7 which also showed the higher Δ*K* value (46.8). Overall we observed a high admixture of individuals with only 2 individuals assigned with a probability *Q* > 0.80 to one cluster. The average percent of assignments of populations to the 7 clusters revealed that some populations were preferentially represented by few clusters (Figure 1a). In the Luz valley one cluster increased with altitude from 11.1 % in population L4 to 24.3 % in population L14 and the average percent of assignment of the red cluster in population L16 was 28.1 %. In Ossau valley the green cluster was specific of the high altitude populations (43.4 and 41.1 % for O12 and O16 populations respectively) and the yellow cluster represented essentially the O4 population (35.5 %).

Using the SNP dataset most of the population structure along the altitudinal gradient was captured for *K* = 2. The probability of the data LnP(D) given *K* was higher than for *K* = 3 (Figure 1a) and the Evanno's criterion was largely higher (Δ*K* = 73.2) than for other *K* values (Δ*K* = 14.9 and 14.1 for *K* = 3 and *K* = 4 respectively). The individuals were mainly highly assigned to one cluster with 164 *Q* values higher than 0.80, but the structure between populations was not highly marked (Figure 1b). The average percents of assignment to one cluster were lower than 55.6 % for the populations of Ossau valley (*i.e.* almost equally assigned to one of the two clusters). In Luz valley one cluster differentiated the populations of low altitudes until 800 m (mean *Q* = 37.7 %) from the high altitude populations (mean *Q* = 61.6 %). The clustering for *K* = 4 was also interesting since populations O4 emerged with preferential assignment to one cluster (mean *Q* = 47.1 %). However the overall admixture level was largely higher than for *K* = 2 with only 10 individuals displaying Q > 0.80.

a)



b)

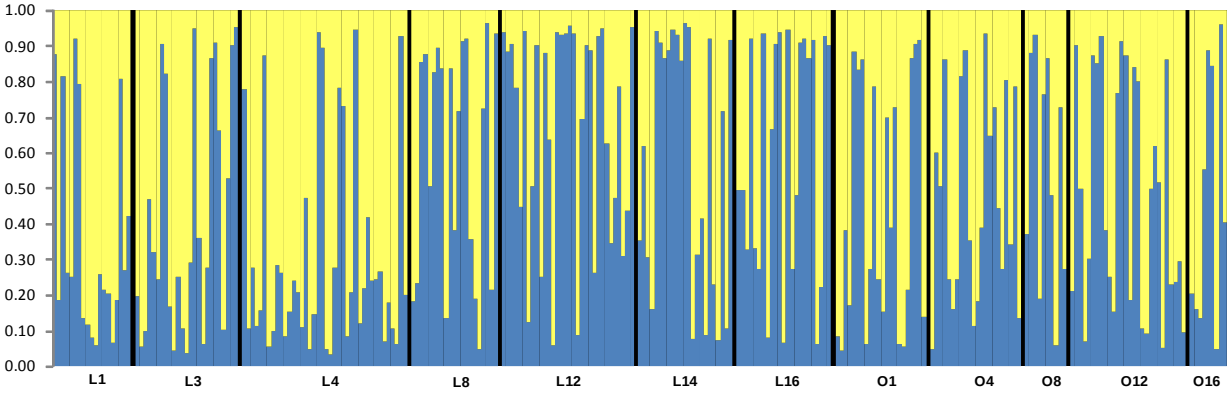


Figure 1: Structure clustering results obtained for the populations of the altitudinal gradient using a) 14 SSRs or b) 90 SNPs

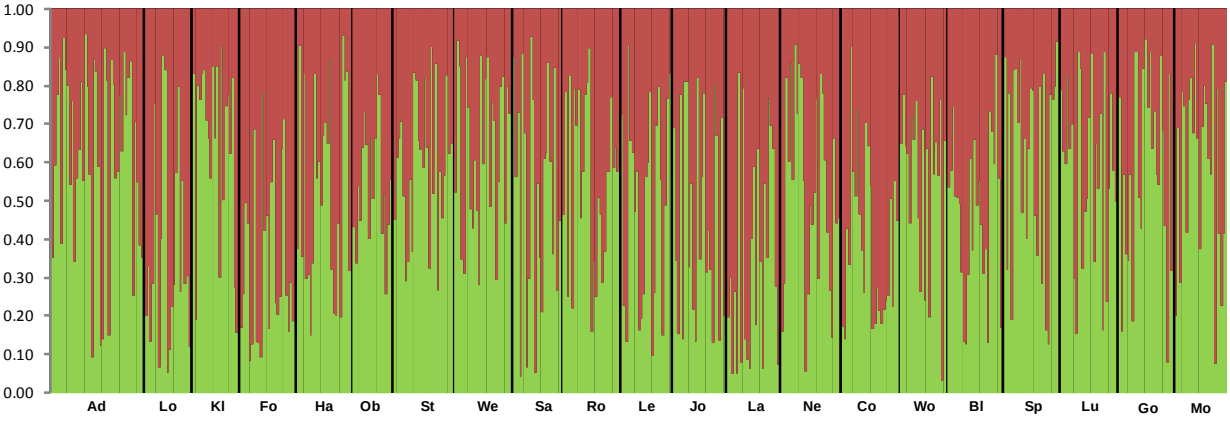


Figure 2: Structure clustering results obtained for the populations of the latitudinal gradient using 90 SNPs

Latitudinal gradient

For the populations of the latitudinal gradient the mean LnP(D) increased from $K = 1$ to $K = 6$ where it reached a peak at -29,574.4 (Figure 2). However because the standard error of LnP(D) between runs was high for $K = 6$, the more likely number of clusters according to Evanno's criterion was $K = 2$ ($\Delta K = 39.1$; Figure 1b). The mean LnP(D) for $K = 2$ was -29,996.5 and -29,706.2 for $K = 4$ which presented also a high ΔK value (22.7). When comparing the results of the assignment probabilities Q of each individual the structure was slightly more informative for $K = 4$ than for K = 2 with populations Ad and Ob represented by different clusters with on average more than 48 % of assignation. Population Mo was also assigned on average to a third cluster with 42.2 %. However the general structure was weak and characterized by a high admixture of individuals to the *K* clusters as witnessed by the 27 *Q* values higher than 0.80.

For the association tests on the populations of the altitudinal gradient we utilized the Q-matrix from the SSR analysis of Structure for $K = 7$ as the genetic structure was less correctly assessed using the SNP dataset. However we also conducted the association test with the Q-matrix from the SNP analysis of Structure for $K = 2$ in order to observe the effect of the population structure on the accuracy of the analyses. In the same way we used the SNP Q-matrix for $K = 2$ as covariate in the association tests on the populations of the latitudinal gradient but we also performed the analyses without accounting for population structure. Moreover we also conducted analyses after removing populations Ob and Ad which showed the higher pairwise differentiations (Alberto *et al.* 2010c) to test if these populations drive the detection of associations.

Association tests

Altitudinal gradient

Analyses conducted with the SSR Q-matrix revealed six trait-marker associations still significant after corrections with the permutation procedure (p-adj < 0.05; Table 2). Two loci, SNP_112 and SNP_103 were associated with GV_LU and explained respectively 8.9 and 9.5 % of the phenotypic variance. SNP_112 is in a gene coding for a Glutathione S-transferase (*GST*) while the gene of SNP_103 codes for an unknown protein. Variation at SNP_134 explained 6.7 % of the variance of BV_LU and belongs to a ribosomal protein (*S11*). We also found three loci (SNP_27, SNP_92 and SNP_102) displaying moderate associations with SV_LU and representing between 4.5 and 5.1 % of the phenotypic variance. These loci are located in genes coding respectively for a *Constans*-like zinc finger protein (*COL*), a

Dihydroflavonol-4-reductase (*DFR*) and an unknown protein. However, we found no concordance between the analyses conducted with BV_LU and GV_LU. The SNPs associated with the breeding value of the trees did not show any significant association with the genotypic value of the tree (p-adj > 0.99). However we found two SNPs from an unknown protein, SNP_102 and 103, associated with SV_LU07 and BV_LU respectively but displaying low linkage disequilibrium ($r^2 = 0.33$).

Table 2: List of SNPs showing significant associations with phenotypic traits in the altitudinal and latitudinal gradient.

Trait	Marker	Gene (Annotation)	PEV	*p-adj*
Altitudinal gradient				
GV_LU	**SNP_112**	Glutathione S-transferase (*GST*)	9.5%	0.0004
GV_LU	**SNP_103**	Unknown protein	8.9%	0.0049
BV_LU	**SNP_134**	Ribosomal protein (*S11*)	6.7%	0.0488
SV_LU07	**SNP_102**	Unknown protein	5.1%	0.0119
SV_LU07	**SNP_92**	Dihydroflavonol-4-reductase (*DFR*)	4.7%	0.0230
SV_LU06	**SNP_27**	*Constans*--like zinc finger protein (*COL*)	4.5%	0.0414
Latitudinal gradient				
BBS	**IND_70**	No annotation	7.5%	0.0001
BBS	**SNP_135**	Ribosomal protein (*S11*)	3.5%	0.0009
BBS	**SNP_12**	Gigantea (*GI*)	3.0%	0.0096
BBS	**SNP_146**	Cysteine protease (*CysP*)	2.7%	0.0205
BBS	**SNP_153**	Ribosomal protein (*L18a*)	3.0%	0.0060
BBS	**SNP_151**	Ribosomal protein (*L18a*)	2.6%	0.0355
SV_BBS	**SNP_13**	Auxin induced protein (*AIP*)	3.4%	0.0197
SV_BBS	**SNP_153**	Ribosomal protein (*L18a*)	3.3%	0.0019
SV_BBS	**SNP_151**	Ribosomal protein (*L18a*)	3.1%	0.0041
SV_BBS	**SNP_146**	Cysteine protease (*CysP*)	3.1%	0.0077

PEV: percentage of phenotypic variance explained by the marker.

p- adj: p-value of the association calculated after 10,000 permutations.

The comparison with the analyses conducted with the SNP Q-matrix showed that only three loci associated using the SSR Q-matrix were still significant (SNP_112, SNP_27 and SNP_102) whereas two additional loci, SNP_72 and SNP_138, were significantly associated with respectively BV_GERM and BV_LU. The population structure had a strong effect on the detection of associations for these two loci since the p-adj values varied from 0.07 to 0.69 in the analysis with the SSR Q-matrix. On the other hand the associations involving SNP_92, SNP_103 and SNP_134 were largely non-significant in the analysis with the SNP Q-matrix

(p-adj > 0.76). The effects of markers changed also with on average higher percentages of variance explained, amounting to 13.0 % of BV_LU for SNP_138.

Latitudinal gradient

We found a total of 10 significant associations (p-adj < 0.05) for the latitudinal gradient involving 7 different markers (Table 2). The percentage of phenotypic variance explained by each marker varied between 2.6 % for SNP_151 and 7.5 % for IND_70 both associated with BBS and respectively located in a gene coding for a ribosomal protein (*L18a*) and a gene with no known function. The other loci associated with BBS explained at most 3.5 % of the phenotypic variance and were located in genes coding for a ribosomal protein (*S11*), a homologous of Gigantea (*GI*) and a Cysteine protease (*CysP*). We also found a second locus, SNP_153, located in the *L18a* gene and presenting an almost complete linkage disequilibrium with SNP_151 ($r^2 = 0.97$). These two markers also presented significant associations with SV_BBS and explained 3 % of the phenotypic variance. SNP_146 associated with BBS (gene *CysP*), and SNP_13 in a gene coding for an Auxin induced protein (*AIP*) contributed to 3.1 and 3.4 % of the variance of SV_BBS.

When comparing the results with the association tests conducted without taking in account the population structure we did not observed any difference in term of significant association. The percentages of phenotypic variance explained were almost equal, varying from 2.5 % for SNP_153 to 7.0 % to IND_70. Results were almost identical when removing populations Ad and Ob, with five out of the seven markers involved in associations in the global analyse still significant but the higher percentage of variance explained was reduced to 4.2 % for IND_70 with BBS (data not shown).

The linear regressions conducted between either the provenance values or the mean BBS for the markers showing significant associations revealed no clinal variation with allelic frequencies except for SNP_12 with the mean BBS ($r^2 = 0.20$; $p = 0.044$) and SNP_134 with PV_LU ($r^2 = 0.50$; $p = 0.022$).

Mode of inheritance

For two markers, SNP_112 and SNP_134, the minor allele was only present at the heterozygous state while for SNP_92 and SNP_103 the homozygous genotypic class for the minor allele was represented only by one individual. Among the 14 significant associations between markers and traits for which the dominance and additive effects could be calculated we found a majority of non-additive effects (Table 3). Interestingly the 6 additive effects all concerned associations detected in the populations of the latitudinal gradient (Figure 3a for

SNP_12). Whereas SNP_103 and IND_70 presented strong underdominance (d/a = -2.038 and -1.537 respectively), SNP_27 was largely overdominant (d/a = 8.410; Figure 3b).

Table 3: Genotypic counts and mode of inheritance of the SNPs showing significant associations with phenotypic traits.

Trait	Locus	N	N_{AA}	N_{AB}	N_{BB}	MAF	P_{AA}	P_{AB}	P_{BB}	2a	d	d/a
Altitudinal Gradient												
GV_LU	**SNP_112**	135	115	20	0	0.07	-0.651	3.618	-	-	-	-
GV_LU	**SNP_103**	135	119	15	1	0.06	0.261	-2.628	5.825	5.564	-5.671	-2.038
BV_LU	**SNP_134**	124	115	9	0	0.04	-0.259	1.878	-	-	-	-
SV_LU07	**SNP_102**	241	189	47	5	0.12	0.009	-0.128	-1.418	1.427	0.576	0.807
SV_LU07	**SNP_92**	258	236	21	1	0.04	-0.069	0.097	2.978	3.047	-1.358	-0.891
SV_LU06	**SNP_27**	257	167	11	79	0.33	0.013	0.831	-0.208	0.221	0.928	8.410
Latitudinal Gradient												
BBS	**SNP_70**	377	129	180	68	0.42	1.783	1.539	2.691	0.908	-0.698	-1.537
BBS	**SNP_153**	443	160	217	66	0.39	2.069	1.631	1.470	0.599	-0.138	**-0.460**
BBS	**SNP_12**	430	122	201	107	0.48	2.189	1.721	1.383	0.805	-0.064	**-0.160**
BBS	**SNP_135**	439	210	119	110	0.39	2.019	1.664	1.355	0.665	-0.023	**-0.069**
BBS	**SNP_146**	444	186	220	38	0.33	1.661	1.700	2.526	0.865	-0.394	-0.911
BBS	**SNP_151**	448	165	214	69	0.39	2.048	1.640	1.507	0.541	-0.138	-0.509
SV_BBS	**SNP_13**	369	327	39	3	0.06	0.081	-0.350	-1.254	1.336	0.236	**0.354**
SV_BBS	**SNP_146**	441	184	219	38	0.33	-0.071	-0.029	0.546	0.617	-0.266	-0.861
SV_BBS	**SNP_153**	440	160	215	65	0.39	0.210	-0.069	-0.251	0.460	-0.049	**-0.212**
SV_BBS	**SNP_151**	445	165	212	68	0.39	0.194	-0.056	-0.258	0.452	-0.024	**-0.105**

N: total number of trees assessed

N_{AA}, N_{AB}, N_{BB}: Number of trees for the two homozygote genotypes (AA and BB) and heterozygote genotypes

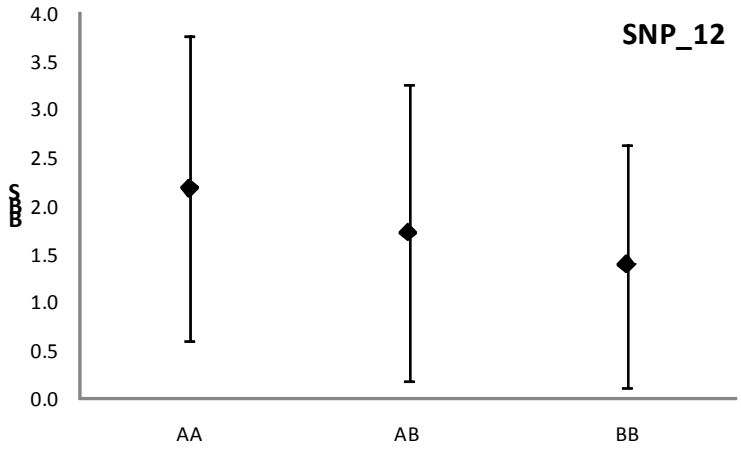
P_{AA}, P_{AB}, P_{BB}: mean phenotypic value of each genotypic class

$a = ½ (P_{AA} - P_{BB})$

$d = P_{AB} - ½ (P_{AA} - P_{BB})$

MAF: minor allele frequency, corresponding to the frequency of allele B.

a)



b)

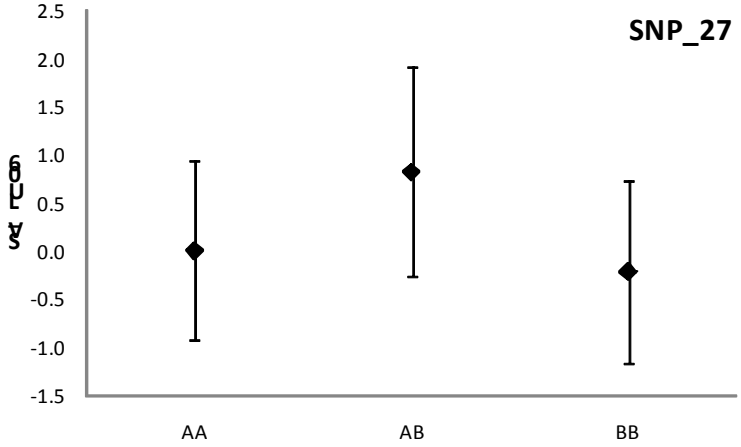


Figure 3: Marker effects on the phenotypic value for each genotypic class of a) SNP_12 (additive effect) and b) SNP_27 (overdominance effect).

For each genotypic class the mean phenotypic value is represented whit bars indicating the standard error.

DISCUSSION

Population structure

Because population structure can be the main cause of false-positive associations we attempted to verify if the markers or populations used for assessing the structure could modify the results of the association tests. To answer this question we used two types of genetic markers carrying different levels of information (altitudinal gradient) and we also tested the effect of removing populations that differentiated markedly (outlier populations) or detecting associations without accounting for structure (latitudinal gradient). Despite the general low level of structure between populations of the altitudinal gradient we found discrepancies between the analyses using either SSRs or SNPs. Whereas Structure based on SSRs identified clusters that were mainly represented in a few populations, the SNP based analysis did not identify these populations. Hence the lower polymorphism of SNPs in comparison to SSRs was not entirely compensated by the higher number of markers. The consequences in terms of detection of associations were noticeable since only three associations out of six were common to both analyses.

For the latitudinal gradient, only SNPs were available to assess the population structure. The overall population structure was however weaker than in the Pyrenees and excluding populations highly differentiated did not change the associations detected. Previous results based on isozyme markers also showed that oak populations are weakly differentiated through the whole distribution range of the species (Zanetto & Kremer, 1995), due to the high outcrossing rate and high gene flow connecting large populations. We thus conclude that most of the neutral population structure was captured with the SNP Q-matrix. Due to the low population differentiation, we further showed that the associations detected did not change when the population structure was not taken into account. We would however suggest using a microsatellite survey in these populations to provide a more accurate estimation of the population structure and the confirmation of the observed associations.

Associations for bud burst in sessile oak

The number of positive associations found between markers and bud phenology variation confirms that association mapping studies based on candidate gene approach is promising in forest tree species. In a recent association study involving 228 SNPs genotyped on 117 candidate genes for cold hardiness in *Pseudotsuga menziesii*, Eckert *et al.* (2009) found 30 trait-marker associations involving 15 unique SNPs and 10 different traits. Here we report significant associations between 16 markers and bud burst in natural populations of oaks

sampled along two environmental gradients. In previous studies on forest trees the proportion of phenotypic variance explained by each marker was usually in the range of 1-5 % (Thumma *et al.* 2005; Gonzalez-Martinez *et al.* 2007; Gonzalez-Martinez *et al.* 2008; Ingvarsson *et al.* 2008; Eckert *et al.* 2009). Similar levels were observed for most of the trait-marker associations detected in our analyses but we also found higher percentages of variance explained in four associations involving different markers and ranging from 6.1 % to 9.5 %. The cumulative effect across several loci could however represent a larger portion of the phenotypic variance. Based on six SNPs associated with cold damage to stem Eckert *et al.* (2009) showed that the cumulative effect accounted for 17 % of the total phenotypic variance. In our case, cumulating the effects of all markers for a given trait lead to a proportion of variance explained varying between 12.5 % (for SV_BBS) and 18.4 % (for GV_LU). This value should however be considered with caution as statistical dependence between loci was not accounted for. No common SNPs showed positive associations in the two environmental gradients. A likely interpretation for the absence of common SNPs is the rather high number of genes that are involved in bud burst. Derory *et al.* (2010) detected 19 QTLs of bud burst in one single mapping pedigree, suggesting that the number would be much larger in natural populations. The high heritability of the trait reported in a companion paper (Alberto *et al.* 2010b), may therefore be due to a large number of genes, each accounting with moderate contribution to the trait. As a result, one could anticipate that different sets of genes may be associated to the trait under different environmental selection pressures. Indeed populations of the latitudinal cline are exposed to very diverse bioclimatic conditions, where interactions between oceanic and continental influences can generate the temperature gradient. Biotic interactions are also likely to vary more across the northern and southern populations than along altitude. Therefore the balance between the selection pressures may modify the contributions of the different loci controlling bud burst in different ways between the populations of the latitudinal and altitudinal gradients. Another explanation would be the low statistical power of the association tests, particularly for the populations of the altitudinal gradient. Long & Langley (1999) have indeed shown by simulations that approximately 500 individuals may be necessary to detect a SNP explaining 5 % of the phenotypic variance. The three associations detected between markers and GV_LU in the populations of the altitudinal gradient comprised only 135 individuals and the homozygous class of the minor allele was only represented by one individual, if not any (Table 3). It would thus be necessary to increase the number of genotype-phenotype comparisons on these populations to increase the power of detection and confirm the observed associations.

None of the markers associated with bud burst variation were also detected as submitted to selection in a previous analysis using multi-locus scans (Alberto *et al.* 2010c) except one, SNP_12. This marker was detected under balancing selection in the altitudinal gradient and is associated with BBS in the latitudinal gradient. However, we found common genes displaying associations and exhibiting large F_{ST} values or clinal variation of allelic frequencies. One SNP located in the *L18a* gene was detected as outlier under directional selection in both gradients, while two adjacent SNPs were associated with BBS in this study. The LD between the outlier and the loci showing associations was however low (r^2 = 0.10 for both). In the *DFR* gene, SNP_92 was associated to SV_LU07, and two different loci located less than 750 pb away in the same gene exhibited clinal variation of allele frequencies in one gradient. These results indicate that most loci involved in the variation of polygenic traits may not be systematically detected by methods based on the variance of allelic frequencies (F_{ST}). Ingvarsson *et al.* (2008) also observed that two markers in the *PhyB2* gene associated with bud set in *Populus tremula* and showing clinal variation with latitude were not more differentiated than neutral markers. The results suggest that gene effects, and not only gene frequencies, may contribute significantly to associations. Outlier loci in terms of differentiation that harbour alleles with low additive (or dominance) values may not show significant associations.

Functional role of genes involved in significant associations

Although we did not found evidence of congruent trait-marker associations between both gradients, some genes showed convergent signals indicating their potential implication in the regulation of bud burst. First, *S11* was commonly associated to bud burst in both gradients for two SNPs located 342 bp apart but showing weak LD in each gradient. A likely interpretation to these results is that the causal mutation is in the close vicinity of both markers and that linkage disequilibrium was conserved between the causal mutation and different SNP in each gradient.

Among the 10 genes showing significant trait associations, Glutathione S-transferase (*GST*) exhibited the highest contribution to variation of apical bud phenology. This protein belongs to a large family of enzymes which are mainly involved in detoxification in plants and animals (Dixon *et al.* 2002). In plants however their functional diversity is not yet fully explored and they are reported to be involved in response to pathogen infection and oxidative stress, cell signaling and regulation of apoptosis (Basantani & Srivastava 2007). Some proteins identified as *GST*s show particular catalytic functions and are involved in plant response to phytohormones like auxins and cytokinin. These phytohormones are well known

to interact during the development of buds. Whereas auxins repress the outgrowth of axillary buds by apical dominance, cytokinins can promote bud activation but are down-regulated by auxin in the stem (Shimizu-Sato *et al.* 2009). The implication of *GST* in the regulation of bud burst via the phytohormone regulation pathways would enlarge the functional diversity of this wide family of enzymes. In addition an auxin induced protein (*AIP*) was also associated to the variation of bud burst, reinforcing the potential role of the phytohormone pathways in the control of bud burst. We also found two genes (*GI* and *COL*) related to the regulation of the flowering pathway in *Arabidopsis* and associated to bud burst regulation. Our results would therefore indicate that *GI* and *COL* are controlling the regulating signals of vegetative bud development in addition to sexual buds as suggested by Horvath (2009). These genes belong to the circadian-clock system which regulates flowering according to photoperiod but it has been shown to be sensitive to temperature conditions (Cao *et al.* 2006, Penfield & Hall 2009, Mikkelsen & Thomashow 2009). Cysteine protease (*CysP*) is involved in diverse processes as protein processing (limited proteolysis) before deposition of storage proteins in developing seeds and subsequent breakdown of storage proteins during mobilization of seed reserves during germination and seedling growth (Grudkowska & Zagdańska 2004). *DFR* catalyses the biosynthesis of flavonoids which are mainly known for their functions in protecting against UV light (UV-B screening pigments), in defense against pathogenic microorganisms but may also be involved in regulating plant growth and enzyme activity (Forkmann & Martens 2001). Finally we found two ribosomal proteins (*S11* and *L18a*) associated with bud burst and previously detected as outlier for directional selection in both gradients (L18a) or showing clinal variation in one gradient (S11). Ribosomal proteins represent a large family of several members generally duplicated several times (249 genes representing 80 protein types in *Arabidopsis*; Barakat *et al.* 2001). Their implication in the specific regulation of bud burst is questionable but their expression was shown to be correlated with leaf growth and cell cycle in *Populus deltoids* (Matsubara *et al.* 2006) and upregulated in germinating seed and growing axillary shoot in *Arabidopsis* (Tatematsu *et al.* 2008). Recently, Van Minnebruggen *et al.* (2010) have shown how duplication of the gene coding for a ribosomal protein (*RPL5*) can influence the development of leaves in *Arabidopsis* mutants. More interestingly, Liu *et al.* (2010) reported recently that phosphorylation of ribosomal proteins during bud dormancy may play important roles in regulating their functions. In our case we do not know the exact localisation of the polymorphism associated to bud burst variance but we can imagine that it could influence the regulation of these genes, and thus the development of buds. Finally it is comforting that in a recent association study conducted in Douglas fir populations for cold

related traits, Eckert *et al.* (2009) also reported an association between a ribosomal protein *RPL31a* and bud burst phenology. These congruent results emphasize the potential involvement of ribosomal proteins in growth traits and bud phenology.

Finally we have shown that association study is a promising approach to detect trait-marker associations in natural populations of sessile oak. We report here 16 significant associations between bud burst variation and genotypes at 10 candidate genes. These associations explain noticeable percentages of variance compared to previous studies. However, despite the interesting number of associations detected, we suspect a lack of statistical power to detect repeatable associations, in particular for the populations of the altitudinal gradient. It seems thus necessary to increase the sample sizes and the number of markers in the future, to assess the genetic architecture of bud burst regulation in natural populations.

REFERENCES

Alberto F, Niort J, Derory J, Lepais O, *et al.* (2010a) Population differentiation of sessile oak at the altitudinal front of migration in the French Pyrenees. *Molecular Ecology* (accepted)

Alberto F, Bouffier L, Louvet JM, Delzon S, Kremer A (2010b) Adaptive responses for seed and leaf phenology in natural populations of sessile oak along altitudinal gradient. *Journal of Evolutionary Biology* (submitted)

Alberto F, Derory J, Kremer A (2010c) Population genomics of clines: Nucleotide diversity of bud burst candidate genes in sessile oak populations from latitudinal and altitudinal gradients. *Genetics* (submitted)

Barakat A, Szick-Miranda K, Chang IF, *et al.* (2001) The organization of cytoplasmic ribosomal protein genes in the arabidopsis genome. *Plant Physiology* **127**, 398-415.

Basantani M, Srivastava A (2007) Plant glutathione transferases - a decade falls short. *Canadian Journal of Botany* **85**, 443-456.

Bradbury PJ, Zhang Z, Kroon DE, *et al.* (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633-2635.

Cao SQ, Jiang ST, Zhang RX (2006) The role of GIGANTEA gene in mediating the oxidative stress response and in Arabidopsis. *Plant Growth Regulation* **48**, 261-270.

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963-971.

Derory J, Scotti-Saintagne C, Bertocchi E, *et al.* (2010) Contrasting correlations between diversity of candidate genes and variation of bud burst in natural and segregating populations of European oaks. *Heredity* (in press)

Dixon DP, Davis BG, Edwards R (2002) Functional divergence in the glutathione transferase superfamily in plants - Identification of two classes with putative functions in redox homeostasis in *Arabidopsis thaliana*. *Journal of Biological Chemistry* **277**, 30859-30869.

Ducousso A, Guyon JP, Kremer A (1996) Latitudinal and altitudinal variation of bud burst in western populations of sessile oak (*Quercus petraea* (Matt) Liebl). *Annals of Forest Sciences* **53**, 775-782.

Ducousso A, Louvet JM, Jarret P, Kremer A (2005). Geographic variations of sessile oaks in French provenance tests. In *Proceedings of the Joint Meeting of IUFRO working groups Genetic of oaks and Improvement and Silviculture of oaks*, ed. R. Rogers, A. Ducousso and A. Kanazashi, FFPRI (Forestry and Forest Products Research Institute) Scientific Meeting Report 3, Tsukuba, Japan, pp 128-138

Eckert AJ, Bower AD, Wegrzyn JL, *et al.* (2009) Association Genetics of Coastal Douglas Fir (*Pseudotsuga menziesii* var. *menziesii*, Pinaceae). I. Cold-Hardiness Related Traits. *Genetics* **182**, 1289-1302.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology* **54**, 357-374.

Forkmann G, Martens S (2001) Metabolic engineering and applications of flavonoids. *Current Opinion in Biotechnology* **12**, 155-160.

Gonzalez-Martinez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB (2007) Association genetics in *Pinus taeda* L. I. Wood property traits. *Genetics* **175**, 399-409.

Gonzalez-Martinez SC, Huber D, Ersoz E, Davis JM, Neale DB (2008) Association genetics in *Pinus taeda* L. II. Carbon isotope discrimination. *Heredity* **101**, 19-26.

Grudkowska M, Zagdanska B (2004) Multifunctional role of plant cysteine proteinases. *Acta Biochimica Polonica* **51**, 609-624.

Horvath D (2009) Common mechanisms regulate flowering and dormancy. *Plant Science* **177**, 523-531.

Ingvarsson PK, Garcia MV, Luquez V, Hall D, Jansson S (2008) Nucleotide polymoirphism and phenotypic associations within and around the phytochrome B2 locus in European aspen (Populus tremula, Salicaceae). *Genetics* **178**, 2217-2226.

Kremer A, Le Corre V, Petit R J, Ducousso A (2010) Historical and contemporary dynamics of adaptive differentiation in European oaks. In : DeWoody A, Bickham J, Michler C, Nichols K, Rhodes G, Woeste K (eds) *Molecular approaches in natural resource Conservation* . Cambridge University Press (in press), pp

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative trait using RFLP linkage maps. *Genetics* **121**, 185-199.

Liu CC, Lu TC, Li HH*, et al.* (2010) Phosphoproteomic identification and phylogenetic analysis of ribosomal P-proteins in Populus dormant terminal buds. *Planta* **231**, 571-581.

Long AD, Langley CH (1999) The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Research* **9**, 720-731.

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, Cottrell J, Csaikl UM*, et al.* (2002) Comparison of levels of genetic diversity detected with AFLP and microsatellite markers within and among mixed *Q. petraea* (Matt.) Liebl. and *Q. robur* L. stands. *Silvae Genetica* **51**, 72-79.

Matsubara S, Hurry V, Druart N*, et al.* (2006) Nocturnal changes in leaf growth of Populus deltoides are controlled by cytoplasmic growth. *Planta* **223:** 1315-1328.

Mikkelsen MD, Thomashow MF (2009) A role for circadian evening elements in cold-regulated gene expression in Arabidopsis. *Plant Journal* **60**, 328-339

Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. *Trends in Plant Science* **9**, 325-330.

Penfield S, Hall A (2009) A Role for Multiple Circadian Clock Genes in the Response to Signals That Break Seed Dormancy in Arabidopsis. *Plant Cell* **21**, 1722-1732.

Price AL, Patterson NJ, Plenge RM*, et al.* (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**, 904-909.

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000a) Association mapping in structured populations. *American Journal of Human Genetics* **67**, 170-181.

Pritchard JK, Stephens M, Donnelly P (2000b) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

Sewell M, Neale D (2000) Mapping quantitative traits in trees. In: Jain S and Minocha S (eds). *Molecular Biology of woody plants*. Forestry Sciences, Volume 64. Kluwer acdameic publishers, Dordrecht, The Netherlands, pp407-423

Shimizu-Sato S, Tanaka M, Mori H (2009) Auxin-cytokinin interactions in the control of shoot branching. *Plant Molecular Biology* **69**, 429-435.

Tatematsu K, Kamiya Y, and Nambara E (2008) Co-regulation of ribosomal protein genes as an indicator of growth status: comparative transcriptome analysis on axillary shoots and seeds in Arabidopsis *Plant Signaling and Behavior* **3**, 450-452;

Thumma BR, Matheson BA, Zhang DQ*, et al.* (2009) Identification of a Cis-Acting Regulatory Polymorphism in a Eucalypt COBRA-Like Gene Affecting Cellulose Content. *Genetics* **183**, 1153-1164.

Thumma BR, Nolan MR, Evans R, Moran GF (2005) Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* **171**, 1257-1265.

White TL, Adams WT, Neale DB (2007) *Forest Genetics*, Cabi Publishing, Wallingford, UK

Zanetto A, Kremer A (1995). Geographical structure of gene diversity in *Quercus petraea* (Matt.) Liebl. I. Monolocus patterns of variation. *Heredity* **75**, 506-517.

Zhao KY, Aranzana MJ, Kim S*, et al.* (2007) An Arabidopsis example of association mapping in structured samples. *Plos Genetics* **3**, e4.

DISCUSSION GENERALE

SYNTHESE DES PRINCIPAUX RESULTATS

Les résultats obtenus dans cette thèse ont à la fois permis de confirmer les observations faites dans plusieurs études antérieures et d'amener de nouveaux éléments pour la compréhension de la biologie de cette espèce.

Forte diversité génétique du chêne

Le Chapitre I de cette thèse avait pour objectif d'étudier l'impact de l'histoire démographique des populations naturelles de chêne sessile sur la diversité génétique neutre le long d'un gradient altitudinal. Cette étude a permis d'analyser l'impact de l'altitude sur les niveaux de diversité et de différenciation, d'un point de vue à la fois rétrospectif par rapport à la colonisation du gradient suite au dernier épisode glaciaire, mais également prospectif dans le cadre de la réponse de ces populations aux changements climatiques en cours. Un autre aspect plus préliminaire était de vérifier que les signatures moléculaires de la sélection sur les gènes candidats du débourrement (Chapitres III et IV) ne seraient pas perturbées par une empreinte démographique trop importante. L'utilisation de 16 marqueurs microsatellites nucléaires a permis de mettre en évidence de forts flux de gènes entre les populations de différentes altitudes, préférentiellement des basses vers les hautes altitudes, qui limitent la différenciation génétique (F_{ST} global = 2.3 %). De nombreuses études sur la diversité d'espèces ligneuses le long de gradients altitudinaux ont montré une diminution avec l'altitude (45 % des études répertoriées dans Oshawa & Ide 2008). Dans notre cas nous n'avons pas remarqué de diminution de la diversité génétique avec l'altitude et nous avons confirmé que les populations naturelles de chêne sessile possèdent un niveau de diversité neutre élevé. De par leurs grandes tailles et les flux de gènes élevés qui les connectent, les populations maintiennent une forte diversité génétique. La différenciation observée le long du gradient altitudinal (F_{ST} global = 2.3 %) est du même ordre de grandeur que celle qui existe entre des populations réparties sur l'ensemble de l'aire de répartition de l'espèce, estimée également à partir de marqueurs microsatellites (Mariette *et al.* 2002). Ce résultat conforte donc l'idée que le chêne sessile possède une diversité génétique importante à l'intérieur des populations qui est maintenue même dans les environnements contraignants de hautes altitudes. De manière surprenante, la seule population où l'on observe une légère baisse de la diversité génétique et un degré plus fort de différenciation qui semblent indiquer de la dérive génétique (population O4), est située à faible altitude. Cette population se trouve sur un versant nord du piémont

pyrénéen et dans une station particulièrement humide (Louvet JM, communication personnelle). Ces conditions écologiques et topographiques sont assez singulières pour le chêne sessile, ce qui semble d'ailleurs affecter la régénération du peuplement (*cf.* faible taux de germination Chapitre II). Par conséquent, le chêne sessile présente une forte capacité d'adaptation à des conditions de températures très contrastées qui n'affectent pas la démographie des populations, mais semble être plus sensible aux conditions hydrologiques et d'exposition. Au final, les faibles niveaux différenciation et la forte diversité génétique, mesurés à l'aide de marqueurs neutres indiquent que ces populations naturelles de chênes sessile représentent un cadre idéal pour l'étude de l'adaptation de cette essence forestière pour le débourrement.

Hétérogénéité des tailles de populations

Si une diminution de la diversité n'est pas observée à travers l'indice de diversité de Nei H_e (Nei 1984), l'utilisation d'une méthode basée sur un modèle de coalescence (Migrate) a permis de détecter des différences au niveau des tailles efficaces de populations entre groupes d'altitudes et vallées, sans que les niveaux de diversité en soient affectés. Ce type de méthode est donc intéressant pour estimer des différences de tailles de populations qui ne sont pas détectés par l'indice de diversité de Nei (1984) H_e. Les faibles tailles efficaces des populations de hautes altitudes s'accompagnent d'une faible structuration génétique qui apparaît de manière indépendante à hautes altitudes dans les deux vallées. Ce résultat peut être lié à des écarts phénologiques de floraison entre les populations de différentes altitudes (Schuster *et al.* 1989, Premoli *et al.* 2003) ainsi qu'à de l'isolement par la distance et le relief à l'intérieur des vallées (Fournier *et al.* 2006, Herrera & Bazaga 2008), qui limiteraient les flux de gènes vers les populations d'altitudes extrêmes.

Rôle évolutif de l'hybridation

Par ailleurs nous avons montré que l'hybridation en populations naturelles avec les chênes pédonculés (*Quercus robur*) et pubescents (*Quercus pubescens*) augmente en fréquence avec l'altitude. Cette observation témoigne du rôle évolutif de l'hybridation au sein du complexe des chênes blancs européens. En effet Petit *et al.* (2004) ont proposé que l'introgression de chêne pédonculé par le chêne sessile pouvait constituer un mécanisme de dispersion et de colonisation de nouveaux habitats. Dans notre cas en plus de contribuer au maintien de la diversité le long du gradient altitudinal (Premoli 2003), il semblerait que l'hybridation puisse

également jouer un rôle adaptatif en créant de nouvelles combinaisons d'allèles dans les populations soumises à des contraintes écologiques particulières. Il est intéressant de noter que l'hybridation a lieu dans des stations qui présentent des caractéristiques propres à l'espèce qui s'hybride avec le chêne sessile. Ainsi, la population L8 où l'on observe majoritairement de l'hybridation avec le chêne pubescent est située sur un escarpement rocheux exposé au sud qui confère un caractère xérique à cette station. Les populations de hautes altitudes (L16 et O16) sont principalement hybridées avec le chêne pédonculé, qui est connu pour sa plus grande résistance au froid que le chêne sessile. L'hybridation pourrait donc s'avérer avantageuses pour l'adaptation aux conditions extrêmes aux limites de la niche écologique d'une espèce, comme l'ont suggéré Lewontin & Birch (1966) et plus récemment Rieseberg *et al.* (2003).

Adaptation locale et forte héritabilité du débourrement

La forte différenciation du chêne sessile pour le débourrement mise en évidence le long du gradient altitudinal dans le Chapitre II montre que l'adaptation pour ce trait est principalement reliée à la température. La répétition des clines entre les deux vallées et également entre les 2 années de mesures indique clairement que l'adaptation locale est corrélée au gradient altitudinal. De plus la tendance clinale observée est concordante avec les résultats de plusieurs études le long de gradients latitudinaux (Liepe 1993, Deans & Harvey 1995, Ducousso *et al.* 1996) qui montrent également un débourrement tardif des provenances des milieux froids. Cette étude a également confirmé la forte héritabilité du débourrement ($h^2 = 0.87$) en populations naturelles rapporté dans les précédentes études chez le chêne (Jensen 1993, Baliuckas & Pliura 2001, Scotti-Saintagne *et al.* 2004) et d'autres espèces forestières (revue dans Howe *et al.* 2003). Le fort contrôle génétique de ce trait en fait donc un sujet idéal pour l'étude de l'adaptation génétique et de son déterminisme moléculaire.

Forte variabilité génétique intra-population

Malgré l'adaptation locale très marquée la variabilité génétique intra-population du débourrement est élevée. Pour un trait si fortement relié à la fitness on pourrait s'attendre à ce que la sélection naturelle favorise une valeur optimale de débourrement dans chaque population. Jump *et al.* (2009) ont proposé l'hypothèse de fluctuations importantes des pressions de sélection en conditions naturelles comme explication des forts niveaux de diversité génétique observés pour des traits adaptatifs. Dans notre cas, nous avons identifié plusieurs pressions de sélections qui pourraient agir pour maintenir la diversité génétique du

débourrement : les gels tardifs au printemps et les pressions des insectes herbivores et des champignons pathogènes. Il existe certainement d'autres facteurs qui peuvent influencer la valeur sélective du trait en conditions naturelles (stress hydrique, compétition pour la lumière avec d'autres espèces végétales). Les fluctuations temporelles et spatiales de l'intensité de ces pressions de sélection, qui peuvent également s'opposer en direction, ont permis le maintien de tels niveaux de variabilité génétique à l'intérieur des populations naturelles. Par ailleurs, la modification de l'intensité des différents facteurs biotiques et abiotiques le long du gradient altitudinal pourrait également expliquer la réduction de la variabilité génétique du débourrement à hautes altitudes (Desprez-Loustau *et al.* 2009, Agrawal *et al.* 2004).

Héritabilité et différenciation de la germination

Cette étude est également la première à s'intéresser au niveau de variabilité de la germination chez une espèce ligneuse. La majorité des études qui ont été menées jusqu'à présent portent sur des espèces herbacées pour lesquelles l'héritabilité de la germination est très faible ($h^2 <$ 0.17 chez *Arabidopsis thaliana*, Donohue *et al.* (2005) ; $h^2 = 0.03$ chez *Lobelia inflata*, Simons & Johnston (2000)). Cette faible héritabilité s'accompagnait d'une forte variabilité phénotypique qui est considéré comme l'exemple typique de la théorie de « bet-hedging » (voir discussion du Chapitre II et références à l'intérieur). Chez le chêne l'héritabilité de la germination est relativement forte ($h^2 = 0.51$). Pour autant la variabilité génétique intra-populations est très forte et largement supérieure à celle du débourrement. La stratégie de diversification de la date de germination semble donc être une caractéristique favorisée par la sélection naturelle, qu'elle ait lieu par plasticité phénotypique chez des espèces herbacées, par effet maternel ou par maintien de variabilité génétique comme chez le chêne. La germination étant déterminante pour le recrutement de la population, la forte variabilité génétique pourrait s'expliquer par une stratégie de diversification par rapport à des conditions environnementales incertaines afin de maximiser la survie (Simons & Johnston 2006).

Déterminisme génétique des traits phénologiques

Par ailleurs nous avons profité de ce dispositif pour tester si le déterminisme génétique de la germination des graines était corrélé à celui du débourrement, ces deux traits présentant tous deux une phase de dormance précédant la croissance et l'expression de gènes identiques qui laissaient supposer une régulation commune (Rohde 2000, Derory *et al.* 2006, Ruttink *et al.* 2007, Rohde & Bhalerao 2007). L'absence de corrélation tant au niveau des provenances que

des descendances semblerait indiquer que le déterminisme génétique de la phénologie des deux traits est différent. Cependant, cette indépendance entre la phénologie des deux traits peut s'expliquer par des différences morphologiques et physiologiques entre la graine et le bourgeon. Dans le cas des graines il existe certains facteurs tels que les interactions avec l'endosperme (tissu maternel), ou les contraintes physiques imposées par les téguments qui peuvent interagir avec le processus de levée de dormance du méristème et moduler ainsi la germination (Rohde & Bahlero 2007). Dans le cas des bourgeons les contraintes physiques imposées par les écailles sont certainement moins fortes et on peut penser que le débourrement soit plus intimement lié à la levée de la dormance. Par conséquent on ne peut pas totalement exclure que le déterminisme génétique de la levée de dormance soit proche pour les deux traits phénologiques mais que des différences morphologiques et physiologiques entre les deux organes masquent cette relation.

Diversité nucléotidique et déséquilibre de liaison

Le Chapitre III a permis de décrire la structure de la diversité nucléotidique de gènes candidats du débourrement dans des populations situées le long d'un gradient altitudinal et latitudinal. La diversité nucléotidique chez le chêne est environ du même ordre de grandeur que celle d'autres essences forestières précédemment étudiées (Brown *et al.* 2004, Pyhäjärvi *et al.* 2007, Ingvarsson 2008). La forte diversité génétique mise en évidence pour les marqueurs isozymes et microsatellites est donc confirmée au niveau nucléotidique pour des gènes potentiellement impliqués dans la variation de traits adaptatifs. Cette caractéristique de l'espèce de posséder une forte diversité génétique à rôle potentiellement adaptatif représente une ressource pour faire face aux changements climatiques en cours.

Le déséquilibre de liaison mis en évidence dans le Chapitre II est faible et décroit rapidement, ce qui est un résultat classique chez des espèces allogames à fortes tailles de populations telles que les espèces forestiers (Savolainen & Pyhäjärvi 2007; Ingvarsson 2008; Eckert *et al.* 2009a). Le nombre important de recombinaisons se produisant en populations naturelles semble donc avoir fortement réduit les liaisons génétiques entre locus pour l'ensemble du génome du chêne. Ce résultat serait bien sûr à confirmer sur un ensemble plus large de gènes (voir *Perspectives*). Cependant il donne des perspectives intéressantes pour la détection d'événements de sélection ayant entrainé un balayage sélectif (et donc une augmentation locale du déséquilibre de liaison).

Signatures de sélection

La détection de SNPs présentant des niveaux de différentiation s'écartant de l'attendu neutre a montré que 15 gènes semblent être soumis à une forme de sélection directionnelle ou balancée dans au moins un gradient. Parmi ces gènes, *GALA* et *L18a* sont soumis dans les deux gradients à la même forme de sélection tandis que *GI* et *ECP* sont soumis à différentes formes de sélection pour différents SNPs selon le gradient. Ce dernier résultat pourrait s'expliquer notamment par le rôle multifonctionnel de ces deux gènes et par des pressions de sélection qui agiraient de manière différente dans les populations des deux gradients. Nous avons également trouvé des relations clinales entre fréquences allèliques et facteurs environnementaux (altitude ou latitude) pour 18 gènes candidats qui semblent liées à la sélection plutôt qu'à de l'isolement par la distance ou de l'admixture entre populations (Vasemägi 2006). Enfin des résultats convergents entre méthodes et/ou entre gradients ont été observés pour 8 gènes qui méritent de plus amples investigations.

Associations entre gènes et débourrement

Finalement, dans le Chapitre IV nous avons mené la première étude d'association réalisée chez le chêne pour le débourrement. La structure des populations estimée à partir de marqueurs microsatellites (Chapitre I) ou SNPs (Chapitre III) a été incorporée dans le modèle. Les résultats ont montré que la structure diffère sensiblement entre les deux types de marqueurs pour les populations des Pyrénées, ce qui modifie les associations détectées entre traits et marqueurs. La structure génétique des populations du gradient latitudinal semble être moins marquée et influence dans une moindre mesure les associations détectées. Au total 16 associations significatives entre débourrement et SNPs ont été découvertes. Ces associations concernent 13 marqueurs situés dans 10 gènes candidats : deux gènes impliqués dans le contrôle circadien de la floraison chez *Arabidospis* (*GI* et *COL*), deux gènes potentiellement liés au voies de régulation des phytohormones (*GST* et *AIP*), deux protéines ribosomales (*S11* et *L18a*), une enzyme de la voie de biosynthèse des flavonoïdes (*DFR*), une Cystéine protéase (*CysP*) et deux protéines inconnues. Les pourcentages de variance expliquée varient entre 2.6 et 9.5 %, ce qui est plutôt élevé par rapport aux associations précédemment mise en évidence chez d'autres espèces forestières (Thumma *et al.* 2005, Gonzalez-Martinez *et al.* 2007, Gonzalez-Martinez *et al.* 2008, Ingvarsson *et al.* 2008, Eckert *et al.* 2009b, Thumma *et al.* 2009). Les gènes impliqués dans ces associations ont pour certains été détectés dans le Chapitre III, soit comme outliers pour le F_{ST} (*GI* et *L18a*) soit avec une variation clinale des

fréquences allèliques (*S11* et *DFR*). Le rôle précis de ces gènes dans la régulation du débourrement reste à déterminer, mais des implications fonctionnelles dans des processus physiologiques ayant attrait à la régulation de la floraison (*GI* et *COL*) et de la croissance (*DFR*, *AIP*, *S11* et *L18a*) semblent valider ces associations.

QUESTIONS SOULEVEES

Disparités des signatures de sélection

Un des résultats surprenants de cette thèse est le manque de concordance entre les marqueurs qui montrent différentes signatures de sélection (Chapitre III), et les marqueurs qui présentent une association avec le débourrement (Chapitre IV). Dans le Chapitre III nous avons évoqué les différences entre les marqueurs qui montrent une différenciation particulière (F_{ST} outliers) et les marqueurs qui présentent des clines de leurs fréquences allèliques. Ce résultat peut s'expliquer par le fait que la différenciation ne soit pas forcément liée à l'adaptation locale au gradient de température. Réciproquement les variations clinales des fréquences allèliques n'induisent pas forcément une différenciation élevée. Des différences clinales de fréquence allèliques peuvent coexister avec une faible différenciation quand l'intensité de sélection intra-population est faible. Cette situation se caractérise par le maintien d'une diversité intra-population importante réduisant de la sorte la différenciation. Les disparités des signatures entre variations de fréquences allèliques (F_{ST} outliers ou variations clinales) d'une part et associations avec les traits d'autant part peuvent s'interpréter par la prise en compte des effets des gènes dans la détection de l'association alors que seules les fréquences des gènes interviennent dans les méthodes basées sur F_{ST}.

Un résultat plus intriguant *a priori* est le fait de détecter des signatures de sélection ou une association qui impliquent différents SNP d'un même gène sur les populations d'un même gradient. C'est le cas par exemple de la protéine ribosomale *L18a* pour laquelle le SNP_152 présente à la fois une variation clinale avec la latitude et une signature de sélection directionnelle dans les deux gradients, tandis que deux autres SNPs (SNP_151 et SNP 153) sont associés à la variance du débourrement dans le gradient latitudinal. Les marqueurs montrant une association sont en déséquilibre de liaison presque complet ($r^2 = 0.97$) tandis qu'ils ne sont pas significativement liés au SNP_152 ($r^2 = 0.10$). C'est également le cas pour *DFR* dont les fréquences allèliques aux marqueurs SNP_89 et SNP_90 sont corrélées respectivement à la latitude et à l'altitude, tandis que le SNP_92 est associé au débourrement dans le gradient altitudinal ($r^2 = 0.26$ avec le SNP_90). Que des SNP différents d'un même gène soient détectés par des méthodes différentes ou dans des clines différents peut

s'interpréter par les variations du déséquilibre de liaison. Tout laisse à penser que les SNP détectés ne sont pas les SNP causant fonctionnellement la variation du trait, mais sont en déséquilibre plus ou moins fort avec ceux-ci.

Réponses différentes le long des 2 gradients

Pour deux gènes, *GI* et *S11*, un même marqueur a montré une signature de sélection directionnelle (SNP_12 pour *GI*) ou une variation clinale de fréquence (SNP_134 pour *S11*) et a également été associé à la variance du débourrement mais dans un gradient différent. Dans les populations étudiées le déséquilibre de liaison est faible et décroit rapidement (Chapitre III). Par conséquent on peut supposer que le locus soumis à sélection se trouve à proximité des marqueurs qui montrent des signatures de sélection, mais se trouve lié à différents marqueurs en fonction de l'histoire de recombinaison des populations des deux gradients.

Les différences observées dans le Chapitre III quant aux marqueurs montrant des signatures de sélection peuvent s'expliquer par la nature hétérogène des conditions climatiques et des pressions de sélection entre les deux gradients, et également entre populations d'un même gradient. Alors que le gradient altitudinal représente une mosaïque d'habitats rapprochés et donc bénéficiant de conditions climatiques proches (hormis la température), le gradient latitudinal fait intervenir certainement intervenir des variations des conditions écologiques (précipitations, composition spécifique) plus importantes. Les adaptations locales et clinales de certains marqueurs détectées entre les deux gradients dans le Chapitre III peut donc s'expliquer par l'hétérogénéité des conditions écologiques au sein des deux gradients de température.

L'adaptation locale pour le débourrement très marquée le long des deux gradients de température ne s'est pas non plus traduite par des associations communes entre débourrement et marqueurs (ou même gènes) dans le Chapitre IV. Plusieurs raisons peuvent être invoquées pour expliquer ce résultat. Tout d'abord la puissance des tests d'association réalisés est certainement trop faible pour identifier tous les marqueurs pouvant avoir un effet sur le trait, en particulier pour les populations du gradient altitudinal où les effectifs varient entre 135 et 260 individus au maximum (*cf.* Chapitre IV - Tableau 1). Cependant les approches théoriques qui se sont intéressé à l'architecture génétique des traits adaptatifs ont montré que la différenciation observée entre populations peut être fortement expliquée par l'effet des combinaisons d'allèles aux différents locus qui contrôlent le trait (Le Corre & Kremer 2003, Latta 2004). On peut donc supposer que la sélection ait favorisé des combinaisons d'allèles à des locus différents entre les deux gradients mais produisant dans les deux cas une forte

différenciation génétique pour le débourrement, ce qui conduirait à des signatures différentes dans les deux clines.

Le choix des gènes candidats

Pour identifier les bases moléculaires de l'adaptation pour des traits l'approche basée sur des gènes candidats semble adaptée afin de diminuer les efforts de séquençage et de cibler les gènes soupçonnés de porter les polymorphismes causaux. Cette approche a surtout été développée chez les espèces non-modèles pour lesquelles la quantité de ressources génomiques est limitée (Gonzalez-Martinez *et al.* 2006). Chez les arbres forestiers elle a permis l'analyse de la diversité nucléotidique afin de rechercher des traces de sélection (Brown *et al.* 2004, Ingvarsson 2005, Namroud *et al.* 2008, Eveno et al. 2008, Wachowiak *et al.* 2009) et plus récemment de tester des corrélations statistiques avec des traits d'intérêt via des études d'associations (Thumma *et al.* 2005, Gonzalez-Martìnez *et al.* 2007, Gonzalez-Martìnez *et al.* 2008, Ingvarsson *et al.* 2008, Eckert *et al.* 2009b, Thumma *et al.* 2009). Cependant l'efficacité de cette approche dépend largement des critères utilisés pour la sélection des gènes étudiés. Dans notre cas, la quasi-totalité des gènes candidats provenaient d'une étude transcriptomique (Derory *et al.* 2006) qui a permis de mettre en évidence des gènes différentiellement exprimés au cours du débourrement. Ces gènes candidats expressionnels ont été confirmés dans leur rôle par leur colocalisation avec des QTLs de développement quand ils ont pu être cartographiés (Derory *et al.* 2010). Tout d'abord il est possible que certains gènes n'aient pas été retenus durant la réalisation de la banque d'EST. On peut aussi se demander si les gènes impliqués dans la variabilité du trait sont forcément exprimés durant les différents stades du débourrement. Le critère d'expression indique seulement qu'ils participent aux processus physiologiques ayant lieu au cours du débourrement, mais pas forcément que leur variabilité est reliée à celle du trait. En effet si la date de débourrement est conditionnée par l'état de dormance des bourgeons, on peut penser que certains gènes peuvent être responsables de niveaux de dormance plus ou moins profonds entre individus et donc de la variabilité du débourrement. Il serait donc intéressant d'étudier également les gènes qui sont exprimés durant la phase de paradormance et d'endodormance. Ce projet est actuellement en cours au sein de l'UMR BIOGECO. Le deuxième critère utilisé pour le choix des gènes candidats a été leur colocalisation avec un QTL du débourrement. Cependant même si l'intervalle de confiance des 19 QTLs du débourrement a pu être réduit par une répétition de mesures sur différentes années et différents tests (Derory *et al.* 2010), le nombre de gènes inclus dans ces intervalles de confiance reste très important. Le critère

fonctionnel peut donc paraître l'un des plus pertinents puisqu'il se base sur des études qui ont pu mettre en évidence l'implication de certains gènes dans d'autres traits phénologiques tels que la levée de dormance des graines ou la floraison, chez des espèces modèles comme *Arabidopsis* ou *Populus* (voir discussions des Chapitre III et IV). Une annotation des contigs d'EST issus de l'étude transcriptomique (Derory *et al.* 2006) par homologie avec le génome de *Populus* est en cours au laboratoire et devrait donner de plus amples informations sur la fonction des gènes déjà séquencés chez le chêne (Frigerio JM, communication personnelle).

PERSPECTIVES

A. Expérimentations de laboratoire

Vers un retour au séquençage

Le faible déséquilibre de liaison mis en évidence sur l'ensemble des 9 fragments de gènes candidats étudiés dans le Chapitre III est un résultat déterminant pour la poursuite des recherches d'associations pour le débourrement chez le chêne. En effet un faible déséquilibre de liaison signifie à la fois que les marqueurs détectés en associations ont de fortes probabilités d'être les mutations causales ou d'être *a priori* situées à proximité. D'un autre côté le faible nombre d'associations détectées peut s'expliquer par la faible représentativité du polymorphisme des gènes candidats étudiés. En effet entre seulement 1 et 7 SNPs par gène ont pu être génotypés en utilisant la technique illumina. Cette technique a été développée à l'origine pour des études sur l'humain dont le génome est nettement moins polymorphe que celui du chêne (5 à 8 SNPs toutes les 10 kb ; Zhao et al. 2003). De plus elle a servi généralement pour des études de génotypage haut-débit de mutations dont l'effet (délétère par exemple dans le cas de maladies génétiques) avait été identifié au préalable. Nous avons également rencontré des problèmes lors de la mise au point de la puce illumina puisque la construction d'amorces n'était possible que sur de zones totalement exemptes de polymorphisme afin d'optimiser l'hybridation sur la matrice d'ADN. Par conséquent bien que cette technique ait permis le génotypage d'un grand nombre d'individus pour un nombre également important de marqueurs, il semblerait préférable pour la poursuite des études de diversité et d'association de privilégier les technologies de séquençage nouvelle génération. En outre le séquençage d'un grand nombre de gènes permettrait de confirmer la diversité nucléotidique importante et le faible déséquilibre de liaison observés sur les 9 fragments de gènes. Finalement ces données offriraient également l'opportunité de réaliser des études de type « genome-wide », c'est-à-dire représentatives de l'ensemble du génome, afin d'estimer

de manière plus précise les paramètres démographiques des populations (diversité, tailles efficaces des populations) et les signatures de la sélection.

Dans le cadre de cette thèse le séquençage de la pleine longueur de trois gènes candidats a été réalisé sur un sous ensemble des populations du gradient altitudinal (environ 200 individus). Ces trois gènes présentaient un intérêt par rapport aux trois critères de sélection évoqués dans le Chapitre III. Le but était à la fois de compléter les données des fragments de gènes sur la diversité nucléotidique et le déséquilibre de liaison, de détecter des signatures de sélection (F_{ST} outliers et variations clinales) et de réaliser une étude d'associations avec le débourrement. Cependant le séquençage de ces gènes s'est révélé problématique pour diverses raisons. Tout d'abord le clonage de ces gènes par fragments d'environ 800-1000 pb s'est avéré difficile malgré de nombreux essais, produisant un nombre trop faible de colonies ayant inséré les fragments pour être exploité. Nous nous sommes donc tournés vers le séquençage direct en diploïde des fragments qui permet d'obtenir simultanément la séquence des deux haplotypes de l'individu. Cette technique s'est avérée beaucoup plus efficaces puisqu'entre 171 et 216 individus ont pu être séquencés en fonction des fragments de gènes. Cependant l'analyse des séquences a révélé un nombre anormalement élevé d'individus présentant des génotypes homozygotes pour tous les sites polymorphes détectés. Ce résultat ne pouvait provenir que d'une amplification préférentielle d'un des deux haplotypes des individus concernés. Une mutation de la séquence au niveau du site d'hybridation de l'amorce a pu empêcher la réaction d'amplification de se produire, créant ainsi un nombre d'allèles « nuls » (non séquencés) élevé. Le nombre estimé d'individus présentant un allèle nul (*i.e.* homozygotes pour tous les sites polymorphes) représentait pour certains gènes plus de la moitié individus séquencés. Par conséquent le choix a été pris d'exclure ces données des analyses réalisées dans les Chapitre III et IV. Nous nous sommes par ailleurs assuré d'éviter le même problème avec les données des gènes candidats et des SNPs génotypés par la méthode illumina en calculant des écarts de la fréquence des hétérozygotes par rapport à l'attendu sous l'hypothèse d'Hardy-Weinberg (indice de consanguinité F_{IS}). Ce résultat illustre les problèmes d'ordre technique qui peuvent être rencontrés lors du séquençage de fragments d'ADN chez une espèce aussi variable que le chêne sessile.

Estimation de l'effet des combinaisons allèliques et modularité de la réponse des populations

Les valeurs de différenciation des SNPs, y compris des « F_{ST} outliers », se sont avérées très largement inférieures à celles des traits (Q_{ST}). Ces différences peuvent s'interpréter par la contribution dominante des associations multiallèliques à la différenciation du trait par rapport à celle des effets propres des allèles (Le Corre et Kremer, 2003). Dans l'immédiat, les données obtenues pourront être utilisées afin de tester en « vraie grandeur » la contribution des associations allèliques. Les études théoriques indiquent que la différenciation d'un trait polygénique est partiellement créée par la covariance des allèles aux différents locus qui le contrôlent (Latta 2004, Le Corre & Kremer 2003). Cette hypothèse est souvent évoquée dans les études de comparaison entre différentiation pour des marqueurs génétiques et différentiation des traits pour expliquer l'écart couramment observé ($F_{ST} < Q_{ST}$). Il serait donc particulièrement intéressant de pouvoir tester cette hypothèse à partir du jeu de données utilisé dans les Chapitres III et IV. Pour cela il est prévu de calculer les corrélations entre fréquences allèliques aux différents marqueurs afin d'observer si l'on détecte des combinaisons d'allèles montrant une covariance importante. Les locus ayant montré des signatures de sélection (F_{ST} outliers, clines des fréquences allèlique et associations) étant privilégiés pour appliquer ce type d'approche. Un test d'association utilisant des variables synthétisant l'information des covariances de l'ensemble des locus (issus d'analyses en composantes principales par exemple) peut être également envisagé pour mettre en évidence l'effet de ces combinaisons d'allèles sur la différenciation du trait.

De manière plus générale, le jeu de données obtenu se prête à évaluer la modularité de la réponse des populations à un gradient environnemental, qui correspond à la combinatoire des allèles dans un système multigénique. Dans ce contexte on peut donc s'attendre à avoir différentes combinaisons allèliques aboutissant au même résultat phénotypique. Une telle situation aboutirait à une forte disparité des signatures moléculaires, ce qui est effectivement observé dans les populations étudiées (cf *Réponses différentes le long des 2 gradients*).

B. Potentiel de réponse du chêne sessile pour le débourrement

Prise en compte de la plasticité phénotypique

La forte variabilité génétique du débourrement et sa forte héritabilité mises en évidence dans le Chapitre II représentent une base pour l'action de la sélection naturelle. Cependant, la rapidité des changements climatiques prédite par l'IPCC (2007) risque fort de poser des difficultés aux arbres pour que l'adaptation génétique puisse être efficace (voir discussion du

Chapitre II). Dans ce contexte il paraît nécessaire de s'intéresser également au rôle que pourrait jouer la plasticité phénotypique du débourrement. En effet cette réponse immédiate des arbres à ajuster leur phénologie en fonction des conditions environnementales de températures pourra certainement permettre dans un premier temps un ajustement des dates de débourrement des individus. Il est généralement admis que la plasticité phénotypique, bien que rapide, soit une réponse limitée en intensité. Cependant la part de la variabilité phénotypique du débourrement observée entre populations le long du gradient altitudinal (6,5 jours d'avance par °C ; Vitasse *et al.* 2009) est expliquée en grande partie par la plasticité phénotypique mesurée par des tests de transplantations réciproques (5,9 jours d'avance par °C ; Vitasse 2009), ce qui concorde avec l'amplitude de la différentiation génétique entre populations observée dans le Chapitre II (entre 1,3 et 1,5 jours.°C à PIE et TOUL respectivement). Par conséquent la plasticité phénotypique du débourrement pourrait entrainer une réponse immédiate du débourrement des populations (acclimatation) et permettre ainsi le laps de temps nécessaire à l'action de la sélection (adaptation) sur la variabilité génétique présente et apportée par les flux de gènes. Néanmoins il faudra considérer dans ce contexte l'importance de la balance entre les températures de chilling qui permettent de lever l'endodormance des bourgeons, et les températures de forcing qui permettent le développement du bourgeon au stade quiescent. Il semblerait que les températures de chilling soient pout l'instant suffisantes chez le chêne pour que le déterminisme environnemental du débourrement soit principalement gouverné par les températures printanières de forcing (Vitasse 2009). Si à l'avenir les besoins en températures de chilling n'étaient plus satisfaits par des hivers devenus trop doux, la réponse des arbres pourraient être plus difficilement prévisible.

Dans le cadre de cette thèse l'étude de la plasticité phénotypique du débourrement était prévue par la transplantation à l'automne 2008 de 1194 plants issus du test de provenances de Pierroton dans les Pyrénées, au site de l'Ayré situé à 1789 m d'altitude dans la vallée de Luz. Le débourrement a été mesuré sur les plants au printemps 2009. Néanmoins les conditions hivernales ont été particulièrement froides cette année et un épais manteau neigeux a recouvert les plants jusqu'en mai. Le débourrement des provenances a eu lieu dans un délai de quelques jours après que la neige ait fondu. Ce résultat montre bien l'effet de la compensation des cumuls de températures de chilling et de forcing sur la phénologie du débourrement. Dans notre cas la neige a certainement maintenu les plants à des températures proches de zéro jusqu'à la fonte, leur permettant ainsi une accumulation très importante de températures de

chilling. Les besoins en froid des plants étant totalement comblés quelle que soit la provenance le débourrement s'est fait de manière quasi synchrone. Cette expérimentation est actuellement toujours en cours et nous avons préféré attendre une année de mesure supplémentaire dans des conditions permettant de mettre en évidence les réponses des différentes provenances pour exploiter les résultats.

Prise en compte de l'épigénétique

Finalement les phénomènes de régulation épigénétique (voir discussion Chapitre II) ouvrent de larges perspectives non seulement pour l'étude de l'adaptation du chêne sessile pour le débourrement, mais également dans un contexte plus général pour la compréhension des mécanismes évolutifs qui gouvernent l'adaptation des espèces à leur milieu. Les observations faites par Skroppa & Kohlmann (1997) chez *Picea abies* révèlent en effet que les conditions environnementales durant la maturation des graines influencent ensuite la phénologie du débourrement et de l'arrêt de croissance chez les plants et que ce phénomène semble perdurer dans le temps. Johnsen *et al.* (2005) a montré que cet effet se traduisait par la régulation de l'expression aux niveaux de gènes impliqués dans la sensibilité à la température. Si les mécanismes moléculaires responsables de cette modulation aux niveaux de l'expression des gènes et du phénotype ne sont pas clairement identifiés chez *Picea abies* et les arbres en général, une littérature croissante s'est développée chez l'espèce modèle *Arabidopsis* et d'autres plantes à fort intérêt économique (coton, maïs…). Ces études ont montré que des phénomènes de méthylation au niveau de l'ADN et des protéines Histones agissent comme des marqueurs qui permettent la régulation de l'expression des gènes (revue dans Lukens & Zhan 2007). Ces phénomènes sont influencés par l'environnement et semblent bénéficier d'une stabilité dynamique, *i.e.* qu'ils peuvent perdurer dans le temps et être transmis entre générations tout en gardant un certain potentiel de réversibilité (Saze 2008).

Il serait donc possible que les mécanismes évolutifs qui gouvernent la variabilité des traits adaptatifs ne soient pas entièrement localisés au niveau de la séquence primaire de l'ADN (marqueurs génétiques) mais également au niveau des marqueurs épigénétiques qui régulent l'expression des gènes. Bien que l'implication de tels phénomènes chez le chêne sessile soit spéculative, il parait crucial de mener des études pour déterminer la part de la variabilité du débourrement attribuable à de tels mécanismes. La mise en évidence et la quantification de ces mécanismes épigénétiques permettraient d'ajuster les modèles de génétique quantitative à travers desquels est évaluée la valeur génétique des individus, par exemple si les effets de

provenances leurs sont majoritairement imputables. De manière plus générale ils permettraient de mieux appréhender les processus évolutifs ayant permis l'adaptation de cette espèce par le passé et ainsi de mieux pouvoir prédire sa capacité de réponse aux changements climatiques en cours et à venir.

CONCLUSION

Les travaux effectués durant cette thèse ont permis de comprendre les changements évolutifs qui ont accompagné l'installation du chêne sessile au cours de l'holocène dans des contextes géographique et historique bien définis. Nous avons montré qu'il existe une forte différenciation entre les populations situées le long de gradients liés à la température, qui découle de leur adaptation depuis la recolonisation post glaciaire. Cette adaptation s'accompagne d'une homogénéité de la diversité génétique neutre et d'une empreinte relativement peu marquée de la démographie. Ce travail a montré que des signatures moléculaires de cette adaptation sont perceptibles dans les séquences gènes candidats. Ces signatures sont cependant entachées d'un bruit de fond lié à la profondeur de la prospection génomique réalisée (peu de SNP par gènes, fragments parties de gènes)., auquel il peut être remédié dans des expérimentations ultérieures.

L'augmentation du nombre de gènes candidats et de la longueur des régions séquencées devrait permettre d'accroître la probabilité de détecter les mutations causales de la variabilité du débourrement.

Enfin toutes les signatures moléculaires possibles n'ont pas encore n'ont pas encore été explorées. Le développement de nouvelles méthodes d'analyses (notamment multi-variables) prenant en compte les combinaisons entre allèles à différents locus devrait également améliorer les modèles d'analyse statistique des traits polygéniques. Ces méthodes seront explorées à brève échéance. Enfin pour mieux appréhender le potentiel d'adaptation du chêne sessile aux changements climatiques il serait important de déterminer le rôle de la plasticité phénotypique dans la capacité de réponse des arbres. Finalement l'investigation des phénomènes épigénétiques semble être une perspective intéressante pour la compréhension des processus évolutifs qui régissent l'adaptation du chêne sessile pour le débourrement.

REFERENCES

Agrawal AA, Conner JK, Stinchcombe JR (2004) Evolution of plant resistance and tolerance to frost damage. *Ecology Letters* **7**, 1199-1208.

Baliuckas V, Pliura A (2003) Genetic variation and phenotypic plasticity of Quercus robur populations and open-pollinated families in Lithuania. *Scandinavian Journal of Forest Research* **18**, 305-319.

Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15255-15260.

Deans JD, Harvey FJ (1995) Phenologies of 16 European provenances of sessile oak growing in Scotland. *Forestry* **68**, 265-273.

Derory J, Leger P, Garcia V, *et al.* (2006) Transcriptome analysis of bud burst in sessile oak (Quercus petraea). *New Phytologist* **170**, 723-738.

Derory J, Scotti-Saintagne C, Bertocchi E *et al.* (2010) Contrasting relationships between the diversity of candidate genes and variation of bud burst in natural and segregating populations of European oaks. *Heredity* (in press).

Desprez-Loustau ML, Vitasse Y, Delzon S, *et al.* (2010) Are plant pathogen populations adapted for encounter with their host? A case study of phenological synchrony between oak and an obligate fungal parasite along an altitudinal gradient. *Journal of Evolutionary Biology* **23**, 87-97.

Donohue K, Dorn L, Griffith C, *et al.* (2005) Environmental and genetic influences on the germination of Arabidopsis thaliana in the field. *Evolution* **59**, 740-757.

Ducousso A, Guyon JP, Kremer A (1996) Latitudinal and altitudinal variation of bud burst in western populations of sessile oak (Quercus petraea (Matt) Liebl). *Annales Des Sciences Forestieres* **53**, 775-782.

Eckert AJ, Wegrzyn JL, Pande B, *et al.* (2009a) Multilocus Patterns of Nucleotide Diversity and Divergence Reveal Positive Selection at Candidate Genes Related to Cold Hardiness in Coastal Douglas Fir (Pseudotsuga menziesii var. menziesii). *Genetics* **183**, 289-298.

Eckert AJ, Pande B, Ersoz ES, *et al.* (2009b) High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (*Pinus taeda* L.). *Tree Genetics & Genomes* **5**, 225-234.

Eveno E, Collada C, Guevara MA, *et al.* (2008) Contrasting patterns of selection at Pinus pinaster Ait. drought stress candidate genes as revealed by genetic differentiation analyses. *Molecular Biology and Evolution* **25**, 417-437.

Fournier N, Rigling A, Dobbertin M, Gugerli F (2006) Random amplified polymorphic DNA (RAPD) patterns show weak genetic differentiation between low- and high-elevation types of Scots pine (Pinus sylvestris L.) in dry continental valleys in the Alps. *Annals of Forest Science* **63**, 431-439.

Gonzalez-Martinez SC, Krutovsky KV, Neale DB (2006) Forest-tree population genomics and adaptive evolution. *New Phytologist* **170**, 227-238.

Gonzalez-Martinez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB (2007) Association genetics in *Pinus taeda* L. I. Wood property traits. *Genetics* **175**, 399-409.

Gonzalez-Martinez SC, Huber D, Ersoz E, Davis JM, Neale DB (2008) Association genetics in *Pinus taeda* L. II. Carbon isotope discrimination. *Heredity* **101**, 19-26.

Herrera CM, Bazaga P (2008) Adding a third dimension to the edge of a species' range: altitude and genetic structuring in mountainous landscapes. *Heredity* **100**, 275-285.

Howe GT Aitken SN, Neale DB, Jermstad KD, Wheeler NC, Chen THH, (2003). From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. *Can. J. Bot.* **81:** 1247-1266.

IPCC (2007). Climate Change 2007: Impacts, Adaptation and Vulnerability. *Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. University Press, Cambridge, United Kingdom and New York, NY, USA., Cambridge.

Ohsawa T, Ide Y (2008) Global patterns of genetic variation in plant species along vertical and horizontal gradients on mountains. *Global Ecology and Biogeography* **17**, 152-163.

Ingvarsson PK (2005) Nucleotide polymorphism and linkage disequilbrium within and among natural populations of European Aspen (Populus tremula L., Salicaceae). *Genetics* **169**, 945-953.

Ingvarsson PK (2008) Multilocus patterns of nucleotide polymorphism and the demographic history of Populus tremula. *Genetics* **180**, 329-340.

Ingvarsson PK, Garcia MV, Luquez V, Hall D, Jansson S (2008) Nucleotide polymoirphism and phenotypic associations within and around the phytochrome B2 locus in European aspen (Populus tremula, Salicaceae). *Genetics* **178**, 2217-2226.

Jensen JS (1993) Variation of growth in Danish provenance trials with oaks (*Quercus robur* L. and *Quercus petraea* Mattuschka Liebl.). *Annals of Forest Science* **50,** 203s-207s

Jump AS, Marchant R, Penuelas J (2009) Environmental change and the option value of genetic diversity. *Trends in Plant Science* **14**, 51-58.

Johnsen O, Fossdal CG, Nagy N, *et al.* (2005) Climatic adaptation in Picea abies progenies is affected by the temperature during zygotic embryogenesis and seed maturation. *Plant Cell and Environment* **28**, 1090-1102.

Latta RG (2004) Gene flow, adaptive population divergence and comparative population structure across loci. *New Phytologist* **161**, 51-58.

Le Corre V, Kremer A (2003) Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. *Genetics* **164**, 1205-1219.

Lewontin RC, Birch LC (1966) Hybridization as a Source of Variation for Adaptation to New Environments. *Evolution* **20**, 315-336.

Liepe K (1993) Growth-chamber trial on frost hardiness and field trial on flushing of sessile oak (Quercus petraea Liebl). *Ann. For. Sci.* **50**, 208s-214s.

Mariette S, Cottrell J, Csaikl UM, *et al.* (2002) Comparison of levels of genetic diversity detected with AFLP and microsatellite markers within and among mixed *Q. petraea* (Matt.) Liebl. and *Q. robur* L. stands. *Silvae Genetica* **51**, 72-79.

Namroud MC, Beaulieu J, Juge N, Laroche J, Bousquet J (2008) Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology* **17**, 3599-3613.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Petit RJ, Bodenes C, Ducousso A, Roussel G, Kremer A (2004) Hybridization as a mechanism of invasion in oaks. *New Phytologist* **161**, 151-164.

Premoli AC (2003) Isozyme polymorphisms provide evidence of clinal variation with elevation in *Nothofagus pumilio*. *Journal of Heredity* **94**, 218-226.

Pyhajarvi T, Garcia-Gil MR, Knurr T, *et al.* (2007) Demographic history has influenced nucleotide diversity in European Pinus sylvestris populations. *Genetics* **177**, 1713-1724.

Rohde A, Howe GT, Olsen JE, *et a*l. (2000) Molecular Aspects of bud dormancy in trees. In: *Molecular biology of woody plants*. Jain, S.M., Minocha, S.C., (eds.). Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 89-134.

Rohde A, Bhalerao RP (2007) Plant dormancy in the perennial context. *Trends in Plant Science* **12**, 217-223.

Ruttink T, Arend M, Morreel K, *et al.* (2007) A molecular timetable for apical bud formation and dormancy induction in poplar. *Plant Cell* **19**, 2370-2390.

Rieseberg LH, Raymond O, Rosenthal DM, *et al.* (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**, 1211-1216.

Savolainen O, Pyhajarvi T (2007) Genomic diversity in forest trees. *Current Opinion in Plant Biology* **10**, 162-167.

Saze H (2008) Epigenetic memory transmission through mitosis and meiosis in plants. *Seminars in Cell & Developmental Biology* **19**, 527-536.

Schuster WS, Alles DL, Mitton JB (1989) Gene flow in limber pine - Evidence from pollination phenology and genetic differentiation along an elevational transect. *American Journal of Botany* **76**, 1395-1403.

Scotti-Saintagne C, Bodenes C, Barreneche T, *et al.* (2004) Detection of quantitative trait loci controlling bud burst and height growth in Quercus robur L. *Theoretical and Applied Genetics* **109**, 1648-1659.

Simons AM, Johnston MO (2000) Plasticity and the genetics of reproductive behaviour in the monocarpic perennial, Lobelia inflata (Indian tobacco). *Heredity* **85**, 356-365.

Simons AM, Johnston MO (2006) Environmental and genetic sources of diversification in the timing of seed germination: Implications for the evolution of bet hedging. *Evolution* **60**, 2280-2292.

Skroppa T, Kohlmann K (1997) Adaptation to local conditions after one generation in Norway spruce. *Forest genetics* **4**: 171-177.

Thumma BR, Matheson BA, Zhang DQ, *et al.* (2009) Identification of a Cis-Acting Regulatory Polymorphism in a Eucalypt COBRA-Like Gene Affecting Cellulose Content. *Genetics* **183**, 1153-1164.

Thumma BR, Nolan MR, Evans R, Moran GF (2005) Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* **171**, 1257-1265.

Vasemagi A (2006) The adaptive hypothesis of clinal variation revisited: Single-locus clines as a result of spatially restricted gene flow. *Genetics* **173**, 2411-2414.

Vitasse Y (2009) Déterminisme environnemental et génétique de la phénologie des arbres de climat tempéré. Thèse de doctorat, Université Bordeaux I, 410 p.

Vitasse Y, Delzon S, Dufrene E, *et al.* (2009) Leaf phenology sensitivity to temperature in European trees: Do within-species populations exhibit similar responses? *Agricultural and Forest Meteorology* **149**, 735-744.

Wachowiak W, Balk PA, Savolainen O (2009) Search for nucleotide diversity patterns of local adaptation in dehydrins and other cold-related candidate genes in Scots pine (Pinus sylvestris L.). *Tree Genetics & Genomes* **5**, 117-132.

Zhao ZM, Fu YX, Hewett-Emmett D, Boerwinkle E (2003) Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. *Gene* **312**, 207-213.

ANNEXES

Molecular Ecology (2009) doi: 10.1111/j.1365-294X.2009.04137.x

Species relative abundance and direction of introgression in oaks

O. LEPAIS,*†‡ R. J. PETIT,*† E. GUICHOUX,*† J. E. LAVABRE,*†§ F. ALBERTO,*† A. KREMER*† and S. GERBER*†

*INRA, UMR 1202 BIOGECO, 69 route d'Arcachon, F-33612 Cestas cedex, France, †Université de Bordeaux, UMR 1202 BIOGECO, 69 route d'Arcachon, F-33612 Cestas cedex, France

Abstract

Successful hybridisation and subsequent introgression lead to the transfer of genetic material across species boundaries. In this process, species relative abundance can play a significant role. If one species is less abundant than the other, its females will receive many heterospecific gametes, increasing mate-recognition errors and thus hybridisation rate. Moreover, first-generation hybrids will also more likely mate with the more abundant species, leading to asymmetric introgression. These predictions have important fundamental consequences, especially during biological invasions or when a rare species threatened by extinction is surrounded by individuals from a related species. However, experimental tests in nature of the importance of the relative abundance of each species on hybridisation dynamics remain scarce. We assess here the impact of species relative abundance on hybridisation dynamics among four species from the European white oak species complex. A total of 2107 oak trees were genotyped at 10 microsatellite markers and Bayesian clustering methods were used to identify reference trees of each species. We then used these reference trees to simulate purebred and hybrid genotypes to determine optimal threshold for genetic assignment. With this approach, we found widespread evidence of hybridisation between all studied oak species, with high occurrence of hybrids, varying from 11% to 31% according to stand and sampling strategies. This finding suggests that hybridisation is a common phenomenon that plays a significant role in evolution of this oak species complex. In addition, we demonstrate a strong impact of species abundance on both hybridisation rate and introgression directionality.

Keywords: frequency-dependent process, genetic assignment, hybridisation, microsatellites, *Quercus*, species delimitation

Received 10 November 2008; revision received 16 January 2009; accepted 26 January 2009

Introduction

Interspecific mating associates heterogeneous genomes, giving rise to new allelic combinations (Rieseberg & Carney 1998). When hybridisation is successful, first-generation hybrids may mate with parental species, producing back-crossed individuals. This leads to gene introgression with transfer of genetic material across species boundaries (Anderson 1949; Martinsen *et al*. 2001; Kim *et al*. 2008). Hybridisation and introgression imply some contact between species so that mating can occur. It has long been argued that local species abundance will impact hybridisation dynamics (Hubbs 1955; Mayr 1963). The rationale is that in species where females exert male choice through prezygotic isolation, hybridisation rate will increase when species relative abundances become sharply unbalanced, because the females belonging to the rare species then receive too many heterospecific gametes and are more likely to make mate-recognition errors (Wirtz 1999; Chan *et al*. 2006). Such a mechanism, sometimes called Hubbs' principle, has been hypothesised in animals (reviewed by Rhymer & Simberloff 1996; Wirtz 1999) and in plants (reviewed by Rieseberg 1997). Differences

Correspondence: Gerber Sophie, UMR BIOGECO, Genetic Team, 69 route d'Arcachon, F-33612 Cestas, France. Fax: +33(0)557122881; E-mail: sophie@pierroton.inra.fr

‡Present address: School of Biological and Environmental Sciences, University of Stirling, Stirling FK9 4LA, Scotland, UK.

§Present address: Estacion Biologica de Donana, CSIC, Integrative Ecology Group, Pabellon del Peru, Avda. M. Luisa S/N, E-41013 Sevilla, Spain.

in species proportion could have consequences beyond the first hybrid generation. This is because first generation hybrids (F_1) will also be more likely to mate with the more abundant species, producing backcrossed individuals that will be more similar to the common species (Anderson & Hubricht 1938; Rieseberg 1997). The validity of Hubbs' prediction is interesting to check because it has important practical and fundamental consequences. For instance, if the minority species is represented by only few individuals that produce a high proportion of hybrids, the species might become locally extinct, by pollen swamping and dilution of the genome of the rare species, although its genes will persist at least temporarily in hybrid individuals (Levin *et al*. 1996; Rhymer & Simberloff 1996). Another situation where species proportion can be highly unbalanced is when a colonising species spreads in an area already occupied by a related species. In this case, the invading species is initially rare, and matings with the local species are likely. Genetic material of the local species incorporated into the invading species can then reach high frequency as the invading population experiences rapid demographic growth, resulting in asymmetric introgression of neutral genes (Currat *et al*. 2008). Clearly, species relative abundance can have important consequences on hybridisation dynamics, affecting both hybridisation rates and the direction of introgression. Although some researchers have acknowledged the fact that species proportion can play an important role in introgression dynamics, only few have experimentally demonstrated its reality in nature (e.g. Buggs 2007; but see Burgess *et al*. 2005; Prentis *et al*. 2007; Field *et al*. 2008; Zhou *et al*. 2008). Additional empirical surveys addressing this issue with different organisms are therefore needed.

Hybridisation has been intensively studied in the genus *Quercus* (Arnold 2006). In particular, hybridisation and introgression are suspected to play a role in postglacial recolonisation of Europe by oaks (Petit *et al*. 2003). Detailed studies of mating system of the two species involved (*Quercus robur* and *Quercus petraea*) in controlled crosses (Steinhoff 1993; Steinhoff 1998; Kleinschmit & Kleinschmit 2000) or in natural populations (Bacilieri *et al*. 1996; Streiff *et al*. 1999) have shown that prezygotic and postzygotic barriers exist, but few studies have focused on the consequences of species abundance on hybridisation dynamics within this species complex. In one recent study, hybridisation rate between two oak species (*Q. petraea* and *Q. pyrenaica*) seemed unrelated to species relative abundance, but the number of investigated stands was limited (Valbuena-Carabaña *et al*. 2007). While oak species are only weakly genetically differentiated, they present important morphological and ecological differences. In forests where several oak species are found in sympatry, species are often clustered according to their ecological requirements (Bacilieri *et al*. 1995). Thus, relative proportions of oak species are expected to vary between stands as a result of local ecological conditions as well as stand history (including forest management). These species represent therefore a good model to test the hypothesis that species proportion affects hybridisation and introgression.

In this study, we adopted a blind (i.e. no a priori classification) approach (Duminil *et al*. 2006) to assign oaks to species and identify hybrids using microsatellite markers and Bayesian clustering methods. We analysed several populations from the four most common species of the European white oak complex in France. We first applied a clustering analysis to all trees studied and then used the results to identify reference trees of each species. These were used to generate artificial genotypes of known ancestry (pure species, hybrids and backcrosses) to determine objective and optimal thresholds for genetic assignment. We analysed several populations and stands with different species composition. This allowed us to test whether relative species abundance influences hybridisation dynamics in this species complex. The specific aims of this paper are (i) identifying hybrid individuals, (ii) estimating the pattern of hybridisation across species and populations, and (iii) testing the effect of parental species proportions on hybridisation rate and introgression.

Materials and methods

Species description

Four oak species were included in this study: *Quercus robur* L. (pedunculate oak), *Q. petraea* (Matt.) Liebl. (sessile oak), *Q. pubescens* Willd. (pubescent or downy oak) and *Q. pyrenaica* Willd. (Pyrenean or rebollo oak). *Quercus robur* and *Q. petraea* are widely distributed in Europe. *Quercus pyrenaica* is found along the Atlantic coast from Morocco and northwestern Spain to western France. *Quercus pubescens* is localised around the Mediterranean Basin with a northern latitudinal limit up to 50 degrees. Distribution range and local species presence are governed by climatic and edaphic factors (Rameau *et al*. 1989). In brief, *Q. pubescens* grows on limestones and in thermophilous stations, whereas *Q. pyrenaica* prefers sandy acidic soils. *Q. robur* is found on rich and deep soils and can support flooding, unlike the other oak species, while *Q. petraea* is found on poorer and dryer soils. Whereas the other three oak species are postpioneer species capable of colonising open land, *Q. petraea* is a late-successional species that grows in stable and well-established forest environment. Thus in the Aurignac region, composed of small forests and woodlands (see below), *Q. petraea* is found in the centre of the stands (Gonzalez *et al*. 2008). The species are traditionally identified during the growing season by examining leaf morphology. *Quercus robur* leaves have short petioles, several secondary veins and their basal parts are typically lobated (Kremer *et al*. 2002). *Quercus petraea* leaves have a longer petiole, no secondary veins and a regular leaf shape. *Quercus*

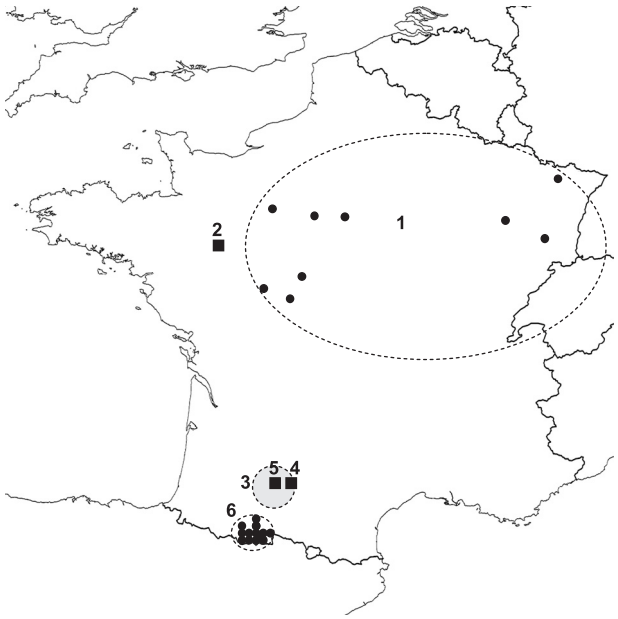


Fig. 1 Location map of the intensively studied stands (squares) and the other sampled populations (encircled) in France (see Table S1 for more details). 1, ONF (National Forest Office) populations; 2, Petite Charnie stand; 3, Aurignac region; 4, Briouant stand; 5, Paguères stand; and 6, Pyrenean populations.

pubescens is similar to *Q. petraea* but the leaves have a higher number of lobes and the abaxial part is densely hairy (Dupouey & Badeau 1993; Curtu *et al*. 2007). *Quercus pyrenaica* leaves are hairy on both sides and have a particular leaf shape with numerous lobes and deep sinuses.

Sampling strategy

A total of 2107 oak trees belonging to the species complex described above were sampled in 53 populations in France (Fig. 1, Table S1, Supporting information). This material had been sampled in the frame of several studies with different objectives, so the sampling strategies are contrasted. The large size of the combined data set should help improve assignment tests (Pritchard *et al*. 2000; Waples & Gaggiotti 2006). In three regions, 10–79 individuals were collected from many populations in France: in the south (Aurignac and Pyrenean stands) and in the north (ONF stands), representing a total of 889 individuals in 50 populations (see Table S1 for more details). In the other areas, stands were more intensively sampled with two stands exhaustively collected, regardless of leaf morphology (Petite Charnie and Briouant) and a third one regularly sampled along a grid (Paguères).

ONF populations consisted in high forests composed mostly of *Q. robur* and *Q. petraea*. Oaks showing typical species morphology were sampled whenever possible. Pyrenean populations were sampled in two valleys at an altitude ranging form 100 to 1600 m. Only *petraea*-like individuals were collected in this study. The Petite Charnie stand has been intensively studied for a long time (Bacilieri *et al*. 1995; Streiff *et al*. 1998; Streiff *et al*. 1999) and only *Q. robur* and *Q. petraea* have been described in this stand, which is part of a continuous high forest. In Aurignac, oak trees showing typical morphology of all three locally abundant oak species (*Q. robur*, *Q. petraea* and *Q. pubescens*) were collected. We sampled one to three individuals by stand (2.5 on average) in 29 forest fragments located within a radius of 30 km around Paguères stand. Briouant and Paguères are two coppice stands localised with Aurignac populations in the long term Ecological Research (LTER-Europe) site 'Vallées et Coteaux de Gascogne'. Paguères includes *Q. robur*, *Q. pubescens* and few *Q. petraea* oaks whereas in Briouant *Q. pyrenaica* is the most frequent species, followed by *Q. robur*, *Q. pubescens* and only few *Q. petraea*.

Two leaves per tree were sampled and kept at 4 °C until stored at –80 °C in the laboratory or immediately dried in silica gel and kept at room temperature. Global positioning system coordinates and morphological species identification using the morphological criteria described above were recorded for each collected tree. Moreover, a detailed morphological analysis was available for the trees from the Petite Charnie (Bacilieri *et al*. 1995) and Briouant (Viscosi *et al*. 2009). Either a discriminant function based on two morphological characters (Kremer *et al*. 2002) was used to distinguish *Q. robur* and *Q. petraea* in the ONF stands or 10 morphological characters were measured to perform a morphological analysis in the case of Pyrenean populations (E. Guichoux, unpublished data and F. Alberto, unpublished data, respectively). When species status was uncertain, oaks were recorded as undetermined species.

Genetic analyses

DNA isolation was performed with a cetyltrimethyl ammonium bromide (CTAB) protocol as previously described (Lepais *et al*. 2006) except for the ONF populations for which the QIAGEN DNeasy Plant Mini Kit was used following the manufacturer's instructions. Ten microsatellite loci selected for their relatively high degree of genetic differentiation between species (Scotti-Saintagne *et al*. 2004; P. G. Goicoechea, unpublished data) were analysed using a multiplex protocol (Lepais *et al*. 2006). Briefly, two polymerase chain reaction were carried out with an MJ Research DNA Engine Tetrad2 thermocycler to amplify the 10 microsatellites: QpZAG110 (Steinkellner *et al*. 1997), QrZAG11, QrZAG112, QrZAG39, QrZAG96, QrZAG7, QrZAG87, QrZAG65, QrZAG5, QrZAG20 (Kampfer *et al*. 1998). Amplified fragments were analysed with an Amersham MegaBace1000 capillary sequencer and individual genotypes were determined with the Fragment Profiler software version 1.2 using the same parameters for all populations.

Table 1 Number of simulated individuals (rows) assigned to the different species or hybrid classes (columns) and computed efficiency, accuracy and global performance of the assignment method (at the bottom). Correct assignments are highlighted in bold

Simulated/assigned	Rob	Pet	Pyr	Pub	Hyb RobPet	Hyb RobPyr	Hyb RobPub	Hyb PetPyr	Hyb PetPub	Hyb PyrPub	Total
Rob	**996**				3	1					1000
Pet		**988**			7			4	1		1000
Pyr			**992**			3		2		3	1000
Pub				**972**			4		14	10	1000
F_1_RobPet		3			**27**						30
bc_RobPet	19	1			**38**	2					60
bc_PetRob		20			**38**			2			60
F_1_RobPyr			1			**28**				1	30
bc_RobPyr	15				2	**40**	3				60
bc_PyrRob	1		16			**41**		1		1	60
F_1_RobPub				1			**29**				30
bc_RobPub	17				4	1	**38**				60
bc_PubRob				11	1		**44**		1	3	60
F_1_PetPyr			1					**29**			30
bc_PetPyr		16			2	1		**38**	3		60
bc_PyrPet			23			2		**32**		3	60
F_1_PetPub							1		**29**		30
bc_PetPub		19	1		1			4	**35**		60
bc_PubPet				29			2		**27**	2	60
F_1_PyrPub								1	2	**27**	30
bc_PyrPub			16			1		4	1	**38**	60
bc_PubPyr			1	25				1	2	**31**	60
Total	1048	1047	1051	1038	123	120	121	118	115	119	4900
Efficiency (percentage)	99.6	99.8	99.2	97.2	68.7	72.7	74.0	66.0	60.7	64.0	
Accuracy (percentage)	95.0	94.4	94.4	93.6	83.7	90.8	91.7	83.9	79.1	80.7	
Performance (percentage)	94.7	93.2	93.6	91.0	57.5	66.0	67.9	55.4	48.0	51.6	

Hyb, hybrids; F_1, first generation hybrids; bc, backcrosses; Rob, *Q. robur*; Pet, *Q. petraea*; Pub, *Q. pubescens*; Pyr, *Q. pyrenaica*.

Admixture analyses

Bayesian clustering of the genetic data was performed using Structure version 2.1 (Pritchard *et al*. 2000; Falush *et al*. 2003). To determine the optimal number of groups (K), we ran Structure with K varying from 1 to 10, with 10 runs for each K value, to find the K value with the highest posterior probabilities. We also used the ΔK statistics to evaluate the change in likelihood (Evanno *et al*. 2005). Our parameters were 50 000 burn-in periods and 100 000 Markov chain Monte Carlo repetitions after burn-in with admixture and correlated allele models without any prior information. For the most likely number of clusters ($K = 4$), we calculated the average result over 10 runs to get the final admixture analysis.

Hybrid simulation and genetic assignment

For each of the four species, we selected at random 65 individuals that had high probabilities (admixture coefficient, $Q > 0.90$) to belong to each of the four corresponding clusters identified in the admixture analysis. This allowed us to estimate allelic frequencies of the four species. We then simulated pure species and hybrid genotypes using these allele frequencies and the R statistic software (R Development Core Team 2005). We simulated 1000 genotypes for each species, 30 F_1 hybrids and 60 backcrosses for all combinations of possible crosses between each pair of species. The number of simulated hybrids is somewhat arbitrary but reflects the expected hybrid percentage observed in real populations (see Results section). We analysed these simulated data set with the Structure software, with $K = 4$ and the same parameters as before, to test the performance of the software to distinguish between pure species and hybrids, and to determine thresholds to assign individuals to these categories to reach a high correct classification rate. We then assigned individuals with the determined threshold (see Results section) and computed efficiency (the proportion of correctly assigned individual), accuracy (the proportion of true hybrids or purebreds assigned in each hybrid or purebred classes) and overall performance (the product of efficiency and accuracy) of the assignment procedure (Vaha & Primmer 2006).

Distance-based analyses

Using the individual tree assignment results, we computed Cavalli-Sforza and Edwards genetic distances (DS;

Cavalli-Sforza & Edwards 1967) between each pair of species or hybrid classes in each population (provided there were a minimum of 10 individuals) with the Populations software (Langella 1999). The resulting distance matrix was used to build an unrooted neighbour-joining tree using the R package APE (Analysis of Phylogenetics and Evolution; Paradis *et al.* 2004).

Hybridisation characteristics and direction of introgression

We further analysed the three intensively sampled stands (Briouant, Petite Charnie and Paguères) to characterise introgression between species. We first performed global analyses to check if there was a difference in the contribution of each species to hybridisation. For $K = 4$, each individual is characterised by a vector of four admixture coefficients. In each stand, we defined two groups of individuals: purebred (whatever their species) and hybrids. We then computed the average of each of the four individual admixture coefficient within groups, resulting in a vector of four averaged admixture coefficients for purebred and a vector of averaged admixture coefficients for hybrids. These two vectors characterised the global genetic composition of purebreds and hybrids in each stand. The null expectation was that each species would contribute to the hybrid gene pool in proportion to its abundance in the stand; that is, the global genetic composition of purebreds should be the same as the global genetic composition of hybrids. To test this hypothesis, we compared the differences between averaged admixture coefficients in purebred and in hybrids using a Student *t*-test. We then investigated the effect of species abundance on differences in genetic composition between hybrids and pure categories. We computed the difference between hybrids and purebreds of each averaged admixture coefficients, considered as an estimate of hybrid excess. This measure of hybrid excess was correlated to the corresponding species relative abundance and tested with a linear model using the R package effects (Fox 2003) to estimate the confidence interval of the linear regression.

We then performed a detailed analysis to test for an effect of parental species relative abundance on introgression directionality. In each stand, we grouped hybrid individuals in one of the six plausible hybrid classes (each characterised by their two parental species). We first computed the average admixture coefficient of each hybrid class in each stand. The genetic composition of each hybrid class is characterised by a vector of four averaged admixture coefficients, among them, the two corresponding to the parental species have a high value while the other have a very low value. We then computed parental species relative abundance for each hybrid class (ratio between the number of oaks of the most abundant parental species and the total number of oaks of the two parental species) in each stand

Table 2 Number (and percentage) of pure species and hybrid oaks as assigned by the Structure software in the different studied stands and populations

Population	Sampling strategy	*N*	Rob	Pet	Pub	Pyr	Hyb RobPet	Hyb RobPub	Hyb RobPyr	Hyb PetPub	Hyb PetPyr	Hyb PubPyr	Total species	Total hybrids
Briouant	Exhaustive, stand	807	240 (29.7%)	3 (0.4%)	83 (10.3%)	235 (29.1%)	28 (3.5%)	35 (4.3%)	48 (5.9%)	7 (0.9%)	40 (5.0%)	88 (10.9%)	561 (69.5%)	246 (30.5%)
Petite Charnie	Exhaustive, stand	262	128 (48.9%)	84 (32.1%)	—	—	15 (5.7%)	11 (4.2%)	12 (4.6%)	7 (2.7%)	5 (1.9%)	—	212 (80.9%)	50 (19.1%)
Paguères	Partial grid-based, stand	149	87 (58.4%)	1 (0.7%)	28 (18.8%)	—	12 (8.1%)	16 (10.7%)	—	2 (1.3%)	—	3 (2.0%)	116 (77.9%)	33 (22.1%)
Aurignac	Partial, 29 populations	75	24 (32.0%)	14 (18.7%)	29 (38.7%)	—	4 (5.3%)	1 (1.3%)	1 (1.3%)	1 (1.3%)	—	1 (1.3%)	67 (89.3%)	8 (10.7%)
Pyrenees	Partial, 12 populations	288	1 (0.3%)	223 (77.4%)	4 (1.4%)	—	31 (10.8%)	3 (1.0%)	1 (0.3%)	9 (3.1%)	15 (5.2%)	1 (0.3%)	228 (79.2%)	60 (20.8%)
ONF	Partial, 9 populations	526	117 (22.2%)	321 (61.0%)	2 (0.4%)	—	24 (4.6%)	11 (2.1%)	4 (0.8%)	26 (4.9%)	21 (4.0%)	—	440 (83.7%)	86 (16.3%)
Total		2107	597 (28.3%)	646 (30.7%)	146 (6.9%)	235 (11.2%)	114 (5.4%)	77 (3.6%)	66 (3.1%)	52 (2.5%)	81 (3.8%)	93 (4.4%)	1624 (77.1%)	483 (22.9%)

N indicates the number of sampled oaks; Hyb, hybrids; Rob, *Q. robur*; Pet, *Q. petraea*; Pub, *Q. pubescens*; Pyr, *Q. pyrenaica* and all hybrid classes between these species by pairs.

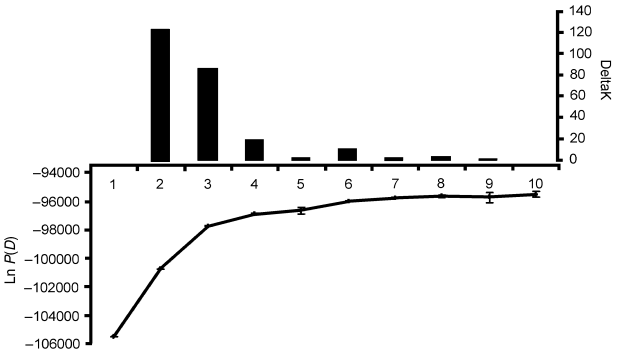


Fig. 2 Estimated number of populations (*K*) derived from the Structure clustering analyses. Mean and standard deviation probabilities of the data over 10 replicated runs (below) and Δ*K* (above) are plotted as a function of the number of clusters (*K* from 1 to 10).

and plotted it against the averaged admixture coefficient of hybrid class that corresponds to the most abundant parental species. If hybridisation is strictly bidirectional or restricted to the first generation (F_1 only), one would expect that the hybrids have an average admixture coefficient value of 0.5. However, if hybridisation is not restricted to the first generation and hybrids themselves can reproduce freely with their parental species, one would expect relative parental species abundance to affect hybrid genetic composition, that is, hybrids would be genetically more similar to the more frequent parental species.

Results

Admixture analysis

The likelihood of the partition of the data increased sharply from $K = 1$ to $K = 3$ and then increased only slightly from $K = 3$ to $K = 6$, where it reached a plateau (Fig. 2). The statistics Δ*K* indicates that $K = 2$ corresponds to the optimal number of groups, but the statistics also gives some support for $K = 3$ or even for $K = 4$ or $K = 6$. We thus report admixture results for $K = 2$, $K = 3$, $K = 4$ and $K = 6$ to compare them (Fig. 3). For $K = 2$, one cluster corresponds to *Quercus robur* (green) and the second to the three remaining morphological species. When adding a third cluster ($K = 3$), *Quercus petraea* is grouped into a specific cluster (yellow) while *Quercus pubescens* and *Q. pyrenaica* are grouped together in the third cluster (pink). For $K = 4$, we get different solutions depending on the run. In seven out of the 10 runs, each species is grouped in one cluster ($K = 4$, Fig. 3: *Q. robur* in the green cluster, *Q. petraea* in the yellow, *Q. pubescens* in the blue and *Q. pyrenaica* in the violet). The other solutions for $K = 4$ (not shown) group *Q. pubescens* and *Q. pyrenaica* in the same cluster while partitioning *Q. robur* and *Q. petraea* in three clusters. Finally, for $K = 6$, only one solution was found: *Q. pubescens* and *Q. pyrenaica* were distinguished as before but *Q. robur* and *Q. petraea* occupied two clusters each. This substructure in *Q. petraea* and *Q. robur* follows a north–south trend with one intraspecific cluster (dark green for *Q. robur* and brown for *Q. petraea*) more frequent in the northern populations while the other (light green for *Q. robur* and orange for *Q. petraea*) is more frequent among southern populations. The genetic distances between the intraspecific clusters are 10-fold smaller than the distances between clusters corresponding to different species, giving strength to the $K = 4$ clustering solution (Fig. S1, Supporting information).

Performance of assignment methods

Distribution of admixture coefficients (*Q*) of simulated individuals (Fig. S2, Supporting information) shows that a threshold value of 0.90 allows separating pure species from hybrids (including F_1 and backcrosses) with the lowest misclassification rate. We thus classified each individual with $Q > 0.90$ as pure species and $Q < 0.90$ as hybrids. However, individuals with $Q < 0.90$ for one cluster but $Q < 0.10$ for each of the three remaining clusters (2.1% of simulated individuals) were supposed to have the majority of their

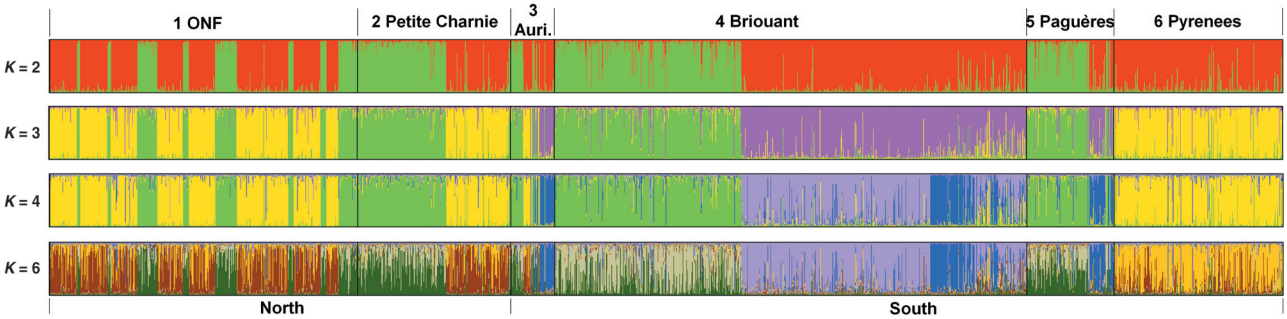


Fig. 3 Structure clustering results obtained for 2, 3, 4 and 6 clusters (*K*). Each individual is represented by a thin vertical line partitioned into *K* coloured segments proportional to its membership in the corresponding genetic cluster. Black lines separate individuals from different populations as indicated at the top, classified according to their latitude, indicated at the bottom. Within populations, individuals are grouped according to their species morphological aspect as determined in the forest (information not used in the clustering analysis).

genome from one species without any significant influence from other species, and they were thus also classified as pure species (changing this rule did not affect the main conclusions of this work, results not shown). For hybrids, we considered that the two species with the highest assignment probability correspond to the hybrid parental species, whatever the probabilities of the third cluster (i.e. the existence of tri-hybrid individuals was ruled out). Note however, that among assigned hybrids 4.9% show a significant contribution ($Q > 0.10$) from a third cluster. Nevertheless, this assignment strategy provides high efficiency and accuracy (Table 1). The overall performance of the method varies from 94.7% to 91.0% depending on the species. Only 0.4% of pure simulated *Q. robur* individuals are wrongly assigned to a hybrid class but the proportion reaches 2.8% for pure *Q. pubescens*. The overall performance is lower for hybrid identification. The majority of simulated F_1 hybrids are correctly assigned to their hybrid class but simulated first-generation backcrosses often fall into the corresponding pure species category (Table 1, Fig. S2). This results in a decrease in the accuracy of pure species identification and in the efficiency of hybrid assignment, as 32% of these backcrosses are wrongly assigned to a pure species class. However, these wrongly assigned individuals are always classified into their parental species class (the species to which the hybrid is backcrossed). Moreover, 2.7% of F_1 and 6.9% of backcrosses are assigned to another hybrid class. Overall, this strategy should result in a conservative approach to hybrid identification (high accuracy at the expense of a decreased efficiency).

Hybridisation between oak species across populations

We assigned all individuals from natural populations using the method indicated above. Among the 1624 trees assigned to pure species, 226 (14%) showed signs of slight introgression (less than 0.90 probability to belong to their own species but less than 0.10 probability to belong to any other species). Among the 483 assigned hybrids (23%), 96 (20%) have a probability, higher than 0.10, to belong to a third species. Those individuals that escape the strict 0.90 threshold rule are far more numerous than in the case of simulated individuals (2.1% and 4.9% in simulated genotypes, respectively, as described above). This result indicates that in real populations, interspecific crosses may be more complex than the ones modelled in simulations. First, the existence of third-generation or later-generation hybrids could explain the high percentage of slightly introgressed trees in nature. Second, hybridisation involving more than two species seems to happen in natural populations.

Overall, we detected a high occurrence of hybrids in all studied populations (Table 2). The percentage of hybrids was higher in the intensively studied stands (Briouant, Petite Charnie and Paguères), ranging from 19.1% to 30.5% (23.9% on average) compared with 10.7% to 20.8% (15.9% on average) in populations where we sampled a limited number of individuals per stand (Aurignac, Pyrenees and ONF) (Table 2).

We identified hybrids between all pairs of species investigated, in particular in Briouant where the four species co-occur (Table 2). Additionally, we detected a number of hybrids involving a species present in the population and another species not identified during field work. This finding is particularly remarkable in the well characterised Petite Charnie stand where only pedunculate and sessile oaks had been described but where hybrids involving *Q. pubescens* and *Q. pyrenaica* were detected using molecular markers (Table 2). A similar finding was made in populations from the Pyrenees and in the ONF stands where hybrids with *Q. pyrenaica* (not known in these areas) were observed. To test if these results can be explained by assignment error, we used the results from the simulated data set (Table 1). Among 2000 simulated pure *Q. robur* and *Q. petraea* trees, six individuals were wrongly assigned to *Q. pubescens* or *Q. pyrenaica* hybrids (0.3%). Out of 150 simulated *Q. robur* × *Q. petraea* hybrids, we wrongly assigned four trees considered to represent *Q. pubescens* or *Q. pyrenaica* hybrids (3%). Assuming that we only have *Q. robur* and *Q. petraea* species and their hybrids in Petite Charnie, we expect to falsely assign less than one individual from the 212 pure species trees to *Q. pubescens* or *Q. pyrenaica* hybrids and less than 1.5 tree from the 50 hybrids to *Q. pubescens* or *Q. pyrenaica* hybrids (Table 2). Thus in total, if the Petite Charnie stand was only composed by *Q. robur* and *Q. petraea* and their hybrids, we would expect less than three erroneous assignments to *Q. pubescens* or *Q. pyrenaica* hybrids. By contrast, we identified 35 hybrid types involving these species (Table 2), a figure that cannot be explained by assignment errors alone.

Analyses of genetic distances between groups confirmed species and hybrid identification. Pure species oaks identified in each population group together in the same common node (Fig. 4). Furthermore, hybrids involving the same pair of species, whatever their geographical origin, share a common node or are localised in the same part of the tree. This is clearly the case for *Q. robur* × *Q. petraea*, *Q. robur* × *Q. pubescens* and *Q. robur* × *Q. pyrenaica* hybrids (Fig. 4).

Genetic composition of species and hybrids

We computed the average of each of the four admixture coefficients for the two categories (pure species and hybrids). In the three intensively studied stands, the overall genetic composition differed between pure species and hybrids (Fig. 5). The fact that the genetic composition of the pure species category differs from that of the hybrid category indicates that the four species are not involved proportionally in the formation of hybrids and backcrosses. In Petite Charnie, *Q. robur* and *Q. petraea* genes seem to be equally

Fig. 4 Phylogenetic neighbour-joining tree based on Cavalli-Sforza and Edwards genetic distances (Cavalli-Sforza & Edwards 1967) between pure species and hybrids as assigned by the Structure software in the different populations. Only groups with more than 10 individuals were used to build the tree, the scale line represents a genetic distance of 0.05. Large branches represent pure oak species with colours corresponding to Fig. 3 at $K = 4$. Thinner branches illustrate hybrid groups with each colour corresponding to a specific hybrid type. Labels at the tip of the branches indicate the corresponding species or hybrid type (Rob, *Quercus robur*; Pet, *Q. petraea*; Pub, *Q. pubescens*; Pyr, *Q. pyrenaica*; and hyb, hybrid) and populations' names are given in the subscript (Bri, Briouant stand; PC, Petite Charnie stand; Pag, Paguères stand; Auri, Aurignac populations; ONF, ONF populations; Pyr, Pyrenean populations).

represented in species and hybrid trees but *Q. pubescens* and *Q. pyrenaica* genes are significantly overrepresented among hybrids ($P < 0.001$ and $P < 0.01$, respectively). In Briouant, *Q. robur* genes are far less present in the hybrid category than in the pure species category ($P < 0.001$) whereas *Q. petraea* and *Q. pubescens* genes are significantly more frequent among the hybrid category ($P < 0.001$ and $P < 0.001$, respectively). In Paguères, we also found that *Q. robur* genes are under-represented among hybrid trees ($P < 0.001$), whereas *Q. petraea* and *Q. pyrenaica* genes are over-represented among hybrids ($P < 0.001$ and $P < 0.001$, respectively).

Species frequency-dependent hybridisation and introgression

Differences in genetic composition between hybrids and pure-bred individuals suggest that genes of the more abundant species are under-represented in hybrids (Fig. 5). To formally test this hypothesis, we have plotted the species relative abundance in each stand against the difference in its genetic composition in hybrids vs. purebreds (Fig. 6). There is a clear negative relationship (Fig. 6, $R^2 = 0.83$, $F_{1,10} = 52.86$, $P < 0.001$). This result comforts our observation that abundant species are proportionally less involved in hybridisation

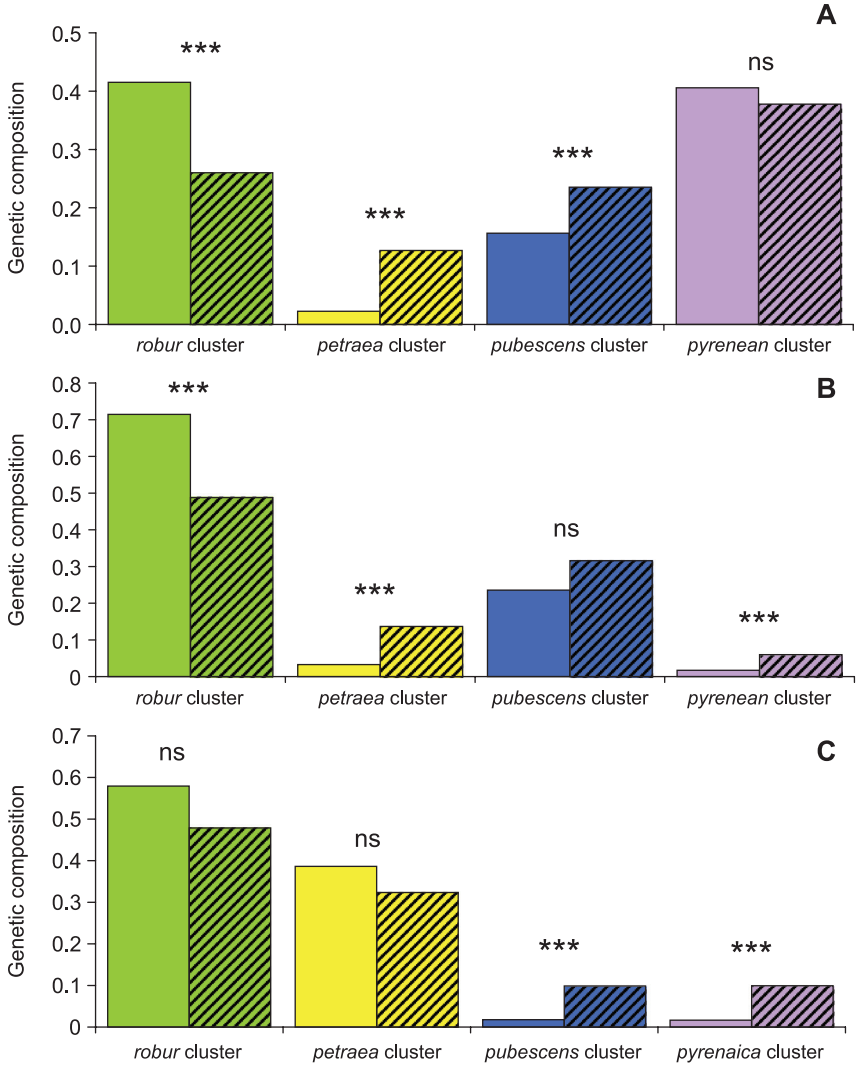


Fig. 5 Comparisons of genetic composition (averaged admixture coefficients from each of the four clusters) for pure species (plain colours) and hybrids (dashed colours) in Briouant (A), Paguères (B) and Petite Charnie (C) stands. Differences were tested with a Student's *t*-test (***: $P < 0.001$, NS: not significant).

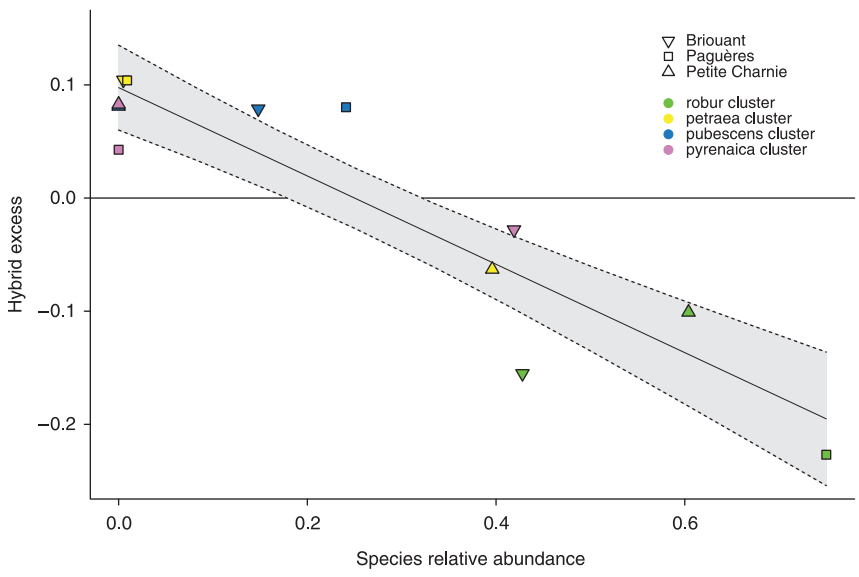


Fig. 6 Change in admixture coefficient between hybrids and purebreds as a function of the corresponding species relative abundance in the stand. The continuous black line indicates no difference between averaged admixture coefficients for hybrids and purebreds. A positive value indicates over-representation of the corresponding cluster in hybrid individuals whereas a negative value indicates over-representation of the corresponding cluster in purebred oaks. Dashed lines and grey shading indicate the confidence interval of the linear regression (large black line; $R^2 = 0.83$, $F_{1,10} = 52.86$, $P < 0.001$). The shapes of the symbols represent the different stands (down-pointing triangle, Briouant; square, Paguères; up-pointing triangle, Petite Charnie) and colours represent clusters (green, *Q. robur* cluster; yellow, *Q. petraea* cluster; blue, *Q. pubescens* cluster; and purple, *Q. pyrenaica* cluster).

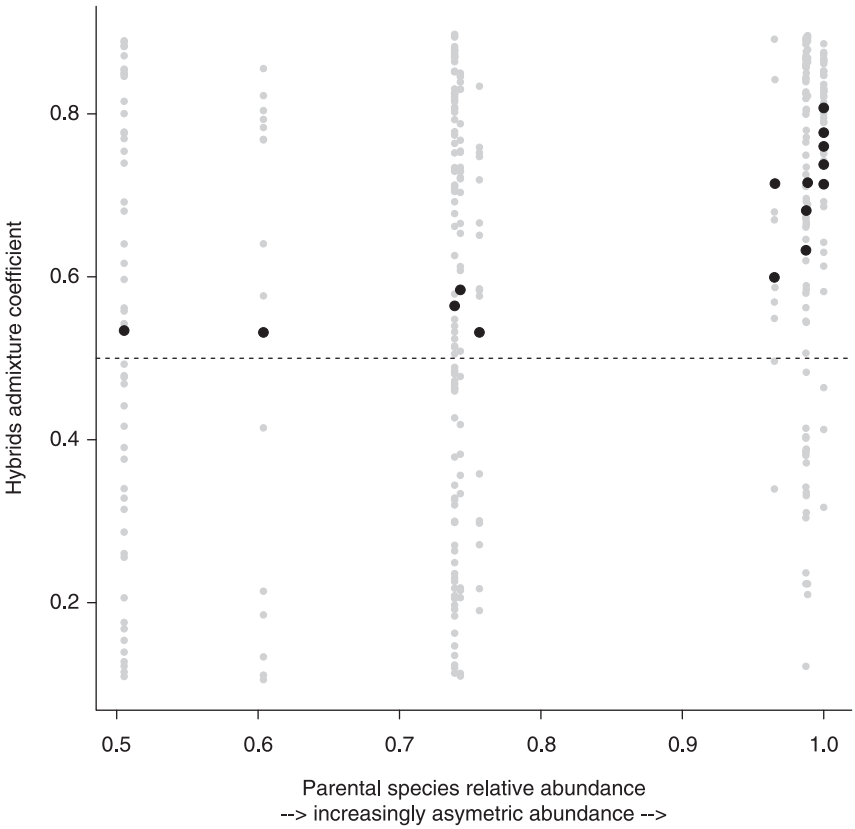


Fig. 7 Effect of parental species relative abundance on hybrid admixture coefficients. Small grey points represent admixture coefficient of each hybrid individuals whereas large black points represent the averaged admixture coefficient for each hybrid class in each stand. For each hybrid class, we used the admixture coefficient corresponding to the most abundant parental species. The horizontal dashed line gives the expected admixture coefficient if introgression was not directional.

than minority species. We then performed a detailed analysis of genetic composition of hybrid classes by using admixture coefficients (Fig. 7). Hybrid individuals admixture coefficients have a large distribution, indicating that hybridisation is not restricted to the first generation (i.e. numerous hybrids had admixture coefficient between 0.65 and 0.9, values that are unlikely for F_1 hybrids, see Fig. S2). Moreover, the averaged admixture coefficient of hybrid classes showed that some classes have an intermediate admixture value, pointing to balanced bidirectional introgression, whereas others hybrid classes have a genetic composition closer to one of the parental species (Fig. 7), indicating directional introgression. Hence, bidirectional introgression seems to take place when parental species are equally represented, whereas directional introgression appears to predominate when parental species differ greatly in abundance (Fig. 7).

Discussion

Our work has addressed the effect of species relative abundance on natural hybridisation and introgression. There are surprisingly few such studies in natural populations. We showed that relative species abundance affects both hybridisation rates and introgression directionality. Previous studies have reported hybridisation patterns between pairs of oak species (Muir *et al*. 2000; Muir & Schlötterer 2005; Valbuena-Carabaña *et al*. 2005, 2007; Gugerli *et al*. 2007) or have studied more species but in one restricted area (Curtu *et al*. 2007). Our extended analyses of 2107 oaks belonging to four species and several populations provide new insights into hybridisation and introgression dynamics within the European white oak species complex. Such large sample sizes should provide accurate estimates of allelic frequencies in the different oak species for use in species delineation and hybrid identification (Waples & Gaggiotti 2006). Using genetic clustering and simulations, we assigned the species or hybrid origin of each sampled oak. We found that hybrids (*sensu lato*: including introgressed individuals) are common in all studied populations, supporting previous claims that hybridisation is ongoing among these oak species (Gugerli *et al*. 2007). Moreover, intensive sampling in three stands allowed us to demonstrate the importance of stand species composition in hybridisation patterns and introgression dynamics.

From clustering to assignment analysis

In the clustering analyses, we found stable results for $K = 6$, highlighting not only differences between species but also a geographical structure within *Quercus robur* and *Quercus petraea*. Such a result might be due to a geographical gradient in allele frequencies, as demonstrated for allozyme data in

Q. petraea (Zanetto & Kremer 1995; Kremer & Zanetto 1997; Le Corre *et al*. 1998). Using more loci on a wider sampling area covering the distribution range of the species could improve the understanding of these subspecific genetic patterns. In any case, it is clear that intraspecific differences are subsidiary to species differences, and thus intraspecific variation does not compromise species identification. The leaf morphology of a subset of the individuals had been previously analysed (Viscosi *et al*. 2009), showing a clear concordance between genetic cluster and morphological features in these oak species.

We then tested the performance of species assignment and hybrid identification using data-based simulations. Our results show that classes of pure and admixed individuals detected with Structure had been reconstructed with good accuracy and efficiency. However, our 10 microsatellites were not able to differentiate first from second-generation hybrids, an objective that has been shown to require more than 48 loci in cases of low genetic divergence, such as the one observed in these oaks (Vaha & Primmer 2006). Note that our estimates of hybrid abundance are conservative since the threshold we selected ($Q = 0.90$) to distinguish pure species from hybrids should slightly underestimate hybrid proportions and minimise assignment error rate among hybrid classes. Altogether, the results indicate that assignment methods, if used with caution, can be efficient to delimitate species across broad geographical ranges, without prior morphological information, as already shown by Duminil *et al*. (2006). They further indicate that assignments are still relevant when more than two species are present and when an intraspecific geographical structure is detected.

Widespread occurrence of hybrids in the European white oak species complex

Our genetic assignment analysis also confirms that sympatric species from the European white oaks complex do hybridise. Overall hybrid frequencies differ among areas (11–30%, Table 2) with more hybrids detected in intensively sampled stands (19–30%) than in less intensively sampled populations (11–21%). Sampling a small proportion of individuals in a stand can lead to an underestimation of hybridisation if oaks with typical leaf morphology are preferentially sampled. In a detailed multivariate analysis of leaf morphology, hybrid individuals were on average morphologically intermediate between parental species (Viscosi *et al*. 2009). Hence, some (but not all) hybrid oaks could be characterised by an intermediate leaf morphology and intentionally (or not) avoided during sampling (Lexer *et al*. 2006). Estimated hybridisation rates based on non-exhaustive sampling should thus be taken with caution.

The hybrid frequencies found in our populations are comparable with, although slightly higher than, previously found in other studies using comparable approaches. An analysis of three stands in Spain comprising *Q. petraea* and *Q. pyrenaica* detected between 6% and 22% of hybrids depending on the stand (Valbuena-Carabaña *et al*. 2007). Likewise, genetic assignment in a four-oak-species stand in Romania detected between 2% and 16% hybrids depending on the species pairs (Curtu *et al*. 2007). These estimates suggest that hybridisation is not a rare event in oaks and that it is a contemporary process. We were able to identify hybrids between all species pairs studied, indicating that no strict reproductive barriers exist. However, the frequency of the different hybrid classes varies among stands, suggesting that local conditions can affect the outcome of hybridisation. The simultaneous analysis of forests located far apart, with material from all four species included as reference, allowed us to detect hybridisation between species pairs in situations where one of the parental species is locally absent. In the Petite Charnie stand, for instance, only *Q. robur* and *Q. petraea* oaks have been described so far (Bacilieri *et al*. 1995; Streiff *et al*. 1998; Streiff *et al*. 1999) but we identified 13% of *Q. pubescens* and *Q. pyrenaica* hybrid types in this stand (Table 2, Fig. 4), compared with only 6% of *Q. robur* × *Q. petraea* hybrids. This finding highlights the importance of including all species potentially connected by gene flow when studying hybridisation with genetic assignment methods. A separate analysis of the Petite Charnie stand, for example, would have resulted in the detection of only two clusters without any chance to identify *Q. pubescens* and *Q. pyrenaica* hybrids.

The presence of hybrids in the absence of one parental species has also been demonstrated in American red oaks (Dodd & Afzal-Rafii 2004), pinyon pines (Lanner & Phillips 1992) and *Aesculus* tree species (DePamphilis & Wyatt 1989; Thomas *et al*. 2008). Two hypotheses can explain such observations: hybridisation by long-distance pollen dispersal or past local extinction of one of the two parental species (Buggs 2007; Thomas *et al*. 2008). Massive deforestation during the last 3000 years by human exploitation and land clearing for agriculture render difficult to estimate original species distribution ranges and thus the possibility of local extinction of *Q. pubescens* and *Q. pyrenaica* to explain the occurrence of their hybrids. Occasional long-distance hybridisation is not unlikely in these highly outcrossing wind pollinated species. The nearest *Q. pubescens* or *Q. pyrenaica* populations are localised some tens of kilometres from Petite Charnie. Because *Q. pubescens* and *Q. pyrenaica* are more drought tolerant and thermophilous than *Q. robur* and *Q. petraea*, dispersal by long-distance pollen hybridisation could be a mechanism to speed up their northern migration facing climate warming.

Frequency-dependent hybridisation and introgression

Species relative abundance is one of the factors that can affect hybridisation pattern and introgression dynamics

(Anderson & Hubricht 1938; Nason *et al*. 1992; Burgess *et al*. 2005). Our detailed analysis of three stands differing in species composition allowed us to estimate the relative species abundance and its impact on the outcome of hybridisation.

Hybridisation rate

We found a deficit of hybrids involving locally dominant species (e.g. *Q. robur* and *Q. pyrenaica* in Briouant and *Q. robur* and *Q. petraea* in Petite Charnie), whereas less frequent or rare species tend to be over-represented among hybrids (Figs 5 and 6). Several hypotheses could account for this observation. First, dominant species are expected to be well adapted to local environmental conditions; their hybrids may therefore have a lower competitive ability. Limited hybrid formation between dominant species in a stand would then be caused by differential selection between hybrid and parental species. Second, if these hybrids were selected against, the strength of reproductive barriers between dominant species could increase as a result of reinforcement (Dobzhansky 1937; Butlin 1987). This would lead to a higher reproductive isolation and a lower hybridisation rate between dominant species, compared with species that came more recently in contact, for which reinforcement would not have time to develop. Comparative analyses of open-pollinated progenies with contrasted species abundance situations would be useful to test the hypothesis of reinforcement. Third, rare species could be over-represented among hybrids because of their difficulty to mate with other rare conspecific partners. Such minority species should receive abundant heterospecific pollen, which would increase hybridisation rate (Rieseberg & Gerber 1995). Relative species abundance and underlying causal factors such as local environment and forest management could have a major influence on hybridisation rate. However, this prediction should be tested by manipulating the proportion of pollen from several species received by female flowers using controlled crosses experiments.

Direction of introgression

As we were unable to differentiate F_1 from backcrosses using direct genetic assignment, we computed the mean admixture coefficients of the different hybrid classes in each stand to get some insight into the genetic composition of hybrid individuals compared to their parental species. A mean admixture coefficient of 0.5 would imply that only first-generation hybrids exist or that each parental species mates in the same proportion with hybrids, producing a balanced number of each type of backcrosses. On the contrary, if the backcrosses were biased towards one of the parental species, we should observe a mean cluster value between 0.5 and 0.9 because a majority of the hybrids would be closer to the successfully backcrossing species. Clearly, the observed distribution of individual admixture coefficients in hybrids indicates that backcrosses are more numerous than F_1, as the majority of hybrids showed admixture coefficient between 0.65 and 0.90 (Fig. 7). These results show that hybridisation is not restricted to the formation of F_1 but instead involves further generations of backcrosses between pure species and F_1 hybrids.

Our results show that the direction of introgression strongly depends on the relative frequency of the parental species in the studied stands (Fig. 7). Knowledge of mating system of oak hybrids are lacking, with the exception of one study using controlled crosses on a fertile *Q. robur* × *Q. petraea* hybrid (Olrik & Kjaer 2007). In our study, we found that the direction of the backcrosses was predominantly towards the more numerous species. Additional analyses of hybrid reproductive behaviour would greatly improve our understanding of the hybridisation dynamic in this species complex. However, it is already clear that interspecific gene flow is a widespread and ongoing process among oak species. Since the species remain morphologically and ecologically distinct (Kremer *et al*. 2002; Petit *et al*. 2003), this observation indicates that collective evolution (*sensu* Morjan & Rieseberg 2004) takes place within these species in the face of extensive interspecific gene flow. It would be interesting now to study if collective evolution can simultaneously take place higher in the hierarchy, within groups of closely related species, as first suggested by Pernès (1984). The European white oaks would seem to be good candidates to test this idea, in view of the high rate of interspecific gene flow they experience. In any case, our results indicate that the rate of exchange between species belonging to the same species complex should not be viewed as a fixed parameter but as a variable one that depends on several factors such as the local composition of the community.

Acknowledgements

We thank Jean-Marc Louvet, Jérôme Willm, Maya Gonzalez and Alain Cabanettes for sampling assistance and sharing their field knowledge. We are grateful to Patrick Léger, Valerie Léger, Pierre-Yves Dumolin and Franck Salin for technical assistance. O.L. is grateful to Martin Lascoux for his invitation at the Evolutionary Biology Centre of Uppsala University. We thank Richard Abbott, Alex Buerkle and three anonymous reviewers for their suggestions that greatly improved the manuscript. Genotyping presented in this publication was performed at the Genotyping and Sequencing facility of Bordeaux (grants from the Conseil Régional d'Aquitaine n°20030304002FA, n°20040305003FA and from the European Union, FEDER n°2003227). Experiments were funded by the Interregional Project Aquitaine/Midi-Pyrénées: 'Évolution de la biodiversité des forêts sous l'effet des changements globaux (changements d'usage et changements climatiques)', by the French Research Agency (ANR) through the QDIV project: 'Quantification of the effects of global changes on plant diversity' (n°ANR-05-BDIV-009-01), and

by the European Union supported project (QLRT-1999-30690) OAKFLOW 'Intra- and interspecific gene flow in oaks as mechanisms promoting genetic diversity and adaptive potential', as well as by the Office National des Forêts ('Traçabilité géographique et identification taxonomique du bois de chêne des forêts domaniales françaises').

References

Anderson E (1949) *Introgressive Hybridization*. Wiley & Sons, New York.

Anderson E, Hubricht L (1938) Hybridization in Tradescantia. III. The evidence for introgressive hybridization. *American Journal of Botany*, **25**, 396–402.

Arnold ML (2006) *Evolution through Genetic Exchange*. Oxford University Press, USA & Oxford, UK.

Bacilieri R, Ducousso A, Kremer A (1995) Genetic, morphological, ecological and phenological differentiation between *Quercus petraea* (Matt) Liebl and *Quercus robur* L. in a mixed stand of Northwest of France. *Silvae Genetica*, **44**, 1–10.

Bacilieri R, Ducousso A, Petit RJ, Kremer A (1996) Mating system and asymmetric hybridization in a mixed stand of European oaks. *Evolution*, **50**, 900–908.

Buggs RJA (2007) Empirical study of hybrid zone movement. *Heredity*, **99**, 301–312.

Burgess KS, Morgan M, Deverno L, Husband BC (2005) Asymmetrical introgression between two *Morus* species (*M. alba*, *M. rubra*) that differ in abundance. *Molecular Ecology*, **14**, 3471–3483.

Butlin R (1987) Speciation by reinforcement. *Trends in Ecology & Evolution*, **2**, 8–13.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis. Models and estimation procedures. *American Journal of Human Genetics*, **19**, 233–257.

Chan C, Ballantyne KN, Aikman H *et al*. (2006) Genetic analysis of interspecific hybridisation in the world's only Forbes' parakeet (*Cyanoramphus forbesi*) natural population. *Conservation Genetics*, **7**, 493–506.

Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. *Evolution*, **62**, 1908–1920.

Curtu AL, Gailing O, Finkeldey R (2007) Evidence for hybridization and introgression whithin a species-rich oak (*Quercus* spp.) community. *BMC Evolutionary Biology*, **7**, 218.

DePamphilis CW, Wyatt R (1989) Hybridization and introgression in buckeyes (*Aesculus*, Hippocastanaceae): a review of the evidence and a hypothesis to explain long-distance gene flow. *Systematic Botany*, **14**, 593–611.

Dobzhansky T (1937) *Genetics and the Origin of Species*. Columbia University Press, New York.

Dodd RS, Afzal-Rafii Z (2004) Selection and dispersal in a multispecies oak hybrid zone. *Evolution*, **58**, 261–269.

Duminil J, Caron H, Scotti I, Cazal SO, Petit RJ (2006) Blind population genetics survey of tropical rainforest trees. *Molecular Ecology*, **15**, 3505–3513.

Dupouey JL, Badeau V (1993) Morphological variability of oaks (*Quercus robur* L., *Quercus petraea* (Matt) Liebl, *Quercus pubescens* Willd) in northern France: preliminary results. *Annales des Sciences Forestières*, **50**, 35s–40s.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software Structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Field DL, Ayre DJ, Whelan RJ, Young AG (2008) Relative frequency of sympatric species influences rates of interspecific hybridization, seed production and seedling performance in the uncommon *Eucalyptus aggregata*. *Journal of Ecology*, **96**, 1198–1210.

Fox J (2003) Effect display in R for generalised linear models. *Journal of Statistical Software*, **8**, 1–27.

Gonzalez M, Deconchat M, Balent G, Cabanettes A (2008) Diversity of woody plant seedling banks under closed canopy in fragmented coppice forests. *Annals of Forest Science*, **65**, 511.

Gugerli F, Walser JC, Dounavi K, Holderegger R, Finkeldey R (2007) Coincidence of small-scale spatial discontinuities in leaf morphology and nuclear microsatellite variation of *Quercus petraea* and *Q. robur* in a mixed forest. *Annals of Botany*, **99**, 713–722.

Hubbs CL (1955) Hybridization between fish in nature. *Systematic Zoology*, **4**, 1–20.

Kampfer S, Lexer C, Glossl J, Steinkellner H (1998) Characterization of $(GA)_n$ microsatellite loci from *Quercus robur*. *Hereditas*, **129**, 183–186.

Kim M, Cui M-L, Cubas P *et al*. (2008) Regulatory genes control a key morphological and ecological trait transferred between species. *Science*, **322**, 1116–1119.

Kleinschmit J, Kleinschmit JGR (2000) *Quercus robur — Quercus petraea*: a critical review of the species concept. *Glasnik Za Smske Pokuse*, **37**, 441–452.

Kremer A, Zanetto A (1997) Geographical structure of gene diversity in *Quercus petraea* (Matt.) Liebl. II. Multilocus patterns of variation. *Heredity*, **78**, 476–489.

Kremer A, Dupouey JL, Deans JD *et al*. (2002) Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak stands. *Annals of Forest Science*, **59**, 777–787.

Langella O (1999) *Populations*, Version 1.2.28. Available from URL: http://www.pge.cnrs-gif.fr/bioinfo/populations/index.php.

Lanner RM, Phillips AM III (1992) Natural hybridization and introgression of pinyon pines in northwestern Arizona. *International Journal of Plant Sciences*, **153**, 250–257.

Le Corre V, Roussel G, Zanetto A, Kremer A (1998) Geographical structure of gene diversity in *Quercus petraea* (Matt.) Liebl. III. Patterns of variation identified by geostatistical analyses. *Heredity*, **80**, 464–473.

Lepais O, Leger V, Gerber S (2006) Short note: high throughput microsatellite genotyping in oak species. *Silvae Genetica*, **55**, 238–240.

Levin DA, Francisco-Ortega J, Jansen RK (1996) Hybridization and the extinction of rare plant species. *Conservation Biology*, **10**, 10–16.

Lexer C, Kremer A, Petit RJ (2006) Shared alleles in sympatric oaks: recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. *Molecular Ecology*, **15**, 2007–2012.

Martinsen GD, Whitham TG, Turek RJ, Keim P (2001) Hybrid populations selectively filter gene introgression between species. *Evolution*, **55**, 1325–1335.

Mayr E (1963) *Animal Species and Evolution*. Harvard University Press, Cambridge, Massachusetts.

Morjan CL, Rieseberg LH (2004) How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology*, **13**, 1341–1356.

Muir G, Schlötterer C (2005) Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci

differentiating two hybridizing oaks (*Quercus* spp.). *Molecular Ecology*, **14**, 549–561.

Muir G, Fleming CC, Schlötterer C (2000) Species status of hybridizing oaks. *Nature*, **405**, 1016–1016.

Nason JD, Ellstrand NC, Arnold ML (1992) Patterns of hybridization and introgression in populations of oaks, manzanitas and irises. *American Journal of Botany*, **79**, 101–111.

Olrik DC, Kjaer ED (2007) The reproductive success of a *Quercus petraea* × *Q. robur* F_1-hybrid in back-crossing situations. *Annals of Forest Science*, **64**, 37–45.

Paradis E, Claude J, Strimmer K (2004) APE. Analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.

Pernès J (1984) *Gestion des ressources génétiques. Tome 2: Manuel*. Agence de Coopération Culturelle et Technique, Paris, France.

Petit RJ, Bodenes C, Ducousso A, Roussel G, Kremer A (2003) Hybridization as a mechanism of invasion in oaks. *New Phytologist*, **161**, 151–164.

Prentis PJ, White EM, Radford IJ, Lowe AJ, Clarke AR (2007) Can hybridization cause local extinction: a case for demographic swamping of the Australian native *Senecio pinnatifolius* by the invasive *Senecio madagascariensis*? *New Phytologist*, **176**, 902–912.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

R Development Core Team (2005) R: a languaage and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from URL: http://www.R-project.org, edn.

Rameau JC, Mansion D, Dumé G (1989) *Flore forestière française: guide écologique illustré, 1: Plaines et collines*. Institut pour le Développement Forestier, Paris, France.

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*, **27**, 83–109.

Rieseberg LH (1997) Hybrid origins of plant species. *Annual Review of Ecology and Systematics*, **28**, 359–389.

Rieseberg LH, Carney SE (1998) Plant hybridization. *New Phytologist*, **140**, 599–624.

Rieseberg LH, Gerber D (1995) Hybridization in the Catalina Island Mountain Mahogany (*Cercocarpus traskiae*). *RAPD Evidence. Conservation Biology*, **9**, 199–203.

Scotti-Saintagne C, Mariette S, Porth I *et al*. (2004) Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. Petraea* (Matt.) Liebl.]. *Genetics*, **168**, 1615–1626.

Steinhoff S (1993) Results of species hybridization with *Quercus robur* L. & *Quercus petraea* (Matt.) Liebl. *Annales Des Sciences Forestieres*, **50**, 137s–143s.

Steinhoff S (1998) Controlled crosses between pedunculate and sessile oak: results and conclusion. *Allgemeine Forst und Jagdzeitung*, **169**, 163–168.

Steinkellner H, Fluch S, Turetschek E *et al*. (1997) Identification and characterization of $(GA/CT)_n$-microsatellite loci from *Quercus petraea*. *Plant Molecular Biology*, **33**, 1093–1096.

Streiff R, Labbe T, Bacilieri R *et al*. (1998) Within-population genetic structure in *Quercus robur* L. & *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. *Molecular Ecology*, **7**, 317–328.

Streiff R, Ducousso A, Lexer C *et al*. (1999) Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. Petraea* (Matt.) Liebl. *Molecular Ecology*, **8**, 831–841.

Thomas DT, Ahedor AR, Williams CF *et al*. (2008) Genetic analysis of a broad hybrid zone in *Aesculus* (Sapindaceae): is there evidence of long-distance pollen dispersal? *International Journal of Plant Sciences*, **169**, 647–657.

Vaha JP, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*, **15**, 63–72.

Valbuena-Carabaña M, González-Martínez SC, Sork VL *et al*. (2005) Gene flow and hybridisation in a mixed oak forest (*Quercus pyrenaica* Willd. and *Quercus petraea* (Matts.) Liebl.) in central Spain. *Heredity*, **95**, 457–465.

Valbuena-Carabaña M, González-Martínez SC, Hardy OJ, Gil L (2007) Fine-scale spatial genetic structure in mixed oak stands with different levels of hybridization. *Molecular Ecology*, **16**, 1207–1219.

Viscosi V, Lepais O, Gerber S, Fortini P (2009) Leaf morphological analyses in four European oak species (*Quercus*) and their hybrids: a comparison of traditional and geometric morphometric methods. *Plant Biosystems* in press.

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1439.

Wirtz P (1999) Mother species-father species: unidirectional hybridization in animals with female choice. *Animal Behaviour*, **58**, 1–12.

Zanetto A, Kremer A (1995) Geographical structure of gene diversity in *Quercus. petraea* (Matt.) Liebl. I. Monolocus patterns of variation. *Heredity*, **75**, 506–517.

Zhou R, Gong X, Boufford D, Wu CI, Shi S (2008) Testing a hypothesis of unidirectional hybridization in plants: observations on Sonneratia, Bruguiera and Ligularia. *BMC Evolutionary Biology*, **8**, 149.

This article is a part of O.L.'s PhD thesis focusing on hybridization dynamics between European white oak species. O.L. has a wide interest in application of molecular markers for studying the ecology, evolution and history of species. R.J.P. is a population geneticist with broad interest in evolution, phylogeography and mating system of trees. E.G. is a PhD student working on the characteristics of oak species used by the barrel industry. J.L. collaborated with O.L. during her Master; she is currently doing a PhD on the spatial and temporal variability of the mutualistic interaction between *Taxus baccata* L. and its frugivores' community. F.A. is a PhD student working on the adaptation of *Quercus petraea* (Matt.) Liebl. along an altitudinal gradient in the Pyrenean Mountains. A.K. has long standing interests in the evolution of temperate and tropical forest trees with particular emphasis on population differentiation at various levels where diversity is expressed (from genes to phenotypes). S.G. is a geneticist interested in population genetics and gene flow studies in forest trees, she supervised O.L.'s thesis.

Supporting information

Additional supporting information may be found in the online version of this article:

Fig. S1 Neighbour-joining tree illustrating the net nucleotide genetic distances, as computed by the STRUCTURE software, between clusters at K = 6.

Fig. S2 Admixture coefficients distribution for simulated individuals: (A) pure species, (B) first generation hybrids (F1), (C) second generation hybrids (backcrosses) and (D) averaged distribution of pure species, first and second generation hybrids.

Table S1 Details of the sampled populations.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Heredity (2009), 1–11

www.nature.com/hdy

ORIGINAL ARTICLE

Contrasting relations between diversity of candidate genes and variation of bud burst in natural and segregating populations of European oaks

J Derory[1,2], C Scotti-Saintagne[1,2,4], E Bertocchi[1,2], L Le Dantec[3], N Graignic[1,2], A Jauffres[1,2], M Casasoli[1,2], E Chancerel[1,2], C Bodénès[1,2], F Alberto[1,2] and A Kremer[1,2]

[1]INRA, UMR1202 BIOGECO, Cestas, France; [2]Université de Bordeaux, UMR1202 BIOGECO, Cestas, France; [3]INRA, UPR419, Unité de Recherche sur les Espèces Fruitières, Villenave d'Ornon, France and [4]INRA, UMR745 ECOFOG, Kourou, French Guiana

Nucleotide diversity was assessed within nine candidate genes (in total 4.6 kb) for the time of bud burst in nine sessile oak (*Quercus petraea*) populations distributed in central and northern Europe. The sampled populations were selected on the basis of their contrasting time of bud burst observed in common garden experiments (provenance tests). The candidate genes were selected according to their expression profiles during the transition from quiescent to developing buds and/or their functional role in model plants. The overall nucleotide diversity was large ($\pi_{tot} = 6.15 \times 10^{-3}$; $\pi_{silent} = 11.2 \times 10^{-3}$), but population differentiation was not larger than for microsatellites. No outlier single-nucleotide polymorphism (SNP), departing from neutral expectation, was found among the total of 125 SNPs. These results contrasted markedly with the significant associations that were observed between the candidate genes and bud burst in segregating populations. Quantitative trait loci (QTLs) for bud burst were identified for 13 year*site seasonal observations in a cloned mapping pedigree. Nineteen QTLs were detected, and QTLs located on linkage groups 2, 5 and 9 contributed repeatedly to more than 12% of the phenotypic variation of the trait. Eight genes were polymorphic in the two parents of the pedigree and could be mapped on the existing genetic map. Five of them located within the confidence intervals of QTLs for bud burst. Interestingly, four of them located within the three QTLs exhibiting the largest contributions to bud burst.
Heredity advance online publication, 9 December 2009; doi:10.1038/hdy.2009.170

Keywords: nucleotide diversity; candidate gene; QTL; *Quercus petraea*

Introduction

Spatial distribution and abundance of long-lived species, such as forest trees, is closely related to phenological events, and especially bud burst (Chuine and Beaubien, 2001). Changes in the date of bud burst may modify the length of the growing season, the flowering time and reproductive success of trees. The fact that bud burst is a fitness-related trait is confirmed by the important clinal variation that is observed in provenance tests of many tree species (Savolainen *et al*., 2007). As the date of bud burst is mainly driven by temperature for temperate tree species (Chuine and Cour, 1999), global warming is expected to have a major impact on the phenology of trees. An increase in the length of the growing season has already been observed for trees in Europe and North America (Menzel and Fabian, 1999; Penuelas and Filella, 2001). Changes in the date of bud burst will modify exposures to late frost (Scheifinger *et al*., 2003), or to outbreaks of phytophagous insects (Van Asch and Visser, 2007). The adaptive response of trees to these rapid climatic changes will depend on the levels of genetic variation within natural populations. In this contribution, we explore the distribution of genetic variation of bud burst in European white oaks, by focusing on the trait variation itself and on the diversity of the underlying putative genes controlling the trait. Large latitudinal and altitudinal clinal variations were observed in provenance tests (Ducousso *et al*., 1996). In a recent contribution, we showed that the geographical pattern of bud burst variation was resulted from local adaptation in response to the warming following the last glaciation (Kremer *et al*., 2009). The clinal pattern observed along latitude and altitude is therefore the consequence of natural selection mediated by the temperature gradient, which has occurred during the last 15 000 years, and is not imprinted by the earlier differentiation existing among the refugial populations (Kremer *et al*., 2002)

The between-population variation was shown to be associated to the maintenance of important genetic variation within populations (Scotti-Saintagne *et al*., 2004). Previous quantitative trait locus (QTL) detection analysis suggested that the important genetic variation was likely due to a large number of genes, with rather small-to-moderate effects (Scotti-Saintagne *et al*., 2004). Technical limitations or reduced pedigree size in forest trees did not permit to implement QTL cloning. Indeed, positional cloning (Tanksley *et al*., 1995) and insertion mutagenesis (Bechtold *et al*., 1993) are not available for

Correspondence: Dr A Kremer, INRA, UMR BIOGECO 1202, 69, rte d'Arachon, Cestas F-33610, France.
E-mail: antoine.kremer@pierroton.inra.fr
Received 18 August 2009; accepted 27 August 2009

tree species. The research of the underlying genes residing within the QTLs was therefore oriented toward a candidate gene (CG) approach (Pflieger *et al*., 2001). In a recent study, accumulation of transcripts in apical buds was monitored during spring time in order to identify differentially expressed genes before and after bud flush of sessile oak, leading to the identification of expressional CGs that were either up- or downregulated during the transition from quiescent to developing buds (Derory *et al*., 2006). In this study, we compare the relations between the diversity of the CGs and the variation of the time of bud burst in natural and segregating populations in order to validate the expressional CGs. In natural populations, we assessed nucleotide diversity and differentiation in populations showing contrasting timing of bud burst. Molecular imprint of natural selection for bud burst was tested in two different ways: (i) by comparing the differentiation of CG sequences with control neutral markers (for example, microsatellites) and (ii) by comparing the differentiation of CG sequences between geographical and phenological groups of populations. In segregation populations, we refined the QTL position of the timing of bud burst by compiling observations across environments and years, and compared their position with the position of CG on the genetic map following earlier analysis conducted by Scotti-Saintagne *et al*. (2004) and Casasoli *et al*. (2006). Our overall objective was to strengthen the selection of CGs based on their variation in natural and segregation populations.

Materials and methods

Sequencing, genotyping and diversity analysis of microsatellites and CGs

Plant material: The study material used for the assessment of nucleotide polymorphisms in CGs and SSRs loci come from a common garden experiment (provenance test) installed in four different plantations in the Central Part of France (from the West to the East). Nine oak populations were sampled within the provenance tests comprising 112 in total (Ducousso *et al*., 1996). The plantations were installed between 1990 and 1995, and each provenance was represented by 10 repetitions of 24 trees in each test (Table 1). Bud burst was assessed in spring after 3 years of plantation following a grading system ranging between 0 (quiescent bud) to 5 (fully developed leaves and elongated shoot). Population differentiation (Q_{ST}) for bud burst, as assessed in these tests, amounted to 0.55 (Ducousso *et al*., 2005). The nine sampled populations were chosen according to their contrasting bud burst recorded in the tests (from early to late flushing) and their geographic distribution. Within this study, the populations were clustered in two different ways: according to the time of the bud burst as recorded in the provenance test (phenological groups) and according to their geographic origin (geographic groups) (Table 1). Within each of the nine sampled populations, six trees were selected at random for sequencing DNA of the CGs.

CGs and microsatellite loci: Nucleotide polymorphisms were assessed in partial fragments of 9 CGs and 15 microsatellites loci (Table 2). CGs were selected according to differential expression of expressed sequence tags (ESTs) before and after bud flush as assessed by cDNA macroarray experiments and real-time reverse-transcriptase PCR (Derory *et al*., 2006) or according to their functional role as assessed in model plants (Table 2). EST sequences were blasted in gene databases such as NCBI (http://www.ncbi.nlm.nih.gov/BLAST/), in order to search for homologs in oak species (Derory *et al*., 2006). A gene product was assigned to each EST, on the basis of sequence similarity to proteins with known function in the trEMBL and Swiss-prot databases, using BLASTX with an *e*-value $\leqslant 10^{-10}$ (http://www.expasy.ch/sprot/). The fragments sequenced in this study corresponded to the nine following genes. Galactinol synthase (GALA), Dof Affecting Germination 2 (DAG2), Alpha-amylase/subtilisin inhibitor (ASI) and metal-nicotianamine transporter (YSL1) were drastically downregulated during the transition from quiescent to developing buds (Derory *et al*., 2006), and thus constitute relevant expressional CGs for molecular signals regulating bud burst transition. The functional role of these genes was investigated in studies conducted on other plants. Indeed, GALA was reported to be involved

Table 1 Descriptive data of the sampled populations

Name	*Country*	*Longitude (°)*	*Latitude (°)*	*Altitude (m)*	*Bud burst score in provenance tests*[a]				*Phenological group*[b]	*Geographical group*[c]
					Test 1	*Test 2*	*Test 3*	*Test 4*		
Mölln	Germany	10.75	53.62	36	0.58	1.18	0.91	1.48	L	N
Sprakensehl	Germany	10.6	52.8	115	0.71	1.11	1.02	1.64	L	N
Cochem	Germany	7.05	50.08	400	1.86	2.17	2.02	3.22	E	C
La Neuville-en-Hez	France	2.33	49.4	70	2.15	1.87	1.93	3.05	–[d]	–[d]
Johanneskreuz	Germany	7.83	49.4	460	1.15	–[e]	1.13	2.14	I	C
Saint Jean	France	6.72	48.82	227	1.02	1.23	1.00	1.51	L	S
Westhoffen	France	7.45	48.6	400	2.77	2.7	2.63	3.21	E	S
Haslach	France	7.38	48.55	265	2	1.81	1.86	2.88	I	S
Klostermarienberg	Austria	16.57	47.41	310	3.28	3.8	4.15	4.08	–[d]	–[d]

[a]Results available in Ducousso *et al*. (1996).
[b]Phenological groups: E, early flushing; I, intermediate; L, late flushing.
[c]Geographical groups: S, Southern; C, Central; N, Northern.
[d]Populations Klostermarienberg and La Neuville-en-Hez not assigned to geographical nor phenological groups because of their eccentric location.
[e]Population Johanneskreuz not present in test 2.

Table 2 Summary data of candidate genes

Candidate gene	*Annotation*	*Coding (bp)*	*Non-coding (bp)*	*Amplified region*	*No. of SNPs*	*No. of singletons*	*Size of indels (bp)*	*Reference for PCR primer pairs*	*Accession no.*
GALA	Galactinol synthase	210	295	E/I/E	18	7	28	Derory *et al*. (2006)	CR327918
ASI	Alpha-amylase/ subtilisin inhibitor	327	0	E	19	3	3/1	Derory *et al*. (2006)	CR627933
DAG2	Dof Affecting Germination 2	360	0	E	2	0		Derory *et al*. (2006)	CR627781
H3	Histone H3	126	248	E/3′-UTR	5	2	2/1 (3 ×)	Casasoli *et al*. (2006)	CR627839
PM23	Seed maturation protein PM23	390	714	E/I/E/I/E	41	3	12/5/4/3 (3 ×)/ 2 (2 ×)/1 (4 ×)	Derory *et al*. (2006)	CR627986
AUX-REP	Auxine-repressed protein	259	188	E/I/E	7	3	2/1	Casasoli *et al*. (2006)	CR627732
YSL1	Metal-nicotianamine transporter YSL1	315	0	E	8	0		Derory *et al*. (2006)	CR627947
GA3	GA3-β-hydroxylase	342	310	E/I/E	3	0	6	Forward: 5′-TCCCTAAGCGTATGTGGTCC-3′ Reverse: 5′-CCCTCTCGTGTTGTTTTGGT-3′	AJ006453
GA20	GA20-oxydase	425	129	E/I/E	4	0	27	Forward: 5′-TCAGCTCAACCAGACTCATC-3′ Reverse: 5′-TTGAAAGTGCCATGAAGGTG-3′	AJ420192
Total		2754	1884		107	18	23		

Abbreviations: E, exon; I, intron; indel, insertion–deletion; UTR, untranslated region.

in tolerance to drought, high salinity and cold (Pukacka and Wojkiewicz, 2002; Taji *et al*., 2002) in *Arabidopsis thaliana*. DAG2 was shown to act as a transcription factor specifically involved in the maternal control of seed germination (Gualberti *et al*., 2002). In buds, as in seeds, it may potentially act on dormancy release. ASI is expressed in germinating seeds of rice, barley and wheat; the protein is multifunctional and is involved in hydrolysis of starch metabolism and in seed defense against pathogens (Furtado *et al*., 2003; Nielsen *et al*., 2004). In oaks, ASI was highly expressed in buds only at the quiescent stage, suggesting that hydrolysis of storage starch or glycogen is repressed in the quiescent bud (Derory *et al*., 2006). The observed reduction in the expression of this gene on bud swelling would indicate the onset of starch mobilization at this developmental stage. Lastly, three CGs putatively encoding for an auxin-repressed protein (AUX-REP), a seed maturation protein (PM23) and a histone (H3) were also selected because of their downregulation in the expression study from quiescent to developing buds. Additional CGs were selected based on their functional role known from other plants. As gibberellins are known to be implied in bud flushing (Or *et al*., 2000; Falusi and Calamassi, 2003), we included Gibberellin 20-oxidase (GA20) and Gibberellin 3-β-hydroxylase (GA3) (Perez-Flores *et al*., 2003; Calvo *et al*., 2004; Israelsson *et al*., 2004), which are involved in the synthesis of GA1, the active form of gibberellins in plants.

In addition, we selected 15 microsatellites loci as 'control markers' that would picture the existing geographic structure for neutral markers (Mariette *et al*., 2002). Genotyping for microsatellites was done on the same individuals as for CGs. The following microsatellite loci were scored on the sampled trees: QrZAG11, QrZAG39, QrZAG96, QrZAG112, QpZAG110, QrZAG5, QrZAG7, QrZAG20, QrZAG65, QrZAG87, QpZAG9, QpZAG15, QpZAG46, QpZAG36 and MSQ13 (Dow *et al*., 1995; Steinkellner *et al*., 1997; Kampfer *et al*, 1998).

DNA extraction and PCR amplification: Total genomic DNA was extracted from buds using a slightly modified protocol of Saghaï-Maroof *et al*. (1984). Primers and PCR conditions to amplify CGs were those used either in Derory *et al*. (2006) or in Casasoli *et al*. (2006). Primer pairs for GA3 and GA20 were designed from published sequences in GenBank (accession numbers AJ006453 and AJ420192, respectively) and their characteristics are presented in Table 1. The amplified fragments of CG covered the domain of the protein or extended to the 3′-UTR (untranslated region) gene region (Table 2).

Microsatellite loci were amplified in four different sets using multiplex PCRs as described in Lepais *et al*. (2006) for two sets of primers and two other sets (MH Pemonge, unpublished results) were used and comprised QpZAG9, QpZAG15, QpZAG46 and QpZAG36 and MSQ13, respectively (Dow *et al*., 1995; Steinkellner *et al*., 1997).

CG sequencing and SSR genotyping: PCR products were cloned into the pCR4-TOPO vector using a TA cloning kit from Invitrogen (Carlsbad, CA, USA). Clones were sequenced using a DYEnamic ET Dye Terminator Kit (Amersham Biosciences Inc., Little Chalfont, Buckinghamshire, UK) on a MegaBACE 1000 automated DNA Sequencer (Amersham Biosciences Inc.). Three different clones of each fragment were sequenced to ensure a high sequencing quality. Microsatellites loci were showed as described in Lepais *et al*. (2006) on a MegaBACE 1000 automated DNA Sequencer (Amersham Biosciences Inc.) using a combination of three different dyes.

SNP detection: The overall sequences per locus were aligned, and polymorphic sites were automatically identified using an informatic pipeline described by Le Dantec *et al*. (2004). Every polymorphic site was then manually verified with CodonCode Aligner v.1.5.1 (CodonCode Corporation, Dedham, MA, USA). In

order to distinguish true polymorphisms from scoring errors, each polymorphic site was visually checked on the chromatograms and further validated for Phred scores (quality threshold above 30).

Diversity and differentiation: For CGs, genetic diversity statistics, including the number of single-nucleotide polymorphisms (SNPs), insertion–deletions (indels), synonymous and replacement mutations, were calculated using DnaSP 4.00.5 software (Rozas *et al*., 2003). Two measures of DNA polymorphism were computed: π, the average number of pairwise nucleotide differences per site in the sample (Nei, 1987), and S, the number of segregating nucleotide sites. These parameters were computed with DnaSP considering SNPs and indels, at three different levels: the whole sequenced region, non-coding regions (including introns, 3′ and 5′-UTRs) and coding regions. The number of haplotypes and the haplotypic diversity (Hd) was also calculated using the DnaSP software (Rozas *et al*., 2003). Tajima's *D*-statistics (Tajima, 1989) and Fu and Li's *D*-statistics and F-statistics (Fu and Li, 1993) were estimated to detect deviations from neutrality using the DnaSP software. Differentiation indexes were estimated, on the whole sequence and at each polymorphic site as F_{ST} (Weir and Cockerham, 1984), using Arlequin ver 3.01 (Excoffier *et al*., 2005). We further tested whether SNPs deviated significantly from neutral expectations using the Bayesian approach for detecting outlier loci (Beaumont and Balding, 2004). The method is based on the posterior distribution of a locus-specific parameter (α_i) related to population differentiation. Population differentiation is modeled as the co-ancestry coefficient by $F_{ST_{ij}}/(1-F_{ST_{ij}})$, where $F_{ST_{ij}}$ is the probability that two randomly chosen alleles of locus i have a common ancestor within that the population j. The model accounts for different F_{ST} values for different subpopulations. Hence, $F_{ST_{ij}}$ is subdivided into a locus-specific effect (α_i), due to mutation or selection, and into a population-specific effect (β_j), such as population sizes and/or migration rates, and γ_{ij} an interaction effect between a population and a locus, for example, local adaptation of a given allele within a given population. The three components are estimated by logistic regression as: $\mathrm{Log}(F_{ST_{ij}}/(1-F_{ST_{ij}}) = \alpha_i + \beta_j + \gamma_{ij}$. Under neutral expectations, α-values are expected to be 0, whereas positive values would indicate directional selection and negative values balancing selection. At $P = 5\%$, α_i is significantly positive (directional selection) if its 2.5% quantile is positive, and is significantly negative (balancing selection) if its 97.5% quantile is negative. Similarly, a significant positive effect of β_j (reduced population size or gene flow) would be detected if 2.5% of its posterior distribution is positive, and negative (very large population size or gene flow) if its 97.5 quantile is negative. Posterior distributions of the parameters were computed with BAYESFST using MCMC simulations (Beaumont and Balding, 2004). The mean and standard deviation of prior distributions were set by default at 0 and 1 for α_i, and -2 and 1.8 for β_j, respectively. The Bayesian detection of outlier loci (Beaumont and Balding, 2004) was preferred to the frequentist approach (Beaumont and Nichols, 1996), as the latter is less robust when sample sizes are low and does not account for population effects on F_{ST} and is limited to the infinite allele mutation (and stepwise mutation model) model that is not appropriate for biallelic loci as SNPs (Eveno *et al*, 2008).

For SSRs, the following diversity and differentiation statistics were assessed: H_O (observed heterozygosity), H_S (mean population diversity), H_T (overall diversity), F_{ST} (differentiation) and F_{IS} (fixation index) (Nei, 1987). H_O, H_S and H_T were estimated using Fstat v2.9.3 software (Goudet, 2001), and F_{ST} and F_{IS} were estimated using Arlequin ver 3.01 (Excoffier *et al*., 2005).

Mapping CGs and detection of QTLs for bud burst

Plant material and assessments of bud burst: QTLs for bud burst were detected within the full-sibling family 3P*A4 comprising 278 full siblings (Scotti-Saintagne *et al*., 2004). Full-sibling seedlings were sown in spring 1995 and raised in a seed bed (field test 1) in the nursery of the INRA Research Station at Pierroton (near Bordeaux France). The full siblings were further vegetatively propagated during two successive campaigns in 1997 (6.2 cuttings per full sibling on average) and 1998 (10.6 cuttings per full sibling), and cuttings were installed in two experimental plantations in the fall of 1998 (field test 2) and spring 2000 (field test 3). Details about the two experimental plantations are given in Scotti-Saintagne *et al*. (2004). Both field tests 2 and 3 are located at the INRA Froot Tree Domain located at Bourran (near the city of Agen in the southwest of France).

Bud burst was recorded in 1999 in the seed bed (field test 1) and over 6 years in each of the two cutting plantations (2000, 2003, 2004, 2005, 2006 and 2007 for field test 2; and 2002, 2003, 2004, 2005, 2006 and 2007 for field test 3). In total 13 observations of bud burst were made over the three test sites. Depending on the number of cuttings available per full sibling, the number of observations per full sibling reached a maximum of 224 for each full sibling of the mapping pedigree. Assessments of bud burst consisted in recording bud development scores (ranging from 0 to 5) every 2–3 days starting when the first tree reached score 1. These data used for the detection of QTLs were the number of days necessary to reach stage 3.

QTL detection and mapping of CGs: Multiple Interval Mapping (Jansen and Stam, 1994; Zeng, 1994) was used to identify QTLs, their position (L in cM) and their contribution to the variance of the trait (percentage of the variation in %) using the MultiQTL V2.5 software package (http://www.multiqtl.com). Confidence intervals at 95% were estimated after 1000 bootstrap resampling (Visscher *et al*., 1996). The position of a QTL was defined as the range between the location of the QTL found with the observed data and the mean of 1000 bootstrap samples. Confidence intervals at 95% were estimated after 1000 bootstrap resampling and were positioned relative to the mean of the bootstrap samples.

The QTL detection was done separately on each parental map in two different steps. The one-QTL (per linkage group (LG)) model was first used and then followed by a model assuming two QTLs. When both models showed the existence of QTLs for a given LG, the two-QTL model was preferred to the one-QTL model when the presence of two QTLs was significant compared with the null hypothesis of no QTL

(H2 > H0) and with the null hypothesis of one QTL (H2 > H1). Otherwise, the one-QTL model was preferred.

To declare the presence of a QTL, type (I) error was set to 0.05 at the genome level. Type I errors obtained at the chromosome level (Doerge and Churchill, 1996) by MultiQTL were transformed at the genome level following the method of Scotti-Saintagne *et al*. (2004). An overall analysis using all 13 observations of bud burst was conducted using the multi-environment model. By bulking the information over all field tests and years, the multi-environment approach allows to improve the QTL detection by decreasing the environmental variance.

CGs were mapped using either SSCP (single-strand conformation polymorphism; Orita *et al*., 1989) or SNP variation (Casasoli *et al*., 2006). A sample of 57–135 full siblings of the mapping pedigree was genotyped for the ESTs of the CGs. MapMaker V.2 software (Lander *et al*., 1987) was used for linkage analysis using a LOD threshold of 6.0 as a grouping criterion and the Kosambi (1944) function to estimate genetic distances.

Results

Nucleotide diversity

For eight out of the nine studied loci, the sequence of 42–52 samples was obtained, except for GA20, where only 21 sequences were obtained. The regions analyzed covered a total number of 4.6 kb, corresponding to 2.7 kb of coding sequence and 1.9 kb of non-coding sequence (introns + 3′-UTR) (Table 2). The fragment sizes ranged from 315 bp for YSL1 to 1104 bp for PM23. A total of 107 SNPs, 18 singletons and 23 indels were detected. In total, 148 mutations (including SNPs, singletons and indels) were detected, suggesting that on average there is one mutation every 31 bp. The indels varied in size from 1 to 28 bp, and were mostly located in intronic regions. The average total nucleotide diversity (6.15×10^{-3}) varied from 1.09×10^{-3} for locus GA3 to 14.7×10^{-3} for locus GALA (Table 3). The level of diversity in non-coding regions (8.15×10^{-3}) is two-fold higher than in coding regions (4.81×10^{-3}). In coding sequences, synonymous regions were more variable than replacement regions (12.67×10^{-3} vs 2.40×10^{-3}), except for locus GA20. Interestingly, the ratio of replacement and silent nucleotide diversity was high for PM23 (0.51) and ASI (0.39), and notably different from the other seven genes. Average haplotypic diversity was 0.76, corresponding on average to 13.2 haplotypes per locus for the sample sizes analyzed (Table 3). With the exception of one locus (QrZag112), total genetic diversity of microsatellites varied between 0.70 and 0.94, and the mean number of alleles between 10 and 20 (data not shown), as previously reported in other genetic surveys (Mariette *et al*., 2002).

Population differentiation

The overall differentiation for CGs and microsatellites was extremely low (Tables 4a and 4b) and highly variable across loci due to the low sample size. There is a slightly higher differentiation for genes, mainly due to DAG2. Differentiation computed among geographical or phenological groups remained at extremely low levels, regardless if it was microsatellites or CGs. We explored in more details the comparison of geographical with phenological differentiation by computing F_{ST} at each single SNP for CGs and at each allele for the microsatellites, and comparing the distribution of F_{ST} values between the two categories of markers and clustering of populations.

The mean value of F_{ST} for SNPs and microsatellite alleles was located at 0 (Figure 1), and their distributions were slightly skewed to positive values. The overall range of distribution was larger in the case of SNPs than in the case of SSR alleles, as a result of the lower overall diversity of SNPs in comparison with SSR alleles. There is no trend of larger differentiation between phenological groups than between geographical groups. However, a detailed analysis of F_{ST} values of SNPs showed that 15 SNPs of GALA (out of 18) exhibited values larger than 0.07 among phenological groups and only 2 SNPs larger values than 0.07 among geographical groups. The Bayesian F_{ST} test did not show any significant SNP (locus-specific effect (α_i)) deviating from neutral expectations among the nine populations. Two different runs of BAYESFST were conducted. This first analysis was based

Table 3 Patterns of nucleotide variation

Locus	*N*	*SNP*												*No. of haplotypes*	*Hd*
		Total				*Non-coding*		*Coding*							
								Total		*Synonymous*		*Replacement*			
		S	π_{total}	π_{silent}	$\pi_{replacement}/\pi_{silent}$	*S*	π	*S*	π	*S*	π	*S*	π		
GALA	44	25	14.7	20.66	0.17	17	17.4	8	11.3	6	39.01	2	3.46	25	0.916
ASI	44	22	12.68	23.8	0.39	0	—	22	12.68	9	23.8	13	9.27	21	0.902
DAG2	42	2	1.56	5.43	0.06	0	—	2	1.56	1	5.43	1	0.34	3	0.528
H3	52	7	3.56	4.65	0.09	4	4.2	3	2.3	2	8.34	1	0.41	12	0.819
PM23	50	44	8.2	9.49	0.51	33	9.3	11	6.2	4	10.45	7	4.84	30	0.959
AUX-REP	49	10	2.93	5.24	0	9	7	1	0.2	1	0.68	0	0	9	0.642
YSL1	48	8	8.24	24.7	0.11	0	—	8	8.24	6	24.7	2	2.73	9	0.867
GA3	46	3	1.09	1.88	0	2	1.9	1	0.4	1	1.65	0	0	5	0.605
GA20	21	4	2.39	5.09	0.11	3	9.1	1	0.4	0	0	1	0.58	5	0.6
Total	396	125	6.15	11.22	0.16	68	8.15	57	4.81	30	12.67	27	2.4	13.22	0.76

Abbreviations: Hd, haplotypic diversity; π, nucleotide diversity ($\times 10^{-3}$); *S*, number of polymorphic sites; SNP, single-nucleotide polymorphism.
Silent sites include synonymous polymorphisms in coding regions and polymorphisms in non-coding regions.

Table 4a Population differentiation of candidate genes (F_{ST} values)

Candidate gene	*Overall F_{ST}*[a]	*Geo F_{ST}*[b]	*Phe F_{ST}*[c]
GALA	0.037	0.033	0.047
ASI	−0.026	**0.044**	−0.033
DAG2	0.180	0.084	−0.078
H3	−0.007	0.029	−0.038
PM23	−0.021	−0.003	0.017
AUX-REP	0.014	0.042	0.020
YSL1	0.062	0.008	0.002
GA3	0.012	0.066	0.010
GA20	0.124	−0.134	0.040
CG mean	0.042	0.019	−0.001

Abbreviation: CG, candidate gene.
The values in bold indicate significant F_{ST} value ($P<0.05$).
[a]Overall F_{ST}, differentiation among all nine populations.
[b]Geo F_{ST}, differentiation between geographical groups of populations (see Table 1 for composition of groups).
[c]Phe F_{ST}, differentiation between phenological groups of populations (see Table 1 for composition of groups).

Table 4b Population differentiation of microsatellites (F_{ST} values)

Locus	*Overall F_{ST}*[a]	*Geo F_{ST}*[b]	*Phe F_{ST}*[c]
QpZAG15	−0.012	0.001	−0.031
QpZAG46	0.013	−0.013	0.045
QpZAG9	0.023	0.018	−0.034
QpZAG110	0.016	−0.008	0.008
QrZAG11	−0.041	0.007	−0.008
QrZAG112	−0.029	**0.068**	0.004
QrZAG39	0.019	0.007	−0.041
QrZAG96	0.055	−0.038	0.013
MsQ13	−0.025	0.013	−0.028
QpZAG36	0.007	−0.001	0.027
QrZAG20	−0.005	0.012	−0.005
QrZAG5	0.030	**0.071**	0.020
QrZAG65	0.013	0.012	0.038
QrZAG7	0.050	0.000	−0.013
QrZAG87	0.039	0.021	0.034
SSR mean	0.013	0.011	0.012

[a]Overall F_{ST}, differentiation among all nine populations.
The values in bold indicate significant F_{ST} value ($P<0.05$).
[b]Geo F_{ST}, differentiation between geographical groups of populations (see Table 1 for composition of groups).
[c]Phe F_{ST}, differentiation between phenological groups of populations (see Table 1 for composition of groups).

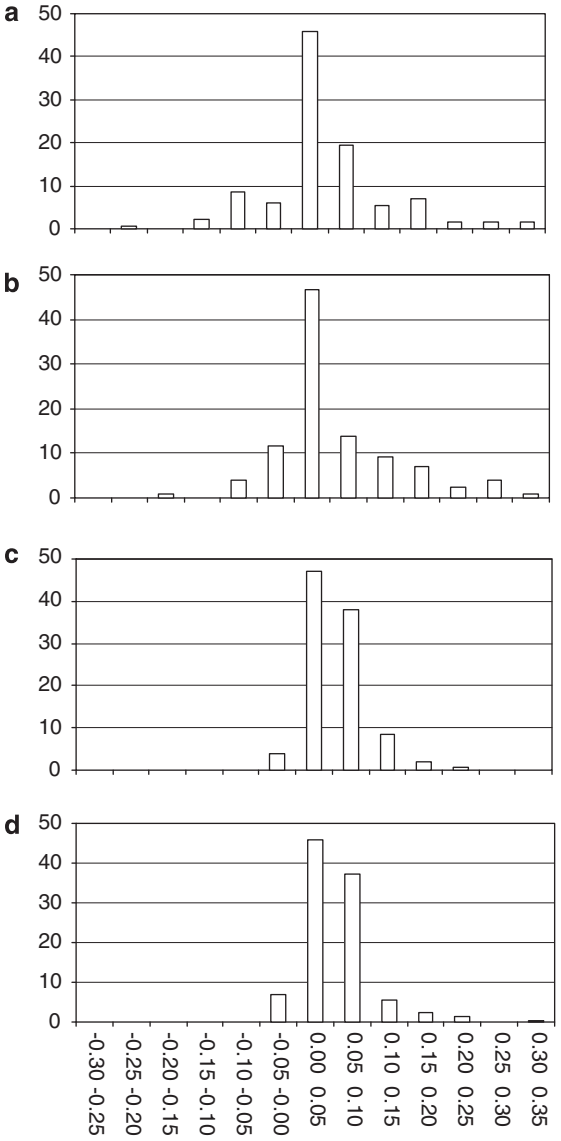


Figure 1 Distribution of F_{ST} values. (**a**) Candidate genes. Differentiation among geographic groups. (**b**) Candidate genes. Differentiation among phenological groups. (**c**) Microsatellites. Differentiation among geographic groups. (**d**) Microsatellites. Differentiation among phenological groups.

on 10 000 drawings from the posterior distribution, and the second on 2000 drawings. Both came to the same conclusion that none of the SNPs did show any significant deviation from neutral expectations. The Bayesian F_{ST} test was then further carried out across the phenological and geographical groups using the same methodology (two independent analysis). No SNP showed either positive or negative significant deviations. However, BAYESFST runs indicated significant negative population-specific effects, for example, for which 97.5 of the posterior distribution was lower then the prior mean −2. Significant negative values of β_i were obtained for all populations except Cochem, Klostermarienberg and Mölln. We further did the computation by changing the prior standard deviation of β_j to 1 and 3, and obtained the same results. None of the analysis did indicate any significant γ-effect (an interaction between a gene and a population).

As the F_{ST} analysis strongly suggested a very low population differentiation, we considered the overall sample of trees as belonging to one single random mating population. Under these hypotheses, we calculated Tajima's *D*-statistic and Fu and Li's *D*-statistics and F-statistics. None of these tests showed any significant deviation from neutral expectations (data not shown). As an example, among the nine genes, four showed positive Tajima's *D*-test, four negative and one close to zero, but none was significant. AUX-REP and ASI were the two genes showing Tajima's *D*-tests closest to significance and both were negative, suggesting a potential directional selection.

QTL mapping of bud burst

The QTL detection resulted in 19 QTLs located on 11 female LGs and as many on 11 male LGs (Table 5). As not all the QTLs are located in homologous regions in the

Table 5 Distribution of QTLs of bud burst

Linkage group	*Number of QTLs*	*LOD score*	*PEV*[a] *min–max*	*QTL1*[b]	*QTL2*[b]
1 F	2	40	1.6–11.8	17.7	38.9
1 M	2	68	0.6–12.3	4.2	47.0
2 F	2	178	4.7–16.1	11.1	43.6
2 M	2	100	0.3–15.0	21.5	101.35
3 F	2	58	0.2–13.3	13.7	23.2
4 F	2	51	0.9–12.0	23.2	40.0
4 M	2	75	1.2–7.9	0	75.6
5 F	2	62	1.0–10.9	23.9	44
5 M	2	101	1.8–15.4	19.0	52.3
6 F	2	36	0.0–6.2	27.2	55.3
6 M	2	51	1.5–21.2	23.7	49.5
7 F	2	64	1.8–10.4	41.3	79.4
7 M	1	18	0–3.4	18.8	
8 M	2	83	2.5–11.0	7.4	31.9
9 F	1	32	0.9–6.7	8.3	
9 M	2	195	1.0–31.1	8.7	25.2
10 F	2	86	1.3–10.5	25.0	70.4
10 M	2	70	0.8–9.7	12.8	35.2
11 F	1	33	0.7–5.3	62.9	—
11 M	1	33	0.2–9.4	54.3	—
12 F	1	13	0–3.0	18.7	—
12 M	1	45	0.5–7.8	0.6	—

Abbreviations: F, female; M, male; LOD; PEV, percentage of variation; QTL, quantitative trait locus.
[a]PEV explained by the QTLs (minimum and maximum value assessed over 13 site*year observations).
[b]Position of QTL (in cM).

male and female LGs, one could conclude that there are more than 19 QTLs controlling bud burst in oaks. Standard deviations of QTL positions were substantially reduced in comparison with the previous analysis (Scotti-Saintagne *et al*., 2004; Casasoli *et al*., 2006), as a result of the multiyear and multisite observations taken into account by the multi-environment option. On average, the standard deviation amounted to 5 cM and reached <0.02 cM in the extreme cases (LG2).

We further monitored the distribution of the contribution of the 38 QTLs across all 13 observations (sites*-years) representing a total of 286 QTL detections, by estimating the percentage of the variation of the mean clonal value explained by the QTL for each observation (Figure 2). The distribution of the percentage of the variation values follows an L-shaped curve. Of the 286 detections, 265 (about 90%) corresponded to QTLs contributing for 12% to the variation of bud burst. The tail of the distribution, for example, the 21 detections with contributions larger than 12%, comprised mainly QTLs located at three LGs (8 detections for LG2, 6 detections for LG9 and 3 detections for LG5). Eight CGs were polymorphic in the two parents of the pedigree and could be mapped on the existing genetic map. Five of them located within the confidence intervals of QTLs for bud burst (using the multi-environment option, Figure 3). Interestingly, these five were located within QTL regions exhibiting the largest contribution to bud burst: LG9 (ASI), LG2 (YSL1) and LG5 (AUX-REP, GA3).

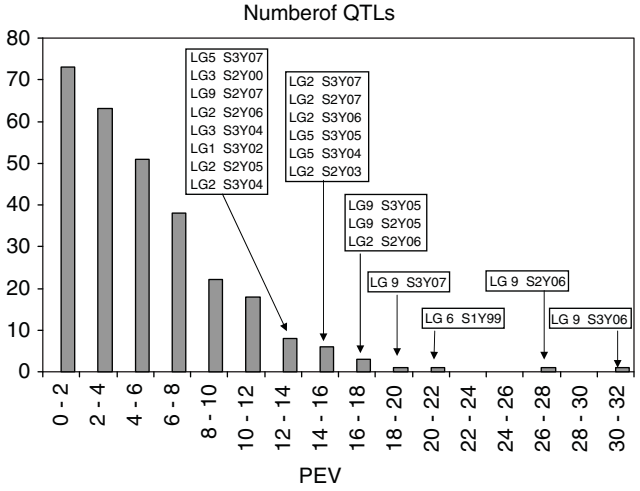


Figure 2 Distribution of the percentage of the variation (PEV) values (percentage of variation explained by the quantitative trait loci (QTLs)) for all QTLs. Tags for QTLs located at the tail of the distribution indicate the linkage groups where the QTL is located and the field Test and Year of the bud burst assessment. For example, LG5 T3Y07 stands for Linkage Group5 (LG5), field Test3 Year 2007.

Discussion

Diversity of CGs of bud burst vs diversity of neutral markers

This study is the second reporting on nucleotide diversity in genes of oaks. Indeed, a recent paper on a Asian oak species (Quang *et al*., 2008), and using genes that were obtained from the same cDNA library than ours (Casasoli *et al*., 2006; Derory *et al*, 2006), made a population survey in 11 genes of *Quercus crispula*. Interestingly, the overall diversity was of similar magnitude in their study ($\pi = 6.93 \times 10^{-3}$) than in ours ($\pi = 6.15 \times 10^{-3}$). These figures are lower than those previously reported on *Populus tremula* (11.1×10^{-3}) (Ingvarsson, 2005), but higher than in pine species, *Pinus taeda* (3.98×10^{-3}) (Brown *et al*., 2004), *Pinus pinaster* (2.41×10^{-3} in Pot *et al*., 2005; 5.51×10^{-3} in Eveno *et al*., 2008) or *Pinus sylvestris* (1.4×10^{-3}) (Dvornyk *et al*., 2002), or other conifers (Gonzalez-Martinez *et al*., 2006; Savolainen and Pyhäjärvi, 2007). Larger diversity in broad-leaves than in conifers is also observed when the comparative analysis is conducted at the level of silent polymorphism: the level of diversity in oaks ($\pi_{silent} = 11.2 \times 10^{-3}$) is higher than earlier reports in pines ($\pi_{silent} = 7.7 \times 10^{-3}$ in *P. sylvestris*, Wachowiak *et al*., 2009; $\pi_{silent} = 8.6 \times 10^{-3}$ in *P. pinaster*, Eveno *et al*., 2008) Although the number of genes is still low, our results confirm earlier findings obtained with other marker systems, suggesting that oak is highly variable species (Kremer and Petit, 1993; Mariette *et al*., 2002). As oaks are outcrossing species exhibiting extensive pollen flow, they compose large populations that contribute to maintain high levels of diversity. However, these results should be taken with caution, as this overall picture overshadows the very large variation observed among genes (15-fold variation from the lower to the largest polymorphic gene) and within genes (between silent and replacement regions). Genes showing the largest diversity were GALA and ASI, and they also showed among the largest ratio of replacement vs silent diversity (Table 3). These two genes are downregulated during the transition between quiescent to developing buds (Derory *et al*., 2006), and their functional role has been investigated in annual plants in germinating seeds, suggesting that similar metabolic pathways may be

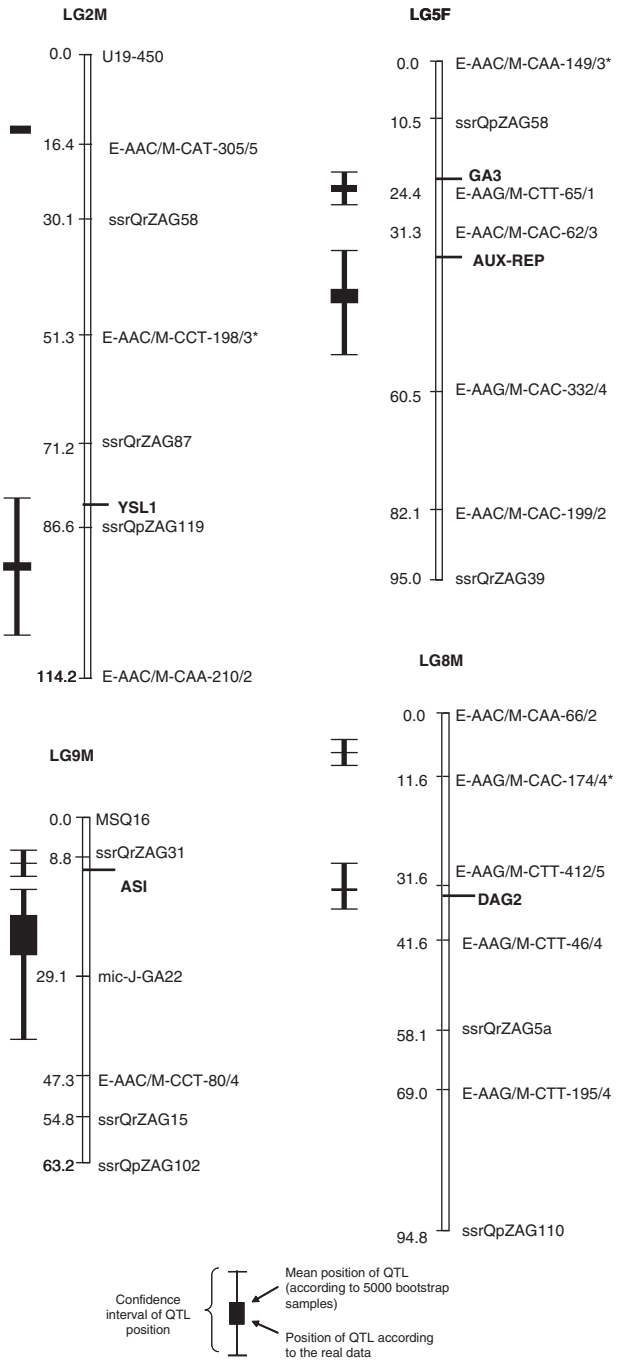


Figure 3 Distribution of quantitative trait loci (QTLs) of bud burst and candidate genes along linkage groups. This figure represents the linkage groups for which QTLs were detected that collocate with at least one of the eight mapped candidate genes. LG2M, linkage group2 male.

active in developing buds and germinating seeds. In *A. thaliana* (Taji *et al*., 2002), galactinol synthase contributes to the accumulation of galactinol under abiotic stress conditions (drought, salinity and cold) and acts as osmoprotectant of the seed. In barley, the expression of ASI in germinating seeds indicates a potential role in defense against pathogens of the developing seed and embryo. The gene is also involved in the inhibition of starch hydrolysis in the peripheral tissues of the seed (Furtado *et al*., 2003). Therefore, both genes appear to have a multifunctional role, which may contribute to the maintenance of a larger diversity due to balancing selection in natural conditions.

Differentiation of CGs vs trait differentiation

Overall population differentiation of expressional and functional CGs of bud burst was of the same level than genomic microsatellites. Furthermore, differentiation among phenological groups was not larger than among geographic groups, and the distributions of the single SNP F_{ST} values for the two subdivisions (phenological vs geographical) overlapped completely. Lastly, F_{ST} values of SNPs did not deviate significantly from neutral expectations based on the Bayesian F_{ST} test. Our results are in line with recent data obtained in various species, where differentiation of CGs was compared with that of neutral markers and of the target trait (in *P. pinaster*, Eveno *et al*., 2008; in *P. sylvestris*, Pyhäjärvi *et al*., 2008; in *Picea abies*, Heuertz *et al*., 2006; and in *P. tremula*, Hall *et al*., 2007; Luquez *et al*., 2007). In all these case studies, the mean differentiation of genes was of the same level than that of the neutral markers and far less than differentiation of the target trait. Recent theoretical developments suggested that this discrepancy may be related to the multi-locus structure of a complex trait. Indeed, Latta (2003) and Le Corre and Kremer (2003) showed that phenotypic differentiation among populations (Q_{ST}) is driven by two major components: covariances among allelic effects at the different loci controlling the trait and variances of allelic effects at each locus. Only the latter component is dependent on F_{ST} of the genes, whereas the former is generated by allelic associations between loci. Interestingly, these authors also showed that the contribution of these components to the phenotypic differentiation of the trait is unbalanced under a wide range of evolutionary scenarios. For tree species, exhibiting extensive gene flow and existing in large populations, differentiation at the trait level is mostly created by allelic associations, rather than changes in allelic frequencies. In other words, diversifying selection, creating population differentiation, is capturing first beneficial allelic associations distributed among the loci contributing to the trait, before modifying allele frequencies at these genes. Hence, strong phenotypic differentiation may coexist with very low differentiation at the genes controlling the trait. Furthermore, the discrepancy is even inflated when the trait is controlled by a large number of loci. Indeed, more loci offer more opportunities for allelic associations to build up. We identified in this study at least 19 QTLs (Table 5) that may be contributing to the trait in only one full-sibling cross. There might be more QTLs in natural populations, suggesting that the high heritability of bud burst in oaks (Scotti-Saintagne *et al*., 2004) may result from the summing of allelic effects over many loci, hence increasing the relative contribution of allelic associations to the overall differentiation of the bud burst. An interesting extension of this study would be to test whether intergenic allelic associations among the CGs were generated by diversifying selection and were responsible for the large phenotypic differentiation that was

observed in the provenance tests as suggested by simulations by Latta (1998) and Le Corre and Kremer (2003). Intergenic allelic associations could be detected by calculating multi-locus F_{ST} taking into account intergenic disequilibria (Kremer *et al*., 1997) but would require larger population samples than those used in this study.

Searching for CGs of bud burst

Our investigations lead to paradoxical conclusions about the link between CGs diversity and bud burst variation. On the one hand, diversity and differentiation statistics show that expressional and functional CGs actually behave as neutral markers, whereas in the same time the collocation of CGs and QTLs was confirmed with stronger confidence in comparison with earlier experiments (Casasoli *et al*., 2006). Mapping positions of five CGs were located within the confidence interval of the three strongest QTLs. There are at least four interpretations to these contrasting results: (1) expressional and functional CGs that were selected for this study are not related to the variation of bud burst; (2) QTL mapping and collocation of CGs is still imprecise; (3) as suggested in the previous paragraph, intergenic allelic associations may be the main component of the phenotypic differentiation of bud burst; (4) the causal mutation of bud burst variation is located outside the genomic region that was explored in this study. With regard to the first and second interpretations, we recently monitored the level of expression of ASI that is located within the confidence interval of the strongest QTL of bud burst within the same mapping population and identified a strong eQTL (expression QTL) located at the same spot, suggesting that the level of expression of ASI is correlated to earliness of flushing (data not published), and reinforcing its putative feature as CGs. Yet, other genes located in the same region may as well be the main source of variation of bud burst. The mean range of the confidence interval of a QTL in this study varies from 2 to 20 cM (Figure 3). Given that the physical size of the oak genome amounts to $\sim$740 Mb/C and the genetic size to 1200 cM (Kremer *et al*., 2007), 1 cM roughly represents 600 000 bp. Hence, the confidence interval of a QTL may comprise from several tens to hundreds genes, among which the gene causing the observed variation of bud burst is. Lastly, we explored nucleotide diversity within a limited region of the different CGs. Although the targeted region derived from EST sequences overlaps the active domain of the corresponding protein (Table 2), no sequence data were available within the promoting regions. As linkage disequilibrium decays quite rapidly in oaks (Quang *et al*., 2008), chances for detecting any causal association locate several hundred base pairs away from the recorded mutation are extremely low.

In conclusion, our results comparing nucleotide diversity of CGs and their collocation with QTL on the genetic maps lead to implement additional investigations in two different directions to confirm their role in bud burst variation. First the exploration of nucleotide diversity needs to be extended across the full length of the genes, including the promoter regions. Second, SNP frequencies should be monitored in a larger set of populations in order to assess for intergenic allelic associations among populations that may account for the overall population differentiation of CGs.

Conflict of interest

The authors declare no conflict of interest

Acknowledgements

We thank Alexis Ducousso for providing the data of the provenance test of *Q. petraea* and his advice for selecting the populations. We acknowledge JM Louvet for his help during the field collections. This work was funded by the European Union (FP5 project TREESNIPS: QLK3-CT-2002-01973).

References

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, Balding DJ (2004). Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol* **13**: 969–980.

Bechtold N, Ellis J, Pelletier G (1993). In planta Agrobacterium mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. *C R Acad Sci III* **316**: 1194–1199.

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.

Calvo AP, Nicolas C, Nicolas G, Rodriguez D (2004). Evidence of a cross-talk regulation of a GA 20-oxidase (FsGA20ox1) by gibberellins and ethylene during the breaking of dormancy in *Fagus sylvatica* seeds. *Physiol Plant* **120**: 623–630.

Casasoli M, Derory J, Morera-Dutrey C, Brendel O, Porth I, Guehl JM *et al*. (2006). Comparison of quantitative trait loci for adaptive traits between oak and chestnut based on an expressed sequence tag consensus map. *Genetics* **172**: 533–546.

Chuine I, Beaubien EG 2001)). Phenology is a major determinant of tree species range. *Ecol Lett* **4**: 500–510.

Chuine I, Cour P (1999). Climatic determinants of budburst seasonality in four temperate-zone tree species. *New Phytol* **143**: 339–349.

Derory J, Leger P, Garcia V, Schaeffer J, Hauser MT, Salin F *et al*. (2006). Transcriptome analysis of bud burst in sessile oak (*Quercus petraea*). *New Phytol* **170**: 723–738.

Doerge RW, Churchill GA (1996). Permutation tests for multiple loci affecting a quantitative character. *Genetics* **142**: 285–294.

Dow BD, Ashley MV, Howe HF (1995). Characterization of highly variable (GA/CT)*n* microsatellites in the bur oak, *Quercus macrocarpa*. *Theor Appl Genet* **91**: 137–141.

Ducousso A, Guyon JP, Kremer A (1996). Latitudinal and altitudinal variation of bud burst in western populations of sessile oak (*Quercus petraea* (Matt.) Liebl.). *Ann Sci Forest* **53**: 775–782.

Ducousso A, Louvet JM, Jarret P, Kremer A (2005). Geographic variations of sessile oak in French provenance tests. In: Rogers R, Ducousso A, Kanazashi A(eds). *Proceedings of the Joint Meeting of IUFRO working groups genetic of oaks and improvement and silviculture of oaks* FFPRI (Forestry and Forest Products Research Institute) Scientific Meeting Report 3 Tsukuba: Japan. pp 128–138.

Dvornyk V, Sirvio 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.

Eveno E, Collada C, Guevara MA, Léger V, Soto A, Díaz L *et al*. (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.

Falusi R, Calamassi C (2003). Dormancy of *Fagus sylvatica* L. buds. III. Temperature and hormones in the evolution of dormancy in one-node cuttings. *Plant Biosyst* **137**: 185.

Fu YX, Li WH (1993). Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.

Furtado A, Henry R, Scott K, Meech S (2003). The promoter of the ASI gene directs expression in the maternal tissues of the seed in transgenic barley. *Plant Mol Biol* **52**: 787–800.

Gonzalez-Martinez SC, Krutovsky K, D. Neale DB (2006). Forest tree population genomics and adaptive evolution. *New Phytol* **170**: 227–238.

Goudet J (2001). Fstat, a program to estimate and test gene diversities and fixation indices. Version 2.9.3.2 (Available at http://www2.unil.ch/popgen/softwares/fstat.htm).

Gualberti G, Papi M, Bellucci L, Ricci L, Bouchez D, Camilleri C *et al.* (2002). Mutations in the Dof zinc finger genes DAG2 and DAG1 influence with opposite effects the germination of *Arabidopsis* seeds. *Plant Cell* **14**: 1253–1263.

Hall D, Luquez V, Garcia VM, St Onge KR, Jansson S, Ingvarsson PK (2007). Adaptive population differentiation in phenology across a latitudinal gradient in European Aspen (*Populus tremula*, L.): a comparison of neutral markers, candidate genes, and phenotypic traits. *Evolution* **61**: 2849–2860.

Heuertz M, De Paoli E, Källmann T, Larsson H, Jurman I, Morgante M *et al.* (2006). Multilocus patterns of nucleotide diversity, linkage disequilibrium, and demographic history of Norway Spruce. *Genetics* **174**: 2095–2105.

Ingvarsson PK (2005). Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European Aspen (*Populus tremula* L., Salicaceae). *Genetics* **169**: 945–953.

Israelsson M, Mellerowicz E, Chono M, Gullberg J, Moritz T (2004). Cloning and overproduction of gibberellin 3-oxidase in hybrid Aspen trees. Effects on gibberellin homeostasis and development. *Plant Physiol* **135**: 221–230.

Jansen RC, Stam P (1994). High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136**: 1447–1455.

Kampfer S, Lexer C, Glössl J, Steinkellner H (1998). Characterization of (GA)n microsatellite loci from *Quercus robur*. *Hereditas* **129**: 183–186.

Kosambi DD (1944). The estimation of map distance from recombination values. *Ann Eugen* **12**: 172–175.

Kremer A, Petit RJ (1993). Gene diversity in natural populations of oak species. *Ann Sci Forest* **50**: 186s–203s.

Kremer A, Zanetto A, Ducousso A (1997). Multilocus and multitrait measures of differentiation for gene markers and phenotypic traits. *Genetics* **145**: 1229–1241.

Kremer A, Kleinschmit J, Cottrell J, Cundall EP, Deans JD, Ducousso A *et al.* (2002). Is there a correlation between chloroplastic and nuclear divergence, or what are the roles of history and selection on genetic diversity in European oaks? *Forest Ecol Manage* **156**: 75–87.

Kremer A, Le Corre V, Petit R J, Ducousso A (2009). Historical and contemporary dynamics of adaptive differentiation in European oaks. In: DeWoody A, Bickham J, Michler C, Nichols K, Rhodes G, Woeste K (eds). *Molecular Approaches in Natural Resource Conservation*. Cambridge University Press (in press).

Kremer A, Casasoli M, Barreneche T, Bodénès C, Sisco P, Kubisiak T *et al.* (2007). Fagaceae: comparative genetic mapping in Fagaceae. In: Kole CR (ed). *Genome Mapping and Molecular Breeding*, Vol. 5. Forest Trees: Springer-Verlag, Berlin. pp 161–187.

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE *et al.* (1987). MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**: 174–181.

Latta RG (1998). Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. *Am Nat* **151**: 283–292.

Latta RG (2003). Gene flow, adaptive population divergence and comparative population structure across loci. *New Phytol* **161**: 51–58.

Le Corre V, Kremer A (2003). Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. *Genetics* **164**: 2005–2019.

Le Dantec L, Chagne D, Pot D, Cantin O, Garnier-Gere P, Bedon F *et al.* (2004). Automated SNP detection in expressed sequence tags: statistical considerations and application to Maritime Pine sequences. *Plant Mol Biol* **54**: 461–470.

Lepais O, Léger V, Gerber S (2006). Short note: high throughput microsatellite genotyping in oak species. *Silvae Genet* **55**: 238–240.

Luquez V, Hall D, Albrectsen B, Karlsson J, Ingvarsson PK, Jansson S (2007). Natural phenological variation in Aspen (*Populus tremula*): The Swedish Aspen Collection. *Tree Genet Genomes* **4**: 279–292.

Mariette S, Cottrell J, Csaikl UM, Goikoechea P, Konig A, Lowe AJ *et al.* (2002). Comparison of levels of genetic diversity detected with AFLP and microsatellite markers within and among mixed *Q-petraea* (Matt.) Liebl. and *Q-robur* L. stands. *Silvae Genet* **51**: 72–79.

Menzel A, Fabian P (1999). Growing season extended in Europe. *Nature* **397**: 659.

Nei M (1987). *Molecular Evolutionary Genetics*. Columbia University, Press: New York, USA.

Nielsen PK, Bonsager B, Fukuda K, Svensson B (2004). Barley α-amylase/subtilisin inhibitor: structure, biophysics and protein engineering. *Biochim Biophys Acta* **1696**: 157–164.

Or E, Belausov E, Popilevsky I, Ben Tal Y (2000). Changes in endogenous ABA level in relation to the dormancy cycle in grapevines grown in a hot climate. *J Hortic Sci Biotechnol* **75**: 190–194.

Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci USA* **86**: 2766–2770.

Penuelas J, Filella I (2001). Phenology: responses to a warming world. *Science* **294**: 793–795.

Perez-Flores L, Carrari F, Osuna-Fernandez R, Rodriguez MV, Enciso S, Stanelloni R *et al.* (2003). Expression analysis of a GA 20-oxidase in embryos from two sorghum lines with contrasting dormancy: possible participation of this gene in the hormonal control of germination. *J Exp Bot* **54**: 2071–2079.

Pflieger S, Lefebvre V, Causse M (2001). The candidate gene approach in plant genetics: a review. *Mol Breed* **7**: 275–291.

Pot D, McMillan L, Echt C, Le Provost G, Garnier-Gér P, Cato S *et al.* (2005). Nucleotide variation in genes involved in wood formation in two pines species. *New Phytol* **167**: 101–112.

Pukacka S, Wojkiewicz E (2002). Carbohydrate metabolism in Norway maple and sycamore seeds in relation to desiccation tolerance. *J Plant Physiol* **159**: 273–279.

Pyhäjärvi T, García-Gil R, Knürr T, Mikkonen M, Wachowiak W, Savolainen O (2008). Demographic history has influenced nucleotide diversity in European *Pinus sylvestris* populations. *Genetics* **177**: 1713–1724.

Quang ND, Ikeda S, Harada K (2008). Nucleotide variation in *Quercus crispula* Blume. *Heredity* **101**: 166–174.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984). Ribosomal DNA Spacer-Length Polymorphisms in Barley: Mendelian Inheritance, Chromosomal Location, and Population Dynamics. *Proc Natl Acad Sci USA* **81**: 8014–8018.

Savolainen O, Pyhäjärvi T (2007). Genomic diversity in forest trees. *Curr Opin Plant Biol* **10**: 1–6.
Savolainen O, Pyhäjärvi T, Knürr T (2007). Gene flow and local adaptation in trees. *Annu Rev Ecol Evol Syst* **38**: 595–619.
Scheifinger H, Menzel A, Koch E, Peter C (2003). Trends of spring time frost events and phenological dates in Central Europe. *Theor Appl Climatol* **74**: 41–51.
Scotti-Saintagne C, Bodénès C, Barreneche T, Bertocchi E, Plomion C, Kremer A (2004). Detection of quantitative trait loci controlling bud burst and height growth in *Quercus robur* L. *Theor Appl Genet* **109**: 1648–1659.
Steinkellner H, Lexer C, Turetschek E, Glossl J (1997). Conservation of (GA)n microsatellite loci between *Quercus* species. *Mol Ecol* **6**: 1189–1194.
Tajima F (1989). Statistical methods for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
Taji T, Ohsumi C, Iuchi S, Seki M, Kasuga M, Kobayashi M *et al*. (2002). Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in *Arabidopsis thaliana*. *Plant J* **29**: 416–426.
Tanksley SD, Ganal MW, Martin GB (1995). Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. *Trends Genet* **11**: 63–68.
Van Asch M, Visser ME (2007). Phenology of forest caterpillars and their host trees: the importance of synchrony. *Annu Rev Entomol* **52**: 37–55.
Visscher PM, Thompson R, Haley CS (1996). Confidence intervals in QTL mapping by bootstrapping. *Genetics* **143**: 1013–1020.
Wachowiak W, Balk PA, Savolainen O (2009). Search for nucleotide diversity patterns of local adaptation in dehydrins and other cold-related candidate genes in Scots pine (*Pinus sylvestris* L.). *Trees Genet Genomes* **5**: 117–132.
Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.
Zeng ZB (1994). Precise mapping of quantitative trait loci. *Genetics* **136**: 1457–1468.

Résumé

Afin d'estimer la capacité de réponse du chêne sessile (*Quercus petraea* Matt. Liebl.) aux changements climatiques en cours, le potentiel d'adaptation de cette espèce pour le débourrement a été mesuré en populations naturelles. Ces populations sont situées le long d'un gradient altitudinal comprenant 12 populations entre 131 et 1630 m, et d'un gradient latitudinal comprenant 21 populations de l'ensemble de l'aire de répartition.

Tout d'abord l'empreinte démographique sur les niveaux de diversité génétique a été estimée sur les populations du gradient altitudinal à partir de marqueurs neutre. Les résultats ont montré que la diversité est maintenue le long du gradient altitudinal grâce notamment à des forts flux de gènes entre populations.

La variabilité génétique du débourrement à été mesurée en tests de provenances pour 10 populations du gradient altitudinal. Les résultats ont montré une forte différenciation ainsi qu'une héritabilité élevée du trait. Une variabilité génétique importante est maintenue à l'intérieur des populations et semble indiquer que de multiples pressions de sélection agissent de manière fluctuante et/ou opposée.

La diversité de gènes candidats pour le débourrement a été étudiée sur les populations des deux gradients environnementaux. Un niveau de diversité nucléotidique relativement fort et un faible déséquilibre de liaison qui décroit rapidement avec la distance ont été observés. Des signatures de sélections ont été mises en évidence sur un ensemble de gènes candidats.

Une étude d'association a été menée entre variabilité du caractère et polymorphisme au sein des gènes candidats sur les populations des deux gradients. Un total de 16 associations significatives a été observé impliquant 10 gènes candidats.

Mots-clés : Adaptation génétique, *Quercus petraea*, débourrement, gènes candidats, gradients.

Abstract

In order to assess the capacity of sessile oak (*Quercus petraea* Matt. Liebl.) to withstand the ongoing climate changes, we estimated its adaptative potential for bud burst within natural populations. These populations are located along two steep temperature gradients: an altitudinal gradient comprising 12 populations located between 131 m 1630 m, and a latitudinal gradient including 21 populations from the species' distribution range.

First the demographic imprint on the overall genetic diversity was assessed on the altitudinal gradient populations using neutral markers. Results showed that genetic diversity was homogeneously distributed along the gradient and maintained at high altitudes.

The genetic variability of bud burst was measured in provenance tests for 10 populations of the altitudinal gradient. We found a high level of genetic differentiation and a high heritability for the trait. A high variability was also observed within populations, indicating that selection pressures may fluctuate in natural conditions.

Genetic diversity of candidate genes for bud burst was assessed on populations from both gradients. A high level of nucleotide diversity was observed, and linkage disequilibrium was low. Selective signatures were observed on few candidate genes.

An association mapping study was performed between bud burst variability and polymorphism at the candidate genes on populations of both gradients separately. A total of 16 associations involving 10 genes were observed.

Keywords : Genetic adaptation, *Quercus petraea*, bud burst, candidate genes, gradients.