

Numéro d'ordre : 2982

Thèse en cotutelle

Présentée à

l'Université Bordeaux 1

Ecole Doctorale Sciences du Vivant, Géosciences et Sciences de l'Environnement

Avec **l'Université des Sciences Agronomiques de Vienne**

par **Jérémy DERORY**

pour obtenir le grade de

Docteur

Spécialité : Biologie Cellulaire et Moléculaire

Génomique et diversité du débourrement chez les chênes

Soutenue publiquement le 24 Mai 2005

Devant la commission d'examen formée de :

Membres du jury :

Richard Michalet, Professeur, Université Bordeaux I

Président

Antje Rohde, Directrice de recherches, Université de Ghent, Belgique

Rapporteurs

Dieter Ernst, Directeur de recherches, GSF, Allemagne

Josef Glössl, Professeur, Université des Sciences Agronomiques, Autriche

Examineurs

Co-Directeur de thèse

Antoine Kremer, Directeur de recherches, INRA, Bordeaux

Co-Directeur de thèse

Remerciements

« Comme tous les genres, le remerciement obéit à des lois. C'est un genre centrifuge, au sens ondulatoire du terme. Comme un cailloux qu'on lance dans la mare, le remerciement fait des cercles, de plus en plus... large...de plus en plus éloigné du centre. (...) Le lauréat remercie d'abord le premier cercle : les notables, les importants, le jury, sans qui la récompense ne lui aurait pas été attribuée ; puis le deuxième cercle, le public (...) »

Daniel Pennac, dans *Merci*.

Ce travail de thèse aujourd'hui quasiment achevé, je dois donc me soumettre à l'exercice imposé et au combien important des remerciements. Cet exercice se justifie pleinement par les contributions diverses et variées qui ont été apportées à ce travail mais les sentiments sont partagés au moment d'écrire. D'un côté, la joie d'exprimer une profonde marque de reconnaissance, d'un autre côté, c'est aussi pour moi une angoisse plus forte que celle ressentie au moment de la rédaction du document puisqu'il faut être à la hauteur de la gratitude et de l'amitié que je porte aux nombreuses personnes qui ont participé de près ou de loin à ce travail.

Je tiens à remercier Antoine Kremer pour m'avoir offert la possibilité de réaliser ce travail de thèse. Mes remerciements sont aussi intenses que la déception qui avait suivie mon travail de DEA, après que des complications aient eu raison de ma volonté de faire une thèse. On m'avait informé quelques mois avant la fin que : « Ce sera une thèse dans la région parisienne ou rien », ce que je n'envisageais pas du tout. Un grand merci au passage à la « régionalisation des écoles doctorales » et au soutien inexistant du responsable de DEA. Au mois de Septembre 2001, j'avais donc fait une croix sur le métier de chercheur que je m'étais imaginé au fil des stages effectués au Laboratoire. La délivrance est venue à la fin du mois d'Octobre par un coup de fil d'Antoine m'annonçant qu'une bourse était disponible sur le sujet présenté aujourd'hui. J'ai cru bon de demander un délai de réflexion mais l'ai rappelé dès le lendemain ou le

surlendemain. Chef, je voudrais ici sincèrement te remercier de m'avoir fait confiance, mais pas seulement pour cette raison. Je n'apprendrais à personne que le monde scientifique est basé sur la compétition et qu'une des conséquences est la pression étouffante exercée par la plupart des directeurs de thèse sur leurs doctorants, pour éviter toute perte de temps. Le succès compétitif est assuré et la créativité est oubliée en chemin. Merci donc pour m'avoir laissé m'exprimer pleinement sur le sujet, tout en prenant soin de canaliser les énergies et de les recadrer au besoin. Merci enfin pour ce que tu m'as appris et pour la simplicité avec laquelle tu abordes tes étudiants.

De nombreuses idées apparurent et prirent forme au cours des discussions avec Christophe Plomion. Je te remercie de m'avoir consacré autant de temps, et de m'avoir aidé dans la réalisation de ce travail avec un tel enthousiasme. Mes débuts scientifiques ont été initiés en grande partie par toi. Après avoir réalisé un stage avec Annie Raffin au mois de Juillet 1999, durant lequel j'ai découvert les joies du travail de terrain et de la génétique quantitative, tu m'as proposé de revenir faire mon mémoire d'ingénieur sur la diversité génétique du pin maritime. C'est durant ce stage que m'est venu l'idée de faire une thèse. Je voudrais te remercier particulièrement pour le soutien incessant que tu as apporté à ce travail de thèse.

Un merci commun à Antoine et à Christophe qui vont me permettre de poursuivre l'aventure pierrotonnaise au cours d'un stage post-doctoral dans le cadre du projet DIGENFOR.

Je remercie également le professeur Josef Glössl pour avoir accepté de superviser cette thèse aux côtés d'Antoine Kremer, et m'avoir accueilli quelques mois, au printemps 2003, dans son Laboratoire de l'Université des Sciences Agronomiques de Vienne. Ce séjour fut au combien enrichissant et m'a permis d'améliorer très nettement ma pratique de l'anglais, jusque là balbutiante. Pour ce qui est de l'usage de la langue germanique, j'en suis malheureusement resté aux formules d'usage : „Entschuldigung, Ich verstehe nicht. “ Oder „Ich spräche ein bisschen Deutsch. “ Merci à Marie-Theres Hauser de m'avoir intégré à son équipe et de m'avoir fait profiter de toute son expérience sur la technique de RT-

PCR quantitative, et merci à elle pour son dynamisme débordant et communicatif.

Un aspect déterminant du travail de thèse, c'est de choisir l'équipe dans laquelle on peut le mieux exprimer sa personnalité pour que la recherche reste toujours un plaisir. Et je ne regrette absolument pas d'avoir effectué ma thèse dans l'équipe de Génétique de la récente UMR Biogeco, le Labo de Pierroton pour les nostalgiques. La croissance régulière de l'équipe n'a pas (encore !) eu raison du mode de fonctionnement, qui permet d'échanger sans aucune contrainte avec l'ensemble des membres du Labo et d'enrichir ainsi sa propre réflexion d'avis extérieurs et bien souvent complémentaires. Je remercie tout particulièrement Manuela Casasoli, Alexandre Jauffrès, Caroline Morera-Dutrey, Noémie Graignic, Jean Marc Louvet, Mimi Bertocchi, Guy Roussel, Bernard Montoussé, Patrick et Valérie Léger et Franck Salin qui ont contribué directement aux résultats de ce travail, ainsi que tous les membres permanents qui m'ont apporté une aide ponctuelle. Je remercie également les membres extérieurs au Labo, qui ont contribué à ce travail : Loïck le Dantec, Virginie Garcia et Jacques Schaeffer.

Je tiens également à remercier les membres de mon comité de thèse, qui s'est réuni à deux reprises et m'a permis de poursuivre le travail dans la sérénité, en apportant un regard critique sur l'avancement des travaux, et en contribuant à apaiser les crises d'anxiété que tout doctorant subit au cours de sa thèse, du genre « pétage de plomb, tout est à jeter, j'ai plus qu'à tout recommencer ». J'exprime donc toute ma gratitude à Isabelle Chuine, Rémy Rageau, Christophe Rothan, Sébastien Duplessis, Christophe Plomion et Antoine Kremer pour leurs critiques et apports constructifs au cours des comités et réunions consacrés à cette thèse. Au même titre que le comité de thèse mais pour la partie expérimentale de cette thèse, je remercie chaleureusement Grégoire le Provost d'avoir su patiemment et formidablement me coacher à la paillasse au début de ma thèse. Je remercie sincèrement Céline Lalanne et Delphine Madur de m'avoir initié au travail de laboratoire.

Pour ce qui est des membres du jury, je commencerais par les remerciements avant soutenance (❶), puis les remerciements que j'espère pouvoir exprimer après la soutenance (❷).

❶ Je n’oublie pas, en effet, les personnes qui ont acceptées d’être membre de la commission d’examen avec, en tout premier lieu, Antje Rohde et Dieter Ernst, qui auront la tâche d’être rapporteurs de ce document. Je remercie dans le même temps le président de cette commission : Richard Michalet, ainsi que mes deux directeurs de thèse : Josef Glössl et Antoine Kremer.

❷ YYYYYYYYYAAA
AAAHHHHHHHHHHHHHHHHHHH
HH.

Les remerciements ne seraient rien sans une mention particulière pour l’amitié. Mes premiers remerciements vont à Manuela et Jérôme. Manuela, qui s’en est retournée au pays où les icônes religieuses sont soit rouges avec quatre roues, soit tout ronds et se font taper dessus tous les dimanches... je crois que j’en oublie une autre dont j’ai oublié le nom. Ce qui ne devait être au départ qu’une collaboration s’est très vite transformé en amitié, qui s’est forgée en partie au cours des rendez vous culinaires réguliers dont Jérôme faisait également partie, quand il n’en était pas directement l’instigateur. J’ai d’ailleurs appris récemment sa technique pour amener tous les visiteurs étrangers du Labo à venir cuisiner chez lui... il est très très fort, mais je ne dévoilerai rien. Manuela, si tu lis cette thèse : « A cet été avec Simone et Alice ! ». La rencontre avec Jérôme coulait de source : deux JD dans le même Labo, c’était un de trop et tout le monde nous l’a bien fait sentir en nous gratifiant régulièrement du prénom de l’autre, difficile donc de passer à côté. Merci à toi pour cette alchimie entre nous et cet amour immodéré de Chico, lé plou granche. Je remercie aussi Hakim, la crapule, de son amitié. Je le connaissais bien avant Pierroton, et les liens se sont renforcés au cours de nos séjours successifs au Laboratoire. Spécial dédicace au p’tit beurre nantais. D’une autre génération mais non moins jeune et pétillant par l’esprit, j’adresse mes remerciements chaleureux à Teresa et Henri pour leur gentillesse et leur clairvoyance.

Mes amis plus lointains, et souvent plus anciens, sont bien présents et l’ont été tout au long de cette thèse malgré les longues (trop longues !) périodes sans se voir. Je les remercie tous affectueusement. Je commencerai par rendre visite au Briançon Crew avec Cérou, Aurèle, Kamel, Jean, Rub et Guy. En redescendant vers le Sud en direction de Sisteron, je m’arrêterai à Gap pour goûter le Génépi

de l'année avec Gino et Fabienne, qui en profiteront pour me présenter Pablo. Il suffira ensuite de poursuivre la route jusqu'à Toulon à travers les Alpes de Haute Provence et le Haut Var, pour se taper un match des « Rouge et Noir » à Mayol avec Ricou, Tiou et Lolo. Ricou nous ramènera avec sa superbe Fiat Uno, achetée récemment. En un saut : voilà Marseille, et Mich se fera un plaisir de préparer des bonites au barbecue, avec Didou, Bru, Christine et Guigo. Je pourrai tout aussi bien partir vers l'Ouest, en direction de Grenoble, depuis Briançon. Dans ce cas, je m'arrêterai voir Fred et Diane à Bourg d'Oisans, qui ont survécu à leur premier hiver là bas et c'est déjà un exploit. Le prochain arrêt se fera à Grenoble où Olive, Paul pour les intimes, et Emeline, à ne pas confondre avec Evelyne parce que ça le fait moins, se feront un plaisir d'organiser une soirée de réveillon inoubliable. Je continuerai ma route vers Lyon, où Fred se prépare activement à son futur rôle de père et Mélanie ne doit pas être de trop pour le coacher. Enfin, mais vraiment si j'en ai la force, j'irai jusqu'à Verdun, ou même un toulonnais né à Paris peut aller s'installer, et ça, c'est quand même pas rien. Heureusement que j'aurai la satisfaction de voir Claire et Salomé parce que, franchement, pour ce qui est de Jérôme, dit la poule ou Jojol, je me taperai pas toutes ces bornes. De retour à Bordeaux, Juju, me donnera sûrement les dernières infos en provenance de la Commanderie, même si je crains qu'il m'en veuille à cause de la défaite de l'OM à Saint Etienne. Heureusement, Séverine, Mathis et Poupon, qui fait un peu partie de la famille Arfi, ne m'en voudront certainement pas. Je ne ferai pas l'effort d'aller voir Sophie et Julien à Rennes puisqu'il vienne ce WE. Voilà donc ce qu'il me reste à faire une fois passée la soutenance.

Ces années passées à Pierroton, et brièvement à Vienne, m'ont également permis de rencontrer de très nombreuses personnes que je remercie chaleureusement pour les moments passés ensemble : Francky, David Poulpy Pot, Céline, Ruud, Jorge, Marie France, Patricia, Fabiano, Josquin, Alex, Manue, Laurent, Cristina, Marie Helène, Antonio, Delphine, Erwan, Marta, Mumu, Naoki, Johannes, Michela, HCl, Agnès, Maria-Martha, Olive, Paola, Caro, Arndt, Sylvain, Corinne, Christian, Sophie, Cécile, Delphine, Mathilde, Berthold, Fabrice, Maurice, David Travolta Chagné, Carlito y Loreto y Antonia y Amelia, Patrick, Manon, Stéphanie, ainsi que tous les membres permanents du laboratoire. Des remerciements particuliers vont aux colocataires de bureaux successifs, qui supportent mes crises régulières lorsqu'un d'entre eux a le malheur de déplacer

un stylo sur mon bureau, ou pire, me piquer mes post-it, mais surtout pour les séances blagues à gogo parfois, souvent même, d'un goût douteux, mais c'est tellement bon. Je remercie donc le premier, l'unique, l'inimitable Jeannot, auxquels pas un membre des Grosses Têtes n'arrivent à la cheville, Philou, Mr « Trois Ecrans » depuis peu, parce qu'un seul ne le stimulait plus assez, Marie-Hélène, la reine du rangement et du secret bien gardé, Christophe AKA Le Jeune, qui a refait la déco de la porte en grande partie, et dont les talents de retoucheur d'images n'échapperont désormais plus à personne et enfin Raoul, arrivé au début de la période de rédaction, et qui m'a définitivement convaincu que les chiliens étaient fous.

Cette thèse n'aurait pas été possible sans la présence et l'amour de Lili qui a la folie et la force de supporter ma vie. Alice, grâce à toi, j'ai de bonnes raisons de vivre et de me projeter dans l'avenir.

Mes parents, mes sœurs et Seb ont su m'apporter, tout au long de ce travail, l'affection, la confiance et l'énergie qui m'ont permis de mener à terme ce travail.

Cette thèse représente donc l'aboutissement de trois années de travail mais aussi le début d'une nouvelle vie professionnelle qui, je l'espère, sera consacrée à la recherche. Je terminerai sur ces mots de Pierre Desproges, dans son *Almanac* : « En cas de morsure de vipère, sucez-vous le genou, ça fait marrer les écureuils. » et vous laisse à des lectures plus sérieuses.

Table des matières

Table des matières	1
Préambule	4
A. Synthèse bibliographique	8
① La phénologie des essences forestières en zone tempérée	9
1.1. Croissance, dormance et variation saisonnière	10
1.1.1. L'activité du bourgeon, moteur de la croissance	10
1.1.2. La dormance, élément clé de la phénologie	11
1.1.3. Le cycle annuel des arbres forestiers en zone tempérée	12
1.2. Bases écophysiologicals et variation de débourrement	14
1.2.1. Effets des variables climatiques sur le débourrement	14
1.2.2. Modèles prédictifs de la date de débourrement	16
1.2.3. Les hormones impliquées dans la dormance	17
1.3. Variabilité du débourrement et changements climatiques	19
1.3.1. Variabilité géographique de la date de débourrement	19
1.3.2. Hypothèses explicatives	20
1.3.3. Impact des changements climatiques sur la date de débourrement	22
② Le débourrement : du phénotype aux gènes	24
2.1. Déterminisme génétique du caractère de débourrement	26
2.1.1. Un contrôle génétique modéré à fort	26
2.1.2. Principes et limites de la cartographie génétique et l'approche QTL	27
2.1.3. Les QTL de débourrement chez les arbres forestiers	29
2.2. Les bases moléculaires du débourrement	31
2.2.1. Dormance de la graine et dormance du bourgeon : des phénomènes similaires ?	31
2.2.2. Stress oxydatif et métabolisme au cours du débourrement	32
2.2.3. La régulation du cycle cellulaire	33
2.3. Vers la mise au point de marqueurs diagnostic de la date de débourrement	35
2.3.1. Le séquençage d'EST et l'analyse du transcriptome	35
2.3.2. L'approche gène candidat	37
2.3.3. Utilisation du polymorphisme de type SNP dans les études d'associations	38
③ Caractéristiques des chênes blancs européens	41
3.1. Le complexe d'espèces des chênes blancs européens	42
3.1.1. Taxonomie et aires de répartition	42
3.1.2. Ecologie et sylvigénèse des chênes sessile et pédonculé	43
3.1.3. Caractéristiques de la reproduction	44
3.2. Hybridation et différenciation des chênes sessile et pédonculé	46
3.2.1. Hybridation	46
3.2.2. Différenciation morphologique	47
3.2.3. Différenciation moléculaire	47
3.3. Point de vue historique et diversité génétique actuelle	49
3.3.1. Modifications des milieux en Europe depuis le dernier cycle glaciaire	49
3.3.2. Zones refuges et voie de recolonisation des chênes	50
3.3.3. Colonisation et diversité génétique actuelle	51
Objectifs de la thèse	53
B. Analyse transcriptomique du débourrement chez le chêne sessile	56
① Matériel et Méthodes	58
1.1. Le matériel végétal	58
1.1.1. Les populations échantillonnées	58
1.1.2. Les stades de débourrement	58
1.1.3. La récolte du matériel	59

1.2. La construction et l'exploitation des banques SSH	61
1.2.1. Extraction des ARN totaux et synthèse des ADNc	61
1.2.2. Construction des banques SSH	62
1.2.3. Clonage des banques et préparation des filtres bactériens à haute densité	63
1.3. Tri différentiel des transcrits et séquençage des EST	65
1.3.1. Hybridation des filtres bactériens à haute densité	65
1.3.2. Analyse de variance des données d'expression	66
1.3.3. Séquençage et traitement bioinformatique des EST	67
1.4. Analyse quantitative de l'accumulation des transcrits au cours du débourrement	69
1.4.1. Les filtres ADNc à haute densité	69
1.4.2. La PCR quantitative en temps réel	70
② Résultats et Discussion	72
2.1. L'étape de tri différentiel et les banques d'EST	72
2.1.1. Sélection des clones différentiellement exprimés	72
2.1.2. Analyse globale et classification fonctionnelle des EST	73
2.2. Suivi d'expression de gènes au cours du débourrement	75
2.2.1. Le regroupement des profils d'expression	75
2.2.2. Transcrits étudiés en RT-PCR quantitative en temps réel	76
2.2.3. Les mécanismes moléculaires régulés durant le débourrement	77
2.3. Sélection de GC	80
2.3.1. Méthodes combinées	80
2.3.2. Des candidats expressionnels et/ou fonctionnels ?	81
③ Conclusions de l'approche transcriptomique	83
C. Cartographie comparée et colocalisation entre GC et QTL	85
① Matériel et Méthodes	87
1.1. Amplification et cartographie des marqueurs EST	87
1.1.1. L'amplification des marqueurs dérivés des EST	87
1.1.2. La cartographie des STS	89
1.2. Cartes de liaison consensus et cartographie comparée des QTL	91
1.2.1. Construction des cartes de liaison consensus chez <i>Q. robur</i> et <i>C. sativa</i>	91
1.2.2. Projection et cartographie comparée des QTL	92
② Résultats et Discussion	95
2.1. L'alignement des douze groupes de liaison grâce aux STS	95
2.1.1. L'amplification et la cartographie des STS	95
2.1.2. L'alignement des cartes génétiques du chêne et du châtaignier	96
2.1.3. Les marqueurs dérivés des EST : un outil de choix pour la cartographie comparée	98
2.2. Le déterminisme génétique du débourrement chez <i>C. sativa</i> et <i>Q. robur</i>	100
2.2.1. Comparaison des positions des QTL chez les deux espèces	100
2.2.2. Un déterminisme génétique similaire	102
2.3. Identification de gènes candidats pour le débourrement	105
2.3.1. Cartographie des gènes candidats expressionnels	105
2.3.2. Des clusters de gènes informatifs ?	107
③ Conclusions de l'approche de cartographie comparée	110
D. Diversité nucléotidique et études d'association de gènes candidats pour le débourrement	111
① Matériel et Méthodes	113
1.1. Du matériel végétal au génotypage des marqueurs SNP	113
1.1.1. Les populations échantillonnées au sein de l'aire naturelle	113
1.1.2. La détection de polymorphisme	114
1.1.3. Le génotypage des SNP	116
1.2. Outils d'analyse de la diversité et des études d'association	118
1.2.1. Analyse des séquences	118
1.2.2. Diversité nucléotidique et différenciation entre populations à certains sites SNP	119
1.2.3. Etudes d'association entre variants alléliques et notes de débourrement	121

② Résultats et discussion -----	123
2.1. Le polymorphisme révélé au sein des gènes candidats-----	123
2.1.1. De la difficulté de mettre en évidence des polymorphismes au sein des séquences-----	123
2.1.2. Les sites polymorphes identifiés au sein des GC-----	124
2.1.3. Diversité nucléotidique-----	125
2.2. Des gènes candidats différenciés et soumis à l'action de la sélection ?-----	127
2.2.1. Différenciation entre les populations-----	127
2.2.2. Des gènes qui s'écartent du modèle évolutif neutre ?-----	128
2.2.3. Le déséquilibre de liaison au sein des GC-----	130
2.3. Etudes d'association-----	132
2.3.1. Choix des locus et des sites SNP-----	132
2.3.2. Diversité et différenciation des sites SNP étudiés-----	133
2.3.3. Des locus SNP en déséquilibre et soumis à l'action de la sélection naturelle ?-----	135
2.3.4. Corrélation entre variants alléliques et notes de débourrement-----	136
③ Conclusions des études de diversité et d'association -----	138
<i>E. Conclusions et perspectives</i> -----	139
① La stratégie mise en place est elle pertinente pour l'étude du déterminisme génétique du débourrement ? -----	140
② Evaluation des techniques employées -----	143
③ Perspectives de cette étude -----	145
<i>Références bibliographiques</i> -----	147
<i>Annexe I : Transcriptome analysis of bud burst in sessile oak (Quercus petraea (Matt.) Liebl.)</i> -----	170
<i>Annexe II : Comparison of QTLs for adaptive traits between oak and chestnut based on an EST consensus map</i> -----	171
<i>Annexe III : Nucleotide diversity and association study in genes involved in bud burst in sessile oak (Quercus petraea (Matt.) Liebl.)</i> -----	172

Préambule



Le débourrement constitue un caractère adaptatif clé des essences ligneuses en milieu tempéré. En effet, au même titre que la phénologie du bourgeon apical et sa dormance, la date de débourrement est une étape importante du rythme biologique de l'arbre contribuant à une synchronie vis-à-vis des conditions du milieu et à la protection contre certains prédateurs, tels que les insectes défoliateurs. Sous l'effet du réchauffement climatique, le suivi de la date de débourrement au cours des dernières années, dans des parcs réservés à ces études, a montré un démarrage de plus en plus précoce de la croissance chez les ligneux forestiers. Les conséquences de ces changements climatiques vont au-delà de la simple question relative à la longueur de saison de végétation, et mettent également en jeu la distribution et la survie même de ces espèces. La capacité d'adaptation de ces dernières est, en grande partie, conditionnée par les niveaux de diversité génétique présents dans les populations naturelles, mais aussi par la plasticité phénotypique des individus. En un mot, la diversité génétique des chênes leur permettra t-elle de s'adapter à ces nouvelles conditions ou allons nous assister à des dépérissement massifs ?

Ainsi, les programmes de recherches visant à décrire la diversité génétique des populations de chênes se sont récemment orientés vers l'étude de la diversité fonctionnelle et adaptative d'un certain nombre de caractères complexes (débourrement, anoxie racinaire...). Auparavant, les études ont largement été consacrées à décrire le niveau et l'organisation de la diversité génétique neutre et à identifier les mécanismes évolutifs qui la maintiennent à travers leur histoire. Un second volet des recherches menées sur les chênes concerne l'étude de la différenciation entre espèces, focalisée, en Europe, sur le complexe formé par le chêne sessile (*Quercus petraea*) et le chêne pédonculé (*Quercus robur*) et, au-delà, sur l'évolution génomique (cartographie génétique, synténie) au sein de la famille des fagacées. La thèse présentée s'inscrit pleinement dans la continuité des ces recherches et vise à localiser les régions génomiques (gènes, QTL) responsables de la variation de la date du débourrement chez les chênes. Elle associe génétique des populations, quantitative et moléculaire, pour localiser ces régions et pour construire des outils d'évaluation de la diversité pour ce caractère dans les populations naturelles. En effet, les méthodes classiquement utilisées pour décrire l'architecture génétique d'un caractère complexe (QTL) ne permettent que de détecter les régions polymorphes dans le croisement utilisé pour la cartographie. La principale question posée dans le cadre de cette thèse peut donc s'énoncer de la manière suivante « Comment détecter également les régions responsables de la différenciation entre populations et non seulement de la variation intra population ? » La stratégie employée dans cette thèse, pour

tenter de répondre à cette question, associe des approches génétiques et moléculaires complémentaires.

Le chapitre A sera consacré à la synthèse bibliographique des connaissances actuelles sur le débourrement des arbres forestiers mais aussi sur les caractéristiques des chênes blancs européens qui ont constitué le matériel expérimental de cette thèse. Dans un premier temps, nous verrons que la phénologie des essences forestières s'inscrit dans un cycle annuel de croissance, particulier aux arbres forestiers. Les bases écophysiologiques de ce phénomène seront ensuite présentées. Enfin, nous aborderons la variabilité du phénomène sous l'influence des facteurs abiotiques, principalement climatiques, et biotiques, tels que les pressions d'insectes défoliateurs et formulerons quelques hypothèses quant aux mécanismes adaptatifs sous jacents. La seconde partie de ce chapitre abordera les connaissances liées à la génétique du caractère avec, en premier lieu, son déterminisme génétique. Nous ferons ensuite la synthèse des bases moléculaires connues à ce jour, en comparant la levée de dormance du bourgeon à la levée de dormance des graines puis en abordant certains processus physiologiques tels que le stress oxydatif, le métabolisme et la régulation du cycle cellulaire. Dans un dernier temps, nous verrons, dans quelle mesure, une approche de type gène candidat pourrait être susceptible de nous mener vers la mise au point de marqueurs diagnostics de la date de débourrement. La dernière partie de cette synthèse bibliographique présentera les principales caractéristiques du complexe d'espèces représenté par les chênes blancs européens, d'un point de vue taxonomique, écologique, sylvicole, reproductif, morphologique, moléculaire et enfin historique sous l'angle de la diversité génétique actuelle.

Les objectifs de la thèse seront clairement détaillés dans un chapitre qui présentera une vision synthétique de la stratégie employée pour mener à bien notre étude. Les chapitres suivants B, C et D seront consacrés à la présentation et à la discussion des résultats issus des différentes expériences mises en place dans le cadre de la thèse.

Ainsi, le chapitre B s'attachera à présenter l'analyse transcriptomique menée afin d'identifier un certain nombre de gènes candidats (GC) pour le débourrement, sur la base de critères d'expression. Les résultats seront discutés à la fois dans le cadre de la stratégie globale de la thèse mais également sous l'angle de la compréhension du processus de débourrement. Quels sont les processus fonctionnels régulés au cours du débourrement ? Peut-on extraire des gènes candidats expressionnels de ces expériences ? Les méthodes employées sont-elles en accord avec l'objectif fixé ? Nous tenterons de répondre à ces différentes questions. Le chapitre C sera consacré à la détection des QTL impliqués dans le débourrement chez le chêne et à la comparaison de leur localisation avec les QTL identifiés chez le

chataignier. Là encore, les questions concernent à la fois la stratégie de la thèse dans son ensemble mais aussi la question plus large de l'évolution de caractères adaptatifs. Quels aspects techniques ressortent de l'utilisation de marqueurs dérivés des EST pour la cartographie comparée ? Quel est l'apport de la colocalisation entre GC et QTL dans le cadre de notre approche ? Les régions génomiques impliquées dans le contrôle du caractère de débourrement sont elles conservées entre les deux espèces ? En dernier lieu, le chapitre D présentera et discutera les résultats issus des études de diversité et d'association mises en place dans le cadre de cette thèse, avec le même souci d'intégration vis-à-vis de la démarche dans son ensemble. La démarche conduite dans cette partie expérimentale était elle en mesure de répondre à nos objectifs ? Pouvons nous préciser non plus les QTL, mais les gènes, voire le polymorphisme, responsables de la variabilité du caractère de débourrement ?

Le dernier chapitre E se proposera d'effectuer une synthèse des différentes parties expérimentales à la fois au regard de la stratégie mais aussi des méthodes employées, en tentant de répondre à la question : la stratégie et les méthodes employées ont-elles été pertinentes pour identifier les régions génomiques impliquées dans la variation du caractère de débourrement ? L'ensemble du travail réalisé sera également mis en perspective pour tenter de faire ressortir des pistes d'améliorations et des perspectives à ces recherches.

Les annexes présenteront les articles rédigés au cours de cette thèse et soumis à publication, ou en cours de rédaction.

A. Synthèse bibliographique

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)

OBSERVATIONS
DU THERMOMETRE,

Faites à Paris pendant l'année M. DCCXXXV.

Comparées avec celles qui ont été faites sous la Ligne;
à l'Isle de France, à Alger, & en quelques-unes
de nos Isles de l'Amérique.

PAR M. DE REAUMUR.

DANS l'Histoire de l'Académie de l'année 1696 on trouve des Observations de M. de la Hire sur le froid & le chaud de cette année : tant que ce célèbre Académicien a vécu, il a continué ses observations du Thermometre, & d'en publier les résultats. Après sa mort, de semblables observations furent faites par M. Maraldi; & après le décès de M. Maraldi, elles l'ont été par M. Maraldi son neveu. Dans un temps où on avoit moins de curiosité qu'aujourd'hui pour les observations physiques, & où l'on ne voyoit peut-être pas autant de quelle utilité elles peuvent être, M. de la Hire crut que s'en étoit assez par rapport au Thermometre, de publier dans les Mémoires de l'Académie, à quelle hauteur s'étoit élevée la liqueur du fien, le jour de chaque année où elle avoit monté le plus, & jusqu'où elle étoit descenduë, le jour où elle s'étoit trouvée le plus bas; c'est-à-dire, qu'il crut qu'il suffisoit au Public, & même à bien des Savants, d'apprendre quel jour de chaque année il avoit fait plus chaud, & quel jour il avoit fait plus froid à Paris. M. Maraldi, qui ont continué les Observations thermométriques après M. de la Hire, s'en font tenu à donner pour chaque année un résultat pareil à celui que ce Savant étoit en usage de donner. Depuis

Mém. 1735.

Z. 11

538 MEMOIRES DE L'ACADEMIE ROYALE
au moins un mois plutôt que dans les années ordinaires. Ce qui l'a rendu tardive, n'est pas précisément de ce que la chaleur n'a pas été considérable; les bleds font souvent moins quand les plus grandes chaleurs de l'été se font sentir; & les chaleurs de l'été n'avancent la maturité du raisin que quand elles agissent sur des terres suffisamment arrosées par les pluies: mais de deux années où les pluies aient été distribuées en égale quantité, & à peu-près de la même manière, & cela dans les mois où elles font les plus nécessaires, de ces deux années, c'est-à-dire, la plus précoce pour les grains & pour les fruits sera celle qui pendant les mois d'Avril, Mai & Juin aura eu une plus grande somme de degrés de chaleur. En 1734 les récoltes se font faites d'un très bon succès; j'ai été curieux de comparer la somme des degrés de chaleur au dessus de la congélation, désignés par ceux de notre Thermometre, qui ont agi sur la surface de la Terre de ce pays, pendant chacun de ces trois mois de 1734, avec la somme des degrés de chaleur qui ont agi pendant chacun de ces trois mois de 1735. Comme les Astronomes prennent un temps moyen pour mesurer la durée du temps vrai, j'ai pris pour chaque jour un degré de chaleur moyenne: & cela en ajoutant les degrés du Thermometre qui ont exprimé la plus petite chaleur du matin, aux degrés qui ont exprimé la plus grande chaleur de l'après-midi; la moitié de cette somme n'a pas pu être prise pour le degré de chaleur moyen. Par exemple, le 3 Avril 1735 à 6 heures du matin, la liqueur du Thermometre étoit à 4 degrés, & le même jour, à 3 heures après midi, elle étoit à 8 degrés; ces degrés ajoutés ensemble, donnent 12 degrés: j'en prends la moitié, 6 degrés, pour l'expression du degré du chaud moyen du 3 Avril. De tous les degrés moyens du mois d'Avril 1735, je fais une somme, & j'en fais de même une de tous les degrés moyens du mois d'Avril 1734. La comparaison de ces sommes me fait voir la différence qui est entre les sommes des degrés de chaleur désignés par les degrés du Thermometre, qui ont agi pendant le même mois de chacune de ces deux années.

Figure 1 : Reproduction partielle de l'article de René Antoine Ferchault de Réaumur paru en 1735 dans les Mémoires de l'Académie Royale des Sciences. (Source : <http://gallica.bnf.fr/>).

① La phénologie des essences forestières en zone tempérée

Du point de vue historique, la phénologie moderne des plantes a vu le jour grâce aux travaux du suédois Carl Linnaeus, également connu sous le nom de Carl von Linné (1707-1778), qui contribua à faire de la botanique une science distincte, jusque là rattachée à la médecine. Il publia en 1735 la première édition de sa classification des choses vivantes, le *Systema Naturae*, dans lequel la taxonomie des plantes était exclusivement basée sur le nombre et l'arrangement des organes reproducteurs. Il créa également des calendriers de floraison pour différentes régions de Suède en prenant soin de noter les conditions climatiques précises des emplacements considérés. En 1735 également, René Antoine Ferchault de Réaumur (1683-1757) soutint dans une publication (Réaumur, 1735) que les blés deviennent matures lorsqu'une certaine somme de température journalière au dessus d'un seuil, accumulée depuis une date fixe, atteint une valeur prescrite (Figure 1). Ce principe constitue la base des modèles phénologiques fondés sur le principe des sommes de température, encore très utilisés aujourd'hui.

La phénologie peut être définie comme « l'étude de la répartition temporelle des phénomènes biologiques récurrents, des causes de leur périodicité au regard des forces biotiques et abiotiques, et des inter-relations entre les phases au sein d'une même espèce ou entre espèces différentes » (Lieth, 1974). Ces phénomènes périodiques permettent de distinguer des états phénologiques qui constituent les stades du cycle de vie d'un organisme. Ces phénomènes cycliques annuels s'observent aussi bien dans le règne animal (migrations, éclosions, levée de diapause...), que dans le règne végétal (débourrement, floraison...). Chez les plantes, ces stades varient selon qu'il s'agisse :

- de fougères, prêles, lycopodes et groupes associés ;
- de Gymnospermes ;
- d'Angiospermes.

Les différents types d'organes reproducteurs sont à l'origine de l'existence de stades phénologiques différents entre ces trois groupes. La synthèse bibliographique se limitera à l'étude des stades phénologiques des arbres forestiers sous climat tempéré, et plus particulièrement à ceux de l'appareil végétatif. La phénologie de la reproduction ne sera pas abordée. Les exemples seront indifféremment extraits d'études portant sur les Gymnospermes (conifères) ou les Angiospermes (feuillus) avec une attention particulière sur les chênes et le caractère de débourrement (éclosion des bourgeons au printemps).

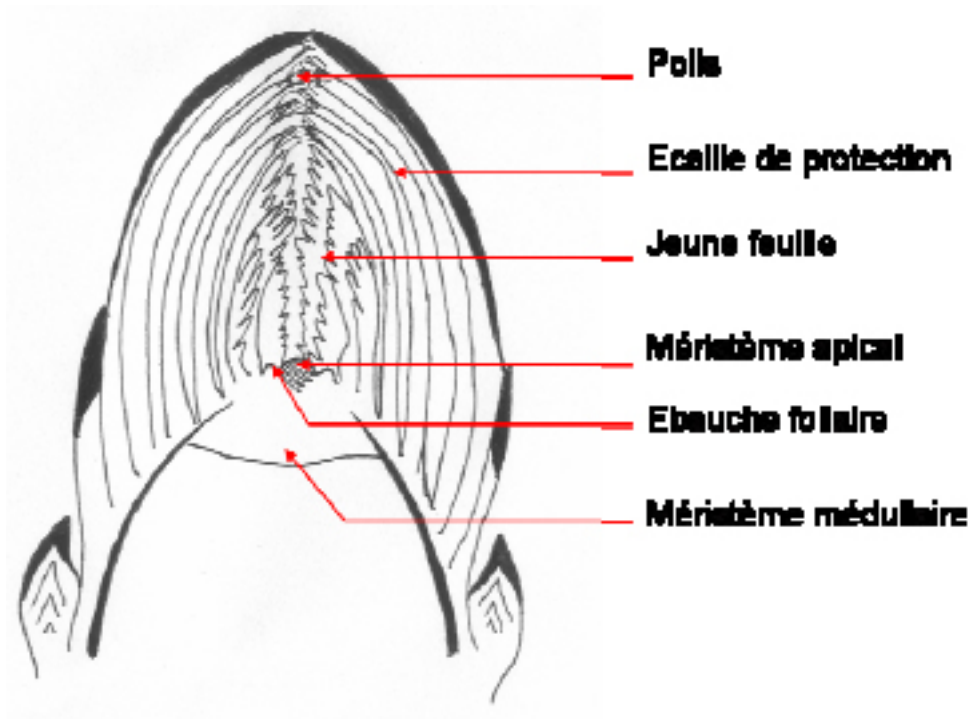


Figure 2 A : Dessin d'une coupe longitudinale de bourgeon apical d'érable à la loupe binoculaire (x 60).

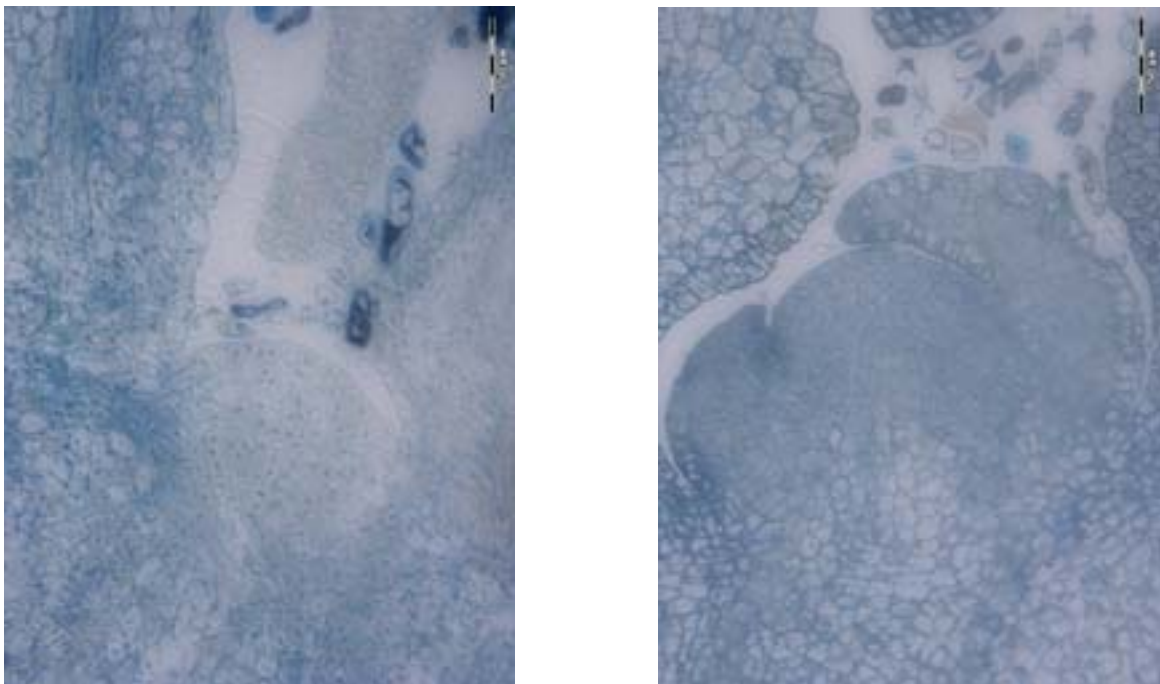


Figure 2 B : Photographies représentant une coupe longitudinale dans un bourgeon de chêne au stade quiescent puis durant le débourrement (Echelle : 50µm). (Tiré de notre étude).

1.1. Croissance, dormance et variation saisonnière

Le cycle végétatif annuel des essences forestières sous climat tempéré peut se résumer à une phase de dormance hivernale suivie, ou précédée, d'une période de croissance estivale. Deux mécanismes antagonistes sont au cœur de ce cycle : la dormance et la croissance. La dormance est une stratégie de survie qui permet à la plante de subsister durant des périodes défavorables à la croissance. La croissance, elle, prend place lors de périodes favorables (saison de végétation). Chez les arbres, l'activité des bourgeons détermine non seulement la saison de végétation mais aussi le port et l'architecture de l'arbre. En ce sens, la dormance possède également une fonction morphogénétique mais que nous ne traiterons pas ici (Rohde *et al.*, 2000).

1.1.1. L'activité du bourgeon, moteur de la croissance

Etymologiquement, le mot bourgeon vient de burjon, terme de botanique d'origine latine qui entre dans la langue française en 1160. Le bourgeon est l'organe central au sein duquel les phases de dormance et de croissance se manifestent et déterminent une croissance périodique du végétal. Le bourgeon peut être défini comme un axe court portant une série très dense de primordiaux foliaires produit par le méristème apical sous-jacent (Figure 2 A). Le méristème apical est formé durant l'embryogenèse. Il constitue la partie centrale du bourgeon où l'activité doit être régulée. On peut y distinguer une zone centrale, composée de cellules indéterminées, et une zone périphérique, source de cellules pour les organes en formation, comme les jeunes feuilles (Figure 2 B). L'anatomie de cette structure principale peut fortement varier entre espèces mais également au sein d'une même espèce en fonction de l'âge, de la position sur le rameau ou des conditions physiologiques (Brown, 1971). Deux types de bourgeons sont à distinguer : les bourgeons apicaux et les bourgeons axillaires. Les bourgeons axillaires proviennent du méristème apical et, le plus souvent, n'ont pas d'activité d'élongation durant l'année où ils sont formés (Richards et Larson, 1981). Les bourgeons apicaux, eux, vont libérer une portion de tige feuillée, ou unité de croissance, au printemps, qui va s'allonger progressivement durant quelques dizaines de jours. Lorsque plusieurs unités de croissance forment la pousse annuelle, on parle de croissance polycyclique.

A l'échelle cellulaire, la croissance du bourgeon peut être décomposée en deux étapes:

- une première phase de prolifération cellulaire et d'organogenèse qui a lieu dans le méristème ;
- une phase d'élongation cellulaire qui prend place immédiatement sous le méristème.

Le laps de temps entre la formation d'un organe et son élongation définit le type de croissance de l'arbre. Lorsque la croissance suit directement la formation des primordiaux, on parle de croissance indéterminée. A l'opposé, la croissance est dite déterminée lorsqu'une période de dormance s'intercale entre la formation du bourgeon et son élongation. Il n'est pas toujours aisé de définir une essence comme appartenant à l'une des deux catégories. En particulier, pour un grand nombre d'espèces, le type de croissance change graduellement avec l'âge, passant d'une croissance indéterminée à une croissance déterminée, comme c'est le cas chez les épicéas et sapins (Jablanczy, 1971). Cette gradation s'observe également au sein du même arbre au cours de la saison de végétation. En effet, les bourgeons en croissance durant la saison de végétation peuvent avoir été formés l'année précédente ou la même année. A la fin de la saison de végétation, la croissance des entrenœuds cesse et les primordiaux foliaires en périphérie du méristème vont se développer en écailles la plupart du temps. L'apparition de ces structures de protection n'empêche en rien la poursuite de l'organogenèse des primordiaux foliaires, bourgeons axillaires et stipules¹ chez les arbres à croissance déterminée. Cependant, toute activité mitotique cesse lorsque le bourgeon rentre en dormance. Un laps de temps de quelques semaines peut se produire entre l'arrêt de croissance et l'entrée en dormance.

1.1.2. La dormance, élément clé de la phénologie

En l'absence de mobilité, la dormance est le mécanisme principal par lequel les végétaux survivent aux conditions environnementales défavorables et synchronisent ainsi leur croissance avec l'environnement. Chez les arbres forestiers, les méristèmes qui formeront les unités de croissance au printemps suivant sont contenus dans les bourgeons écailleux qui constituent donc les organes dormants. La structure même de ces bourgeons permet de préserver les tissus des dégâts causés par le gel par la formation d'écailles à la périphérie et l'apparition de poils à la périphérie des écailles pour une meilleure imperméabilité de la structure. De nombreux termes et définitions ont été utilisés pour décrire la dormance mais ce

¹ Stipule : appendice le plus souvent foliacé ou membraneux, inséré au point où le pétiole se relie à la tige.

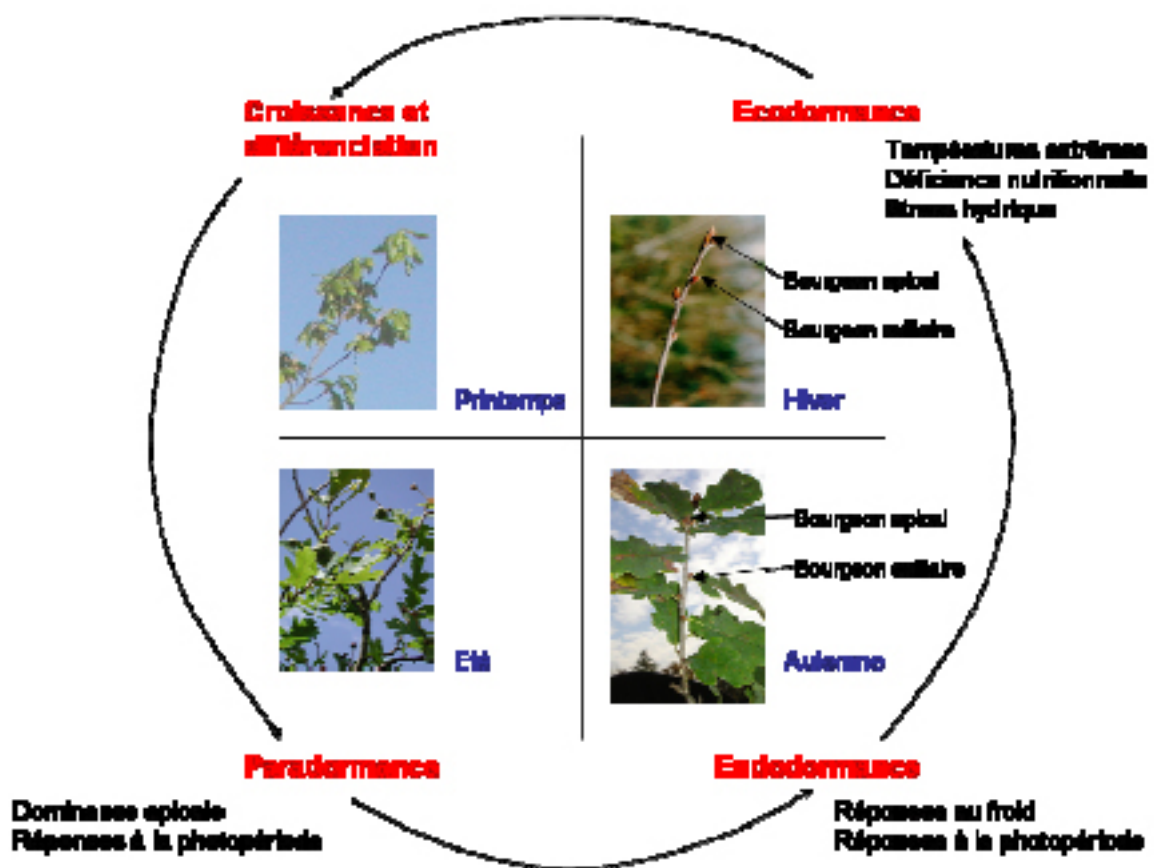


Figure 3 : Diagramme représentant les signaux et saisons correspondant aux trois types de dormance associés aux bourgeons apicaux et axillaires chez le chêne (modifié d'après Horvath *et al.*, 2003).

phénomène ne connaîtra pas de définition idéale avant qu'il ne soit entièrement disséqué et compris. Nous utiliserons ici la définition faite par Lang (1987), basée sur la définition suivante: « La dormance est l'arrêt temporaire de croissance visible de toute structure de la plante contenant un méristème ». A partir de cette définition, Lang a divisé la dormance en trois catégories :

- l'écodormance : l'arrêt d'activité est imposé par des facteurs environnementaux et la croissance est possible lorsque les conditions redeviennent favorables ;
- la paradormance : l'arrêt de croissance est provoqué par un autre organe de la plante, distinct et éloigné du tissu dormant ;
- l'endodormance : l'arrêt de croissance est lié à des facteurs endogènes au tissu dormant et la croissance est impossible, même en présence de conditions favorables.

Ces trois types de dormance se déclinent tout au long de la saison et selon le facteur à l'origine du phénomène (Figure 3). Ainsi, on parle de paradormance apicale si le facteur à l'origine de la dormance est le méristème apical, ou encore d'endodormance photopériodique si la photopériode est le facteur déclenchant de l'endodormance. Ce système est de nature pragmatique puisqu'il se base sur une terminologie descriptive et physiologique et, de ce fait, a beaucoup été utilisé. Du point de vue cellulaire, la dormance est liée à un arrêt des divisions cellulaires au sein du méristème apical. Devitt et Stafstrom (1995) ont émis l'hypothèse que les cellules de bourgeon dormant étaient arrêtées en trois stades du cycle cellulaire : au cours de la phase G1, à la transition entre les phases G1 et S, ou encore entre les phases S et G2. Par la suite, Gutierrez *et al.* (2002) ont montré que, dans la plupart des cas, les cellules contenues dans les bourgeons végétatifs sont arrêtées en phase G1 du cycle cellulaire, avant la phase S.

1.1.3. Le cycle annuel des arbres forestiers en zone tempérée

Vers la fin de la période active de croissance et de photosynthèse, les arbres forestiers se préparent à la saison hivernale et à la saison de végétation suivante par la formation de bourgeons écailleux (phase de bourgeonnement) aux environs des mois d'Août et Septembre. Comme on l'a évoqué précédemment, les organes de l'année suivante sont préformés dans le bourgeon avant l'entrée en phase d'endodormance (cf. Fig. 3). Cette phase de dormance au sein de structures protectrices est indispensable à la préservation des tissus exposés, sans cela,

au gel (Levins, 1969). L'année suivante, au cours du printemps, après que l'endodormance ait été levée par les températures froides, le développement des feuilles, fleurs et bourgeons apicaux est à nouveau possible sous certaines conditions de température et de photopériode, c'est la phase d'écodormance. Du point de vue du développement, l'occurrence de ces différentes étapes est loin d'être triviale. En effet, le débourrement des ébauches foliaires doit avoir lieu le plus tôt possible, afin de maximiser l'activité photosynthétique et, dans le même temps, les feuilles doivent être protégées des gelées tardives qui les endommageraient et donc réduiraient les capacités de photosynthèse, comme cela est le cas chez le chêne (Ducouso *et al.*, 1996). Il en est de même pour la croissance de la tige. Une fois le débourrement réalisé, l'arbre va croître durant quelques semaines puis connaître à nouveau une période de cessation d'activité avant de réaliser une seconde pousse. Sous nos climats tempérés, cette seconde pousse a lieu en Juin pour les essences de feuillus (pousse de la Saint Jean). Sur des individus jeunes, la même succession peut se reproduire en Août et parfois même en Septembre. D'autres espèces ne réalisent qu'un seul cycle dans l'année.

Au final, comme la longueur de la saison de végétation est définie par l'intervalle entre la date de débourrement et la date de bourgeonnement, ces deux événements phénologiques ont une importance capitale pour la productivité d'une essence forestière. Nous verrons dans les chapitres suivants que la survie, l'abondance et la distribution des essences peuvent également être affectées par des modifications de phénologie, et en particulier de la date de débourrement.

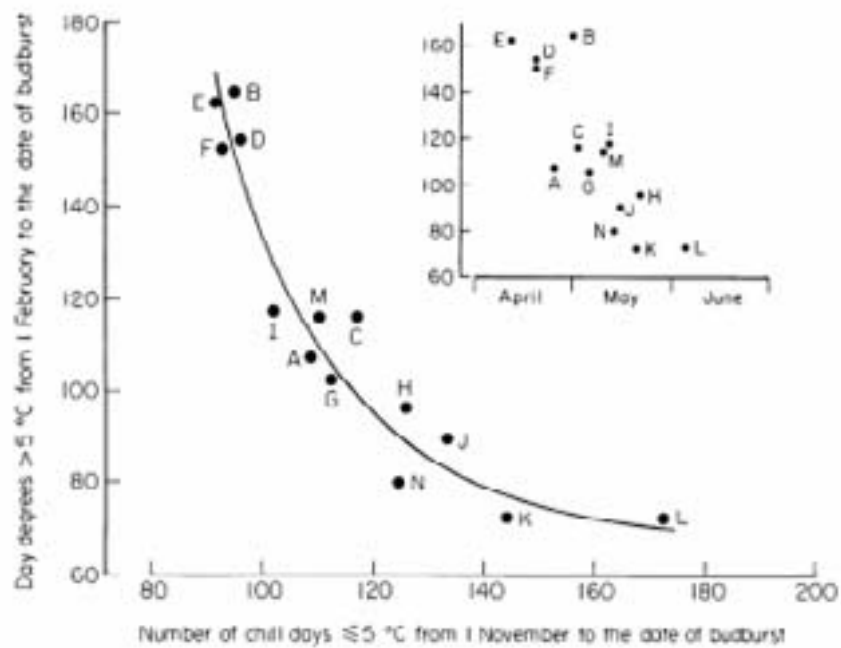


Figure 4 : Relation entre le nombre de jours froids avec une température moyenne $\leq 5^{\circ}\text{C}$ du 1^{er} Novembre à la date de débourrement et les degré jours $< 5^{\circ}\text{C}$ reçus du 1^{er} Février à la date de débourrement, pour quatorze enregistrements des dates de débourrement chez *Picea sitchensis* en Grande Bretagne (enregistrements A à N). La courbe est décrite par l'équation suivante :

$$\text{Degré jours } < 5^{\circ}\text{C} = 67,4 + 4401,8 \exp(-0,042 \times \text{jours froids } \leq 5^{\circ}\text{C})$$

Le graphe en haut à droite montre la relation médiocre obtenue en utilisant la date de débourrement plutôt que le nombre de jours froids en abscisse. (tiré de Cannell et Smith, 1983).

1.2. Bases écophysiologiques et variation de débourrement

Le développement cyclique et annuel des plantes n'est pas une préoccupation récente des scientifiques. En effet, dès 1735, le biologiste français de Réaumur a posé les bases du modèle des sommes de température par l'observation des degrés de chaleur moyens au cours des années 1734 et 1735, et leur effet sur la maturité des blés (Réaumur, 1735). Après avoir observé que les blés étaient plus tardifs en 1735 qu'en 1734, de Réaumur a calculé le degré de chaleur moyen des mois d'Avril, Mai et Juin de ces deux années là en sommant la plus petite température du matin et la plus grande température de l'après-midi puis en divisant cette somme par deux. Il obtint ainsi le degré de chaleur moyen d'une journée qu'il cumula d'Avril à Juin. Ainsi, la somme des degrés de chaleur était de 1260 en 1734 contre 1015 en 1735, « d'où il suit que cette dernière année a dû être plus tardive que l'autre, comme elle l'a été. », conclut de Réaumur.

1.2.1. Effets des variables climatiques sur le débourrement

Comme on vient de le voir, les travaux de de Réaumur ont mis en évidence l'influence des conditions météorologiques sur la phénologie des espèces en zone tempérée. Depuis, de nombreux travaux d'expérimentations et de simulations de la phénologie des arbres forestiers ont permis de mettre en évidence les relations existantes entre la croissance du bourgeon et les conditions météorologiques, même si certains résultats sont encore contradictoires. Ainsi, Sarvas (1974) suggère qu'aussitôt que la dormance est levée, les arbres réagissent aux températures favorables par le développement ontogénétique². Sur le même principe, Cannell et Smith (1983) démontrent que la date de débourrement de branches latérales de jeunes plants de *Picea sitchensis* peut être modélisée en prenant en compte les sommes de degré jours³ ($> 5^{\circ}\text{C}$) et les températures froides hivernales ($\leq 5^{\circ}\text{C}$) (Figure 4). Dans cette étude, les auteurs montrent que, dans les conditions climatiques hivernales britanniques, 140 jours avec une température moyenne $\leq 5^{\circ}\text{C}$ sont nécessaires à la levée de dormance. Ils concluent également que la photopériode n'a qu'un effet négligeable sur la date de débourrement chez *Picea sitchensis*. Enfin, Cannell et Smith mettent l'accent sur l'imprécision de ces modèles

² Ontogénétique : développement biologique de l'individu depuis la conception jusqu'à l'âge adulte.

³ Degré jour : écart, pour un jour donné, entre la température moyenne observée durant la journée et une valeur de température seuil. Pour une période donnée, c'est la somme, jour par jour, de ces écarts.

qui ignorent les autres variables environnementales, telles que la température du sol ou l'amplitude des températures diurnes. Pour Murray *et al.* (1989), il apparaît évident que la somme de température nécessaire au débourrement décroît à mesure que la durée des périodes précédentes de températures froides augmente. Ce principe est illustré sur la Figure 4, où l'on voit très bien que si le nombre de jours froids augmente (axe des abscisses), le nombre de degré jours requis pour mener au débourrement diminue proportionnellement (axe des ordonnées). Cette observation conduit Murray à définir une sorte de somme minimale de température nécessaire au débourrement, lorsque la dormance des bourgeons d'une essence donnée a été entièrement levée (Campbell et Sugano, 1979; Cannell et Smith, 1983, 1986). De cette manière, si les bourgeons ne sont pas entièrement sortis de dormance à la fin de l'hiver, une somme de degré jours plus importante sera nécessaire pour mener au débourrement. Falusi et Calamassi (1990) montrent le rôle prépondérant du froid dans le contrôle de la dormance chez le hêtre (*Fagus sylvatica* L.). Ils observent également que la photopériode semble interagir avec la température mais considèrent ce paramètre comme négligeable dans le contrôle de la dormance chez le hêtre.

Dès 1967 et à partir d'expériences menées en laboratoire, Worrall et Mergen (1967) avaient trouvé que la longueur du jour a un effet sur la levée de dormance mais prétendaient que ce n'était pas le cas en conditions naturelles. D'autres auteurs avancent, au contraire, que les conditions de lumière affectent la phénologie des espèces ligneuses, tels Wareing (1956) et Vegis (1964) qui insistent sur l'importance de la longueur du jour dans le contrôle de la levée de dormance. Farmer (1968) suggère, quant à lui, que la photopériode agit sur la totalité du cycle de la dormance. Campbell et Sugano (1979) observent que le débourrement est influencé à la fois par les températures à la sortie de l'hiver mais aussi par des effets d'interaction entre la photopériode et les températures froides durant la phase de dormance. Ces auteurs notent également une certaine variation entre les données mesurées et modélisées, qu'ils attribuent à la non prise en compte dans le modèle de certaines variables affectant le débourrement. Heide (1993b) démontre que le froid et une photopériode longue sont nécessaires à la levée de dormance et au débourrement chez le hêtre (*Fagus sylvatica* L.), contrairement aux résultats obtenus sur la même essence par Falusi et Calamassi (1990). Il montre également qu'il existe de fortes variations de comportement entre plusieurs essences originaires de l'hémisphère Nord incluses dans son étude (Heide, 1993a).

1.2.2. Modèles prédictifs de la date de débournement

Plus récemment, Chuine et Cour (1999) ont étudiés les effets, en conditions naturelles, des températures de l'été précédant la dormance et des températures froides automnales et hivernales sur le débournement de quatre essences européennes sous climat méditerranéen: *Platanus acerifolia*, *Castanea sativa*, *Quercus pubescens* et *Vitis vinifera*. La photopériode n'était pas prise en compte dans cette expérience puisque les sites comparés subissaient le même régime photopériodique. Les auteurs concluent que les températures estivales de la saison précédente n'ont aucune influence sur le débournement, qui est principalement déterminé par les conditions thermiques depuis la sortie d'endormance hivernale jusqu'au débournement, et, dans une moindre mesure, par les températures froides qui précèdent cette période. L'étude révèle également que des essences telles que *P. acerifolia* ou *Q. pubescens* sont peu influencées par les températures froides. Les auteurs avancent deux explications à cela. D'une part, un besoin élevé en températures froides retarderait d'autant le débournement alors que ces essences rencontrent peu de gelées tardives dans leur aire géographique d'origine. D'autre part, une forte demande en températures froides serait difficilement satisfaite sous un climat où les températures automnales descendent rarement en dessous de 10°C, classiquement considérée comme la limite supérieure de température froide effective. L'étude de Hunter et Lechowicz (1992) aboutissait aux mêmes conclusions quant aux déterminants de la date de débournement. Les températures froides hivernales et la somme de températures printanières suffisent, selon eux, à modéliser avec précision les dates de débournement. La prise en compte de la photopériode ne semble pas améliorer significativement la précision de ces modèles. Kramer (1994) observe même une perte de pouvoir prédictif des modèles de débournement, appliqué à *Fagus sylvatica*, lorsque la photosensibilité est incluse comme paramètre.

La date de débournement des essences forestières sous climat tempéré peut donc être modélisée grâce à trois paramètres majeurs, à savoir les températures chaudes printanières, les températures froides hivernales et la photopériode. Mais il n'existe pas de consensus quant à l'influence relative de ces différents facteurs sur la date de débournement. De plus, de nombreux autres facteurs ne sont pas pris en compte ou sont encore inconnus. Cependant, la température à elle seule (températures froides hivernales et/ou températures chaudes printanières) permet de modéliser avec une bonne précision (variations observées de l'ordre de quelques jours) la date de débournement de la plupart des essences forestières de la zone

tempérée (Chuine et Cour, 1999). On peut donc considérer, en première approche, que le régime thermique précédent le débourrement (périodes hivernale et printanière) est le principal facteur environnemental déterminant la date de débourrement.

1.2.3. Les hormones impliquées dans la dormance

De nombreuses études physiologiques ou écologiques (pour une revue, voir Rohde *et al.*, 2000) ont mis en évidence l'action des hormones végétales dans le contrôle des différents types de dormance. Ce sont principalement l'acide abscissique et les gibbérellines qui sont impliqués dans le contrôle de l'endodormance. Nous ne traiterons pas, dans cette partie, de l'auxine et des cytokinines impliquées dans la paradormance imposée aux bourgeons axillaires par le bourgeon apical⁴.

Historiquement, la découverte de l'acide abscissique dans les années 1960 est étroitement liée aux recherches menées sur la dormance des bourgeons. En effet, la découverte de la « dormine » au sein de bourgeons dormants d'*Acer pseudoplatanus*, puis extraite à partir de feuilles, a permis son isolation et sa caractérisation. Sans équivoque, l'acide abscissique (ABA) joue un rôle important dans la dormance des bourgeons mais aussi le développement et la dormance des graines, comme dans de nombreuses réponses adaptatives à des stress environnementaux (Rohde *et al.*, 2000). Mais, malgré un grand nombre d'études consacrées aux changements saisonniers des concentrations d'acide abscissique dans les bourgeons (Or *et al.*, 2000; Rinne *et al.*, 1994; Weber et Nooden, 1973) ou dans les graines (Feurtado *et al.*, 2004; Nicolas *et al.*, 1996; Pukacka et Wojkiewicz, 2002), le rôle de l'ABA en tant qu'inducteur de la dormance reste à confirmer. Les études menées sur les bourgeons ont montré que la concentration en ABA suit un cycle saisonnier avec un maximum durant la phase d'endodormance.

Les gibbérellines furent isolées pour la première fois en 1926 sur le champignon *Gibberella fujikuroi* par le scientifique japonais Eiichi Kurosawa. L'action de ces hormones sur la dormance et la croissance a principalement été étudiée chez les graines (Debeaujon et Koornneef, 2000; Jacobsen *et al.*, 2002; Schmitz *et al.*, 2001; Yamauchi *et al.*, 2004), où la germination est corrélée à une forte accumulation de gibbérellines (Jacobsen *et al.*, 2002), la voie de biosynthèse de ces dernières étant activée lors de l'imbibition de la graine (Calvo *et al.*, 2004; Yamauchi *et al.*, 2004). Chez les bourgeons, l'application de gibbérellines durant la

⁴ Phénomène de dominance apicale.

période de sortie de dormance stimule très fortement la croissance chez *Fagus sylvatica* (Falusi et Calamassi, 2003). Chez des cultivars de poirier japonais, la concentration en gibbérellines augmente très rapidement après que les bourgeons aient reçus une certaine dose de températures froides (Tamura *et al.*, 2002), ce qui tendrait à prouver que la levée de dormance des bourgeons, comme la levée de dormance des graines, est associée à une augmentation de concentration en gibbérellines, qui ont un effet positif sur la croissance. Plus récemment, Steber et McCourt (2001) ont suggéré que les brassinostéroïdes jouent un rôle dans la stimulation de la germination de la graine chez *Arabidopsis thaliana*. Dans le même temps, les résultats obtenus par Leubner-Metzger (2001) sur la germination de la graine de tabac indiquent que les brassinostéroïdes et les gibbérellines favorisent la germination de la graine à l'aide de voies de transduction du signal et de mécanismes différents.

D'un côté, l'acide abscissique et, de l'autre, les gibbérellines et brassinostéroïdes agissent de manière antagoniste sur la dormance des graines et, vraisemblablement, sur la dormance des bourgeons, même si le rôle exact de ces hormones dans le maintien de la dormance et le déclenchement du débourrement reste à élucider.

1.3. Variabilité du débourrement et changements climatiques

La date de débourrement est donc un caractère adaptatif majeur à considérer chez les arbres forestiers. En effet, ce stade phénologique joue un rôle critique dans l'initiation de la saison de végétation, dont la longueur détermine en grande partie la croissance annuelle de l'arbre. Au-delà de la productivité, la date de débourrement conditionne également les dégâts dus aux gelées tardives qui peuvent être infligés à l'arbre et, par la même, causer un retard de croissance, des défauts de croissance de la tige, voire menacer sa distribution ou sa survie (Chuine et Beaubien, 2001). C'est pourquoi, l'étude des stades phénologiques clés (débourrement, bourgeonnement, floraison...), mesurés depuis de nombreuses années sur diverses espèces, a connu un regain d'intérêt ces dernières années dans le cadre des recherches sur les conséquences probables des changements climatiques passés et à venir.

1.3.1. Variabilité géographique de la date de débourrement

L'installation et le suivi de tests de provenances mesurés, entre autres caractères, pour la date de débourrement, a permis depuis de nombreuses années de décrire et d'analyser les variations de comportement, entre espèces ou au sein de la même espèce chez les essences forestières. Ainsi Steiner (1979) étudie les relations entre les dates de débourrement de plusieurs essences de pins (*P. banksiana*, *P. flexilis*, *P. strobiformis*, *P. nigra*, *P. ponderosa*, *P. resinosa*, *P. strobus*, *P. sylvestris* et *P. virginiana*) et les facteurs géographiques et climatiques. Chez toutes les essences, excepté chez *P. resinosa*, il note que les différences observées sont fortement corrélées avec la latitude, la longitude mais aussi la proximité à l'océan des provenances considérées. Les populations du Nord et des régions continentales sont plus précoces que les populations du Sud et des régions côtières. Il existe également de fortes corrélations entre ces variations de date et les isothermes des températures moyennes du mois de janvier sur l'ensemble de l'aire de répartition des essences. De son côté, Worrall (1983) installe plusieurs provenances d'*Abies amabilis* et *Abies lasiocarpa* à différentes altitudes s'étageant de 430 à 1400 mètres. Dans chaque plantation, les provenances de haute altitude débourrent plus tôt. Selon Worrall (1983), les différences de dates de débourrement entre essences et entre provenances sont dues à la fois aux différences de sommes de température et de seuil de température pris en compte. Chez *A. lasiocarpa*, ces deux paramètres sont faibles et l'essence débourre très tôt. Chez *A. amabilis*, les seuils de

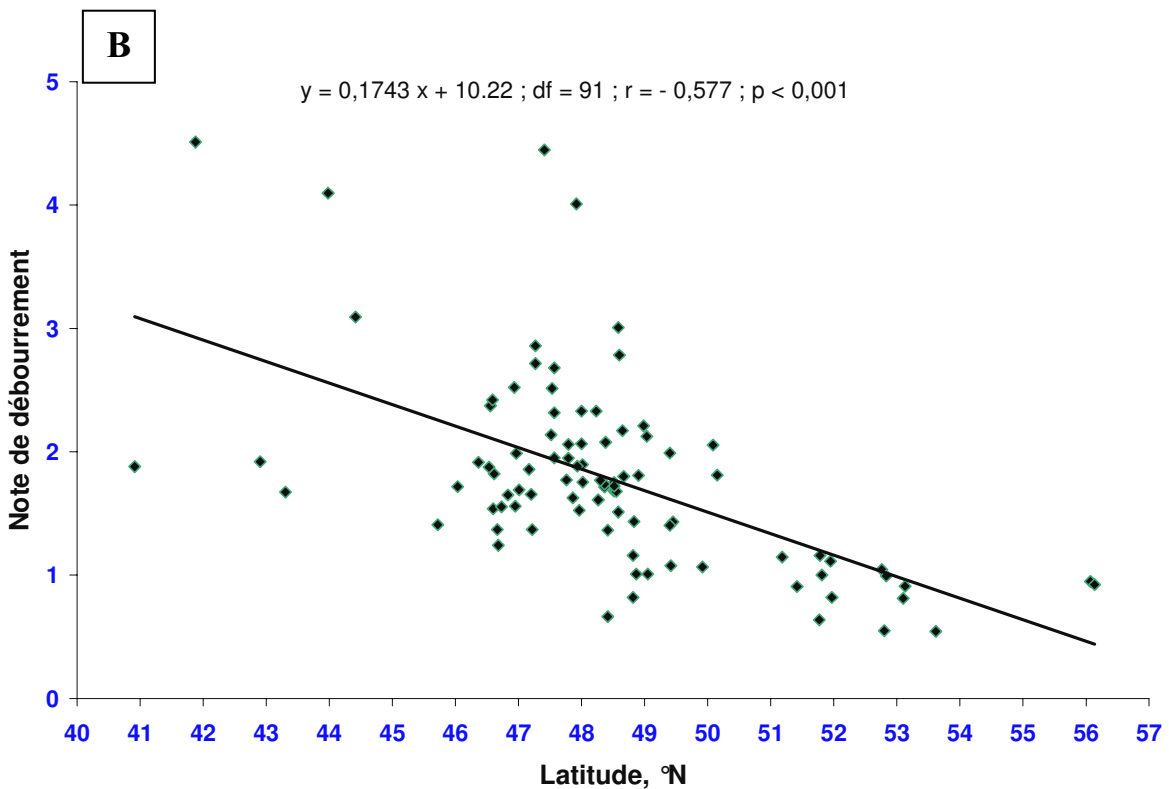
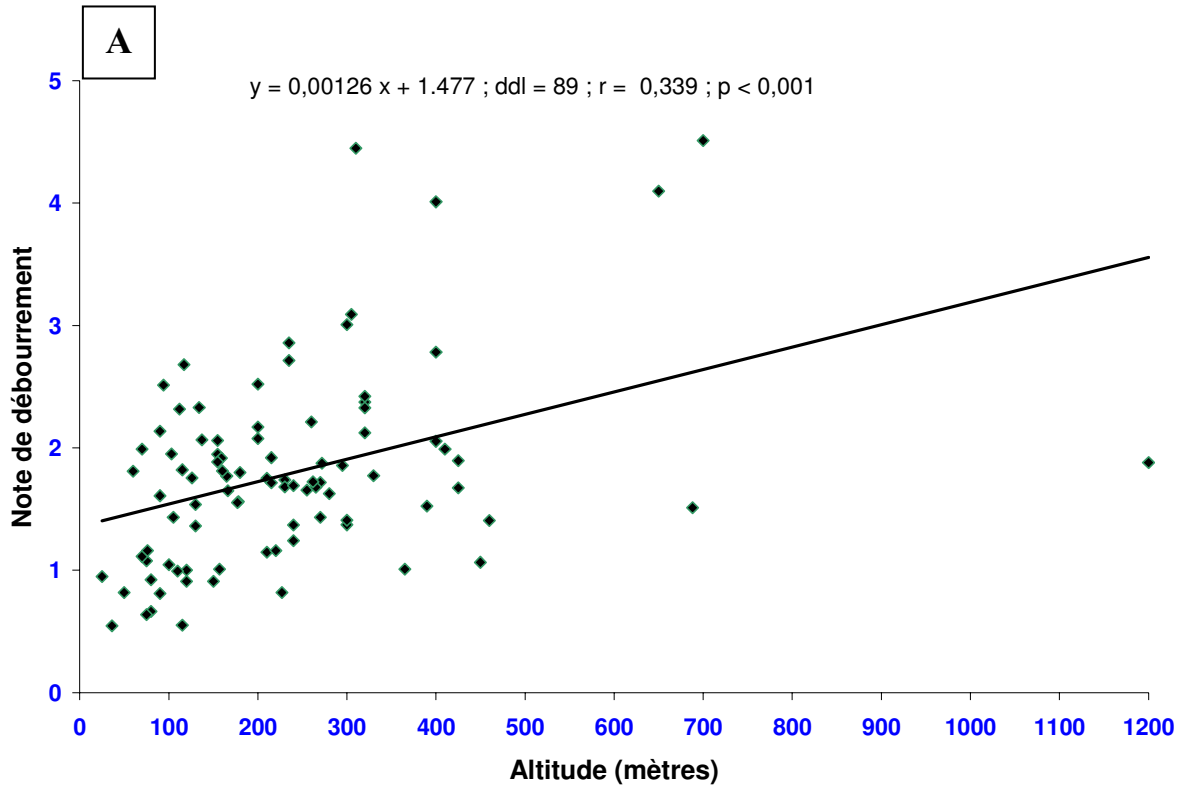


Figure 5: Diagrammes de régression entre : A. altitude et note de débourement pour 90 provenances de chêne sessile ; B. latitude et note de débourement pour 92 provenances de chêne sessile (d'après Ducouso *et al.*, 1996). L'équation de la régression, ainsi que le nombre de degrés de liberté, le coefficient de corrélation et la probabilité associée sont présentés en haut de chaque graphique.

températures efficaces sont significativement plus faibles pour les provenances de haute altitude. Entre les plantations, le débourrement est décalé d'environ 1 jour pour une élévation supplémentaire de 20 m.

Pour le chêne sessile (*Quercus petraea* (Matt.) Liebl.), le comportement de 50 populations, originaires de l'ouest de l'aire de répartition européenne, a été étudié dans quatre tests de provenances (Ducousso *et al.*, 1996) installés en France. Les variations géographiques observées de date de débourrement sont clinales et reliées à l'altitude et à la latitude. En effet, les populations du Sud et de plateau sont plus précoces que celles du Nord et de plaine. La figure 5A présente la corrélation observée entre la note de débourrement et l'altitude de la population testée, pour les 50 provenances étudiées par Ducousso *et al.* (1996) ainsi que 40 autres provenances mesurées depuis (Ducousso, communication personnelle). Le cline altitudinal observé va dans le même sens que celui reporté par Worrall (1983) chez *Abies*, à savoir que les populations d'altitude sont plus précoces que les populations de plaine. En revanche, le cline latitudinal observé (figure 5B) est inverse à celui trouvé sur plusieurs essences de pins par Steiner (1979), sur la plupart des conifères par Wright (1976) mais aussi sur chêne rouge (Ducousso *et al.*, 1997; Schlarbaum et Bagley, 1981). Il est, par ailleurs, parallèle à celui observé sur noyer (Germain, 1992).

Les variations géographiques de la date de débourrement répondent donc à des clinés à la fois latitudinaux et altitudinaux. L'origine de ces tendances n'est pas évidente à définir *a priori* mais reflètent très certainement des adaptations aux conditions de température, voire climatiques en général, et aux insectes défoliateurs.

1.3.2. Hypothèses explicatives

Les différences de comportement, décrites précédemment, ne sont donc pas le fait du hasard mais sont associées à des gradients climatiques ou géographiques. Par ailleurs, ces différences clinales s'observent à différents niveaux au sein d'une espèce et chez différentes espèces. L'ensemble de ces résultats constituent des évidences indirectes mais fortes de l'action de la sélection naturelle sur le débourrement (Howe *et al.*, 2003).

Les pressions de sélection pour la date de débourrement, exercées sur une essence forestière telle que le chêne sessile, sont essentiellement de deux ordres : abiotiques (climatiques) et biotiques (insectes défoliateurs). En effet, le chêne sessile est sensible aux gelées tardives de printemps qui causent des dégâts irréversibles aux tissus (Ducousso *et al.*,

1996). La susceptibilité aux dégâts de gel est fortement corrélée à la date de débourrement et les individus les plus précoces sont les plus touchés. Inversement, les individus plus tardifs sont moins sensibles à ce type de dommages climatiques. Ces résultats confirment ceux obtenus par Liepe (1993) en chambre de croissance, sous conditions contrôlées. La sélection naturelle aura certainement favorisé les génotypes tardifs qui ne souffrent pas de dégâts dus au gel. A l’opposé, les provenances précoces seraient favorisées par la sélection en conditions climatiques plus clémentes afin d’optimiser la longueur de la saison de végétation. La sélection favorisant les génotypes tardifs peu sensibles aux dégâts de gel est donc contre balancée par la sélection favorisant des génotypes précoces, qui auront un avantage sélectif pour la croissance en conditions favorables.

Quelques explications peuvent être avancées quant aux clines observés dans le cas des espèces autres que le chêne pour lesquelles les provenances septentrionales sont plus précoces que les provenances méridionales, quelle que soit la latitude du lieu d’installation des tests. La sélection aurait favorisé les génotypes précoces au Nord leur conférant ainsi un avantage pour la longueur de la saison de végétation, nécessairement limitée par les conditions climatiques à ces latitudes. Les provenances du Sud seraient plus sensibles au gel et la sélection aurait donc favorisé les génotypes tardifs, leur permettant de minimiser les risques liés au gel.

Les différences de date de débourrement peuvent également résulter des pressions de sélection exercées par les insectes défoliateurs. Crawley et Akhteruzzaman (1988) ont montré que les différences de phénologie au sein de populations de chêne pédonculé (*Quercus robur* L.) mènent à des différences substantielles de la structure de la communauté des insectes herbivores associés. Les attaques de certains insectes défoliateurs se produisent à des stades bien particuliers du débourrement de l’arbre. Même si l’arbre ne meurt pas de la défoliation, sa croissance est réduite et sa fitness en est réduite. Il en résulte donc que les arbres qui ne présentent pas le stade de sensibilité aux défoliateurs seront favorisés. Parallèlement, le stade phénologique de l’hôte au moment de l’émergence de la population d’insectes affecte à la fois les performances mais également la colonisation de l’espèce hôte (Tikkanen et Tiitto, 2003). A ce titre, les variations de phénologie peuvent être considérées comme un mécanisme actif de lutte contre les populations d’insectes défoliateurs (Tikkanen et Tiitto, 2003; Tuomi *et al.*, 1989). Ainsi, chez le pin Douglas, des taux de croissance élevés et un débourrement tardif favorisent la résistance de l’arbre aux attaques de vers défoliateurs du bourgeon (Chen *et al.*, 2001). C’est pourquoi les variations des cohortes d’insectes et de leur abondance à travers l’aire de distribution naturelle des essences forestières est potentiellement à l’origine d’un gradient phénologique, au même titre que les conditions climatiques. On notera, par ailleurs,



Figure 6 : Réseau de stations IPG réparties au long de 28 latitudes de la Scandinavie à la Macédoine et de 37 longitudes de l'Irlande à la Finlande au Nord et du Portugal à la Macédoine au Sud. Le réseau compte 57 jardins et couvre différentes régions climatiques de climat froid à chaud et depuis les façades maritimes aux zones continentales de l'Europe.

(Source : http://www.agrar.hu-berlin.de/pflanzenbau/agrarmet/ipg_2.html)

que les changements climatiques, en particulier les printemps plus chauds, peuvent perturber la synchronie de phénologie entre l'hôte et l'insecte comme cela a été observé entre le chêne pédonculé (*Quercus robur* L.) et un papillon défoliateur (*Operophtera brumata*) (Visser et Holleman, 2001). Les pressions de sélection exercées par les insectes défoliateurs peuvent être à l'origine de la forte variabilité de date de débourrement observée chez le chêne favorisant ainsi l'apparition de génotypes qui ne présentent pas le stade de sensibilité aux défoliateurs.

1.3.3. Impact des changements climatiques sur la date de débourrement

La communauté scientifique dans son ensemble a abouti à un consensus concernant les changements climatiques globaux : sans contexte, le climat se réchauffe. Ce réchauffement résulte en grande partie des émissions de gaz carbonique (CO₂) imputables aux activités humaines telles que les processus industriels, la combustion des énergies fossiles ou l'utilisation des espaces naturelles par la déforestation. Alors que le climat terrestre dans son ensemble s'est réchauffé d'environ 0,6°C au cours des 100 dernières années, la décennie des années 90 aura été la plus chaude de tout le XX^{ème} siècle, et peut être même du millénaire, et les années 1998, 2001 et 2002 ont été les plus chaudes jamais enregistrées. Les projections envisagent une augmentation de température comprise entre 1,4°C et 5,8°C d'ici à 2100 (IPCC, 2001).

Bien que la survie des essences forestières ne semble pas menacer à cours terme par les changements climatiques, il est certain que ces derniers vont fortement affecter la productivité, l'abondance et la distribution des espèces ligneuses (Chuine et Beaubien, 2001). L'impact des changements climatiques, observés au siècle précédent, sur la phénologie a déjà été mesuré sur diverses espèces végétales (Penuelas et Filella, 2001; Walther *et al.*, 2002). Dès à présent, la floraison de *Populus tremuloides* au Canada occidental a été avancé de 26 jours tout au long du XX^{ème} siècle (Beaubien et Freeland, 2000). En Europe, la longueur de la saison de végétation s'est accrue d'environ 10,8 jours depuis les années 1960 (Menzel et Fabian, 1999). Ces résultats ont été obtenus à partir de l'analyse des observations faites au sein des Jardins Phénologiques Internationaux (IPG), qui couvrent une grande partie de l'Europe. Ces jardins contiennent à la fois des arbres et des arbustes (Figure 6) mesurés pour une dizaine de stades phénologiques particuliers dont le débourrement du bourgeon. Les auteurs montrent également, à l'aide de simulations basées sur des modèles phénologiques,

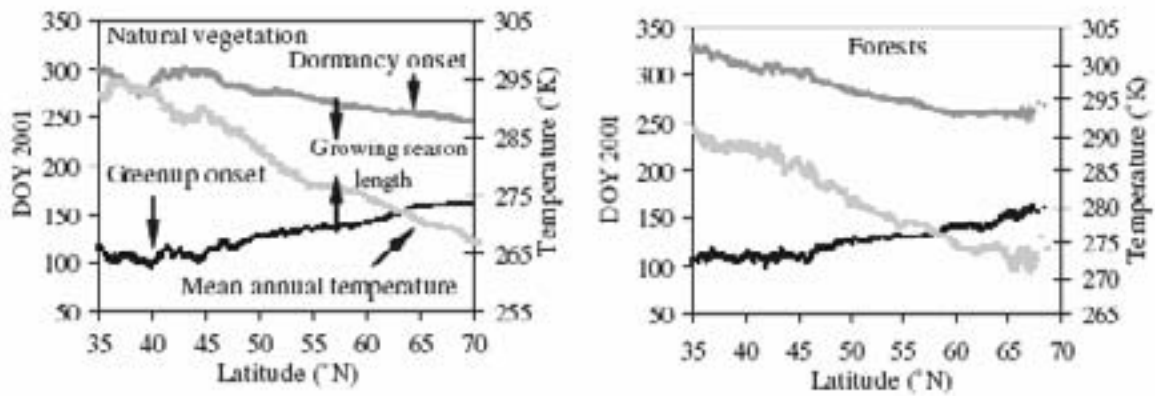


Figure 7 : La phénologie comme une fonction du type de couverture végétal, de la latitude et de la température à la surface terrestre pour l'Europe et l'Asie. Données obtenues à partir de l'analyse de données satellitaires (tiré de Zhang *et al.*, 2004).

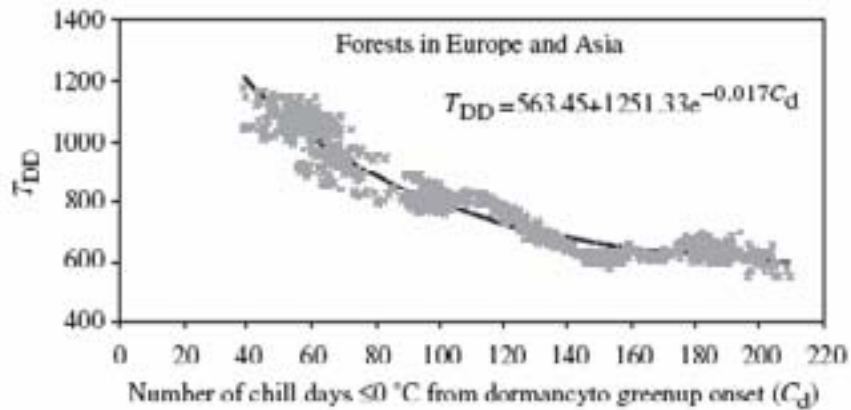


Figure 8 : Relation entre le nombre de degré jours jusqu'au débournement et le nombre de jours froids accumulés pour les forêts en Europe. T_{DD} représente le nombre de degré jours en base 0 depuis le début de dormance jusqu'au débournement (tiré de Zhang *et al.*, 2004).

que le débourrement pourrait être avancé jusqu'à 6 jours, en fonction de l'espèce, pour une augmentation de la température hivernale de l'air de 1 °C (Menzel et Fabian, 1999). Pour les Etats-Unis, White *et al.* (1999) ont mesuré une augmentation de la saison de croissance de 5 jours entre 1900 et 1987 alors que la température moyenne annuelle augmentait de 1°C dans le même temps.

L'augmentation apparente de longueur de la saison de végétation est en accord avec les résultats obtenus par analyse de données satellitaires. Ces études se basent sur l'indice de végétation normalisé (NDVI) qui permet de déterminer les stades phénologiques de couverts végétaux à partir d'images satellites. Ainsi, Myneni *et al.* (1997) ont mesuré une avancée de 8 ± 3 jours de la reprise de croissance printanière et un retard de 4 ± 2 jours de la fin de la saison de végétation entre 1981 et 1991. L'étude menée par Zhang *et al.* (2004), à l'échelle du globe, confirme la dépendance de la phénologie à l'égard de la température et de la latitude (Figure 7). Par ailleurs, ils observent également que les couverts forestiers sont les plus sensibles aux variations de température et mesurent une variation de longueur de saison de végétation de 6,3 jours/°C pour les forêts européennes. Enfin, les auteurs ont testé, à l'échelle du continent européen, la fiabilité avec laquelle le débourrement pouvait être prédit à partir du nombre de jours froids et de l'accumulation de degrés-jours. La relation obtenue est donnée à la figure 8 et le modèle permet d'expliquer 94% de la variation de thermopériode. Bien que l'inférence des stades phénologiques d'après les données satellites nécessite d'être améliorée afin de mieux correspondre aux observations faites au sol (Badeck *et al.*, 2004), ces modèles constituent de formidables outils de prévision des changements de phénologie en réponse aux changements climatiques.

La date de débourrement est donc un caractère clé de la phénologie des essences forestières telles que les chênes blancs européens. Elle s'inscrit dans le cycle annuel d'alternance de phases de croissance et de dormance, qui assure à l'arbre une synchronie vis-à-vis des conditions de milieu, et notamment la phase hivernale très défavorable à la croissance. Le caractère est à la fois sous l'influence de facteurs biotiques (populations d'insectes défoliateurs) et abiotiques (conditions climatiques) et peut être modélisé en première approche à l'aide de la température. Mais au-delà des modèles prédictifs, la nécessité est grande de disséquer les mécanismes moléculaires contrôlant le phénomène de débourrement afin de mieux comprendre l'effet du climat sur ce caractère.

② Le débourrement : du phénotype aux gènes

Les valeurs de débourrement présentent une variation phénotypique continue de type loi normale caractéristique d'un caractère dit quantitatif, ou complexe, placé sous la dépendance d'un grand nombre de gènes et de facteurs environnementaux. La valeur phénotypique (P) peut ainsi se décomposer en une valeur génotypique (G), sous le contrôle d'un certain nombre de gènes, et d'une valeur environnementale (E) induite par les facteurs du milieu. Si l'on néglige les effets d'interaction entre composantes, l'équation $P = G + E$ résume ce principe. La génétique quantitative, grâce à la modélisation mathématique de la transmission génétique des caractères complexes, vise à caractériser les effets des facteurs génétiques impliqués dans l'expression de caractères complexes. Ainsi, à travers l'estimation de paramètres statistiques tels que l'héritabilité, l'écart-type phénotypique ou le coefficient de corrélation génétique, nous pouvons estimer l'importance relative des facteurs génétiques ou environnementaux sur le contrôle du caractère. Cependant, cette approche ne suffit pas à connaître le nombre et l'effet des différents gènes sur l'expression du caractère, la localisation de ces gènes sur les chromosomes ainsi que leur fonction physiologique. Dans un premier temps, l'utilisation de marqueurs moléculaires qui ségrégent dans une descendance issue d'un croisement contrôlé permet la construction de cartes génétiques, qui sont un agencement linéaire de locus le long d'un chromosome, et dont les positions relatives ont été déterminées à partir des taux de recombinaison (unité de distance : le centimorgan). A partir de ces cartes, la décomposition génétique des caractères quantitatifs en composantes mendéliennes permet de déterminer le nombre de locus en cause, leur position sur le génome et leurs effets. Cette détermination est obtenue par la recherche systématique d'associations statistiques entre les variants alléliques aux marqueurs cartographiés et la variation du caractère.

Cependant, cette approche ne permet pas l'identification directe des gènes impliqués dans la variabilité du caractère mais seulement des zones chromosomiques (QTL : Quantitative Trait Locus, Geldermann, 1975) qui les contiennent. Le développement d'outils moléculaires de diagnostic pour le caractère de date de débourrement passe donc par une identification des gènes responsables de la variation génétique observée. Pour répondre à cela, une stratégie complémentaire au clonage positionnel ou à la mutagenèse d'insertion est l'approche gène candidat. Le concept de gènes candidat, qui a vu le jour dans le domaine de la génétique humaine est le suivant : un gène candidat est un gène qui se trouve dans une région génomique impliqué dans une maladie donnée, et dont le produit protéique suggère qu'il soit

le déterminant génétique de la maladie. Une acceptation plus large du terme a vu le jour et le concept de gène candidat peut reposer sur des données d'expression, de fonction ou de position sur des cartes génétiques. Le degré d'implication du gène désigné comme tel dans la variation génétique observée du caractère peut être estimé à l'aide d'études d'associations mais la validation *sensu stricto* ne peut se faire qu'à l'aide d'expériences de transformation génétique ou de complémentation fonctionnelle.

2.1. Déterminisme génétique du caractère de débourrement

La date de débourrement pour un individu donné est donc déterminée, en proportions variables, par le génotype et l'environnement. Les outils statistiques de la génétique quantitative ont permis d'estimer la proportion de variabilité phénotypique imputable aux effets génétiques (héritabilité du caractère). Grâce à la détection de QTL, des locus contrôlant cette variabilité génétique ainsi que leur position respective sur le génome et leurs effets ont été décrits. L'ensemble de ces résultats permet de mieux appréhender le déterminisme génétique de la date de débourrement.

2.1.1. Un contrôle génétique modéré à fort

Sur le plan évolutif mais aussi sélectif (au sein de programmes d'amélioration), la question du contrôle génétique des caractères, du moins la composante additive de la variance génétique, est centrale puisqu'elle détermine l'intensité de la réponse du caractère à la sélection. De ce point de vue, la date de débourrement chez le chêne, et les arbres forestiers en général, est intéressante puisque soumise à un contrôle génétique modéré à fort chez toutes les espèces étudiées à ce jour.

En effet, dès 1978, Eriksson *et al.* (1978) avaient observé que le débourrement chez *Picea Abies* était soumis au contrôle de multiples gènes. Plus récemment, des essais clonaux au champ de peupliers ont permis de montrer que la date de débourrement est parmi les caractères mesurés les plus héritables puisque 98% de la variation phénotypique totale peut être due à des facteurs génétiques, et seulement 2% à des facteurs environnementaux ou autres (Bradshaw et Stettler, 1995). Ce résultat a été confirmé ensuite par Frewen *et al.* (2000) qui estiment l'héritabilité clonale moyenne à 94% pour le débourrement chez *Populus*. Chez le sapin de Douglas (*Pseudotsuga menziesii* (Mirb.) Franco var. *menziesii*), l'héritabilité individuelle ou familiale de la date de débourrement est comprise entre 0,44 et 0,95 (pour une revue, voir Jermstad *et al.*, 2001). Les valeurs sont très similaires chez *Salix* puisque l'héritabilité clonale moyenne est estimée entre 0,43 et 0,72 (Tsarouhas *et al.*, 2003). Chez le pin noir (*Pinus nigra*), l'héritabilité au sens large pour la date de débourrement a été estimée à 0,69 (Matziris, 1994).

Concernant le chêne pédonculé (*Quercus robur*), Jensen (1993) a obtenu une valeur élevée de l'héritabilité avec $h^2=0,87$ mais l'auteur précise que les valeurs d'héritabilité de

cette étude sont à prendre avec précaution et doivent être comparées à des estimations obtenues dans des conditions similaires (mêmes provenance, âge, traitement ...). Ces valeurs sont effectivement beaucoup plus élevées que celles obtenues par Ducouso (communication personnelle) avec une héritabilité estimée à 0,3 chez le chêne sessile et 0,47 chez le chêne pédonculé. Enfin, dans l'étude de Scotti-Saintagne (2004), le caractère de date de débourrement possède une héritabilité estimée entre 0,15 et 0,51, selon le caractère mesuré et le modèle utilisé (variance de dominance nulle ou égale à $\frac{1}{2}$ de la variance additive).

Quelle que soit l'espèce étudiée, le caractère de débourrement présente donc des héritabilités moyennes ($\approx 0,3-0,5$) à fortes ($\approx 0,5-0,8$). Par comparaison à d'autres caractères phénotypiques pour lesquelles des estimations d'héritabilité sont disponibles chez les arbres, le débourrement est le caractère le plus héritable, avec la densité du bois. Ces résultats plaident en faveur d'un contrôle génétique modéré à fort du caractère, qui se trouve donc sous le contrôle d'un certain nombre de gènes à effet majeur.

2.1.2. Principes et limites de la cartographie génétique et l'approche QTL

La décomposition génétique d'un caractère en composantes mendéliennes comme méthode permettant de déterminer le nombre de locus mis en cause, leur position sur le génome et leurs effets, requiert l'existence d'une carte génétique de l'espèce considérée. Une carte génétique est une représentation, à l'aide de marqueurs génétiques, des chromosomes d'une espèce. Les marqueurs utilisés peuvent être morphologiques, biochimiques ou moléculaires et doivent être polymorphes et présenter une ségrégation de type mendélienne. La plupart des marqueurs utilisés actuellement sont moléculaires et il en existe de nombreux soit dominants (AFLP, RAPD, ISSR), soit co-dominants et locus spécifiques (RFLP, SSR, ESTP).

La première étape de la cartographie consiste en la sélection des parents du croisement et la création de la descendance en ségrégation. Les parents sont choisis de sorte à maximiser les chances d'obtenir du polymorphisme moléculaire. La caractérisation moléculaire des individus issus du croisement va ensuite permettre le calcul des fréquences de recombinaisons entre marqueurs. Ce phénomène de recombinaison, aussi appelé « crossing-over », apparaît au cours de la prophase de la première division méiotique par échange de matériel génétique entre les chromosomes homologues maternels et paternels. C'est au sein de l'équipe du généticien Thomas Hunt Morgan (1866-1945) que l'idée apparut de lier le concept de

recombinaison génétique avec la distance physique séparant deux locus. Cette fréquence est effectivement d'autant plus élevée que la distance entre les deux marqueurs sur le chromosome est grande. Une fois la ségrégation mendélienne des marqueurs vérifiée, la construction de la carte génétique consiste donc à détecter les liaisons entre marqueurs et à réunir les marqueurs au sein de groupes de liaison, représentatifs des chromosomes au niveau physique. Les marqueurs sont ensuite ordonnés au sein de chaque groupe et les distances entre eux estimées en centiMorgan.

A partir de cet agencement de marqueurs le long d'un groupe de liaison, il est possible d'estimer le nombre, la position et les effets des locus contrôlant le caractère considéré (Tanksley, 1993). Le principe de la détection de QTL est relativement simple. Il s'agit en effet de rechercher de manière systématique les marqueurs dont les allèles en ségrégation dans la descendance montre une liaison statistique significative avec la valeur phénotypique du caractère mesuré sur ces mêmes individus. Si le principe est simple, en revanche les méthodes statistiques de détection des QTL sont complexes et évoluent rapidement (Doerge, 2002). La détection des QTL peut aboutir, dans certains cas, au clonage du gène impliqué. C'est actuellement le cas de quelques dizaines de gènes, que ce soit dans le règne animal ou végétal (Glazier *et al.*, 2002; Paran et Zamir, 2003). Cependant, dans le cas des arbres forestiers, les effectifs souvent limités (≈ 200 individus) des populations de cartographie ne permettent pas d'estimer de façon précise la position des QTL ou de réaliser leur cartographie fine. Chez les essences de conifères, c'est aussi la taille du génome qui empêche la mise en pratique du clonage positionnel. Au-delà de ces limitations portant sur le clonage du gène impliqué dans la variation du caractère phénotypique, de très nombreux paramètres influencent la détection des QTL et doivent être pris en compte lors de l'interprétation des résultats. Nous retiendrons principalement l'héritabilité et l'effectif de la descendance. En effet, plus l'effectif de la population sera faible, plus la puissance de détection des QTL sera faible, leur localisation imprécise et leurs effets surestimés (Beavis, 1995). Parallèlement, si l'héritabilité est faible, il est nécessaire de répéter plusieurs fois la recherche des QTL dans des environnements contrastés afin d'obtenir une estimation fiable des effets et du nombre minimal de QTL impliqués, ou de multiplier végétativement les individus de la population de cartographie.

Malgré les limites inhérentes à la technique de détection des QTL et celles particulières liées aux arbres forestiers, l'approche QTL permet de connaître le nombre minimal ainsi que les effets des locus mis en cause dans le déterminisme génétique d'un caractère tel que le débourrement et cette approche a largement été utilisée chez les arbres forestiers.

Tableau 1 : Synthèse des études de détection de QTL pour le débourrement menées chez les arbres forestiers.

Espèce	h²	Effectif et type de descendance	Nombre de QTL détectés	Variance phénotypique expliquée	Référence
<i>Pseudotsuga menziesii</i>		78-190 (F ₂) deux sites	4,1/caractère/année (4 années différentes)	1,2-11,6%	Jermstad <i>et al.</i>, 2001a
<i>Salix viminalis</i> x <i>S. schwerinii</i>	0,43-0,72	92 (F ₁)	1,5/caractère/deux âges	12-24%	Tsahouras <i>et al.</i>, 2003
<i>Populus trichocarpa</i> x <i>P. deltoides</i>	0,98	55 (F ₁) répétitions	5	28,7-51,5%	Bradshaw et Settler, 1995
<i>Populus trichocarpa</i> x <i>P. deltoides</i>	0,94	346 (F ₂)	9	5,9-16,8%	Frewen <i>et al.</i>, 2000
<i>Quercus robur</i> L.	0,15-0,51	89-220 (F ₁) répétitions	12/trois sites	3,1-10,7%	Saintagne <i>et al.</i>, 2003
<i>Castanea sativa</i>		153 (F ₁)	7/année (trois années)	3,2-12,2%	Casasoli <i>et al.</i>, 2004

2.1.3. Les QTL de débourrement chez les arbres forestiers

Deux revues récentes (Sewell et Neale, 2000; van Buijtenen, 2001) ont fait le point des principaux résultats de détection de QTL obtenus chez les arbres forestiers. Outre le débourrement, ou plus généralement les caractères liés à la phénologie, les arbres forestiers ont été étudiés pour la croissance (Plomion *et al.*, 1996; Tsarouhas *et al.*, 2002), la qualité du bois (Sewell *et al.*, 2000), la multiplication végétative (Grattapaglia *et al.*, 1995) ou encore l'efficacité d'utilisation de l'eau (Brendel *et al.*, 2002). Les résultats de détection de QTL effectuée chez les arbres forestiers sur le caractère de débourrement sont synthétisés dans le tableau 1.

Concernant le débourrement, la première remarque est que le nombre de QTL détecté est faible alors même que l'héritabilité estimée pour le caractère est forte. Or, on sait que la puissance d'analyse sera d'autant plus forte que l'héritabilité est forte (Beavis, 1995). Le maximum observé est de 13 pour le chêne pédonculé mais l'étude a été faite sur trois sites et 8 sont communs à l'ensemble des trois. Les contributions des QTL à la variance phénotypique varient en fonction des études puisque les valeurs s'échelonnent de 1,2% dans le cas d'un QTL détecté sur *Pseudotsuga menziesii* à 51,5% dans le cas d'un QTL détecté sur le croisement *Populus trichocarpa* x *P. deltoides*. Il faut cependant noter que cette dernière étude a été conduite sur 55 individus et que les effectifs faibles entraînent une surestimation des effets des QTL (Beavis, 1995). La deuxième valeur la plus élevée de variance phénotypique expliquée est trouvée chez le croisement *Salix viminalis* x *S. schwerinii*, où l'effectif est de 92 individus. Donc, en tenant compte de l'effectif de la descendance, les contributions des QTL à la variance phénotypique demeurent faibles pour l'ensemble de ces études. Par ailleurs, comparés aux QTL détectés pour d'autres caractères, les QTL impliqués dans le débourrement sont plus stables entre âges et environnements différents (Casasoli, 2004).

Dans le cas particulier du chêne, la recherche de QTL contrôlant le caractère de débourrement a été effectuée au sein de l'UMR Biogeco dans le cadre de la thèse de Caroline Scotti-Saintagne (2003; 2004). La population de cartographie utilisée était constituée de 278 descendants plein frères avec un parent femelle (3P) originaire du site de Pierroton, et un parent mâle (A4) originaire d'Arcachon. Ces individus ont été propagés végétativement et installés sur trois sites expérimentaux, mesurés durant trois années. Ce dispositif a permis la recherche et la localisation des QTL pour le débourrement, ainsi que pour des caractères de

croissance, non présentés ici. A travers les trois sites expérimentaux, 32 QTL indépendants ont été détectés, qui représentent au moins 12 gènes ou régions chromosomiques uniques contrôlant le caractère. Ces derniers expliquent 3,1 à 10,7% de la variance phénotypique clonale observée. Les auteurs concluent que le débourrement est sous le contrôle de plusieurs QTL à effets faible à modéré, mais qui contribuent à une forte variance génétique additive.

Afin d'atténuer les limites discutées au paragraphe précédent, concernant la détection des QTL chez les arbres forestiers, une solution serait d'effectuer la détection sur plusieurs années et dans plusieurs environnements ou encore de comparer les résultats obtenus pour le même caractère chez des espèces phylogénétiquement proches. C'est cette dernière stratégie qui a été employée chez les Fagacées en comparant les QTL détectés chez le chêne pédonculé (*Quercus robur L.*) et le châtaignier (*Castanea sativa*), dans le cadre de la thèse de Manuela Casasoli (2004) effectuée au sein de l'UMR Biogeco. Une partie du travail a été effectuée en collaboration et les résultats seront présentés dans le chapitre C ainsi que dans l'article en Annexe II de cette thèse.

Mais au-delà de ces recherches sur le déterminisme génétique de la date de débourrement, l'exploitation des connaissances physiologiques à travers l'analyse de mutants, les études d'expression ou l'approche gène candidat permettent d'aller vers une meilleure compréhension des mécanismes moléculaires mis en œuvre au cours du débourrement.

2.2. Les bases moléculaires du débourrement

Les études physiologiques et écologiques menées sur le débourrement ont permis d'accumuler une somme importante d'informations (cf. § 1) exploitées au niveau moléculaire, notamment par le biais d'analyse de mutants ou d'expériences de transformation sur des plantes modèles. Avant toute chose, nous commencerons par présenter les résultats issus d'études portant sur la dormance et la germination des graines qui peuvent être exploités, dans une certaine mesure, pour l'étude du phénomène de débourrement. Nous réaliserons ensuite une synthèse des résultats disponibles sur la levée de dormance du bourgeon, particulièrement concernant les transcrits relatifs au stress oxydatif, au métabolisme et au cycle cellulaire.

2.2.1. Dormance de la graine et dormance du bourgeon : des phénomènes similaires ?

Dès 1956, Wareing (1956) avait émis l'hypothèse que la dormance de la graine et du bourgeon avaient des caractéristiques communes. Plusieurs raisons motivaient et motivent encore cette hypothèse : i/ dans les deux cas une période de températures froides hivernales est requise pour la levée de dormance, ii/ les hormones ont une influence similaire sur la stimulation ou l'inhibition des deux processus, et iii/ l'acide abscissique et les gibbérellines sont intimement liés aux deux phénomènes. Même s'il apparaît que les études relatives aux graines ne peuvent être directement extrapolées aux bourgeons, les gènes impliqués dans la dormance ou sa levée chez les graines sont des candidats de premier intérêt à tester sur les bourgeons (Rohde *et al.*, 2000).

Aalen *et al.* (1994) ont ainsi identifiés des transcrits spécifiquement exprimés au cours de la dormance de la graine chez l'orge (*Hordeum vulgare*) présents à la fois dans la couche à aleurone et dans l'embryon. D'autres auteurs ont isolé des transcrits différenciellement exprimés, au sein d'embryons de caryopses de folle avoine (*Avena fatua*), entre les graines dormantes et non dormantes (Johnson *et al.*, 1995). Certains transcrits peuvent également être spécifiquement exprimés au moment de la levée de dormance, comme ceux identifiés par Bailey *et al.* (1996) au cours de la levée de dormance et la germination de graines de *Trollius ledebouri*. Des travaux similaires menés sur *Pseudotsuga menziesii* ont abouti à l'isolation et la caractérisation de trois transcrits dont l'expression est stimulée au cours de la levée de dormance de graines soumises à un traitement de stratification (Jarvis *et al.*, 1996). Des études

Tableau 2 : Gènes impliqués dans la dormance ou la germination de la graine.

Nom du gène	Organisme	N° d'accèsion	Nombre d'homologues protéiques TrEMBL	Implications	Références
Gibberellin 3β-hydroxylase GA4 et GA4H	<i>A. thaliana</i>	L37126 AF070937	24	GA4, GA4H et GA20ox catalysent les dernières étapes dans la voie de biosynthèse des gibbérellines actives (GA1 et GA4) et sont surexprimés lors de la germination de graine d' <i>Arabidopsis</i> . Par ailleurs, il semblerait que l'étape de 20-oxydation soit plus limitante que celle de 3 β -hydroxylation dans la synthèse des gibbérellines actives.	Chiang <i>et al.</i> , 1995 Yamaguchi <i>et al.</i> , 1998 Israelsson <i>et al.</i> , 2004 Calvo <i>et al.</i> , 2004 Perez-Flores <i>et al.</i> , 2003
Gibberellin 20-oxidase GA20ox	<i>C. maxima</i>	AJ308480	82		
Abscisic acid insensitive 3 ABI3	<i>A. thaliana</i>	AJ002473	6	Le niveau d'expression d'ABI3 est corrélé à la phase de dormance des graines et son rôle potentiel pourrait se situer au niveau post-transcriptionnel.	Giraudat, 1992 Rohde <i>et al.</i> , 1999 Zeng <i>et al.</i> , 2003
Brassinosteroid insensitive 1 BRI1	<i>A. thaliana</i>	AF017056	9	BRI1 et DET2 sont impliqués dans la voie de perception et de transduction du signal des hormones brassinostéroïques et semblent jouer un rôle dans la stimulation de la germination des graines d' <i>A. thaliana</i> .	Li and Chory, 1997 Li <i>et al.</i> , 1996 Steber and McCourt, 2001
Steroid reductase DET2	<i>A. thaliana</i>	U53860	5		
Dof affecting germination DAG1 et DAG2	<i>A. thaliana</i>	AJ224122 AJ237810	1	DAG1 et DAG2 agissent de manière antagoniste et semblent contrôler la germination de la graine chez <i>Arabidopsis</i> en régulant les mêmes gènes.	Yanagisawa, 2002 Gualberti <i>et al.</i> , 2002

plus récentes ont également permis de mieux caractériser au niveau moléculaire l'implication des gibbérellines et de l'acide abscissique ainsi que le rôle central du cycle cellulaire dans la germination des graines (Calvo *et al.*, 2004; Feurtado *et al.*, 2004; He et Gan, 2004; Pawlowski *et al.*, 2004). La revue proposée par Koornneef *et al.* (2002) fait le point des études menées sur la dormance et la germination de la graine, principalement sur *Arabidopsis thaliana* mais aussi sur quelques céréales ou Solanacées telles que le tabac ou la tomate. Les résultats présentés dans cette revue proviennent de recherches réalisées à l'aide de mutants ou par suivi d'expression de gènes au cours de la dormance et de la germination. Le tableau 2 présente une liste de gènes dont le rôle dans le débourrement pourrait être testé au regard de leur implication dans la germination de la graine.

A côté des nombreuses études menées sur la levée de dormance et la germination des graines, certains auteurs se sont intéressés à la levée de dormance et au débourrement des bourgeons.

2.2.2. *Stress oxydatif et métabolisme au cours du débourrement*

Le contraste est grand entre la somme d'études consacrées à la germination des graines et celles relatives au débourrement. Outre les aspects hormonaux précédemment abordés, les premières études moléculaires conduites sur le débourrement l'ont été sur des bourgeons axillaires en « forçant » le débourrement soit par décapitation du bourgeon terminal afin de lever la dominance apicale, soit par application de substances chimiques.

Ainsi, les recherches menées par Wang, dans les années 80 et 90, se sont intéressées aux conséquences métaboliques du débourrement de branches latérales de pommier (*Malus domestica*) provoqué par application de thidiazuron, régulateur de croissance de la famille des diphénylurée ayant une activité de type cytokinine. Les auteurs observent une augmentation d'activité de la superoxyde dismutase corrélée au débourrement induit par le thiadiazuron (Wang *et al.*, 1991c). L'activité de la catalase augmente au cours de la levée de dormance et décroît lorsque le bourgeon se met à croître (Wang *et al.*, 1991d). L'activité de la peroxydase est élevée durant la dormance, décroît lors de la réhydratation du bourgeon, puis augmente au moment du débourrement et décroît à nouveau après l'expansion du bourgeon. L'activité de la polyphénol oxydase est inverse à celle trouvée pour la peroxydase durant le développement du bourgeon (Wang *et al.*, 1991d). Les activités enzymatiques relatives à l'ascorbate et au glutathione ont également été étudiées (Wang *et al.*, 1991a). D'une manière générale, le

débourrement induit par le froid, la chaleur ou le disulfure d'allyle et de propyle est associé à l'activation des systèmes anti-oxydants (Wang et Faust, 1994), menant à l'élimination des radicaux libres. Parallèlement, les enzymes de la glycolyse (glyceraldéhyde-3-phosphate déshydrogénase (GAPDH), pyruvate kinase (PK), isocitrate déshydrogénase (ICDH)) sont activées durant le débourrement tandis que l'activité des enzymes du cycle des pentoses phosphate (glucose-6-phosphate déshydrogénase (G6PDH), 6-phosphate gluconate déshydrogénase (6-PGDH)) est plus forte dans les bourgeons dormants que dans les bourgeons non dormants (Wang *et al.*, 1991b).

Le phénomène de débourrement est donc associé à des modifications des voies métaboliques, en particulier du métabolisme des carbohydrates avec la glycolyse ou le cycle des pentoses phosphates. Un stress oxydatif très prononcé est également associé à la reprise d'activité du bourgeon.

2.2.3. La régulation du cycle cellulaire

Au cours des années 90, les recherches sur le débourrement se sont orientées vers l'activité du cycle cellulaire. Ainsi, Devitt et Stafstrom (1995) ont étudié la régulation du cycle cellulaire au cours de cycles d'alternances croissance et dormance chez des bourgeons axillaires de pois (*Pisum sativum*). Les auteurs ont suivi le niveau d'expression d'un certain nombre de transcrits relatifs au cycle cellulaire (histones H2A et H4, MAP et cdc2 kinases, cycline B). Le niveau de chacun de ces transcrits est faible dans les bourgeons dormants, augmente lorsque la croissance des bourgeons secondaires est stimulée par décapitation du bourgeon terminal, diminue lorsque les bourgeons cessent de croître et deviennent dormants puis augmente à nouveau lorsque les bourgeons se remettent en croissance. Par cytométrie de flux, les auteurs ont également montré que les cellules de bourgeons dormants étaient arrêtées en trois points du cycle cellulaire : en cours de phase G1, aux transitions G1-S et S-G2. Shimizu et Mori (1998) ont réalisé des études similaires qui suggèrent que la majorité des cellules sont arrêtées en cours de phase G1 dans les bourgeons dormants. Ils ont également suivi le niveau d'expression de transcrits codants pour les cyclinB, cyclinD et cdc2, qui passent d'un niveau faible dans les bourgeons dormants à un niveau très élevé après décapitation du bourgeon terminal. D'autres gènes, associés à la dormance, ont été isolés et caractérisés chez le pois (Madoka et Mori, 2000a, b; Stafstrom *et al.*, 1998) mais leur fonction reste à déterminer. Récemment, une revue sur les signaux régulant la dormance du bourgeon a

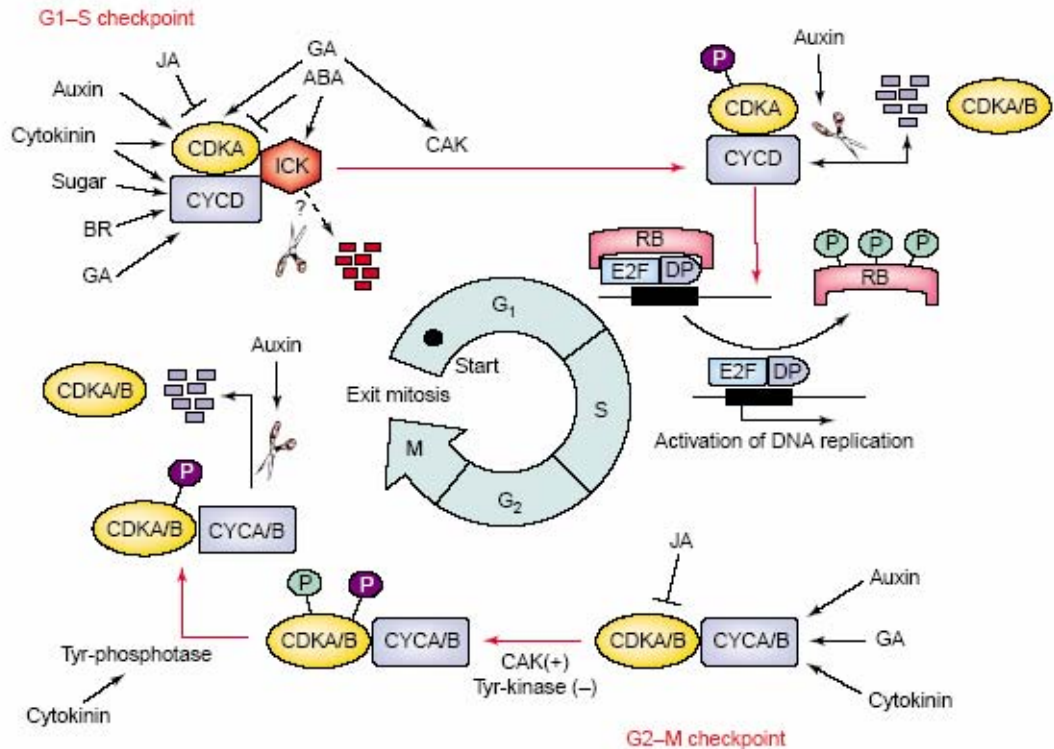


Figure 9: Modèle de transitions des phases G1-S et G2-M du cycle cellulaire chez les plantes. L'activation de la progression du stade G1 implique l'expression de D-type cyclins (CYCD) et de leurs sous-unités catalytiques, les cyclin-dépendent kinases (CDKA) ainsi que la dissociation de la protéine inhibitrice de CDK (ICK1) et la phosphorylation du résidu Thr160 (P en violet) de CDKA. CYCD et CDKA sont régulés positivement par de nombreux facteurs dont l'auxine, la cytokinine, les brassinostéroïdes (BR), le sucre et l'acide gibbéréllique (GA). ICK1 est induite par l'acide abscissique (ABA). La phosphorylation de CDKA est réalisée par la CDK-activating kinase (CAK), induite par l'acide gibbéréllique. Le complexe actif CDKA-CYCD hyper-phosphoryle la protéine de la tumeur rétinoblastome (RB), ce qui inhibe sa liaison avec le facteur de transcription E2F et la protéine de docking (DP), permettant ainsi le remodelage de la chromatine, l'activation de la transcription, la réplication de l'ADN et la transition de phase S. Le complexe SCF (SKP1 – Cullin – F-box-protein) sert d'intermédiaire pour l'ubiquitination et la protéolyse (ciseaux sur le schéma) de ICK1 et CYCD, nécessaire pour enclencher la transition G1-S chez la levure (et potentiellement chez les plantes comme indiqué par le ?). L'initiation de la transition G2-M nécessite l'induction des A- et B-type cyclins (CYCA/B) et l'activité des A- et B-type CDKs (CDKA/B). Auxine, cytokinine et gibbérélline sont toutes impliquées dans l'expression et la stabilité des CYCA/B et CDKA/B. Durant la phase G2, le résidu Thr160 (P en violet) des CDKs est positivement phosphorylé (+) par CAK, et les résidus Thr14 ou Tyr15 (P en vert) des CDKs sont négativement phosphorylés (-) par une tyrosine kinase. Une cytokinin-régulée tyrosine phosphatase (CDC25) retire le phosphate inhibiteur et permet à la transition G2-M de se produire. La phase de mitose requiert une ubiquitin-dépendante protéolyse des B-type cyclins. L'anaphase-promoting complex (APC) régule l'ubiquitination et la protéolyse des CYCA/B. L'auxine apparaît comme impliquée dans la dégradation des cyclines. L'acide jasmonique (JA) inhibe l'activité CDK lors des deux transitions G1-S et G2-M. (Tiré de Horvath *et al.*, 2003)

été présentée par Horvath *et al.* (2003). Outre la présentation d'un modèle de transition des phases G1-S et G2-M (Figure 9), qui met en évidence les déterminants moléculaires de ces deux étapes clés du cycle cellulaire, les auteurs suggèrent que le remodelage de la chromatine joue un rôle dans la régulation de la dormance du bourgeon. En effet, Law et Suttle (2003) ont mis en évidence des modifications de méthylation de l'ADN du méristème au cours de l'induction et de la levée d'endodormance chez la pomme de terre. Le remodelage de la chromatine est un indicateur de modification dans la structure et l'organisation de la chromatine. Or, durant ces dernières années, ce phénomène s'est révélé être un mécanisme clé de l'expression des gènes chez les eucaryotes. Ainsi, la régulation de la dormance chez les bourgeons pourrait être similaire à la régulation de la floraison chez *A. thaliana*, où l'action du locus FLC (FLOWERING LOCUS C) a été mise en évidence en tant que répresseur de la floraison. La vernalisation permet la floraison la saison suivante en provoquant une répression de FLC. Cette répression du locus FLC, nécessaire à la floraison, est maintenue par l'activité de gènes de vernalisation (VRN) qui mettent en jeu des changements des histones à proximité du locus FLC (Bastow *et al.*, 2004; Sung et Amasino, 2004). Un phénomène similaire pourrait se produire dans le bourgeon lors de l'application des températures froides hivernales, qui sont indispensables au débourrement chez un grand nombre d'essences forestières.

Parallèlement à la compréhension fine des processus physiologiques mis en jeu et des régulateurs du débourrement, une approche complémentaire consiste en la recherche de marqueurs moléculaires liés à la variation de la date du débourrement. Une telle approche serait à même de fournir des éléments d'aide à la description et à la compréhension de la physiologie du débourrement.

2.3. Vers la mise au point de marqueurs diagnostic de la date de débourrement

Les études menées sur le débourrement et décrites précédemment se basent principalement sur l'analyse de banques de mutants et des niveaux d'expression de transcrits entre bourgeons (ou graines) dormants et non-dormants. Une approche de tri différentiel à haute résolution, telle que la méthode SSH (Suppression Subtractive Hybridization, Diatchenko *et al.*, 1996) pourraient être la méthode de choix dans le cas où l'on s'attend à ce que peu de signaux apparaissent ou à ce que le signal soit faible en intensité (Rohde *et al.*, 2000), comme ce peut être le cas pour le débourrement. Intégrée dans le cadre d'une approche gène candidat et couplée à des études d'association, qui permettent de relier les variations nucléotidiques au sein des gènes candidats aux variations du caractère de débourrement, la recherche des gènes spécifiquement exprimés à un stade donné au cours du processus de débourrement pourrait aider à la compréhension de ce dernier et à la mise au point de marqueurs diagnostics du caractère.

2.3.1. Le séquençage d'EST et l'analyse du transcriptome

Le séquençage systématique du génome n'est pas réalisable, en l'état actuel, sur des espèces comme les chênes (0,9 pg/C). Mais si l'on cherche à identifier des gènes candidats influençant un caractère d'intérêt, une approche basée sur les ADN complémentaires (ADNc), copie des ARN messagers (ARNm), peut se révéler suffisante. En effet, savoir où et quand un gène est transcrit en ARN et traduit en protéine constitue souvent un indice important pour déterminer sa fonction biologique (Bouchez et Höfte, 1998).

Les ARNm représentent seulement 1 à 3% de la population d'ARN totaux chez les eucaryotes, sont dépourvus d'introns et de séquences intergéniques, et codent généralement pour des protéines exprimées dans les différents tissus ou cellules. Le séquençage d'EST (Expressed Sequence Tag), ou étiquettes de séquences exprimées, constitue une voie parallèle au séquençage systématique des génomes. Cette technique est facilement utilisable et est largement employée chez la totalité des organismes eucaryotes ou procaryotes. En effet, le nombre d'EST séquencées était de 25 788 811 au 23 Février 2005, identifiées chez plus de 790 organismes. Parmi ces EST, 36 étaient issues de *Quercus petraea* et 53 de *Quercus robur* (Source : http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). L'un des principaux avantages du séquençage d'EST est l'accession directe à la partie codante du génome. L'étude

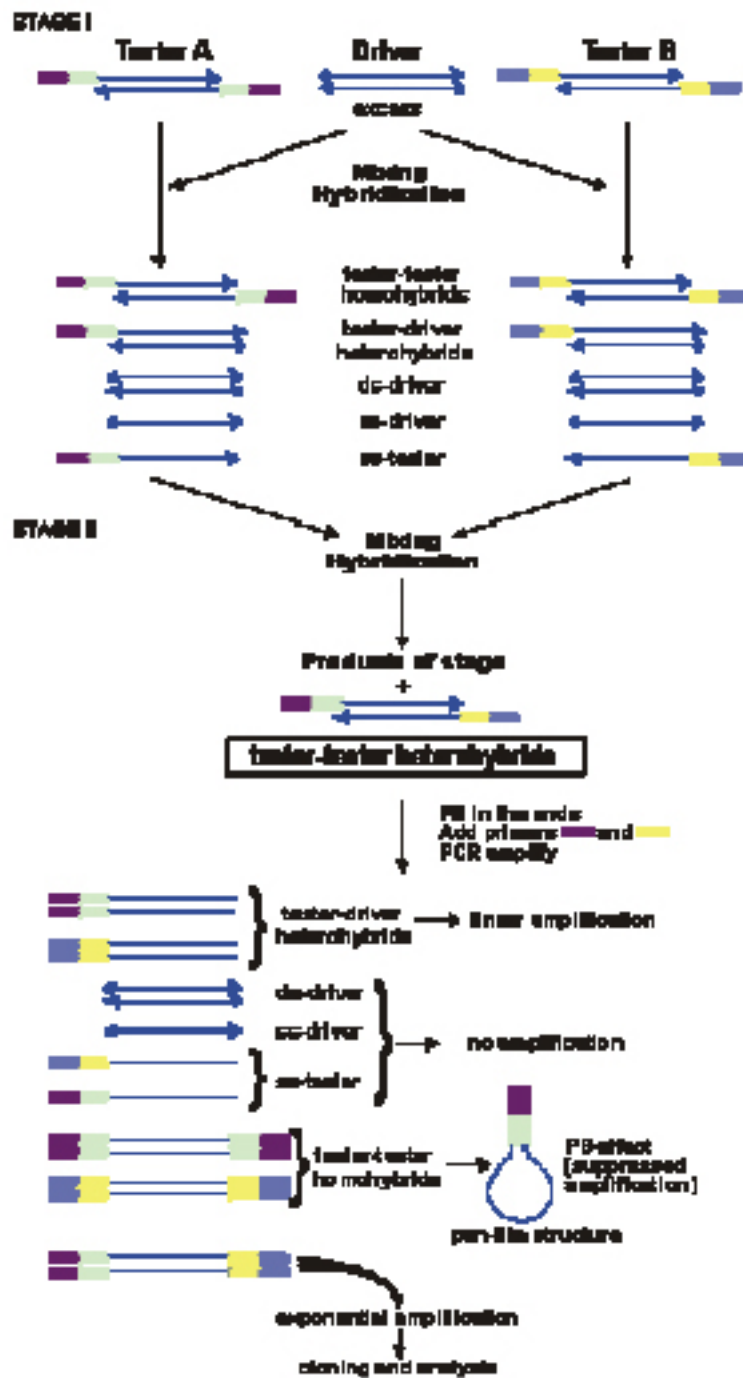


Figure 10 : La méthode SSH permet d'isoler, de manière spécifique, les fragments d'un échantillon (tester) comparé à une référence (driver). La première étape consiste en une synthèse des ADNc suivi d'une digestion enzymatique (RsaI). Les échantillons sont ensuite séparés en deux lots (Tester A et B) et ligués à deux types d'adaptateurs différents (violet/vert et bleu/jaune). Après deux hybridations successives en présence de témoin en excès, l'ensemble des fragments est amplifié par PCR à l'aide d'amorces localisées sur les adaptateurs (violet et bleu) mais l'amplification n'est exponentielle que pour les fragments différentiels et non redondants de l'échantillon (échantillons double brin avec les 2 adaptateurs différents). (Source : <http://www.evrogen.com/t6.shtml>)

de l'expression de ces gènes peut ensuite se faire grâce aux outils de la génomique fonctionnelle, qui s'intéresse à l'ensemble des transcrits (transcriptome) et des protéines (protéome) mais nous ne traiterons pas, dans cette synthèse, des aspects liés au protéome dont l'étude ne sera pas abordée dans la partie expérimentale de cette thèse. Les méthodes récentes d'analyse d'expression permettent la détection de transcrits en combinant la synthèse d'ADNc et la PCR. Certaines de ces techniques font appel à la fois à l'hybridation de fragments d'ADNc et à la PCR (ADNc-RDA : Representational Difference Analysis ; SSH : Suppression Subtractive Hybridization), certaines sont basées uniquement sur la PCR (DDRT-PCR : Differential display ; AP-PCR : Arbitrarily Primed PCR ; ADNc-AFLP : Amplified Fragment Length Polymorphism), d'autres, enfin, se basent sur l'acquisition en masse de séquences codantes (réseaux d'ADN, SAGE : Serial Analysis of Gene Expression ; Northern électronique).

Parmi ces différentes méthodes, la SSH (Diatchenko *et al.*, 1996), en particulier, permet l'amplification sélective de fragments d'ADNc dont l'expression est spécifique d'un état physiologique particulier tout en évitant d'amplifier les fragments communs (Figure 10). Cette technique a également l'énorme avantage de permettre l'isolation de transcrits rares (ratio de 1/1000). Elle a été utilisée avec succès dans l'identification de gènes impliqués dans des processus développementaux. La SSH a ainsi permis l'identification de gènes impliqués dans la maturation florale chez *Dianthus caryophyllus* (Kim *et al.*, 1999; Ok *et al.*, 2003), l'isolation de transcrits répondant à un stress à l'ozone chez *Pisum sativum* (Savenstrand *et al.*, 2000), l'identification d'un facteur de transcription spécifiquement exprimé au moment de la sénescence de la feuille chez *A. thaliana* (Hinderhofer et Zentgraf, 2001), ou encore la détermination des gènes impliqués dans l'association symbiotique entre *Eucalyptus globulus* et *Pisolithus tinctorius* (Voiblet *et al.*, 2001). Malgré le coût et la difficulté de mise en œuvre, c'est une technique d'analyse extrêmement puissante. A côté de la SSH, une technologie récente permet l'étude en masse des profils d'expression des gènes : les puces à ADN, appelées « DNA arrays » en Anglais, ou encore « DNA chips ». La méthode consiste à immobiliser sur un support solide des fragments d'ADN ou des oligonucléotides correspondant à différents gènes ou ADNc connus. L'hybridation est ensuite réalisée entre ces fragments d'ADN et des sondes préparées à partir de populations d'ARN messagers extraits de cellules, tissus ou organismes entiers et convertis en ADNc. Le signal d'hybridation pour chaque spot individuel peut être quantifié automatiquement par mesure radiographique ou par fluorescence, selon la technique de marquage utilisée, et reflète en principe l'abondance relative des ARNm correspondants. Une première méthode, appelée « macroarray », utilise

des filtres de nylons comme support et permet l'étude de milliers, voire dizaine de milliers de transcrits. La seconde, « microarray », se sert d'une lame de verre comme support et permet l'analyse simultanée de centaines de milliers de transcrits. Les ADNc utilisés dans le cadre de l'approche puces à ADN peuvent provenir des méthodes décrites précédemment, telle la SSH. Yang *et al.* (1999) ont ainsi pu démontrer que le couplage des deux méthodes permet l'identification rapide de gènes différentiellement exprimés.

2.3.2. L'approche gène candidat

Un des principaux objectifs de la génétique moléculaire est l'identification et l'isolation des gènes qui gouvernent les caractères d'importance. Deux méthodes classiques permettent de mener au clonage des gènes d'intérêt : le clonage positionnel (Tanksley *et al.*, 1995) et la mutagenèse d'insertion (Bechtold *et al.*, 1993). Or, comme mentionné précédemment, ces deux techniques sont quasi impossibles à mettre en œuvre chez les arbres forestiers, et en particulier les chênes, hormis l'activation tagging. Soit les effectifs des pedigrees de cartographie sont insuffisants pour mener à bien un travail de cartographie fine, soit les outils moléculaires ne sont pas développés sur ces espèces. L'approche gène candidat (GC) constitue donc une stratégie alternative, en particulier dans le domaine de la génétique des plantes (Byrne et McMullen, 1996; Pflieger *et al.*, 2001). Il s'agit de proposer des gènes de fonction connue qui pourrait *a priori* correspondre aux locus majeurs contrôlant le caractère. Ce peut être des gènes de structure ou de régulation. L'hypothèse de travail est que le polymorphisme moléculaire au sein du gène candidat est relié à la variation phénotypique. Cette approche a été largement utilisée dans le domaine de la génétique humaine et animale (Rothschild et Soller, 1997; Tabor *et al.*, 2002; Wayne et McIntyre, 2002) mais aussi en génétique des plantes depuis les années 90 (Byrne et McMullen, 1996; Pflieger *et al.*, 2001). La définition d'un gène candidat diffère selon la discipline concernée. Les physiologistes considèrent comme gènes candidats tous les gènes impliqués dans l'expression d'un caractère, alors que les généticiens considèrent seulement les gènes polymorphes potentiellement impliqués dans la variation du caractère (Pflieger *et al.*, 2001).

Nous prendrons ici une définition large du concept de gènes candidat, basée sur des présomptions de natures différentes :

- des gènes candidats « fonctionnels », proposés sur la base de connaissances sur les mécanismes moléculaires impliqués. Les orthologues de ces gènes

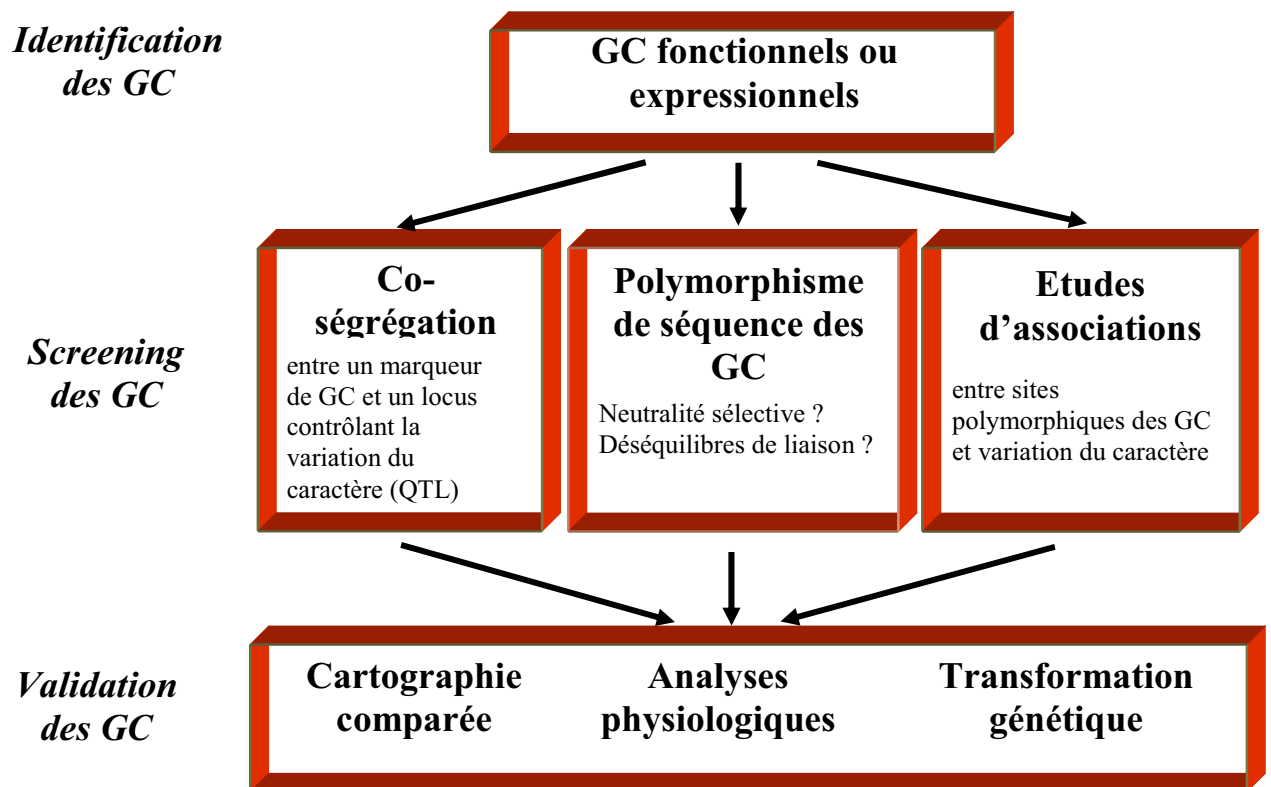


Figure 11 : L'approche gènes candidats (Tiré de Pflieger *et al.*, 2001).

Tableau 3 : Diversité nucléotidique de régions codantes et non codantes.

Organisme	Référence	Codant			Non Codant (Introns + UTR + NTR)		
		bp	Nb SNP	SNP/bp	bp	Nb SNP	SNP/bp
Soja	Zhu <i>et al.</i> , 2003	28693	57	1/503	32160	140	1/230
Orge	Bundock et Henry, 2004	615	10	1/61	1548	70	1/22
Arabette	Kuittinen et Aguadé, 2000	741	5	1/148	1088	31	1/35
Maïs	Guillet-Claude <i>et al.</i> , 2004	—	—	1/82	373	18	1/21

candidats, disponibles chez une espèce, pourront être également considérés comme des gènes candidats pour le caractère dans l'espèce étudiée ;

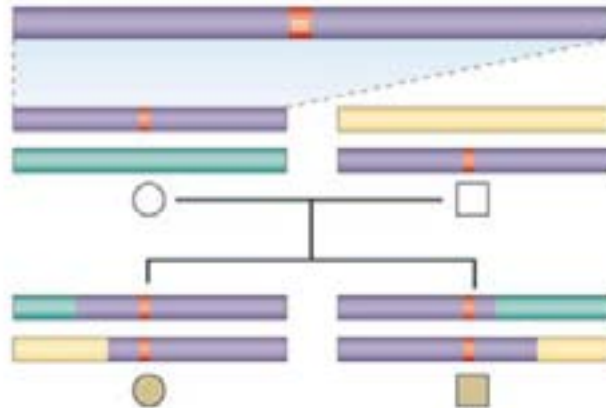
- des gènes candidats « expressionnels », différentiellement exprimés au cours d'un processus développemental (Ok *et al.*, 2003; Vilaine *et al.*, 2003), en réponse à un stress ou entre deux états physiologiques contrastés (Dubos *et al.*, 2003; Savenstrand *et al.*, 2000) ou entre des génotypes contrastés (Wellmer *et al.*, 2004);

L'approche se déroule ensuite en deux étapes supplémentaires après l'identification des gènes candidats (Figure 11) (Pflieger *et al.*, 2001). Tout d'abord, un polymorphisme moléculaire doit être révélé pour : i/ au sein d'un pedigree de cartographie, localiser les GC sur une carte génétique afin d'observer les liaisons génétiques entre les GC et des QTL, ii/ en populations naturelles, estimer les corrélations statistiques entre les variants alléliques des GC et les variations du caractère phénotypique, mais aussi iii/ analyser la structuration de la diversité nucléotidique par l'étude du polymorphisme de séquence des GC. Cette étape permet de préciser l'implication potentielle des différents GC dans le déterminisme du caractère considéré, les deux stratégies employées (pedigree et populations naturelles) étant complémentaires l'une de l'autre. Enfin, dans une troisième étape, si la co-ségrégation et/ou une corrélation statistique a été trouvée, des expériences complémentaires doivent être menées pour confirmer l'implication du GC dans la variation du caractère. C'est l'étape finale de validation qui consistera à étudier la conservation de la position des GC entre espèces proches, mesurer les niveaux d'expression des ARNm du GC et/ou de la protéine et/ou déterminer l'activité enzymatique du GC. La cartographie comparée et les analyses physiologiques fournissent une somme d'arguments pour ou contre l'implication du GC dans le caractère mais ne constituent pas des preuves indubitables du rôle joué par le GC. Les expériences de transformation génétique peuvent démontrer de manière claire si le GC est le gène déterminant la variation phénotypique du caractère.

2.3.3. Utilisation du polymorphisme de type SNP dans les études d'associations

La recherche de polymorphisme s'avère donc nécessaire à la fois pour cartographier les gènes candidats, et rechercher des liaisons statistiques entre les variations nucléotidiques et les variations du caractère. Les marqueurs de type SNP (« Single Nucleotide

Etude de liaison génétique utilisant les descendants d'un croisement (pedigree)



Etude d'association utilisant une population d'individus non apparentés

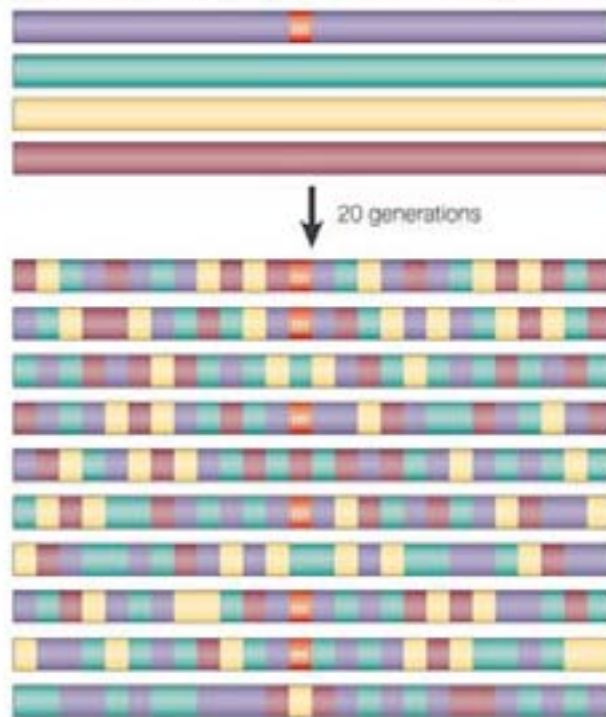


Figure 12 : Analyse de liaison versus étude d'association (Tiré de Cordon et Bell, 2001).

Lorsqu'une mutation fonctionnelle apparaît (*m*), cela se produit au sein d'un haplotype de variants génétiques déjà existants. Dans le cas des analyses de liaison génétiques menées sur des pedigree, dans lesquelles il y a peu d'opportunités pour que des phénomènes de recombinaison se produisent, les régions candidates qui portent la mutation sont très larges et peuvent contenir des centaines voire des milliers de gènes potentiels. A l'inverse, les études d'association s'appuient sur des événements de recombinaisons historiques et les régions candidates sont (théoriquement) extrêmement petites dans des populations allogames en panmixie, recouvrant seulement un gène ou un fragment du gène. Au cours des générations successives, alors que la mutation est transmise, la recombinaison va progressivement la séparer des allèles spécifiques de son haplotype original. Certains variants particuliers de l'ADN peuvent être hérités ensemble durant plusieurs générations. Ce type d'association non aléatoire d'allèles est connu sous le nom de déséquilibre de liaison.

Polymorphism »), ou polymorphisme d'un seul nucléotide, sont à même de répondre à ces deux objectifs. Ils rendent compte de la substitution d'une seule base nucléotidique par une autre à noyau équivalent (transition) ou à noyau différent (transversion). Le rapport transition/transversion est très variable d'une espèce à l'autre et n'est pas lié au règne d'appartenance. Dans la pratique, le marqueur SNP est bi-allélique, même si les quatre bases peuvent être rencontrées au niveau du site SNP (Brookes, 1999; Vignal *et al.*, 2002). Ces marqueurs constituent une forme très abondante de variation dans le génome. Chez l'homme, la proportion de polymorphismes de type SNP a été estimée à environ 90% du polymorphisme de l'ADN (Collins *et al.*, 1998), la densité étant d'environ 1 SNP pour 1000 bp. En comparaison, elle est de 1 SNP / 273 bp chez le soja (Zhu *et al.*, 2003) et 1 SNP / 61 bp chez le pin taeda (Brown *et al.*, 2004) mais ces études ne recouvrent pas la totalité du génome et sont souvent limitées aux zones codantes du génome. Il existe une grande disparité au sein du génome, notamment entre zones codantes et non-codantes (Tableau 3). En tant que marqueurs, ils sont plus stables que les microsatellites et la probabilité qu'ils aient des conséquences au niveau fonctionnel est plus grande lorsqu'ils apparaissent dans des régions codantes ou régulatrices de gènes, contrairement aux microsatellites le plus souvent situés dans des régions non codantes du génome (Brumfield *et al.*, 2003). Enfin, ils bénéficient du développement de techniques de génotypage à très haut débit (Syvänen, 2001). Toutes ces caractéristiques en font des marqueurs de choix pour les études d'associations qui nécessitent de génotyper un grand nombre d'individus.

Le développement des études d'associations s'est essentiellement réalisé dans le domaine de la génétique humaine, en particulier pour l'étude du déterminisme des maladies génétiques complexes (Cardon et Bell, 2001; Long et Langley, 1999). Le principe est d'estimer les corrélations statistiques entre les variants génétiques et la variation du caractère à une échelle populationnelle. Les premières études, désignées comme études cas-témoin (case control studies) concernaient des maladies à déterminisme simple, telle les maladies liées au système immunitaire (Cardon et Bell, 2001). Il existe à présent différentes méthodes dans le domaine des études d'associations basées soit sur les haplotypes, soit sur les marqueurs, soit sur le déséquilibre de liaison (Transmission Disequilibrium Test) mais aucun plan expérimental n'apparaît *a priori* comme idéal (Long et Langley, 1999). L'un des principaux avantages d'une étude d'association par rapport à une approche QTL est que si l'on détecte une association significative entre un site polymorphe d'un gène candidat et la variation phénotypique d'un caractère, les chances de se trouver à proximité du site fonctionnel, ou d'un site en interaction avec celui-ci, sont plus élevées (Figure 12). En effet,

dans le cas de pedigree, la fenêtre de déséquilibre de liaison entre allèles est extrêmement large et il est possible de détecter des QTL liés aux marqueurs même à des distances de recombinaison importantes. En revanche, dans le cas de populations naturelles, les nombreuses générations de recombinaisons ont réduit la fenêtre de déséquilibre de liaison autour des sites fonctionnels. Il est donc essentiel d'estimer la distribution et l'intensité des déséquilibres de liaison au sein des régions candidates afin de prédire l'efficacité des études d'association ultérieures (Gaut et Long, 2003). Néanmoins, l'ensemble des facteurs évolutifs (dérive génétique, migrations, subdivisions de populations, sélection naturelle) va également affecté ou créé du déséquilibre de liaison. C'est pourquoi la structuration de la population d'association doit être testée au préalable, à l'aide d'un ensemble de marqueurs cartographiés non liés physiquement entre eux, non liés aux gènes candidats, et neutres vis-à-vis du caractère. Dans le domaine végétal, les études d'associations ont permis de mettre en évidence, par exemple, que la variation allélique d'un gène de déhydrine était associée avec la tolérance au froid au cours de l'émergence de la plantule chez le niébé (*Vigna unguiculata* [L.] Walp.), une légumineuse tropicale (Ismail *et al.*, 1999). Chez le maïs, Thornsberry (2001) a montré que les polymorphismes au locus *dwarf8* sont associés aux variations de la date de floraison. Chez *Arabidopsis*, les gènes FRI et FLC contribuent de manière épistatique à la variation phénotypique de la date de floraison générant un cline latitudinale pour les haplotypes à ces deux locus (Caicedo *et al.*, 2004).

Pour conclure sur ces deux premières parties de la synthèse bibliographique, le processus de débourrement est donc relativement bien décrit d'un point de vue quantitatif chez différentes espèces forestières. En revanche, les connaissances au niveau moléculaire sont encore largement incomplètes et l'utilisation de techniques fines de biologie moléculaire, ainsi que la mise au point de marqueurs diagnostics de la date d'occurrence, devraient permettre d'aller vers une meilleure compréhension du phénomène. Nous allons à présent dans cette dernière partie bibliographique faire le point des connaissances actuelles sur les chênes, objets de la partie expérimentale de cette thèse.

③ Caractéristiques des chênes blancs européens

La partie expérimentale du travail a été conduite sur le chêne et en particulier sur les chênes blancs européens, pour plusieurs raisons. Tout d'abord parce que les chênaies sont une composante essentielle, à la fois économique et écologique, des forêts européennes, et en particulier de la forêt française. En effet, les chênaies couvrent un total de 4,2 millions d'hectares sur un total de 16,9 M ha, soit environ $\frac{1}{4}$ de la surface boisée en France. Les chênes blancs, principalement représentés par le chêne sessile (*Quercus petraea* (Matt.) Liebl.) et le chêne pédonculé (*Quercus robur*), représentent pour plus des $\frac{3}{4}$ de la surface en chêne susmentionnée. C'est donc une composante majeure de la filière bois du fait du volume et de la valeur des bois produits (entre 30 et 700 euros le m³). La France arrive en tête des producteurs de bois de chêne au niveau européen. L'utilisation principale du bois concerne la menuiserie (ébénisterie, tonnellerie, ...). En second lieu, les chênes constituent des écosystèmes en eux-mêmes, et hébergent des centaines d'espèces de plantes, d'insectes et de microorganismes. En effet, de tous les écosystèmes forestiers tempérés, la chênaie est celui qui accueille la plus grande biodiversité. A ce titre, il est possible que les mécanismes qui affectent la diversité génétique de caractères adaptatifs aient une incidence sur les espèces associées aux chênes.

Dans cette partie, nous présenterons d'abord le complexe d'espèces que constitue les chênes blancs européens en abordant la taxonomie, la répartition géographique, l'écologie ainsi que la reproduction de ces espèces. Ensuite, nous verrons dans quelle mesure le chêne sessile et le chêne pédonculé peuvent s'hybrider et quelles sont les connaissances concernant la différenciation, à la fois au niveau phénotypique et moléculaire, entre les deux espèces. Nous ferons aussi le point sur les données de diversité génétique obtenues sur les chênes blancs européens, en particulier sous l'angle des changements climatiques du Quaternaire et leurs conséquences sur les espèces.



Figure 13 : Caractéristiques morphologiques des quatre principales espèces de chênes blancs européens. (Tiré de Jacamon, 1984).

3.1. Le complexe d'espèces des chênes blancs européens

Les chênes sont des ligneux de taille importante, pouvant atteindre jusqu'à 40 mètres (*Q. petraea*, *Q. robur*) et dont le tronc peut atteindre deux mètres de diamètre chez le chêne pédonculé (*Q. robur*). En arbre isolé, le houppier est large et irrégulier et le feuillage n'est pas très dense. Ils peuvent atteindre plusieurs centaines d'années. En plus de la reproduction sexuée, la multiplication peut être végétative par drageons (*Q. pyrenaica* ...) et rejets de souche (*Q. robur* ...). Les feuilles sont dentées, sinuées-lobées ou entières mais toujours caduques (sauf chez le chêne vert (*Q. ilex*) et le chêne liège (*Q. suber*)) et alternes.

3.1.1. Taxonomie et aires de répartition

Nous nous référons dans cette partie à la monographie de Camus (1934-1954) qui reste encore l'ouvrage de référence malgré son ancienneté et une nomenclature partiellement obsolète mais qui semble cependant confirmée par les données moléculaires. Les chênes appartiennent donc au genre *Quercus* et à la famille des *Fagaceae*, dont font aussi partie le châtaignier européen (*Castanea sativa*) et le hêtre (*Fagus sylvatica*). Le genre *Quercus* est divisé en deux sous-genres : *Cyclobalanopsis* et *Euquercus*. Les chênes blancs européens font partie du second sous-genre et se placent dans deux sections (le sous-genre *Euquercus* en contient sept) : *Lepidobalanus* et *Mesobalanus*. Ces deux sections comprennent environ 150 espèces qui se retrouvent principalement en Amérique du Nord et en Asie. En Europe, si l'on excepte quelques espèces rares ou limitées à certaines régions méridionales (*Q. canariensis* et *Q. faginea* en Espagne, *Q. frainetto* en Italie, *Q. cerris* et *Q. pedunculiflora* dans les Balkans), seules quatre grandes espèces (Figure 13) sont présentes, à savoir :

- section *Lepidobalanus* :

- *Quercus robur* L. (= *Q. pedunculata* Ehr.), le chêne pédonculé ;

- *Quercus petraea* (Matt.) Liebl. (= *Q. sessiliflora* Salisb. = *Q. sessilis* Ehr.), le chêne sessile ou chêne rouvre ;

- *Quercus pubescens* Willd. (= *Q. lanuginosa* Lam.), le chêne pubescent.

- section *Mesobalanus* :

- *Quercus pyrenaica* Willd. (= *Q. toza* Bast.), le chêne tauzin.

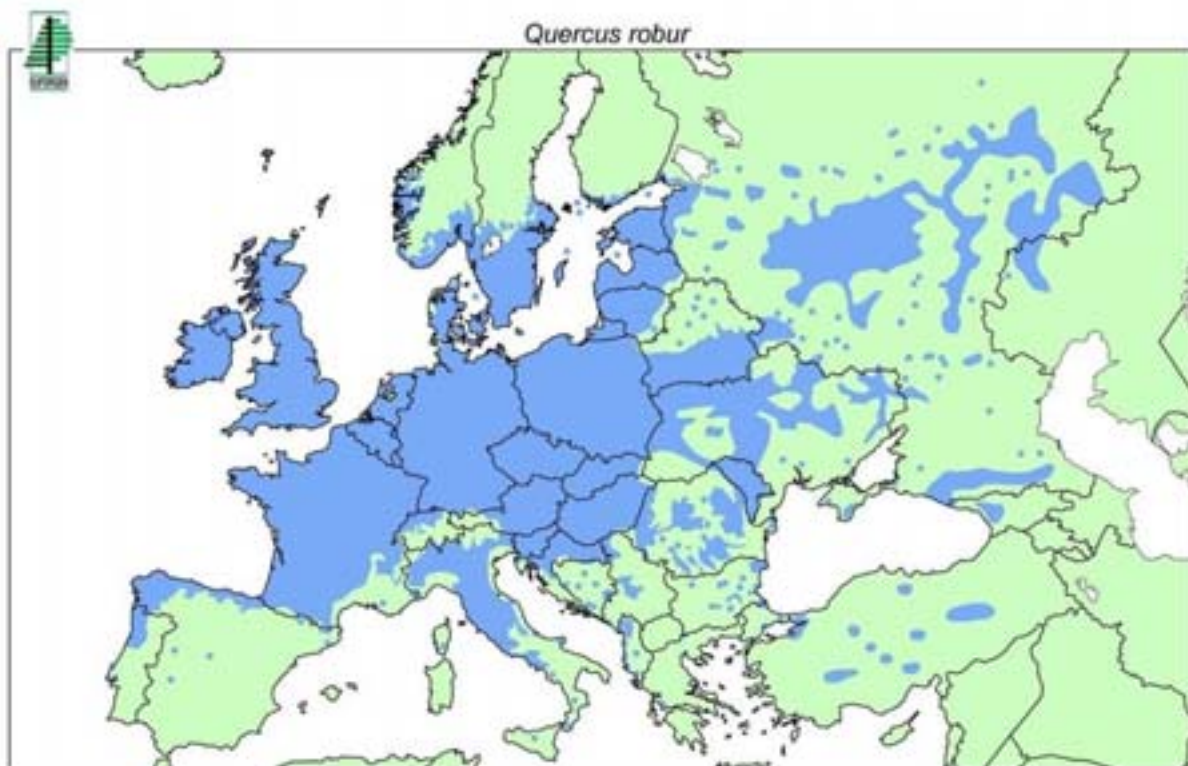


Figure 14: Aires de distribution des chênes pédonculé et sessile. (Tiré de Ducouso et Bordacs, 2004)

Ces espèces forment un complexe caractérisé par de larges possibilités d'hybridation que nous aborderons dans un prochain paragraphe. Deux espèces sont largement prédominantes par leurs aires de répartition, à savoir le chêne pédonculé et le chêne sessile. Le chêne pédonculé possède l'aire de répartition la plus vaste qui s'étend du Portugal à la Scandinavie et de l'Irlande jusqu'à l'Oural (Figure 14). L'aire de distribution du chêne sessile est contenue dans la précédente, limitée à l'Est à la frontière ukrainienne (Figure 14). Le chêne pubescent occupe quant à lui le Sud de l'Europe et le chêne tauzin est limité au Sud-Ouest de l'Europe.

Dans la suite du document, nous ne ferons référence qu'aux deux espèces majeures que sont le chêne sessile et le chêne pédonculé qui ont fait l'objet de la partie expérimentale de cette thèse.

3.1.2. *Ecologie et sylvigénèse des chênes sessile et pédonculé*

Les deux espèces se caractérisent par une très large amplitude écologique mais occupent en moyenne des milieux contrastés, et sont sympatriques dans les zones intermédiaires (Grandjean et Sigaud, 1987). Deux facteurs principaux différencient les deux espèces : la disponibilité en eau et la richesse du sol (Grandjean et Sigaud, 1987). Le chêne pédonculé (*Q. robur*) possède des affinités atlantiques mais accepte les climats continentaux. C'est préférentiellement un arbre de vallée, mais il peut s'élever jusqu'à 1300 mètres en pays basque, et jusqu'à 600 à 900 mètres ailleurs (Lanier, 1994). Il se retrouve sur des sols fertiles et bien approvisionnés en eau, et même sur des sols compacts temporairement inondés. Le chêne pédonculé craint la sécheresse et laisse la place au chêne sessile sur des sols plus pauvres et bien drainés, voire sableux, ainsi que sur sols acides (Becker et Levy, 1990). Le chêne sessile (*Q. petraea*) préfère les climats atlantiques doux. C'est un arbre de collines et de plateaux qui peut s'élever jusqu'à 1600 mètres dans les Pyrénées, 800 à 1000 mètres ailleurs (Lanier, 1994).

Dans les successions forestières, les chênes sessile et pédonculé sont qualifiés d'espèces post-pionnières (Rameau, 1990). Pour illustrer ce concept, nous prendrons l'exemple d'une évolution progressive dans le cas d'une déprise agricole. Une fois la culture arrêtée, la parcelle est colonisée par des espèces pionnières telles que les bouleaux ou les pins. La composition floristique du jeune peuplement va ensuite évoluer vers une phase transitoire composée d'espèces post-pionnières puis une phase optimale avec des dryades⁵. Ces

⁵ Dryades : espèces végétales caractéristiques des milieux arrivés à l'équilibre écologique.

successions naturelles sont modifiées par l'action de l'homme, qui a souvent favorisé les chênes aux dépens d'espèces d'intérêt sylvicole moindre, et a mis en place des peuplements plus ou moins artificialisés. Dans les conditions naturelles, le chêne pédonculé est un post-pionnier nomade. En France, il se rencontre dans quasiment toutes les formations forestières de l'étage collinéen car il fait partie des phases pionnières ou transitoires. Les seuls peuplements climaciques se rencontrent sur alluvions récentes des vallées. Il est alors souvent en mélange avec le frêne. Ailleurs, l'évolution naturelle des milieux l'élimine au profit du chêne sessile. Cette dynamique explique les nombreux dépérissements constatés suite aux années de sécheresse. Le chêne sessile est un post-pionnier typique qui se rencontre dans les phases transitoires les plus tardives. C'est la phase précédant la hêtraie qui est la formation climacique de la plupart des forêts de l'étage collinéen. Le hêtre est toutefois absent des régions à pluviométrie faible (Centre, Est du secteur Ligérien,...) et le climax est alors le chêne sessile.

3.1.3. Caractéristiques de la reproduction

Les chênes fructifient à un âge variable selon les conditions de milieu. En forêt, le chêne sessile commence à se reproduire vers l'âge de 30 à 45 ans (Jones, 1959). Le chêne pédonculé fructifie beaucoup plus tôt et plus régulièrement que le chêne sessile. On peut considérer que les chênes ne se reproduisent pas avant 15 à 20 ans en conditions pionnières (Ducouso et Petit, 1996). Les chênes sessile et pédonculé sont des espèces monoïques. Les inflorescences mâles sont des chatons pendants à la base des pousses de l'année et qui se forment avant le développement complet des nouvelles feuilles. Les fleurs femelles, très petites, sont situées sur des rameaux axillaires et leur développement est synchronisé avec celui des fleurs mâles. Ces espèces sont considérées comme strictement allogames. La floraison a lieu d'Avril à Mai environ à des latitudes moyennes. La phénologie de la floraison peut être très variable localement en fonction des individus et des conditions de milieu.

La dispersion du pollen est anémophile chez les chênes. La structure particulière des fleurs mâles (chatons) facilite la libération des grains de pollen par le vent (Figure 13). La dispersion des graines est barochore et zoochore chez le chêne. Le poids élevé des glands les fait tomber immédiatement sous le houppier. Ils peuvent ensuite être dispersés par deux catégories d'animaux : les petits mammifères (campagnols, mulots, écureuils) et les oiseaux (essentiellement le geai). Les observations montrent que les petits animaux sont peu efficaces

pour la dispersion des glands à cause de la forte prédation qu'ils exercent et parce que les distances parcourues sont de l'ordre de quelques dizaines de mètres seulement. Au contraire, les geais (geai des chênes *Garrulus glandarius* en Europe) semblent parfaitement spécialisés dans le transport et la dissémination des glands. Bien que les glands constituent leur nourriture principale, les geais sont moins destructeurs que les rongeurs. Un seul individu disséminerait en moyenne plusieurs milliers de glands par saison. Les distances de transport semblent varier de la centaine de mètres jusqu'à plusieurs kilomètres.

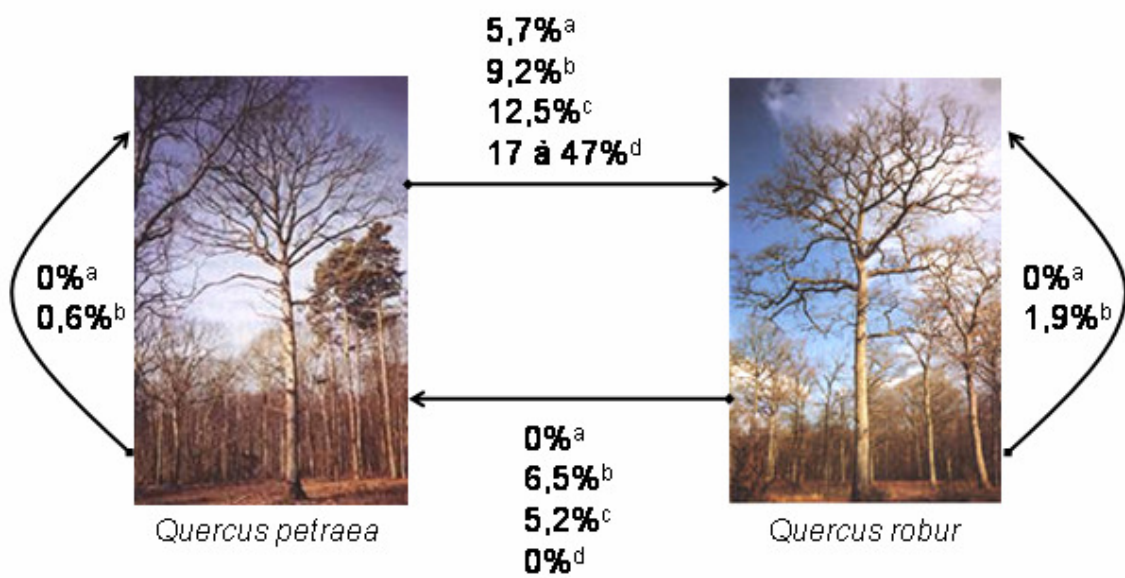


Figure 15 : Taux d'hybridation et d'autofécondation chez les chênes sessile et pédonculé.
^a Aas (1993) ; ^b Steinhoff (1993) ; ^c Kleinschmit et al. (1995) ; ^d Bacilieri et al. (1995).
 (Tiré de Scotti-Saintagne, 2003).

3.2. Hybridation et différenciation des chênes sessile et pédonculé

Linné a défini la notion d'espèce sur des critères morphologiques. Dans son livre « The Origin of Species », Charles Darwin parlait des chênes comme des espèces suspectes. A l'aide de simples critères morphologiques, les botanistes rencontrent de nombreux intermédiaires. En effet, les deux espèces peuvent s'hybrider mais il est possible de les différencier sur le plan phénotypique et, dans une certaine mesure, au niveau moléculaire.

3.2.1. Hybridation

Les deux espèces sont allogames, monoïques et les floraisons sont synchrones à l'intérieur d'un peuplement en mélange (Bacilieri *et al.*, 1995). Il n'y a donc aucun obstacle d'ordre phénologique qui s'oppose à l'hybridation. L'hybridation entre le chêne sessile et pédonculé a été mise en évidence à la fois en conditions contrôlées (Aas, 1993; Kleinschmit *et al.*, 1995a; Steinhoff, 1993) et en conditions naturelles (Bacilieri *et al.*, 1995; Bacilieri *et al.*, 1996; Streiff, 1998). L'hybridation est asymétrique entre les deux espèces avec un sens préférentiel du père sessile vers la mère pédonculé, les résultats issus des différentes études précédemment citées sont résumés à la Figure 15. L'autofécondation apparaît comme un phénomène vraiment très marginal puisque la valeur maximale observée est de 1,9%. La morphologie foliaire des hybrides juvéniles est similaire au phénotype maternel et cette période juvénile peut se prolonger jusqu'à 5 ans (Bacilieri *et al.*, 1995; Kleinschmit *et al.*, 1995b). Les formes intermédiaires obtenues sont fertiles avec des performances de croissance proches de celles du chêne pédonculé et supérieures à celles du chêne sessile dans les sept premières années. En peuplement mélangé adulte, la proportion d'individus hybrides estimée sur la base de critères morphologiques est très inférieure à la valeur mesurée au stade graine à l'aide de marqueurs moléculaires. Cela peut signifier que les critères morphologiques ne sont effectivement pas de bons marqueurs diagnostics ce qui a été vérifié au stade juvénile mais pas au stade adulte, faute du matériel adéquat. Une autre possibilité est qu'il existe une barrière reproductive post-zygotique (stérilité des hybrides) et/ou des pressions de sélection éliminant progressivement les hybrides au cours de la vie du peuplement. Les données disponibles à l'heure actuelle sur la viabilité ou la fertilité des hybrides ne permettent pas de conclure, même si les observations suggèrent fortement que les individus issus de croisements contrôlés interspécifiques soient fertiles (Roussel, communication personnelle).

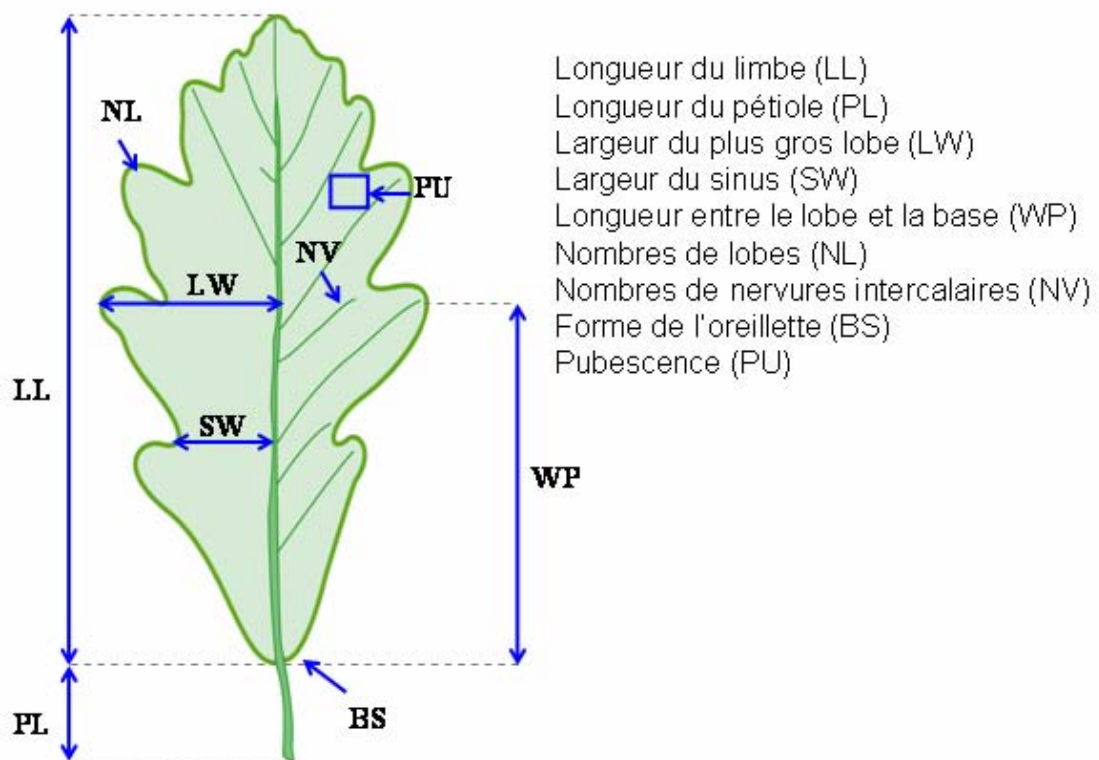


Figure 16: Caractères morphologiques foliaires différenciant le chêne sessile et le chêne pédonculé. (Tiré de Scotti-Saintagne, 2003).

Si les individus hybrides restent difficiles à déterminer avec assurance, la différenciation entre les individus sessiles d'une part et les individus pédonculés de l'autre se fait plus aisément à l'aide de critères morphologiques ou moléculaires.

3.2.2. Différenciation morphologique

Parmi les critères de différenciation phénotypique, les niches écologiques contrastées qu'occupent les deux espèces en sont un que nous avons déjà abordé dans le paragraphe 3.1.2. Nous ne traiterons ici que des critères de différenciation morphologiques.

Le chêne pédonculé tire son nom de la forme du gland qui possède un long pédoncule le reliant à la branche où il s'insère. C'est le premier critère de distinction entre les deux espèces puisque le gland du chêne sessile est dépourvu de pédoncule. Les bourgeons sont ovoïdes chez les deux espèces mais apparaissent souvent pointus chez le chêne sessile, ce critère ne permet cependant pas de distinguer les deux espèces avec certitude. C'est au niveau de la feuille que l'on rencontre le plus grand nombre de caractères différenciant les deux espèces (Figure 16). Plusieurs auteurs ont pu mettre en évidence le caractère discriminant de ces variables par analyse multivariée (Grandjean et Sigaud, 1987; Kremer *et al.*, 2002a). La distribution est typiquement bi-modale avec une zone de chevauchement de phénotypes intermédiaires pouvant correspondre soit à des formes hybrides, soit à des individus extrêmes de l'une ou l'autre des deux espèces. Kremer *et al.* (2002a) ont montré que les caractères les plus discriminants étaient le nombre de nervures intercalaires, la longueur du pétiole, la forme de l'oreillette, la largeur du sinus et la pilosité.

La présence de formes intermédiaires et la difficulté de distinguer les deux espèces au stade juvénile illustrent les limites des marqueurs morphologiques. Les outils de la génétique permettent d'avoir une approche génétique de cette différenciation.

3.2.3. Différenciation moléculaire

Dès les années 80, la différenciation entre les chênes sessile et pédonculé a été étudiée à l'aide de marqueurs moléculaires. Ainsi, ont été testés : les marqueurs de type isozyme, RAPD (Random Amplified DNA Polymorphism), SCAR (Sequence Characterized Amplified Region), protéine, AFLP (Amplified Fragment Length Polymorphism), microsatellite, ADNc (ADN complémentaire), ADNr (ADN ribosomique). Malgré des différences de fréquences

alléliques entre les deux espèces pour la plupart des marqueurs, aucun marqueur diagnostique n'a pu être mis en évidence. L'étude, réalisée par Bodénès (1997), illustre bien l'homogénéité génétique entre les deux espèces. Sur 2800 fragments RAPD analysés, seuls 2 % présentent des différences significatives de fréquence entre les deux espèces. Un résultat similaire a été trouvé par Cervera *et al.* (2000) puisque 1,7% des 115 fragments AFLP analysés étaient différenciellement exprimés. Zoldos (2001) a récemment montré qu'au niveau de 7 séquences hautement répétées du génome (100 à 700 répétitions), les deux espèces présentent exactement le même pattern d'hybridation moléculaire.

Il existe cependant des marqueurs discriminants mis en évidence dans certaines études. Ainsi, dans une étude faite sur 170 marqueurs AFLP, un marqueur présente 71% de différence de fréquence allélique entre le chêne sessile et le chêne pédonculé (Coart *et al.*, 2002). De la même manière, deux des 155 marqueurs AFLP étudiés par Stéphanie Mariette au cours de sa thèse (Mariette, 2001) possèdent un coefficient de différenciation supérieur à 40%. A l'aide des fréquences alléliques, le marqueur le plus discriminant permet de séparer les espèces dans 87% des cas. C'est mieux que les 78% d'assignation correcte de l'espèce obtenu en compilant l'information sur 20 locus microsattellites (Muir *et al.*, 2000). D'après Saintagne *et al.* (2004), les caractères impliqués dans la différenciation entre les chênes sessile et pédonculé sont sous contrôle polygénique et distribués en clusters tout au long du génome, avec la présence de certains hot spots de différenciation. Les différences interspécifiques reposent donc sur la somme de petites différences réparties tout au long du génome à plusieurs locus et non de fixation d'allèles différents à certains locus particuliers (Saintagne *et al.*, 2004). Cette dernière étude donne par ailleurs une description plus quantitative de la différenciation interspécifique, en montrant clairement que la distribution de la différenciation suit une courbe en L, la très grande majorité des marqueurs ne manifestant aucune différence de fréquence entre les deux espèces.

Dans une dernière partie, nous allons nous intéresser à l'histoire de la colonisation des chênes en Europe après la dernière glaciation, en mettant l'accent sur les conséquences de cette recolonisation sur la diversité génétique actuelle.



Figure 17: Etat de la végétation au dernier maximum glaciaire en Europe (-18 000 ans).
 (Source : <http://www.esd.ornl.gov/projects/gen/nerc.html>).

3.3. Point de vue historique et diversité génétique actuelle

Les plus vieux fossiles de chêne ont été identifiés avec certitude dans des dépôts datant de l'Oligocène (-38 millions d'années) mais l'apparition du genre *Quercus* pourrait remonter au début du tertiaire. Au Miocène (-26 millions d'années) et au Pliocène (-7 millions d'années) apparaissent des formes très proches des espèces modernes. Il est probable que celles-ci aient divergées dès le Miocène, c'est-à-dire avant la séparation des continents actuels. Les nouvelles espèces apparurent à l'occasion de changements climatiques de grande amplitude durant le Tertiaire et restèrent confinées à des latitudes méridionales.

3.3.1. Modifications des milieux en Europe depuis le dernier cycle glaciaire

La distribution actuelle des chênes résulte des alternances climatiques périodiques qui se sont amplifiées au Quaternaire. Au total, plus de 17 alternances de périodes glaciaires (de durée variant entre 50 et 100 000 ans) et interglaciaires (de durée variant entre 10 et 20 000 ans) se sont succédées depuis le début du Quaternaire, il y a 2 millions d'années. En comparaison, la période de stabilité climatique actuelle semble être tout à fait exceptionnelle. La dernière période froide en Europe a débuté vers -115 000 ans, interrompue par des périodes de réchauffement de courte durée aux alentours de -105 000, -100 000, -82 000, -50 000, -40 000 et -30 000 ans, cependant trop brèves pour permettre la migration des arbres tempérés hors des zones refuges (Birks, 1986). Le maximum glaciaire a eu lieu entre -20 000 et -18 000 ans. L'histoire de la végétation, durant ces phases glaciaires, peut être reconstituée à partir de l'étude des pollens fossiles, c'est la paléo-palynologie. La végétation est rare en dehors des régions couvertes par la calotte glaciaire (Figure 17). Les principales zones refuges des arbres auraient été la partie Ouest des Balkans, le Sud des Alpes et les massifs italiens ainsi que le Sud de l'Espagne dans une moindre mesure. A partir de -15 000 ans, le réchauffement climatique est brutal, et des changements de végétation vont se produire rapidement.

A travers l'Europe, les communautés herbacées se modifient, passant de steppes et toundras de climat froid et sec à des steppes. La réponse est dans un premier temps plus lente en ce qui concerne la couverture forestière. Cependant, des forêts ouvertes composées de bouleaux et de conifères boréaux semblent faire leur apparition à travers la Russie occidentale très rapidement à la suite du réchauffement (Velichko, 1993). Entre -12 000 et -11 000 ans,

l'analyse des pollens indique qu'une végétation forestière (bouleaux, saules) a reconquis pratiquement toute l'Europe centrale et du Sud à l'exception de l'Europe du Nord, colonisée seulement par quelques bouleaux et pins à la fin de cette période (Huntley, 1988). L'Europe de l'Est est couverte par des forêts de conifères. Une nouvelle période froide et sèche apparaît entre -10 800 et -10 000 ans qui provoque la disparition temporaire de la couverture forestière précédemment installée et son remplacement par des steppes et toundras (Laval *et al.*, 1991; Starkel, 1991; Velichko, 1993). Certains auteurs (Bohncke, 1993) pensent que des espèces telles que les pins sont restées présentes, dans une certaine mesure, et que la végétation présente à cette période n'était pas équivalente à celle présente durant le précédent maximum glaciaire (-18 000 ans). A partir de -9 000 ans, la forêt est à nouveau présente pratiquement partout en Europe. C'est d'abord (-8 000 ans) une forêt ouverte avec de nombreuses clairières herbacées (Huntley et Prentice, 1993; Starkel, 1991) puis la forêt devient dense avec des conifères plus abondants qu'à l'heure actuelle dans sa partie orientale. Jusqu'à -5 000 ans ensuite, les forêts tempérées décidues mixtes sont en expansion. Elles sont principalement composées, par ordre d'arrivée le plus fréquent dans les séries polliniques (Birks, 1986), de *Corylus*, *Ulmus*, *Quercus*, *Tilia*, *Alnus glutinosa*, *Fraxinus*. A partir de -5 000 ans, la forêt dense est en régression au profit de paysages plus ouverts, sous l'action de facteurs naturels mais aussi de l'homme puisque cette régression coïncide avec l'arrivée des populations du Néolithique, qui pratiquent la culture sur brûlis.

3.3.2. Zones refuges et voie de recolonisation des chênes

En réponse aux modifications du climat, les aires de répartition des espèces animales et végétales ont subi des fluctuations importantes. En Europe, les possibilités de reflux vers le Sud ont été limitées par les barrières physiques naturelles que sont les Alpes et les Pyrénées (orientées d'Est en Ouest) ainsi que la Méditerranée, entraînant de nombreuses extinctions. C'est pour cette raison que la flore forestière européenne serait actuellement beaucoup plus pauvre que la flore forestière américaine.

La génétique et la paléobotanique ont été associées pour reconstruire le scénario de la colonisation de l'Europe par les chênes depuis la dernière glaciation. L'exploitation de la banque européenne de données polliniques a ainsi fait apparaître que les Chênes caducifoliés ont été confinés en trois zones géographiques du Sud de l'Europe durant la dernière période glaciaire (Brewer *et al.*, 2002), à savoir un refuge ibérique, un refuge italien et un refuge

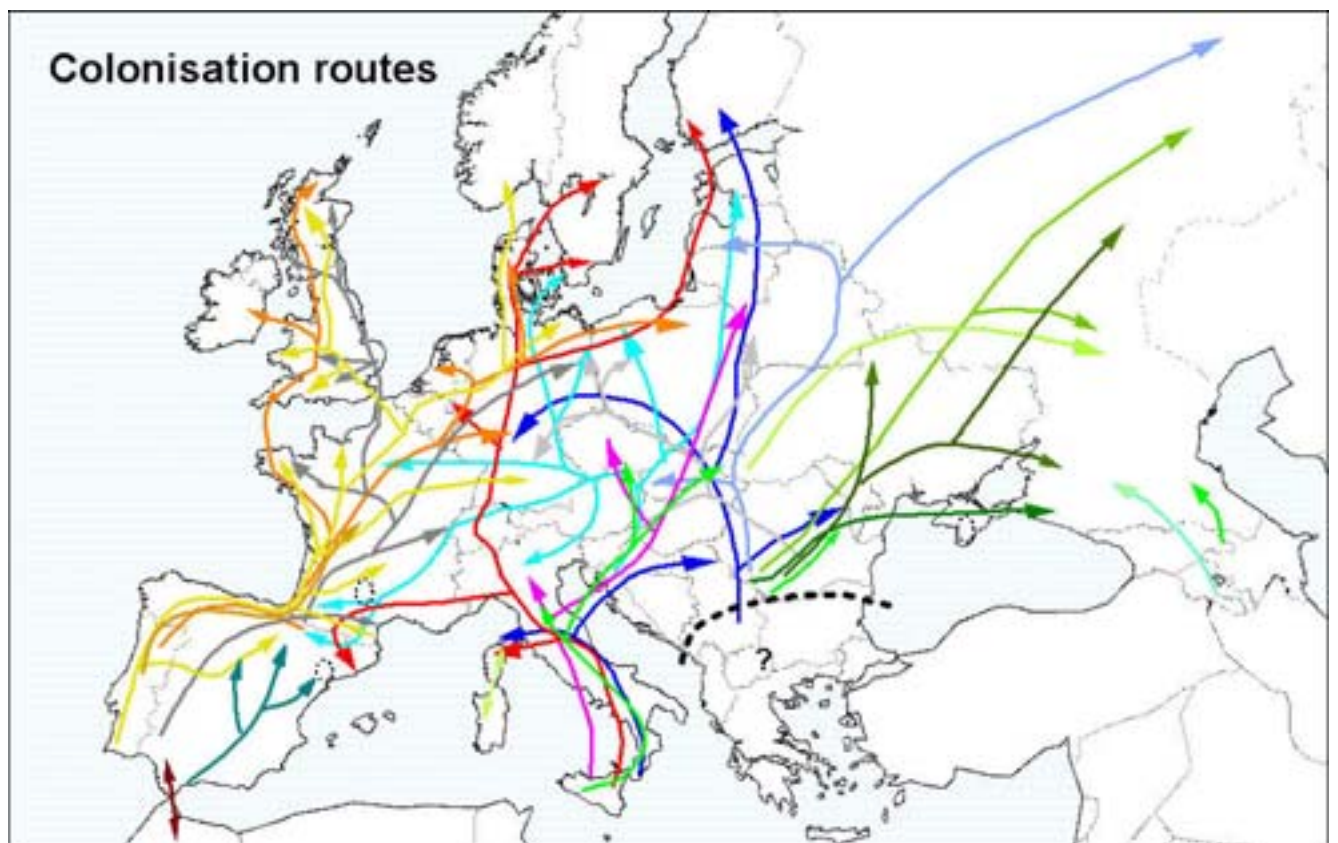
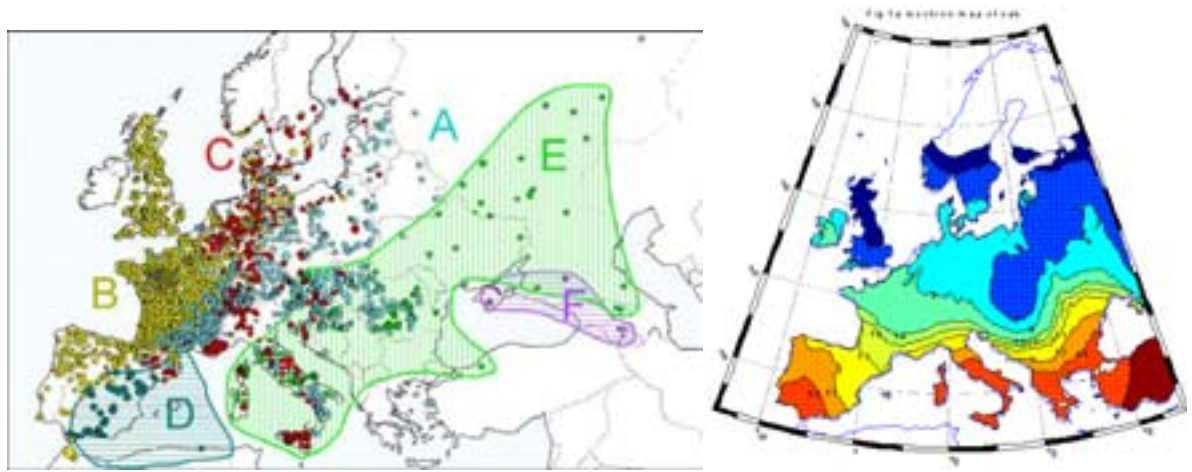


Figure 18 : Voies de recolonisation empruntées par les chênes blancs européens (en bas). Cette carte fait la synthèse de la carte de distribution géographique des variants d'ADN chloroplastique chez les chênes blancs (en haut à gauche) et celle de la distribution des pollens fossiles de chêne en Europe (en haut à droite). (Tiré de Kremer *et al.*, 2002b).

balkanique. Dès le réchauffement climatique (-15 000 ans), les Chênes ont progressivement colonisés l'Europe, en marquant un ralentissement entre -11 000 ans et -10 000 ans, pour finalement occuper la totalité de l'aire actuelle vers -6 000 ans. Par comparaison des données génétiques et palynologiques, il apparaît que les Chênes situés à l'Ouest d'une ligne Toulouse-Cologne-Amsterdam sont principalement issus des refuges ibériques les plus occidentaux (Figure 18). Les îles britanniques ont été exclusivement colonisées à partir des Chênes issus de cette région. A l'inverse, les chênaies de l'Europe centrale et orientale ont été constituées par des migrants provenant très probablement des refuges balkaniques et italiens. La confrontation des données génétiques et palynologiques a donc permis de reconstituer les principales voies de migration empruntées par les Chênes (Kremer *et al.*, 2002b) (Figure 18). L'analyse des cartes polliniques a montré que les chênes ont progressé à la vitesse moyenne de 380 mètres par an, ce qui est tout simplement stupéfiant si l'on compare aux caractéristiques biologiques des chênes. Deux mécanismes de dispersion ont alors été comparés par simulation informatique. Le premier correspond à une diffusion continue de l'espèce, le second est qualifié de dispersion composite et regroupe la diffusion proprement dite ainsi qu'une dispersion par « sauts de puces ». L'introduction d'événements de dispersion rares, mais à longue distance, a un impact considérable sur la vitesse de recolonisation (Le Corre *et al.*, 1997). Il se crée ainsi des îlots très en amont du front de migration, qui suivent leur propre développement démographique pour finalement fusionner quelques générations plus tard avec le front de migration, ce qui accélère considérablement la propagation de l'espèce. Les geais ou les corvidés peuvent être à l'origine de ces dispersions rares à longue distance même si on ne peut exclure l'action de l'homme ou des cours d'eau. Au final, l'hypothèse du modèle de dispersion composite est confortée par l'organisation spatiale actuelle de la diversité de l'ADN chloroplastique (Le Corre *et al.*, 1997).

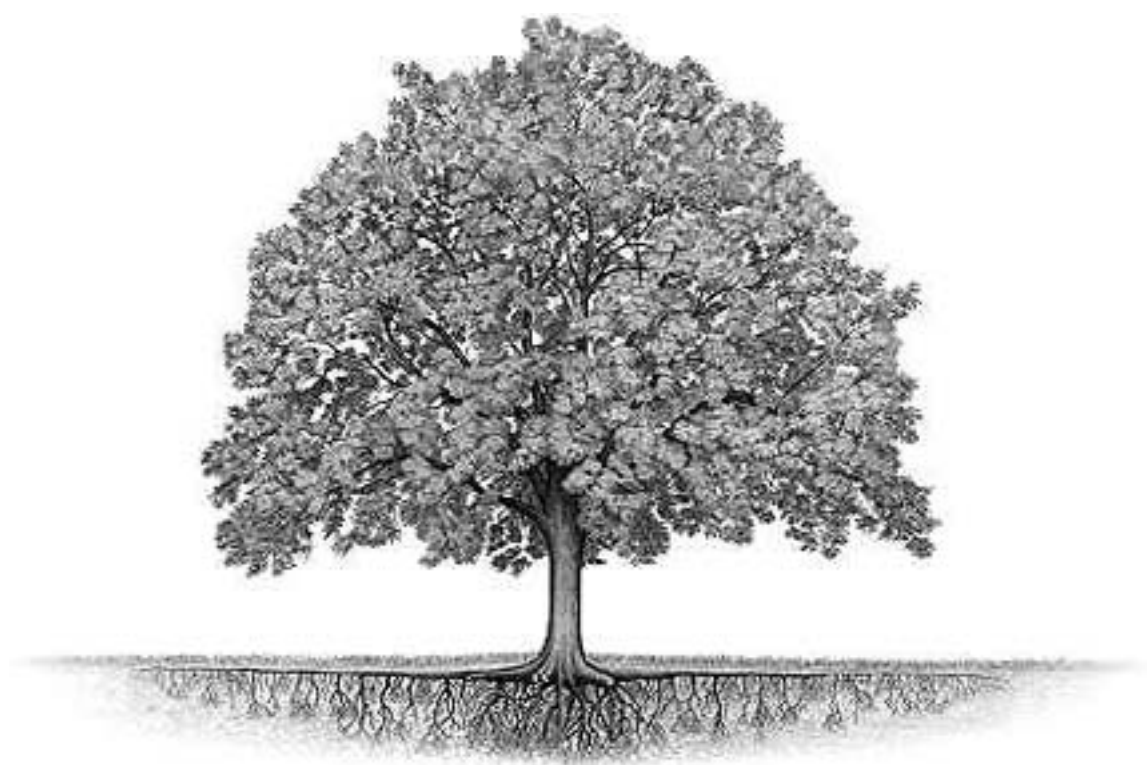
3.3.3. Colonisation et diversité génétique actuelle

La synthèse présentée ici est entièrement extraite de l'article publié par Kremer *et al.* (2002b). A la fin de l'ère glaciaire (-18 000 ans), les peuplements de chênes sessile et pédonculé étaient donc éclatés en trois zones refuges isolées, du moins totalement entre la péninsule ibérique et les deux autres, et partiellement entre l'Italie et les Balkans. L'isolement de plus de 100 000 ans, qui précède, abouti sans doute à une différenciation génétique élevée entre les trois refuges pour l'ensemble des caractères : gènes « neutres », ADN

chloroplastique, caractères adaptatifs. A mesure que le réchauffement radoucit le climat, les chênes migrent vers le Nord puis dans des directions différentes selon l'origine (Figure 18). Cette progression a été rapide notamment grâce aux événements de dispersion à longue distance, qui génèrent des événements de fondation, accentuant ainsi la différenciation génétique entre les forêts. A mesure que les nouveaux peuplements se mettent en place, les flux de pollen deviennent de plus en plus intenses, autorisant l'échange de matériel génétique entre forêts issues de refuges différents. Ces flux de gènes vont progressivement gommés la différenciation génétique pour tous les caractères adaptatifs et les marqueurs neutres. En revanche, les chloroplastes n'étant transmis que par les graines, les flux de gènes n'ont aucun effet sur la structure géographique des variants d'ADN chloroplastique qui restera inchangée jusqu'à nous. Les peuplements en place sont rapidement soumis aux pressions environnementales locales ce qui crée une différenciation génétique pour les caractères adaptatifs. A l'inverse, la diversité des marqueurs neutres n'est pas affecté par ces pressions sélectives et tend à s'homogénéiser progressivement en réponse aux flux de pollen, qui s'accroissent à mesure que l'espace se couvre de chênes. Ce scénario est surtout valable pour le chêne pédonculé. La progression du chêne sessile est sans doute soumise aux mêmes processus mais très largement favorisée par l'hybridation avec le chêne pédonculé. Le caractère systématique de l'hybridation lui permet d'occuper les mêmes forêts que le chêne pédonculé et aboutit à une organisation de la diversité neutre et adaptative identique à celle du chêne pédonculé.

Cette synthèse bibliographique nous a permis de faire le point des connaissances accumulées sur le caractère de débourrement, connaissances à la fois sur la génétique et l'écophysiologie du caractère. Les principaux déterminants climatiques, hormonaux et moléculaires ont été mis en lumière sans pour autant qu'une vision intégrative du phénomène ne voit le jour de manière claire. La partie expérimentale de cette thèse a été mise en place afin d'apporter des éléments de réponse quant aux déterminants génétiques du débourrement, mais également pour mettre au point un outil d'évaluation de ce caractère sur la simple base d'analyses moléculaires relativement faciles à mettre en œuvre et peu coûteuses.

Objectifs de la thèse



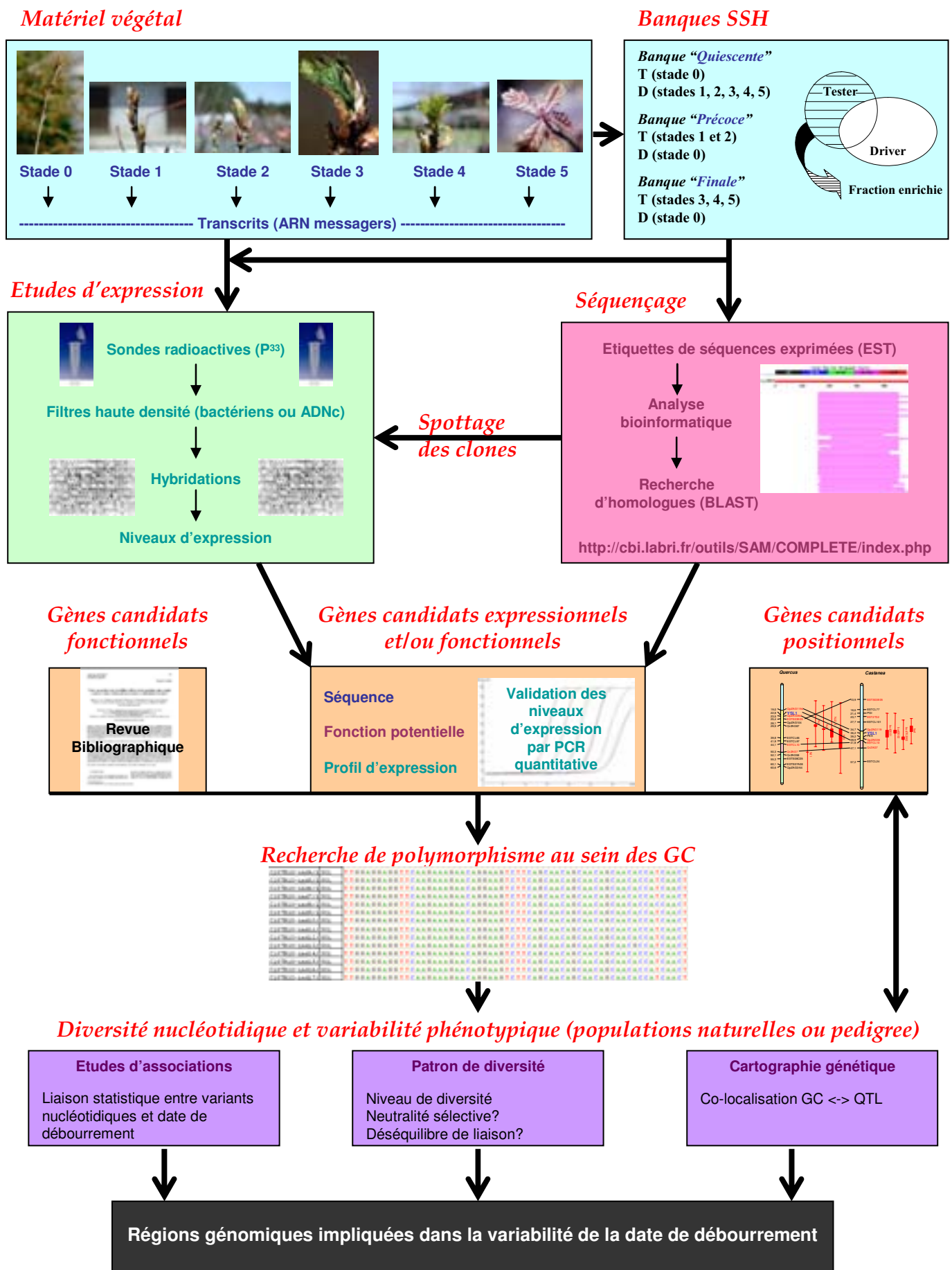


Figure 19 : Synopsis de la stratégie de thèse employée.

Dans le contexte du réchauffement climatique global, les forestiers et les aménagistes s'interrogent sur la capacité d'adaptation des espèces longévives à répondre rapidement à ces changements brutaux. En effet, les mêmes variations de température annoncées par les modèles climatiques vont se produire sur une période beaucoup plus courte qu'au cours du réchauffement postglaciaire. Ces modifications généreront des pressions de sélection qui auront un impact direct sur les dates de débourrement des arbres et sur la longueur de leur saison de végétation et, au-delà, sur la distribution et la survie de ces espèces, modifiant par la même les équilibres écologiques. Or, ces questions sont souvent abordées sous le seul angle climatique et environnemental sans considérer la composante génétique. Les deux composantes majeures de la réponse adaptative concernent le niveau de diversité présent dans les populations soumises à ces nouvelles pressions de sélection, ainsi que la plasticité phénotypique des individus. La connaissance du niveau et de l'organisation de la diversité génétique constitue l'objectif premier de cette thèse. La stratégie adoptée (figure 19) consiste à combiner des informations provenant à la fois : i/ de l'étude de la variation des caractères quantitatifs (date de débourrement) en population naturelle et de leur décomposition en composantes à hérédité mendélienne (les QTL) au travers de la cartographie génétique, et ii/ de la variation nucléotidique dans des gènes candidats identifiés soit dans la littérature soit au travers d'expériences de génomique menées sur le chêne.

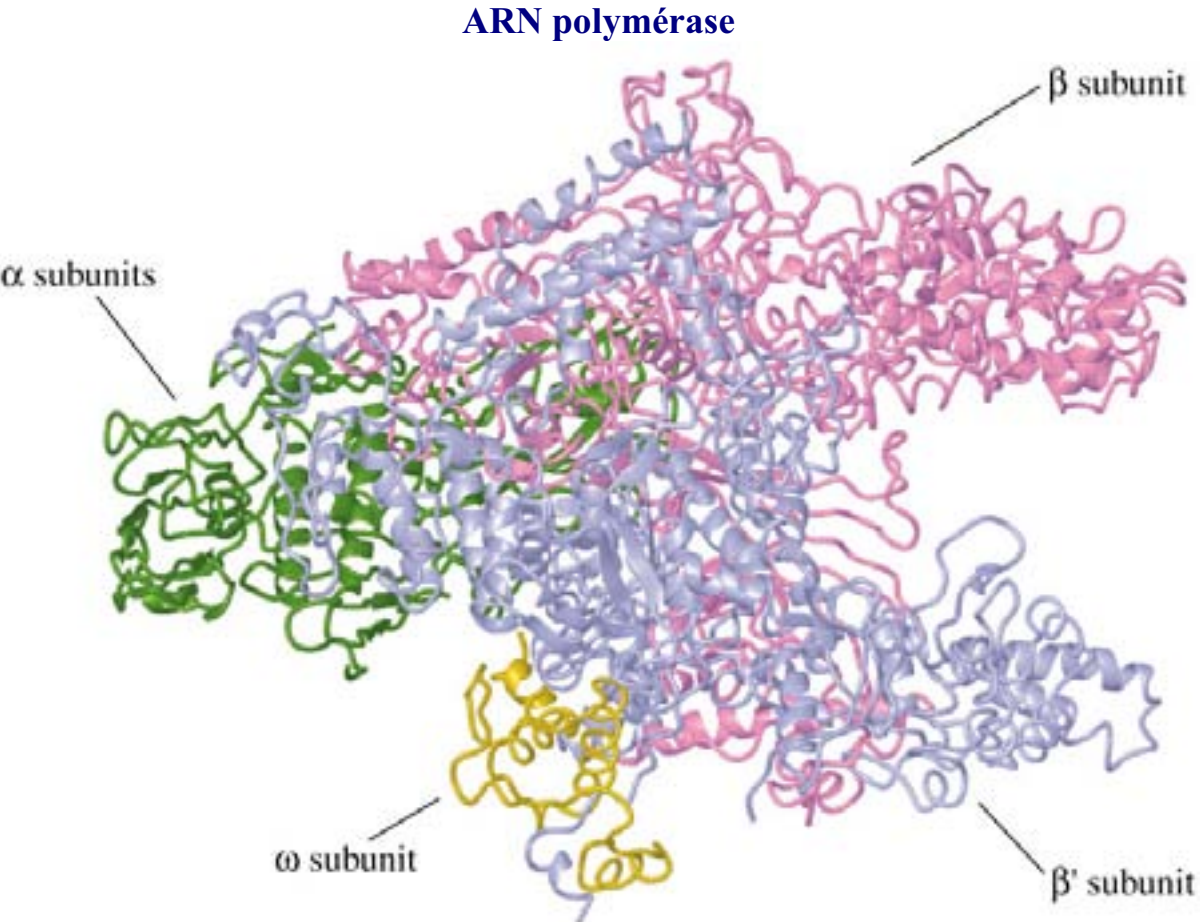
Sur le plan expérimental, la thèse se propose de poursuivre trois objectifs :

1. Identifier un certain nombre de gènes candidats sur la base de leur profil d'expression au cours du processus de débourrement (Chapitre B et Annexe I) ;
2. Cartographier ces gènes candidats sur les cartes génétiques existantes du chêne et du châtaignier afin d'étudier les colocalisations entre GC et QTL (Chapitre C et Annexe II);
3. Etudier en populations naturelles les associations statistiques entre variants alléliques des GC et variation de la date de débourrement (Chapitre D et Annexe III).

En complément des gènes identifiés sur la base de leur profil d'expression, la synthèse bibliographique nous a également permis d'identifier des gènes candidats fonctionnels inclus

dans les objectifs 2 et 3. De plus, préalablement aux études d'association, les données de séquences, issues de l'étape de recherche de polymorphisme au sein des GC, peuvent être exploitées afin d'étudier la structuration de la diversité nucléotidique et de rechercher des « signatures » de la sélection naturelle au sein des gènes candidats. *In fine*, l'intégration de l'ensemble des données d'associations, de cartographie génétique et de structuration de la diversité devrait permettre d'identifier les gènes impliqués dans la variation de la date du débourrement chez les chênes.

*B. Analyse transcriptomique du débourrement chez le
chêne sessile*



La première partie expérimentale de cette thèse est décrite dans le chapitre B qui suit. Cette approche transcriptomique visait principalement à identifier un certain nombre de gènes candidats sur la base de critères d'expression différentielle au cours du processus de débourrement réalisé en conditions naturelles. Le second objectif visé était de préciser les processus fonctionnels fortement régulés au cours du processus, en complément des études déjà menées sur le sujet. Pour ce faire, il a d'abord fallu créer les ressources en étiquettes de gènes exprimés (EST) nécessaires pour mener à bien ce travail. La faible disponibilité de séquences de chêne dans les bases de données a, en effet, imposée ce choix. Le choix de la technique s'est porté sur la construction de banques soustractives et suppressives (SSH) afin d'enrichir préférentiellement en fragments spécifiques d'un stade développemental du débourrement et s'offrir la possibilité d'isoler des transcrits faiblement exprimés. L'utilisation de filtres à haute densité hybridés avec des sondes radioactives spécifiques des différents stades de débourrement devait ensuite nous permettre de répondre aux deux objectifs fixés plus haut, la RT-PCR quantitative en temps réel venant compléter l'étude pour valider les niveaux d'expression des gènes candidats potentiels.

Tableau 4 : Caractéristiques des populations échantillonnées dans le cadre de l’approche transcriptomique.

<i>Région de provenance</i>	<i>Altitude</i>	<i>Données climatiques</i>			<i>Peuplements</i>	
		<i>Référence</i>	<i>Température</i>	<i>Pluviosité</i>	<i>Nombre</i>	<i>Superficie</i>
07-NORD-EST gréseux	180-600m	Phalsbourg (377m)	8,9°C (14,2°C)	806 mm (430 mm)	17	530,91 ha
11-NORD-EST limons et argiles	190-450m	Mirecourt (294m)	9,2°C (14,4°C)	747 mm (392 mm)	28	813,96 ha



Figure 20 : Stades développementaux échantillonnés au cours du débourrement.

① Matériel et Méthodes

Dans toute la suite du paragraphe, lors de l'utilisation de kit de biologie moléculaire, le nom du kit ainsi que du fournisseur sera mentionné. Dans tous les cas, nous avons suivi le protocole fourni par le fabricant et celui-ci ne sera pas détaillé dans le texte.

1.1. Le matériel végétal

1.1.1. Les populations échantillonnées

Dans cette première partie du travail expérimental, nous avons restreint l'échantillonnage à deux populations de chêne sessile (*Quercus petraea* (Matt.) Liebl.), originaires du Nord Est de la France (Tableau 4). La classification des populations utilisées est celle des régions de provenance française contenant des peuplements classés, qui seules peuvent être utilisées pour le reboisement. La numérotation des deux populations se réfère à cette classification. C'est la disponibilité en plants qui nous a amené à sélectionner parmi deux régions de provenance au lieu d'une seule, mais nous pouvons considérer ces deux régions comme unique du point de vue du débourrement. Ainsi, la variabilité exploitée dans cette expérience est celle existante à l'intérieur de la population entre les individus, et non la variabilité existante entre populations géographiques. Environ un millier de plants, âgés de 1 an, a donc été installé en pépinière en février 2002, après une année de croissance en pépinière.

1.1.2. Les stades de débourrement

Les stades de débourrement échantillonnés dans cette étude sont les stades morphologiques, classiquement utilisés pour la notation de la phénologie du débourrement. La figure 20 illustre ces différents stades depuis le stade quiescent (stade 0) jusqu'au développement des entre-nœuds et l'élongation de la tige (stade 5). Au stade 0, le bourgeon est quiescent et aucun signe d'activité extérieure n'est visible ; au stade 1, le bourgeon se renfle ; au stade 2, il y a ouverture du bourgeon ; au stade 3, les feuilles débutent leur expansion ; au stade 4, au moins une feuille est complètement sortie du bourgeon ; au stade 5, les entre-nœuds ont débuté leur élongation. En parallèle, nous avons effectué des coupes

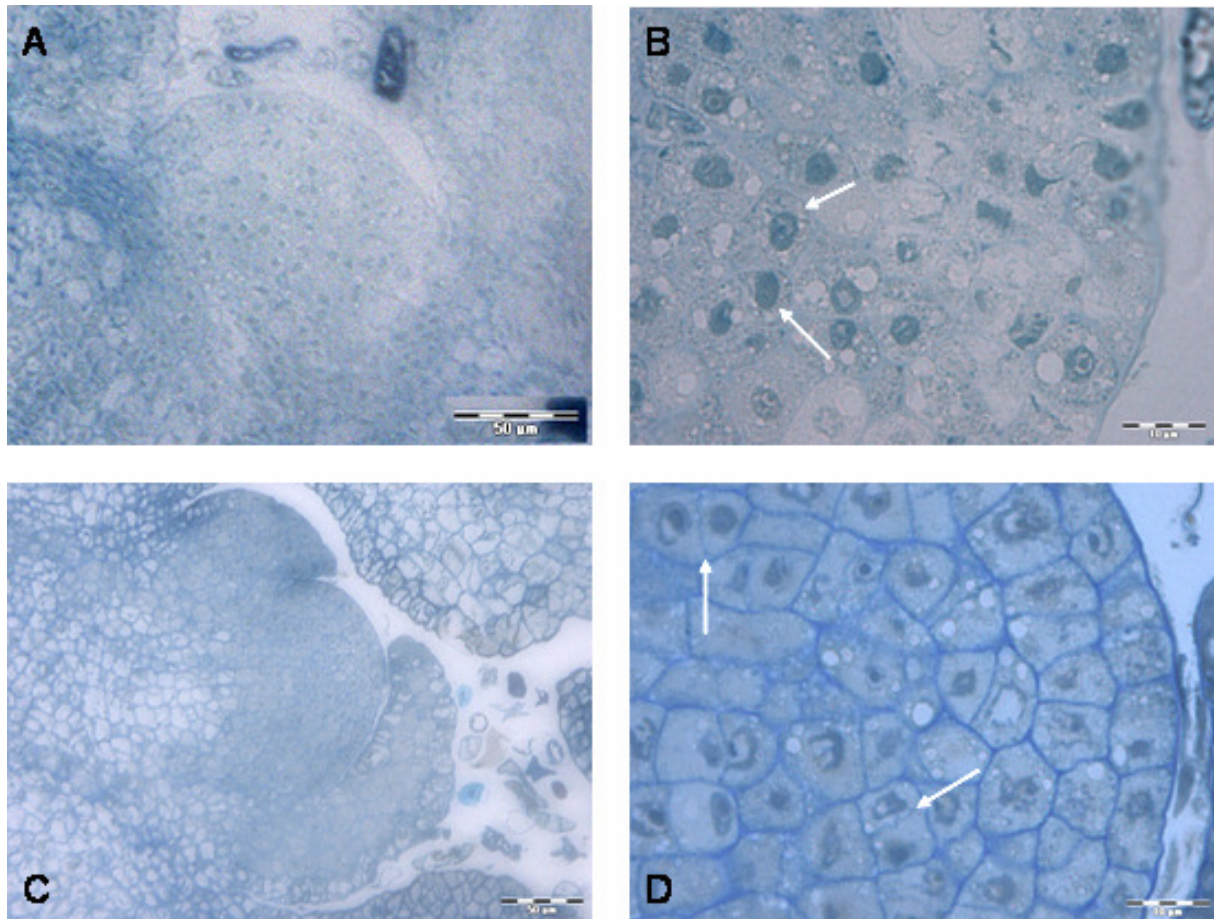


Figure 21 : Coupes anatomiques de méristèmes de bourgeons de chêne sessile au stade 0 (A et B) et au stade 2 (C et D). A. Méristème apical au stade 0 (50µm). On ne trouve pas d'activité mitotique. B. Méristème apical au stade 0 (10µm). On distingue des cellules méristématiques typiques (flèches) : microvacuolisation, nucléoles, cytoplasme à forte densité, volume nucléaire important. C. Organisation d'un méristème apical au stade 2 (50µm). Les primordia foliaires sont visibles de chaque côté du méristème. On observe une activité mitotique, visible en D. Organisation d'un méristème apical au stade 2 (10µm). Cellules en division (flèches). Les vacuoles sont plus visibles qu'au stade 0.

Tableau 5 : Matériel échantillonné aux différents stades de débourrement.

Stade	Nombre	Quantité
Stade 0 (21 Février)	70 Individus	1,6g
Stade 0bis (25 Mars)	70 individus	0,9g
Stade 1 (3-6 Avril)	70 Individus	3,6g
Stade 2 (5-15 Avril)	32 Individus	3,6g
Stade 3 (15-20 Avril)	69 Individus	9,5g
Stade 4 (20-24 Avril)	13 Individus	2g
Stade 5 (24-26 Avril)	67 Individus	31,1g

anatomiques de certains stades de débourrement présentés à la figure 21. Pour ces coupes observées en microscopie optique, les échantillons ont été fixés durant 4 heures à 4°C dans un tampon phosphaté (pH=7,2 / 10mM) contenant 2,5% de glutaraldéhyde puis traités toute la nuit à 0°C dans un tampon phosphaté contenant 1% de tétr oxyde d'osmium. Après trois rinçages à l'eau distillée, les bourgeons ont été fixés 1 h à 4°C à l'obscurité dans une solution aqueuse d'acétate d'uranyle à 2%. Les échantillons ont alors été progressivement déshydratés à température ambiante dans une série de bains contenant des concentrations croissantes d'éthanol puis inclus dans de l'Araldite (résine Epoxy). Des sections de 1µm faites au microtome ont été colorées au bleu de toluidine et observées sous microscope optique Leica DMLB (Leica Corporate, Solms, Deutschland) à l'aide du logiciel d'analyse SIS (Kodak, New-York, USA). Ces coupes nous ont permis de confirmer l'état quiescent des bourgeons récoltés au stade 0. En effet, les cellules du méristème apical ne présentent pas de figure de division à ce stade mais au contraire une organisation typique de cellules méristématiques avec un volume nucléaire important, une micro vacuolisation et un cytoplasme à forte densité. A l'inverse, des coupes de méristème au stade 2 montrent une activité de division cellulaire importante ainsi que le développement de primordiaux foliaires autour du méristème. Nous n'avons pas pu réaliser de coupes anatomiques des autres stades de débourrement par manque de matériel végétal approprié ou de temps pour la mise au point de coupes sur ces stades.

1.1.3. La récolte du matériel

Le suivi de la phénologie a été effectué à partir de l'installation des plants en pépinière (21 Février 2002). Durant le débourrement, la récolte des bourgeons terminaux s'est déroulée tous les jours, dans la matinée, jusqu'au 26 Avril 2002. Seuls les bourgeons terminaux ont été prélevés afin de ne pas « polluer » l'analyse des niveaux de transcrits par des phénomènes liés à la dominance apicale qu'exerce le bourgeon terminal sur les bourgeons axillaires sous-jacents. Ainsi, chaque plant échantillonné à un stade donné était étiqueté et ne faisait l'objet que d'une seule récolte. Le tableau 5 présente les quantités de bourgeons ainsi que le nombre d'individus prélevés pour chacun des stades. Les bourgeons échantillonnés étaient immédiatement plongés dans l'azote liquide avant d'être stockés à -80°C. Concernant le stade 0, nous avons échantillonné à deux reprises : une première fois au moment de l'installation des plants durant la phase de quiescence proprement dite (21 Février) puis une seconde fois au moment de la reprise d'activité du premier plant (25 Mars). Notre intention était ainsi d'avoir

une cinétique plus fine au cours de la quiescence, c'est-à-dire durant la phase de quiescence proprement dite (Stade 0) et juste avant le premier signe visible de reprise de croissance du bourgeon (Stade 0bis). Malheureusement, les quantités de bourgeons récoltés ne nous ont pas permis de conserver ces deux stades distincts et ils ont été réunis dans la suite des expérimentations.

1.2. La construction et l'exploitation des banques SSH

1.2.1. Extraction des ARN totaux et synthèse des ADNc

Nous avons extraits les ARN totaux à partir du protocole de Chang *et al.* (1993), après avoir préalablement testé en parallèle le kit RNA Easy Plant (Macherey-Nagel, Düren, Deutschland). Les ARN extraits avec le kit étaient de faible qualité et difficilement exploitables pour la suite donc nous avons opté pour l'extraction à l'aide du protocole de Chang (1993) adapté par Dubos (2003). La seule modification que nous ayons faite à ce protocole était la quantité de matériel utilisé qui était de 0,25g à 0,5g de matériel frais pour 15 ml de tampon d'extraction, chaque extraction étant répétée trois fois. Les échantillons d'ARN totaux ont ensuite été digérés à 37°C pendant 1 heure en présence de 10 U de DNase RNase Free RQ1 (Promega, Madison, USA) et d'inhibiteur de RNase : Rnasin (Promega), et ce afin d'éliminer les contaminations en ADN génomique. Les protéines et l'ADN génomique dégradé ont été éliminés par purification sur colonne RNeasy (Qiagen, Hilden, Deutschland). Les extraits ont ensuite été analysés par électrophorèse sur gel d'agarose (1,2%, TBE 0.5X, 0,1% BET) et spectrophotométrie. La quantité et la qualité des ARN extraits ont été déterminées par la mesure des densités optiques à 260nm et 280nm. Les rapports de ces deux densités (DO_{260}/DO_{280}) étaient compris entre 1,7 et 1,9 ce qui correspond à une qualité satisfaisante puisque la quantité d'ARN totaux est presque deux fois supérieure à celle des protéines. Les rendements obtenus variaient entre 12 et 100 µg d'ARN totaux par gramme de matériel frais en fonction du stade considéré.

La synthèse des ADNc s'est effectuée de deux manières différentes en fonction de l'utilisation ultérieure (banques SSH ou PCR quantitative en temps réel). Nous décrivons ici le protocole utilisé dans le cadre de la construction des banques SSH (Diatchenko *et al.*, 1996) ou de la synthèse des sondes radioactives, l'autre étant décrit dans le paragraphe relatif à la PCR quantitative. Les ADNc ont été synthétisés à partir du kit SMART cDNA Synthesis kit (BD Biosciences Clontech, Paolo Alto, USA). Ce kit offre l'avantage de pouvoir travailler à partir de très petites quantités d'ARN totaux (1 µg) ce qui se révélait intéressant dans notre cas puisque le matériel végétal était très limité dans le cas du stade 0 (2,5g au total sur les stades 0 et 0bis). Le désavantage de la méthode est l'utilisation d'une phase d'amplification par PCR des ADNc double brin. Elle nécessite une mise au point préalable afin de stopper la réaction PCR en cours de phase exponentielle d'amplification et donc de conserver les proportions relatives des différents transcrits dans le mélange. Cinq µL de produit réactionnel

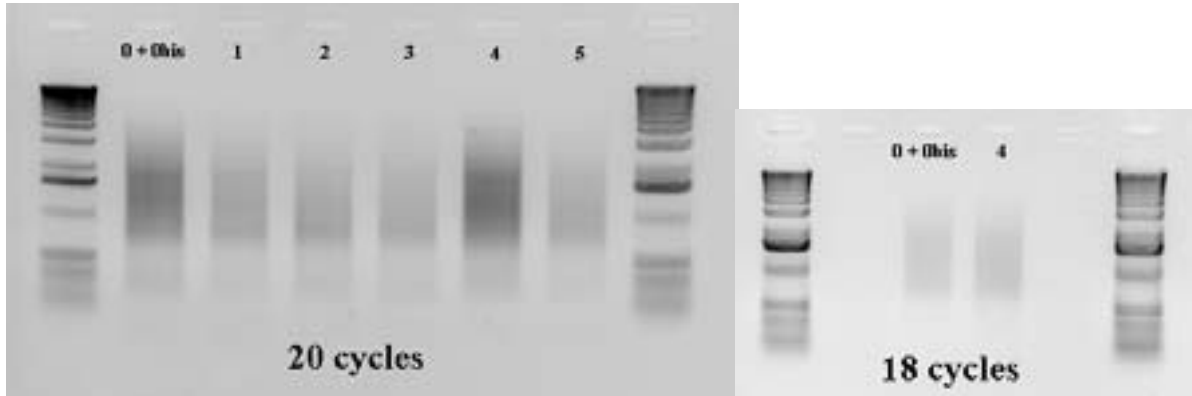


Figure 22 : Electrophorèse de 5µL de produit d'amplification du 2^{ème} brin des ADN complémentaires des stades de débourrement, après 20 cycles. Les stades 0+0bis et 4 ont été réamplifiés par la suite et la réaction stoppée à 18 cycles au lieu de 20.

Tableau 6 : Populations d'ADNc utilisées pour la construction des banques SSH.

<i>Banque</i>	<i>Population Tester</i>	<i>Population Driver</i>
Quiescente (Q)	Mélange des stades 0 et 0 bis (1/2 de chaque)	Mélange des stades 1, 2, 3, 4 et 5 (1/5 de chaque)
Précoce (P)	Mélange des stades 1 et 2 (1/2 de chaque)	Mélange des stades 0 et 0 bis (1/2 de chaque)
Finale (F)	Mélange des stades 3, 4 et 5 (1/3 de chaque)	Mélange des stades 0 et 0 bis (1/2 de chaque)

ont ainsi été prélevés après 15, 17, 19, 21 et 24 cycles. La phase de plateau apparaissait entre 21 et 24 cycles et nous avons opté pour 18 ou 20 cycles selon le stade afin de conserver la proportionnalité (Figure 22). Les ADNc synthétisés ont été purifiés par une extraction au chloroforme alcool isoamylique (proportion 24 :1) et précipités à l'éthanol avant d'être repris dans un volume de 30 µL d'eau stérile. Pour finir, les ADNc synthétisés ont été visualisés sur gel d'agarose (1,2%, TAE 1X, 0,1% BET) pour vérifier leur quantité et leur qualité.

1.2.2. Construction des banques SSH

Nous avons construit 3 banques SSH à partir des populations d'ARN totaux précédemment extraits. La méthode SSH est basée sur l'amplification sélective de fragments différentiellement exprimés entre deux conditions. On appelle « tester » la population qui contient les transcrits spécifiques et « driver » celle qui sert de référence. Les deux populations sont hybridées et les séquences hybrides sont éliminées. De fait, les ADNc restants, non hybridés, sont ceux présents dans la population « tester » mais absent de la population « driver ». Le tableau 6 présente les populations « tester » et « driver » utilisées pour la construction des banques SSH. La stratégie employée nous a permis d'obtenir une première banque précoce (P) contenant les transcrits spécifiques du début du débourrement (Stades 1 et 2), une seconde (banque F) spécifique de la fin du débourrement (Stades 3, 4 et 5) et, enfin, la banque inverse aux deux premières contenant les transcrits spécifiques du stade quiescent (banque Q). Pour cela, les populations d'ARN totaux ont été mélangées et la synthèse des ADNc, à l'aide du kit SMART (BD Biosciences Clontech), s'est faite à partir des populations d'ARN totaux en mélange, et ce pour conserver la proportionnalité des transcrits entre les différents stades lors de la phase d'amplification du 2^{ème} brin par PCR. Nous avons construit les banques SSH à l'aide du kit Clontech PCR-Select cDNA Subtraction (BD Biosciences Clontech) en suivant les instructions données par le fournisseur. Les ADNc sont tout d'abord purifiés sur colonne de chromatographie puis digérés à l'aide de l'enzyme Rsa I. Après avoir été de nouveau purifiés, les ADNc sont ligués à des adaptateurs et hybridés deux fois de suite en présence d'un excès d'ADNc du « driver ». Deux PCR successives permettent ensuite d'amplifier spécifiquement les transcrits de la population « tester ».

1.2.3. Clonage des banques et préparation des filtres bactériens à haute densité

Chaque banque d'ADNc a d'abord été soumise à une élongation de 30mn à 72°C afin de rajouter une adénine à l'extrémité de chaque fragment. Les trois préparations d'ADNc issus de la construction des banques ont été clonés au sein du vecteur plasmidique pCR 4-TOPO puis le vecteur transformé dans TOP10 *E. coli*, à l'aide du kit TOPO TA cloning (Invitrogen, Carlsbad, USA). Nous avons alors étalé 10 à 100µL de produit de transformation sur des boîtes de Pétri contenant un milieu de culture LB Agar ainsi que 50µg/ml d'ampicilline et 80µg/ml de X-gal. Les boîtes ont ensuite été incubées toute la nuit à 37°C pour permettre la croissance des colonies bactériennes. Nous avons, dans un premier temps, sélectionné 48 colonies par banque qui ont été amplifiées par PCR dans 50µL de volume réactionnel puis séquencées, selon le protocole décrit au paragraphe 1.3.3. « Séquençage des clones différentiellement exprimés ». Cette étape a permis d'estimer le taux de redondance des différentes banques. Après cette étape de validation, quelques centaines de clones par banque ont été prélevés sur grandes boîtes de Pétri (22 x 22cm) à l'aide du robot BioPick (BioRobotics, Woburn, USA), disponible sur le site INRA d'Auzeville (Génopole de Toulouse). Les clones repiqués ont été mis en culture, à 37°C toute la nuit, en plaques 384 puits en présence de LB « Freezing » contenant 100µg/ml d'ampicilline et 4,4% de glycérol. L'ajout de glycérol, en faible proportion, directement dans le milieu de culture, permet la mise en congélation immédiate des clones après croissance sans ajout de glycérol. A ce stade, chaque clone a été nommé conformément au numéro de la plaque 384, sa banque d'origine et l'identifiant du puit. Ainsi, le clone QSSH2-J18 provient du puit J18 de la plaque 2 de la banque Q (quiescente).

Chacune des colonies, précédemment individualisée, a alors été placée sur une membrane de nylon (22 x 22cm) à l'aide du robot MicroGrid II (BioRobotics). Le robot place directement les colonies sur le filtre, lui-même installé sur une plaque contenant du LB Agar. La plaque est alors mise en culture à 37°C jusqu'à ce que les colonies apparaissent, en général une nuit est nécessaire. Les cellules bactériennes sont alors lysées en plaçant les filtres durant 4 minutes sur du papier Whatmann imbibé de solution de lyse (0,5M NaOH; 1,5M NaCl). Après avoir traités les filtres en bain-marie à 88°C pendant 4mn et dans une solution de neutralisation (1 M Tris-HCl pH 7,5; 1,5M NaCl) pendant 4mn, l'ADN est fixé aux filtres. Enfin, les filtres sont incubés 30mn à 37°C dans une solution de protéinase K (0,26mg

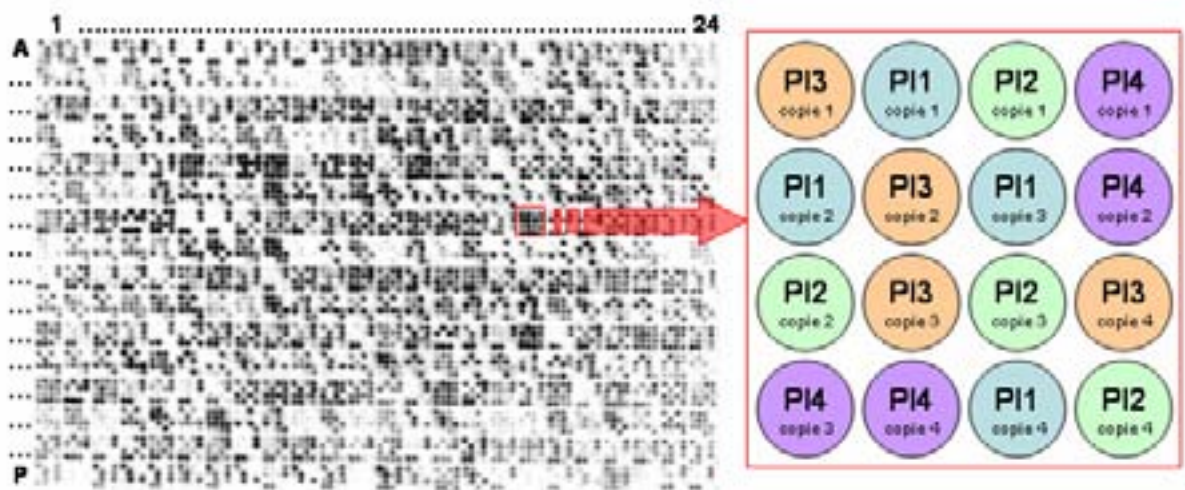


Figure 23 : Schéma des filtres bactériens à haute densité. Le filtre est organisé de la même façon qu'une plaque 384 puits et jusqu'à 16 colonies peuvent être « spottées » dans l'emplacement correspondant à un puit, ce qui nous a permis de placer 4 répétitions de chaque clone sur une même membrane puisque nous avons 4 plaques 384 à installer par banque. (PI = Plaque)

proteinase K/ml; 50 mM Tris-HCl; 50 mM EDTA; 0,1M NaCl; 1% N-Lauroyl Sarcosine; pH 7,4) et rincés dans le 2XSSC. Deux copies de chaque filtre ont été produites par banque, c'est-à-dire deux filtres Q, deux filtres P et deux filtres F. Chaque banque était contenue dans un peu plus de deux plaques 384 et une quatrième plaque de contrôle, composée de Desmine et de milieu LB seul, a été utilisée. La Desmine est une sous unité constitutive des fibres intermédiaires des muscles du squelette et cardiaque. Elle est par conséquent absente de tous les tissus végétaux et nous servira de contrôle positif puisqu'une quantité fixe de transcrit de Desmine est additionné au mélange réactionnel lors de la synthèse des sondes (cf. paragraphe suivant 1.3.1.). A l'inverse, le milieu LB sert de contrôle négatif. Au final, chaque filtre (7 x 10 cm) contenait entre 5000 et 6000 spots correspondant chacun à un clone, chaque clone étant répété 4 fois par membrane (Figure 23). La totalité de la surface de la membrane était également pourvue de contrôles positifs et négatifs provenant de la 4^{ème} plaque 384.

1.3. Tri différentiel des transcrits et séquençage des EST

1.3.1. Hybridation des filtres bactériens à haute densité

Les ARN totaux provenant des pool Q (Stade 0+ 0bis), P (stades 1 + 2) et F (stades 3 + 4 + 5) ont été utilisés pour préparer les sondes complexes à l'aide du kit SMART cDNA Synthesis (BD Biosciences Clontech). Chaque pool d'ADNc a ensuite été purifié au chloroforme alcool isoamylique, comme décrit précédemment. Le marquage des sondes a été réalisé à l'aide du kit Prime-a-Gene Labeling System (Promega, Madison, USA) en présence de 30 μCi de $[\alpha\text{-}^{33}\text{P}]$ dATP, 30 μCi de $[\alpha\text{-}^{33}\text{P}]$ dCTP après que 1 ng d'ADN de Desmine digéré ait été ajouté au pool d'ADNc. La purification des sondes a été effectuée avec le kit QIAquick Nucleotide Removal (Qiagen) et l'activité spécifique de chaque sonde estimée grâce à un compteur à scintillation. Les filtres ont d'abord été pré incubés durant 5 à 6 heures à 65°C sous agitation dans un tampon d'hybridation (5X SSC, 5X Denhardt's, 0,5% SDS, 1 mg/ml ADN de sperme de saumon dénaturé). Les filtres ont alors été incubés pendant une nuit à 65°C dans 20ml de tampon d'hybridation contenant la sonde complexe dénaturée (10mn en bain-marie à ébullition puis 10mn sur glace). A la suite de l'hybridation, les filtres ont été lavés en deux étapes : i/ 5mn à température ambiante dans une solution de 2X SSC, 0,5% SDS, puis ii/ deux lavages successifs pendant 15mn à 65°C dans la même solution. Les filtres ont alors été exposés à un film autoradiographique Imaging screen K (Bio-Rad, Hercules, USA) durant 3 jours. A la fin de la durée d'exposition, les radiographies ont été digitalisées à l'aide du système PhosphorImager (Amersham-Biosciences, Buckinghamshire, UK) avec la résolution maximale de 50 μm puis les signaux d'hybridation détectés et quantifiés grâce au logiciel ArrayVision (Imaging Research Inc., St Catharines, Canada). La quantification de l'intensité du signal pour chaque spot a été réalisée après soustraction du bruit de fond local autour du spot. La normalisation des signaux entre les filtres a été calculée à l'aide des signaux d'intensité des contrôles. En effet, trois cent quatre vingt huit réplicats d'ADN de Desmine étaient présents sur la membrane à des endroits spécifiques. Chaque signal d'intensité de clone a été normalisé en divisant sa valeur par le ratio entre la valeur moyenne d'intensité de la Desmine sur le filtre et la valeur moyenne globale sur tous les filtres. Ce type de filtre peut être hybridés 4 à 5 fois. Il est alors nécessaire de laver les filtres pour éliminer la radioactivité résiduelle. Les filtres sont pour cela incubés dans une première solution (0,4 M NaOH) à 45°C durant 30mn sous agitation. Ils sont ensuite incubés dans une

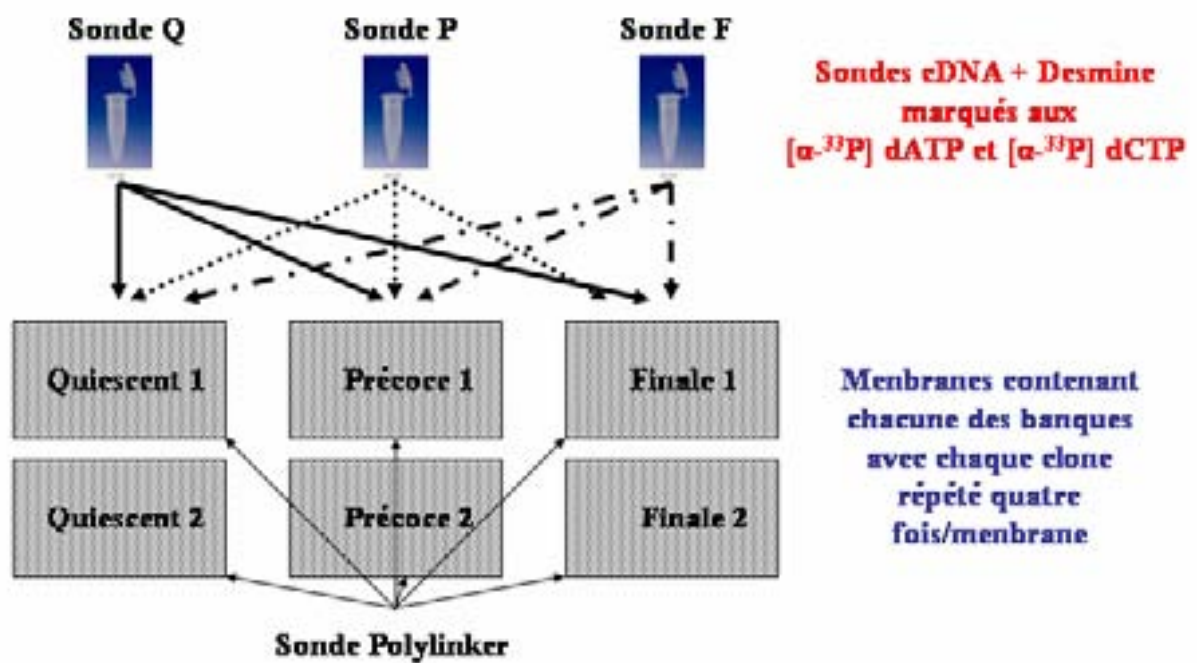


Figure 24 : Schéma d'hybridation des filtres bactériens à haute densité.

seconde solution (0,1X SSC, 0,1% SDS, 0,2 M TrisHCl pH 7,5) à 45°C pendant 15mn sous agitation. Ils se conservent enroulés en film plastique et placés à -20°C.

Au final, chaque clone issu des banques SSH était présent 4 fois sur une même membrane et chaque membrane était dupliquée, mettant à notre disposition 8 répétitions du signal d'intensité pour chaque clone avec chacune des sondes (Figure 24). En plus des hybridations avec les sondes complexes, une dernière hybridation a été réalisée sur chaque membrane à l'aide d'une sonde polylinker marquée comme précédemment et correspondant au vecteur seul afin de déterminer la présence ou l'absence d'un clone donné sur la membrane.

1.3.2. Analyse de variance des données d'expression

L'analyse des données d'hybridation a été effectuée grâce au logiciel SAS (Statistical Analysis System version 6.12, SAS Institute, Cary, NC). Nous avons réalisé une analyse de variance (ANOVA) afin d'estimer les effets de la sonde, du filtre et de l'interaction entre ces derniers pour les valeurs d'intensité de chaque clone, à l'aide de la procédure PROC GLM, selon le modèle linéaire suivant :

$$Y_{ijk} = \mu + F_i + P_j + (FXP)_{ij} + e_{ijk}$$

Où Y_{ijk} = valeur d'intensité de la $k^{\text{ième}}$ répétition du clone,

μ = moyenne d'intensité de toutes les répétitions,

F_i = effet du filtre i ($i=1-2$),

P_j = effet de la sonde j ($j=1-3$),

$(FXP)_{ij}$ = effet d'interaction entre le filtre i et la sonde j ,

e_{ijk} = effet résiduel non pris en compte par les effets précédents.

Après que les clones présentant un effet sonde aient été repérés, nous avons effectué une comparaison de moyenne entre sondes à l'aide de l'option MEANS /SNK, disponible au sein de la procédure PROC GLM, et ceci afin de classer les niveaux d'intensité de chaque clone en fonction du stade de débourement. Ce premier tri nous a permis de sélectionner les clones différenciellement exprimés ($P < 0,05$) entre, au moins, deux conditions (Q, P et F). Les clones ainsi sélectionnés ont été analysés plus finement, et tout d'abord séquencés.

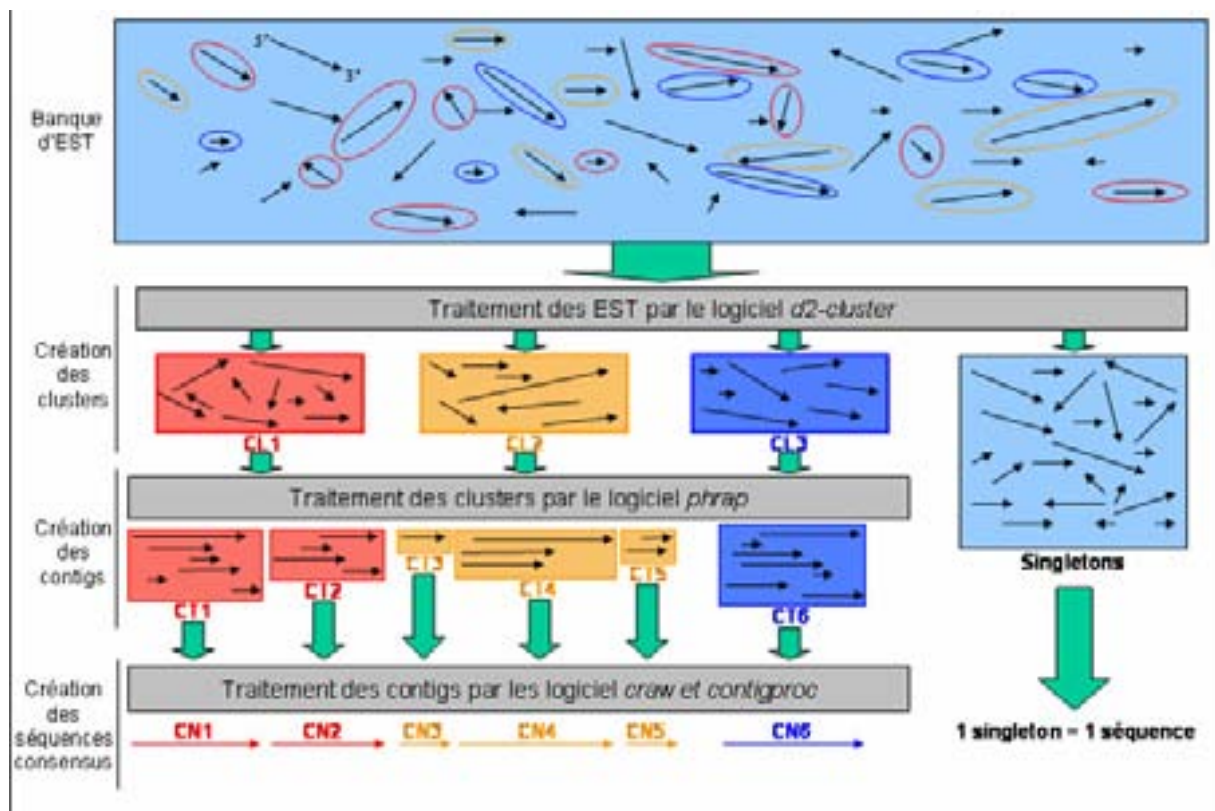


Figure 25 : Schéma du traitement bioinformatique des EST. (Tiré de Boury, 2003)

1.3.3. Séquençage et traitement bioinformatique des EST

Environ 900 clones différentiellement exprimés au cours du débourement, ont été sélectionnés et les inserts de ces clones ont d'abord été amplifiés selon le protocole suivant. L'amplification des ADN plasmidiques a été réalisée à l'aide d'une machine GeneAmp PCR system 9700 (Perkin-Elmer, Wellesley, USA) dans 20 μ L de volume réactionnel, contenant 0,3 unité de *Taq* DNA polymérase (Invitrogen), 1X de tampon PCR (20 mM Tris pH 8.4 / 50 mM KCl / 2 mM MgCl₂), 0,2 μ M d'amorces M13 forward (5'-TGACCGGCAGCAAAATG-3') et M13 reverse (5'-GGAAACAGCTATGACCATG-3'), 200 μ M de chaque dNTP et une petite quantité de cultures bactériennes. Les conditions PCR étaient de 94°C durant 10mn, suivi de 40 cycles à 94°C pendant 1mn, 55°C pendant 1mn puis 72°C durant 1mn. La réaction se terminait par une phase d'élongation finale de 10mn à 72°C. Après avoir été vérifiés sur gel d'agarose (2%, TBE 0,5X, 0,1% BET), les produits PCR obtenus ont été purifiés grâce au kit QIAquick 96 PCR purification (Qiagen) puis séquencés à l'aide du kit DYEnamic™ ET Dye Terminator (Amersham Biosciences Inc.) et révélés sur un séquenceur capillaire Megabace 1000 automated DNA sequencer (Amersham Biosciences Inc.). Les séquences produites ont été analysées à l'aide d'outils informatiques fournis par le Centre de Bioinformatique de Bordeaux afin de créer un jeu de données non redondant d'EST. Ce traitement s'est déroulé en trois étapes principales. Premièrement, un traitement qualitatif des séquences est effectué, qui élimine la séquence du vecteur de clonage et la séquence des adaptateurs ayant servis durant la construction des banques et masque les régions de faible qualité. Dans un deuxième temps, les outils stackPACK™ (Electric Genetics Corporation, Reston, USA) nous ont permis de réaliser le regroupement (clustering) puis l'assemblage et l'alignement de manière à éliminer la redondance entre EST en regroupant les séquences correspondant à un même gène (Figure 25). Enfin, une fonction potentielle a été assignée à chaque EST grâce à la procédure BLAST (Altschul *et al.*, 1997) : un produit de gène a ainsi été assigné à chaque EST, sur la base des similarités de séquences avec des protéines de fonction connues présentes dans les bases de données trEMBL et Swiss-Prot à l'aide de la procédure BLASTX, en retenant les annotations pour lesquelles les e-value étaient $\leq 10^{-10}$. Le jeu de données complet est disponible *via* internet à l'adresse <http://cbi.labri.fr/outils/SAM/COMPLETE/index.php> sous le nom de projet « Quercus petraea bud ESTs ». L'ensemble des transcrits a également été déposé dans la base de données EMBL

(Kulikova *et al.*, 2004) (<http://www.ebi.ac.uk/embl/index.html>) sous les numéros d'accessions CR627501 à CR628310 et CR926157.

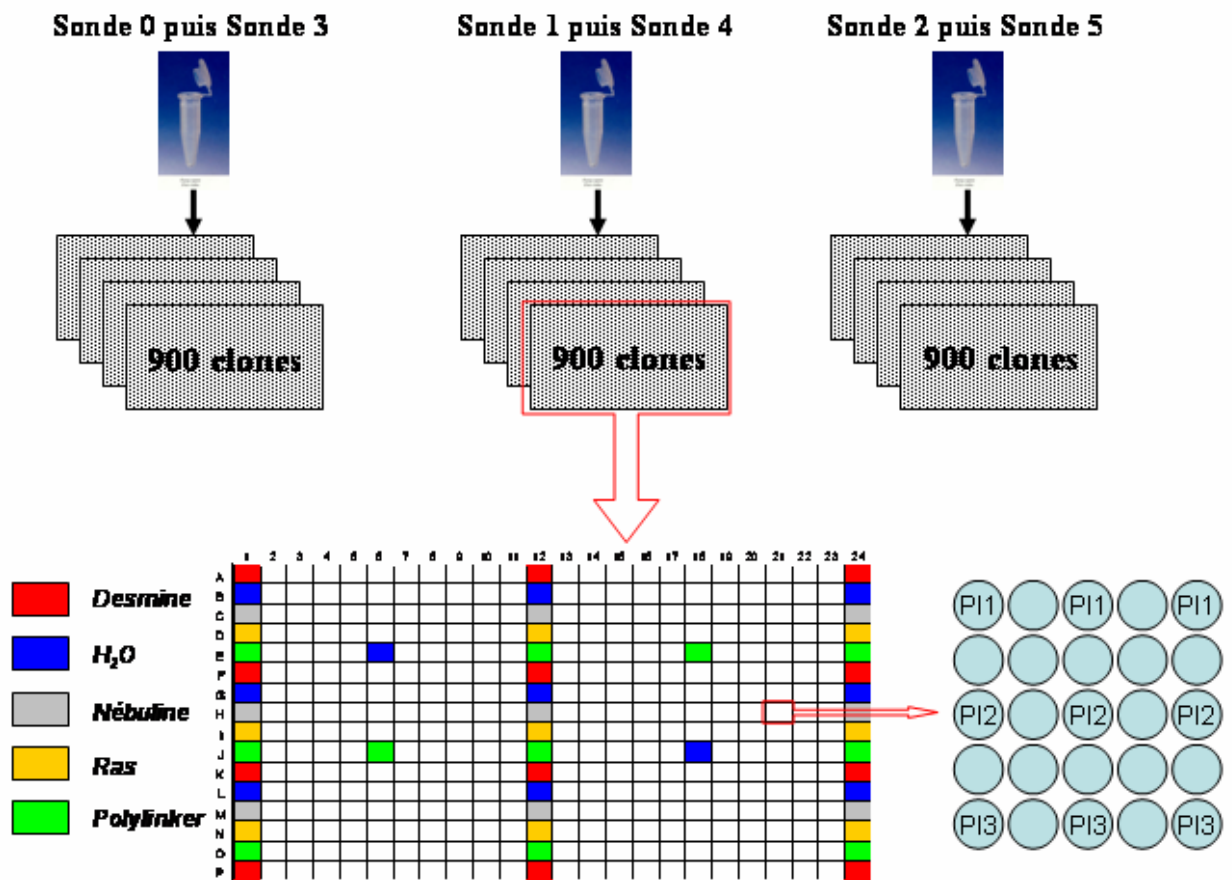


Figure 26 : Schéma d'organisation et d'hybridation des filtres ADNc à haute densité. (P1 = Plaque).

1.4. Analyse quantitative de l'accumulation des transcrits au cours du débourrement

1.4.1. Les filtres ADNc à haute densité

Les niveaux d'expression des clones intéressants, sélectionnés au cours du tri précédent à l'aide des filtres bactériens à haute densité, ont été estimés à chaque stade de débourrement grâce à des filtres ADNc à haute densité. Ces filtres sont plus sensibles que les filtres bactériens utilisés précédemment et permettent une meilleure estimation des quantités de transcrits à chaque stade. Afin de produire les filtres ADNc, les ADN plasmidiques de chacun des quelques 900 clones retenus ont d'abord été amplifiés dans 50µL en suivant le même protocole que pour le séquençage des clones, décrits dans le paragraphe précédent. Les filtres ont été produits par Eurogentec (Liège, Belgique) par dépôt des produits PCR (20 ng/µL) sur membrane de nitrocellulose grâce au robot Microgrid (BioRobotics). Chaque aiguille de 0,4µm de diamètre dépose 500 nl de produit PCR avec un espacement de 500 µm sur des membranes de 7 x 10 cm saturées en présence de NaOH 0,1M à une densité de 33 clones par cm². Chacun des 900 clones était répété 3 fois sur 12 filtres identiques qui nous ont servis pour les hybridations. 156 contrôles (36 d'ADN de desmine, 33 d'ADN de polylinker, 33 de solution d'H₂O, 27 d'ADN de nébuline, 27 d'ADN de RAS ou récepteur tyrosine kinase) ont été installés sur chaque membrane afin de pouvoir normaliser les valeurs d'intensité entre filtres (Figure 26). Les ARN totaux de chaque stade développemental ont été utilisés pour construire des sondes radioactives. Le protocole est décrit dans le paragraphe 1.3.1. Quatre filtres identiques ont été hybridés avec la sonde 0, la sonde 1, la sonde 2, la sonde 3, la sonde 4 et la sonde 5. Les filtres ont ensuite été lavés, exposés et les signaux digitalisés, comme décrit précédemment. Pour éliminer la radioactivité présente sur les filtres, ces derniers ont été incubés à température ambiante dans une solution bouillante (0,1X SSC, 0,5% SDS) durant 15mn. Cette étape a été effectuée deux fois pour s'assurer que la radioactivité résiduelle était totalement retirée du filtre.

Les valeurs d'intensité normalisées (normalisation par la Desmine) ont été analysées par ANOVA pour tester les effets de la sonde, du filtre et de l'interaction entre les deux. Le modèle utilisé est indiqué dans le paragraphe 1.3.2. Des comparaisons de moyenne ont aussi été réalisées pour détecter les différences significatives de niveaux d'expression entre sonde pour chaque cible présente sur le filtre. Les données d'expression ont été organisées selon la méthode du clustering pour pouvoir les interpréter plus aisément. La méthode permet en effet le regroupement des gènes par profil d'expression similaire grâce au calcul de distance

Tableau 7 : Caractéristiques des fragments amplifiés par PCR quantitative en temps réel.

<i>Nom du clone / Fonction potentielle</i>	<i>Longueur EST (longueur produit PCR)</i>	<i>Amorces</i>	<i>T°C hybridation</i>
01B07 S-phase kinase-associated protein 1A	437 bp (235 bp)	F. CGATGAGGAGCTCAAGAACT R. CCATTGGTTCTCCCTGCGAA	53°
06G05 Unknown (Zwh21.1 hypothetical protein)	415 bp (279 bp)	F. ACAACGGGGCAAGTCCATTAGA R. CTATATAATCTGAGGCAGCGG	53°
08A01 Galactinol synthase	298 bp (210 bp)	F. TCAAATTCAATGTTCTCTGGATGG R. GAGGTACTTCAATGCTGGCTTCTT	53°
01D04 60S ribosomal protein L13	288 bp (203 bp)	F. AGCTTCTTCATTGCTCAGGTAA R. CAGGTACAATAGGACGGAGTTTTTC	53°
08C11 Unknown (T19F6.8 hypothetical protein)	453 bp (316 bp)	F. TGAAGCAGGAAGCGTCCAAAGT R. GCCATTCTAGGTGTTGAAGG	53°
07F12 Glyceraldehyde 3-phosphate dehydrogenase	557 bp (280 bp)	F. AAGTCTACGACAGAAACATTA R. CTACCAATTGTCTTGCTCCTCTA	49°
08G04 Unknown (At2g14910 hypothetical protein)	672 bp (393 bp)	F. CAGCATTACTTGGCAGGGATAGGA R. CAACCGGAAAAGGGATGAAGTGT	55°
07B10 DOF zinc finger protein DAG2	574 bp (360 bp)	F. CTACAGCCTCACTCAACCAAGATA R. TCCAAAGCTGAAAGATGAGTAGAG	53°
09C04 Pollen specific protein SF21	621 bp (410 bp)	F. ACACATGCCATAGAAGTAGAGT R. AGCCAGCTCTTGTCACTTAT	53°
07B05 Late embryogenesis abundant protein Lea5	642 bp (380 bp)	F. TGTGGTTCAACAGCAGATTT R. TCTTCTACAGCAGCAACACC	53°
08B04 Alpha-amylase/subtilisin inhibitor (RASI)	495 bp (327 bp)	F. TTCGTGAGCCTTCGTCTTCTC R. ACAGTGCCGGACGTCCTCTTCAA	55°

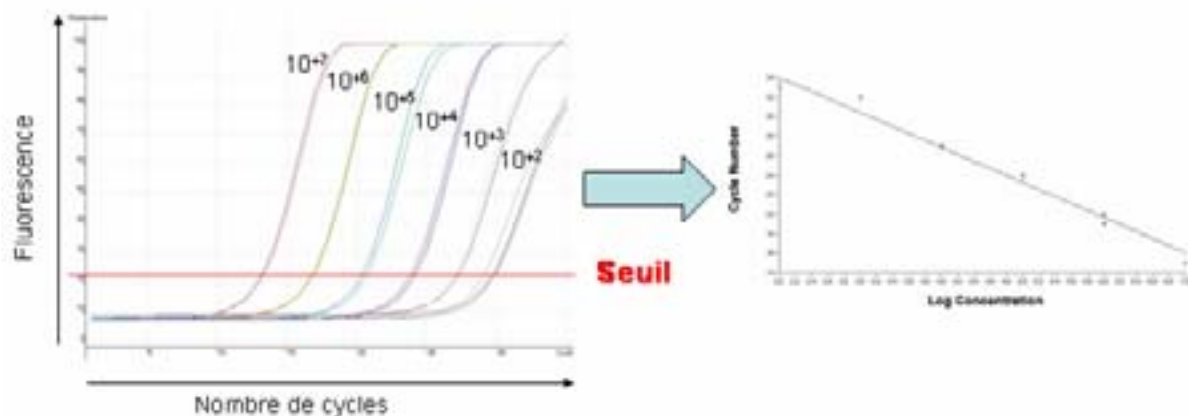


Figure 27 : Principe de détermination de la courbe standard à partir de la gamme de dilution.

euclidienne, représentative de la similarité entre gènes. Nous avons utilisé le logiciel EPCLUST, disponible à l'URL : <http://ep.ebi.ac.uk/EP/>.

1.4.2. La PCR quantitative en temps réel

Les ARN totaux ont été extraits conformément au protocole décrit en 1.2.1. Ils ont ensuite été traités à la RNase free DNase I (Promega) et purifiés à l'aide de colonnes RNeasy (Qiagen). La qualité et la quantité des ARN ont été évaluées par mesures spectrophotométriques et sur gel d'agarose. Des aliquots de 2,5µg d'ARN totaux ont été rétro-transcrits en présence de 200 U de la SuperScript™ RNase H⁻ Reverse Transcriptase (Invitrogen), additionnée de 2µM d'amorces oligo d(T)22 et 0,7mM de dNTP. Le volume final de la réaction était de 15µL et la synthèse des ADNc simple brin s'est effectué à 37°C pendant 1 heure. L'activité de la transcriptase inverse a été stoppée par incubation 5mn à 95°C. Des dilutions au 1/5^{ème} et au 1/10^{ème} ont été préparées à l'aide d'eau distillée, et les réactions de PCR quantitative ont été réalisées avec 1 ou 2µL de ces ADNc dilués, en fonction du niveau d'expression du produit ciblé. Nous avons utilisé une série de dilutions du produit PCR de la cible pour la quantification des quantités de transcrits. Ces échantillons standards ont été obtenus par amplification des ADNc à l'aide de paires d'amorces spécifiques des gènes cibles (Tableau 7). Les réactions PCR ont été réalisées en 35 cycles (45s à 94°C, 45s à la température d'hybridation et 60s à 72°C), après une phase de dénaturation initiale de 10mn à 94°C, et la réaction s'est terminée par une phase d'élongation de 10mn à 72°C. La taille, la qualité et la quantité des fragments amplifiés ont été contrôlées par électrophorèse sur gel d'agarose. La concentration ainsi que la taille des produits PCR obtenus a permis le calcul du nombre de molécules en solution par la formule suivante :

$$([X \text{ ng}/\mu\text{L}] / [\text{Nb pb}] \times 660) \times 6,023 \times 10^{23} = Y \text{ molecules}/\mu\text{L}$$

où $6,023 \times 10^{23}$ est le nombre d'Avogadro, c'est-à-dire le nombre d'entités élémentaires (molécules dans notre cas) dans une mole de matière ; 660 est la masse (en ng) d'une mole de paire de bases. Le calcul du nombre de molécules en solution permet donc de réaliser une gamme de dilutions, typiquement de 10^9 molécules/µL à 10^2 molécules/µL.

Ces dilutions en série, de concentrations connues, permettent de tracer une courbe standard de la concentration initiale d'un échantillon d'ADNc en fonction du nombre de cycles pour atteindre un seuil de fluorescence donné, lequel est déterminé statistiquement par le logiciel d'analyse (Figure 27). Une fois la courbe standard réalisée, il est possible de

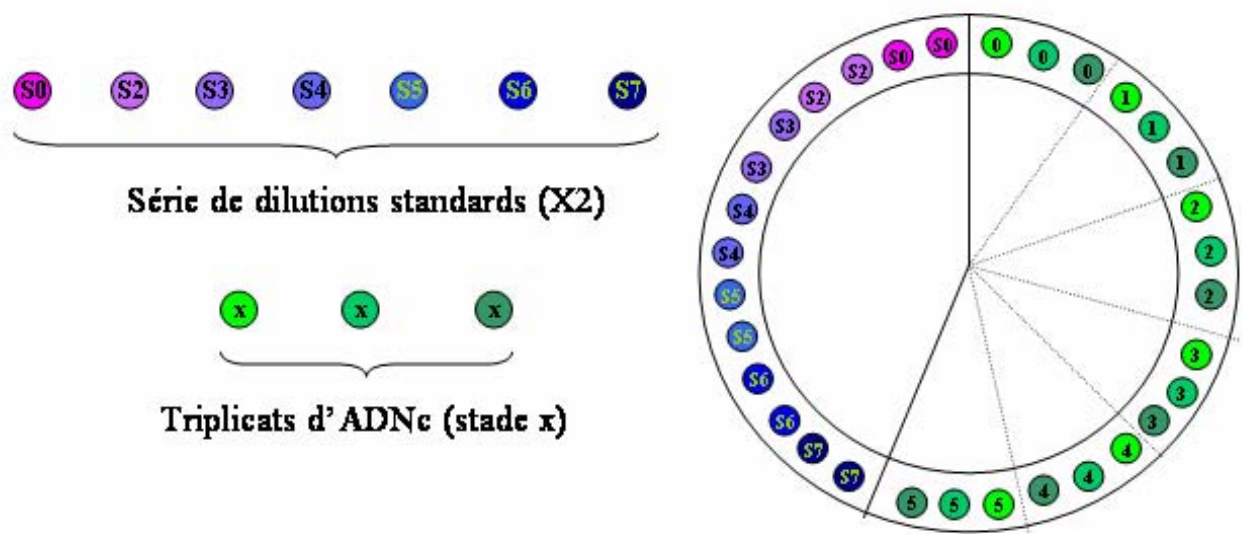


Figure 28 : Disposition des échantillons analysés simultanément durant chaque réaction de PCR quantitative en temps réel.

Tableau 8 : Conditions PCR des fragments analysés en RT-PCR quantitative en temps réel.

<i>Fragment</i>	<i>Temp. Hybrid. testées</i>	<i>Durée Hybrid. testée</i>	<i>[MgCl₂] testées</i>	<i>Conditions Light Cycler</i> T°C Hybrid.-Tps Hybrid.- [MgCl ₂]-Tps Elong.
<i>60S ribosomal protein</i>	53°	5s	2 mM	53°-5s-2mM-10s
<i>G3PDH</i>	49°-50°	5s	2-2,5mM	49°-5s-3mM-10s
<i>Cluster 8</i> <i>Lea 5</i>	53°-54°	5s	1,5-2-2,5-3-4 mM	54°-5s-2,5mM-15s
<i>Cluster 1</i>	53°	5s	1,5 mM	53°-5s-1,5mM-10s
<i>Clone 06G05</i>	53°	5s	1,5-2 mM	53°-5s-2mM-10s
<i>Clone 08A01</i>	53°	5s	2 mM	53°-5s-2mM-10s
<i>Clone 08C11</i>	53°-54°	5s	2-2,5-3-3,5-4 mM	53°-5s-2,5mM-10s
<i>Cluster 24</i>	53°	3s-4s-5s	1,5-2-2,5-3 mM	53°-4s-3mM-16s
<i>Clone 07B10</i>	53°-54°	3s-4s-5s	1,5-2-2,5-3-3,5-4 mM	53°-4s-3mM-15s
<i>Clone 08G04</i>	54°-55°-56°	5s	2-2,5 mM	55°-5s-2,5mM-16s 2µl
<i>Cluster 11</i>	54°-55°	5s	2-2,5 mM	55°-5s-3mM-13s

déterminer la concentration de n'importe quel échantillon en reportant le nombre de cycles atteints à la valeur seuil sur l'axe des ordonnées de la courbe standard. Les échantillons de la gamme de dilution, ainsi que des échantillons d'eau distillée, ont donc été inclus, en double, dans chaque réaction de PCR quantitative pour déterminer une courbe standard propre à chaque réaction. Par ailleurs le niveau d'expression d'un fragment donné a été estimé simultanément sur trois copies des populations d'ADNc testées, chaque copie correspondant à une rétro-transcription indépendante. Le dispositif utilisé dans chaque réaction de PCR quantitative en temps réel est schématisé à la Figure 28. Pour chaque fragment analysé suivant ce dispositif, une réaction supplémentaire a été effectuée avec des populations d'ADNc issues d'une série d'extractions d'ARN indépendantes de celles utilisées en routine afin de valider les niveaux de transcrits observés. Les réactions se sont déroulées à l'aide d'une machine LightCycler 2000 (Roche Applied Science, Penzberg, Germany) dans un volume réactionnel final de 10µL contenant 1µL de LightCycler FastStart DNA Master SYBR Green I, 300nM de chaque amorce (Tableau 7), 1,5 à 3mM de MgCl₂ et 1 ou 2µL d'échantillon d'ADNc. Les conditions relatives à chaque fragment analysé sont présentées dans le Tableau 8. Du point de vue expérimental, ce travail a débuté au sein du groupe de Marie-Theres Hauser au Centre de Génétique Appliquée, dirigé par Josef Glössl, au sein de l'Université des Sciences Agronomiques de Vienne (Autriche). Nous avons travaillé à l'aide d'une machine Rotor-Gene 2000 (Corbett Research, Sydney, Australia) et d'un mix PCR « maison » mis au point par cette équipe (Karsai *et al.*, 2002). Malheureusement, il nous a été impossible de mettre au point ce mix sur machine LightCycler, la principale raison à cela tenant au fait que le LightCycler fait appel à des capillaires de verre d'un diamètre infime alors que le Rotor-Gene permet de travailler en tubes plastique 0,2 ml. Or, la difficulté majeure de mise au point tient au fait que les composants réactionnels se concentrent autour des parois par des phénomènes de tension de surface. Le diamètre des tubes plastiques permet d'éliminer ce type de phénomène par l'utilisation de détergent inclus dans le mix PCR (TRITON X-100, NP40) mais il semblerait que le diamètre des capillaires (1,5 mm de diamètre extérieur) ne permette pas d'éliminer ces phénomènes. Une fois chaque fragment analysé, les quantités de transcrits estimées au cours du débourement pour chacun des gènes ont été relativisées à celles d'un gène de contrôle, dont l'expression est relativement stable durant tout le processus étudié. Après avoir testé la G3PDH dont les niveaux de transcrits variaient de façon non négligeable durant le débourement, nous avons opté pour une protéine ribosomale 60S.

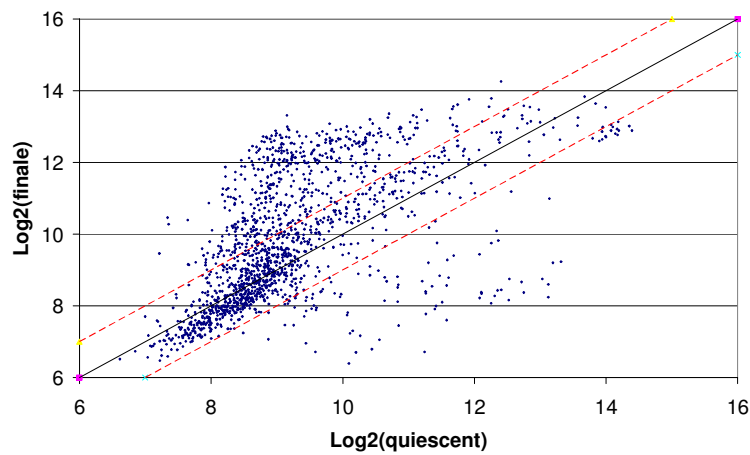
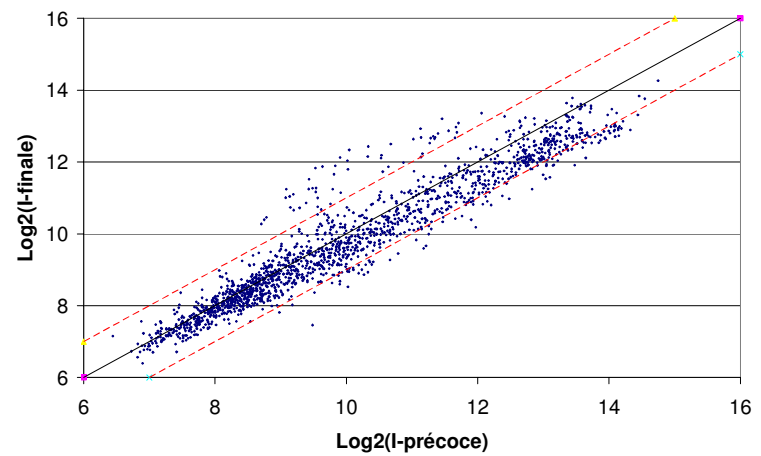
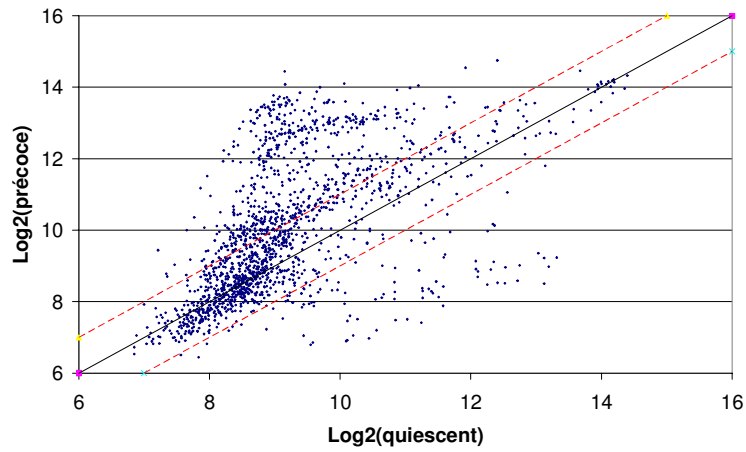


Figure 29 : Nuages de points représentant les niveaux d'intensité des 1800 EST entre les stades complexes quiescent, précoce et final au cours du débourrement. Les droites pointillées en rouge représentent une sur- (ou sous-) expression d'un ratio 2.

② Résultats et Discussion

2.1. L'étape de tri différentiel et les banques d'EST

2.1.1. Sélection des clones différentiellement exprimés

La première étape après la construction des banques a consisté à estimer la redondance des banques. Les 48 premières séquences, issues de chaque banque, ont été analysées afin d'obtenir un jeu de données contenant les transcrits uniques (Contigs + Singletons) et d'estimer la redondance, calculée comme suit :

Redondance = (Nombre de contigs – Nombre d'EST dans les contigs) / Nombre total d'EST. Nous avons ainsi obtenu une valeur de redondance de 35% pour la banque quiescente, 24% pour la banque précoce et 37% pour la banque finale. Ces valeurs sont tout à fait acceptables et nous ont permis d'envisager l'utilisation ultérieure des banques.

Au total, plus de 2400 clones ont été installés sur les filtres bactériens mais seuls 1800 ont été analysés par ANOVA, les autres n'ayant pas poussés correctement sur les filtres. C'est pourquoi, dans un premier temps, nous avons appliqué un critère de présence à chacun des spots copiés sur les filtres en utilisant le signal d'hybridation obtenu avec la sonde polylinker. Les clones considérés présents étaient ceux dont le ratio Intensité du spot/Intensité du bruit de fond > 1. Après analyse de variance, nous avons retenu les clones présentant un effet sonde significatif à 5%. Ensuite, tous les clones différentiellement exprimés entre au moins deux conditions, c'est-à-dire dont les intensités moyennes étaient significativement différentes, à l'aide de la méthode de Student Newman Keuls (SNK), ont été retenus. Et enfin, parmi ceux là, nous avons conservé ceux dont le niveau d'expression variait deux fois ou plus entre au moins deux conditions. La Figure 29 présentent les nuages de points représentant les niveaux d'expression entre deux conditions. On peut voir que les écarts d'expression les plus marqués s'observent entre les stades précoce et quiescent et final et quiescent. En effet, 604 et 513 clones sont surexprimés plus de deux fois, respectivement aux stades précoce et final, par rapport au stade quiescent. Inversement 98 et 109 sont surexprimés (ratio > 2) au stade quiescent, respectivement par rapport aux stades précoce et final. En revanche, seulement 56 clones sont surexprimés plus de deux fois au stade final par rapport au stade précoce et 103 clones sont en situation inverse. Ce résultat n'est pas surprenant et confirme l'efficacité de la méthode SSH dans l'isolation de transcrits spécifique d'un stade même si cette isolation n'est pas exclusive d'un stade puisque les clones isolés sont exprimés aux différents stades. La spécificité est quantitative et non qualitative dans ce cas. Par ailleurs, on peut dire que les

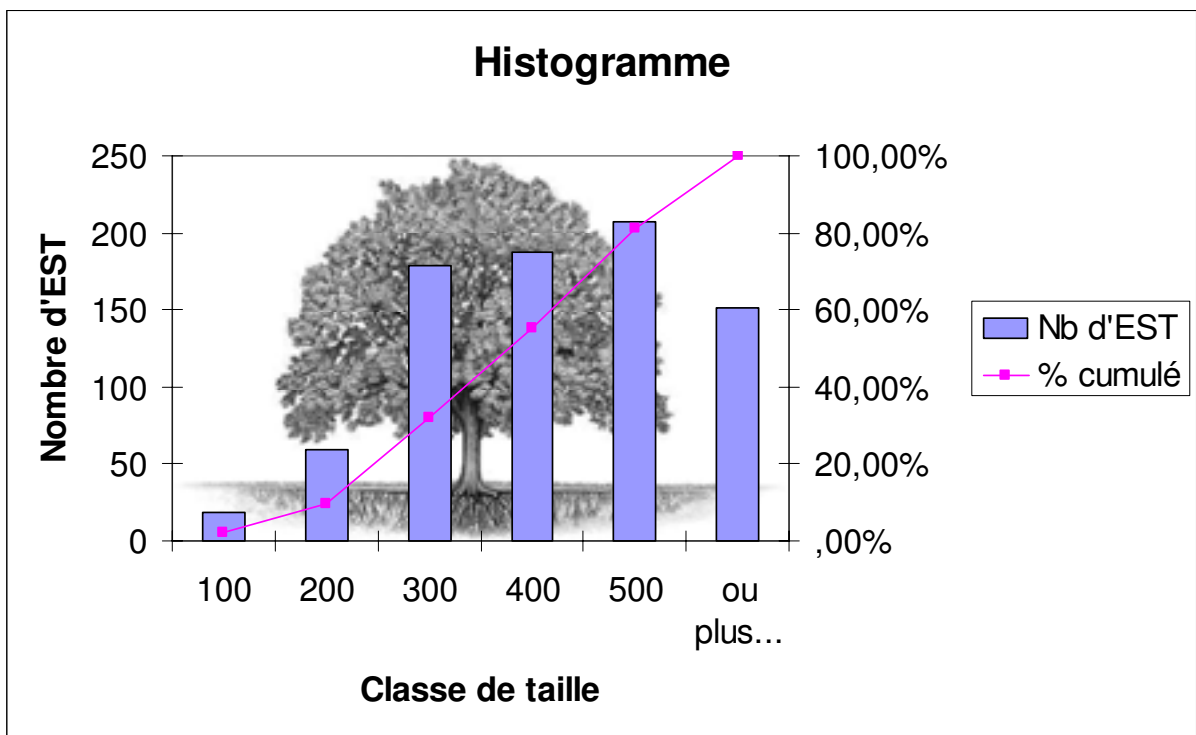


Figure 30 : Répartition par classe de longueur, en pb, des 801 EST séquencées.

banques ayant le mieux enrichies sont les banques précoce et finale puisque entre 500 et 600 clones sont spécifiques de ces deux stades par rapport au stade quiescent. En revanche, 200 clones seulement semblent spécifiques du stade quiescent ce qui laisse supposer que l'enrichissement a moins bien fonctionné ici ou qu'il y a moins d'activité transcriptionnelle à ce stade (on verra que ça n'est pas le cas lors de l'analyse des niveaux d'expression à l'aide des filtres ADNc), ou encore que les transcrits exprimés à ce stade n'en sont pas spécifiques. Enfin, peu de clones sont spécifiques du stade final par rapport au précoce, et inversement, ce qui n'est pas étonnant puisque aucune banque n'avait été construite dans ce sens.

2.1.2. Analyse globale et classification fonctionnelle des EST

Au total, 900 clones ont été retenus après l'étape de tri différentiel, réalisé à l'aide des filtres bactériens puis séquencés. Après l'élimination des séquences du vecteur et des séquences de mauvaise qualité, 801 séquences sont disponibles dont 403, 206 et 192 proviennent respectivement des banques Précoce, Finale et Quiescente. Un total de 302 694 nucléotides a été séquencé, contenant 43,8% de nucléotides GC et 2,4% de nucléotides non assignés. La longueur moyenne d'une EST est de 376 pb, le maximum observé étant de 712 pb et le minimum de 71 pb. Les longueurs les plus représentées sont entre 300 pb et 500 pb (Figure 30), à l'exception de la banque Quiescente qui contient 35% d'EST dont la longueur dépasse 500 pb. Ces 801 EST représentent 233 transcrits uniques, dont 137 contigs incluant 705 EST et 96 singletons. Ces chiffres correspondent à une redondance globale de 71%, nettement supérieure à celles estimées lors de la validation des banques, ce qui indique qu'un séquençage supplémentaire de clones n'apporterait que peu d'éléments nouveaux en termes de nouveaux EST. L'ensemble de ces EST a été déposé dans la base de données nucléotidiques EMBL (Kulikova *et al.*, 2004) sous les numéros d'accèsion CR627501 à CR628310.

Un rôle fonctionnel a été attribué à chacun des transcrits sur la base de la similarité de séquences avec des protéines de fonction connue déposées dans GenBank lorsque la probabilité associée au BLASTX était $\leq 10^{-10}$. Nous nous sommes référés pour cela à la classification fonctionnelle proposée par le MIPS (Munich Information center for Protein Sequences). Trente cinq % des transcrits appartiennent à la catégorie des « Hypothetical proteins » (8%) ou « No hit proteins » (27%). Parmi les trois banques, la catégorie prédominante est « Energy » avec 21% des EST, suivie de « Protein synthesis » et « Cell

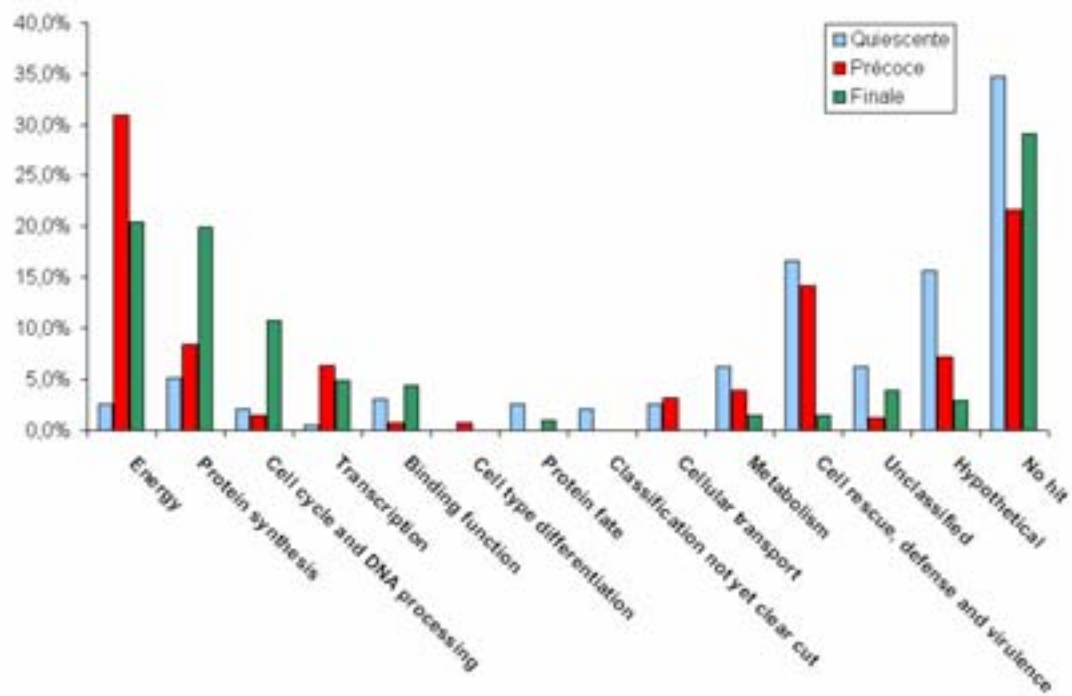


Figure 31 : Catégories fonctionnelles des EST en fonction de la banque d'origine.

rescue, defense and virulence », avec 11% chacune. Les autres catégories contiennent 0,4 à 5% des EST restantes. Mais la comparaison des catégories fonctionnelles entre les trois banques nous éclairent d'avantage (Figure 31). En effet, la catégorie « Energy », par exemple, compte pour 2,6% des EST dans la banque Q, alors qu'elle représente 30,9% et 20,4% des EST, dans les banques P et F respectivement. De la même manière, la catégorie « Protein synthesis » est moins représentée dans la banque Q (5,2%) que dans les banques P (8,4%) et F (19,9%). Une tendance similaire s'observe pour les EST relatives au cycle cellulaire. A contrario, la proportion d'EST classifiées dans « Cell rescue/defense » est plus importante dans la banque Q (16,6%) que dans les banques P et F (14, 1% et 1,5%, respectivement). Ceci est également le cas pour les EST relatives au métabolisme avec 6,2% dans la banque Q, contre 4% et 1,5%, respectivement dans les banques P et F. Enfin, les EST annotées en tant que protéines hypothétiques sont plus nombreuses dans la banque Q (15,5%) que dans les banques P (7%) et F (3%). Afin de s'assurer que l'annotation obtenue n'était pas liée à la taille de l'EST considérée, nous avons réalisé une analyse de variance en testant l'effet de la classe fonctionnelle (3 classes : protéine annotée, hypothétique et inconnue) sur la longueur de l'EST. Il existe un effet significatif de la classe fonctionnelle sur la taille de l'EST. En effet, les EST appartenant à la catégorie « No hit » sont significativement plus courtes que les deux autres catégories (331 pb en moyenne contre 386 pb pour les protéines annotées et 440 pb pour les protéines hypothétiques). Les EST classées dans la catégorie « No hit » sont donc sensiblement plus courtes ce qui pourrait expliquer l'absence de similarités de séquence convenables.

La surreprésentation des protéines hypothétiques dans la banque Quiescente est un résultat prometteur pour l'identification de nouvelles séquences potentiellement impliquées dans le débourrement chez le chêne. Par ailleurs, cette analyse globale montre que les transcrits sont relativement spécifiques d'une banque. En d'autres termes, la méthode SSH a isolé avec succès des transcrits spécifiques d'un stade de débourrement.

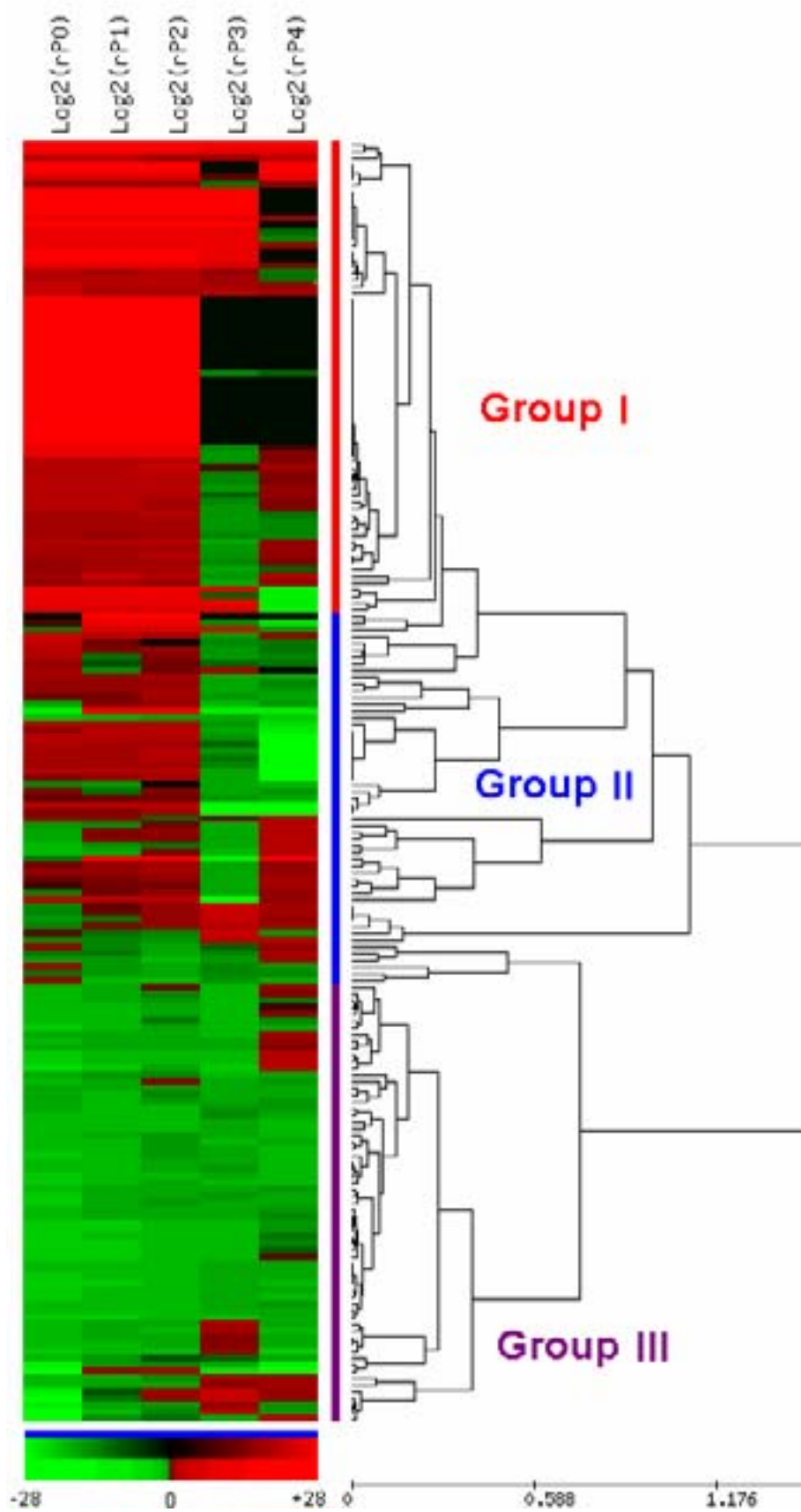


Figure 32 : Regroupement hiérarchique des 190 transcrits uniques séparés en trois groupes d'expression.

2.2. Suivi d'expression de gènes au cours du débourrement

2.2.1. Le regroupement des profils d'expression

Les niveaux d'expression des 900 clones d'ADNc sélectionnés après l'étape de tri différentiel ont été estimés à l'aide des filtres ADNc à haute-densité pour chacun des stades échantillonnés (six stades). Pour chaque hybridation, un critère de présence a tout d'abord été appliqué à chaque spot, en retenant uniquement ceux dont le rapport *Intensité du spot / Intensité du bruit de fond* était supérieur à 2. Ensuite une analyse de variance a été réalisée gène par gène après normalisation grâce à l'intensité de la Desmine sur les filtres. Le plan d'expérience employé ne permet en aucun cas de comparer les niveaux d'expression entre les gènes pour un même stade mais seulement les niveaux d'intensité pour un même gène en fonction du stade, la disponibilité de 12 répétitions par gène pour chaque stade rendant l'analyse de variance statistiquement assez puissante.

Quarante trois transcrits uniques ne présentaient pas d'effet sonde et ont donc été retirés de l'analyse. Les 190 transcrits restants ont été analysés par regroupement hiérarchique agglomératif et peuvent être séparés en trois groupes distincts (Figure 32). Les valeurs reportées dans la figure 32 sont les logarithmes en base 2 des ratios d'intensité par rapport au stade 5. Soixante dix gènes ont été classés dans le groupe I et apparaissent surexprimés aux stades 0, 1 et 2 et sont soit induits, soit réprimés aux stades 3 et 4. Ce groupe de gènes ne présente pas d'expression différentielle entre les stades 0 (quiescent) et 1 (renflement du bourgeon). Il est principalement constitué de gènes dont la fonction est liée au métabolisme, en particulier le métabolisme des carbohydrates avec des EST codant potentiellement pour une galactinol synthase, une glucan endo-1,3- β -glucosidase, une alpha-amylase/subtilisin inhibitor. Ce groupe contient également en proportion importante des gènes relatifs aux mécanismes de défense et protection cellulaire, tels qu'une catalase isozyme 3, des déhydrines, des protéines heat-shock ou encore une LEA5. Ensuite, on trouve des gènes du cycle cellulaire ou codant pour des protéines de liaison, ou encore la G3PDH. Le détail des annotations des transcrits de chaque groupe est donné dans le tableau 2 de l'Annexe I. A l'inverse de la majorité des gènes du premier groupe, les 55 gènes appartenant au groupe II présentent une expression différentielle entre les stades 0 et 1, et sont, dans la plupart des cas, réprimés aux stades ultérieures 3 et 4. Au regard des signaux précoces régulant le débourrement, l'ensemble des gènes de ce groupe est particulièrement intéressant puisque leur expression varie fortement entre le stade quiescent (# 0) et le stade début de débourrement (#

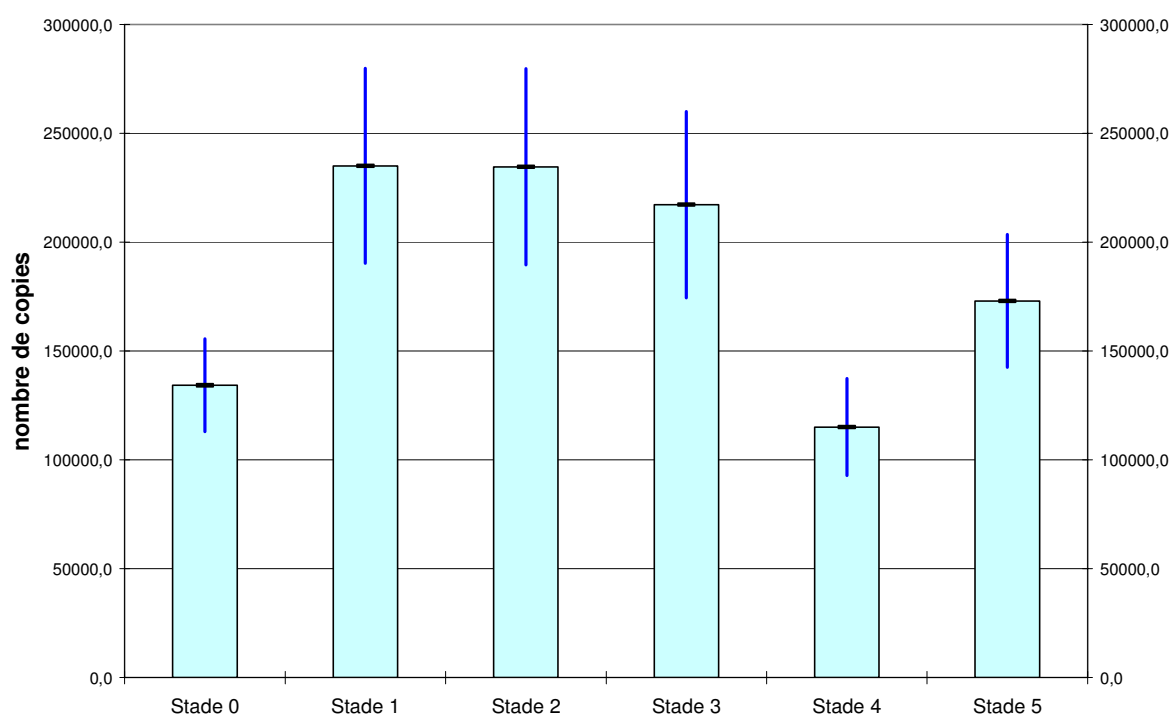


Figure 33 : Niveaux d'expression au cours du débourrement du transcrit codant potentiellement pour une protéine ribosomale L13 60S. Les efficacités des réactions PCR varient entre 90% et 99% et 9 à 12 répétitions ont été utilisées pour chaque stade.

1). Les transcrits relatifs à la synthèse protéique et aux mécanismes de défense et protection cellulaire sont majoritaires dans ce groupe. La plupart des EST présente une forte différence d'expression entre les stades 0 et 1, comme celles codants potentiellement pour des protéines ribosomales ou une protéine metallothionein-like de type 3, ou encore une glutathione S-transférase. Les classes fonctionnelles liées au contrôle transcriptionnel et à l'organisation de la chromatine sont également bien représentées dans le groupe II avec les histones H3 et H4 et des facteurs de transcription tels que DAG2 ou MONOPTEROS. Enfin, 5 autres catégories sont représentées par un ou deux membres dans ce groupe (détail : cf. Annexe I. Tableau 2). Dans le groupe III, les 65 gènes sont réprimés, comparativement au stade 5, durant quasiment tout le débourrement à quelques exceptions près. C'est donc au stade 5, et dans une moindre mesure au stade 4, que ces gènes sont préférentiellement exprimés. Il s'agit principalement de transcrits relatifs aux voies énergétiques, tels que les protéines M du photosystème II, des sous unités de la Rubis-Co, le complexe cytochrome B6-F, une plastocyanine ou encore une ATPsynthase. Des transcrits codant potentiellement pour des protéines impliquées dans la synthèse protéique ou la transcription sont également présents dans ce groupe.

2.2.2. Transcrits étudiés en RT-PCR quantitative en temps réel

Nous avons mesuré les niveaux d'accumulation de transcrits au cours du débourrement pour quelques gènes par la technique de RT-PCR quantitative en temps réel. Ces gènes ont été choisis parmi les plus régulés au cours du phénomène de débourrement et constituent des gènes candidats expressionnels et/ou fonctionnels (cf. § 2.3.2.). Il a fallu, dans un premier temps, rechercher un gène de contrôle dont l'expression était suffisamment stable. Après le test de quelques EST, nous avons opté pour une protéine ribosomale 60S L13 (Figure 33). L'expression de cette protéine est régulée au cours du débourrement mais nous avons validé le profil d'expression sur des extractions d'ARN indépendantes et considérerons donc que ce transcrit constitue notre gène de référence. Les niveaux d'accumulation de transcrits des autres gènes testés sont donc présentés relativement à ceux de ce gène de référence.

Le tableau 8 (page 71) présente les conditions PCR utilisées pour chaque fragment. Il illustre le fait que la phase d'optimisation des réactions PCR a été fastidieuse et a nécessité de tester les concentrations en chlorure de magnésium et en amorces, la température d'hybridation, la durée de la phase d'hybridation et la durée de la phase d'élongation. Le

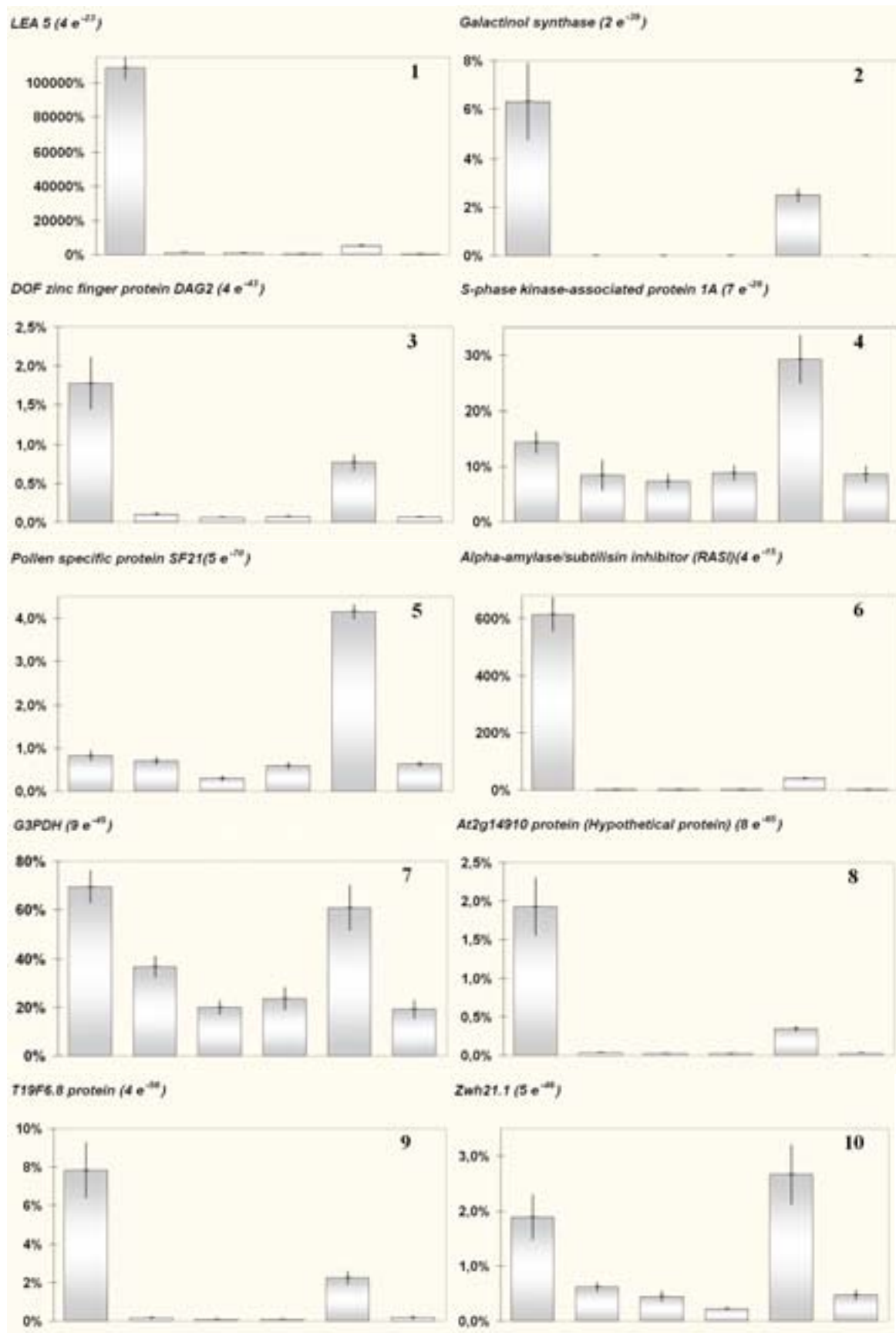


Figure 34 : Niveaux d'accumulation de 10 transcrits au cours du débourrement, estimés par RT-PCR quantitative en temps réel.

principal obstacle à l'utilisation de cette technique réside dans la présence d'artefacts de dimères d'amorces en compétition dans le mélange réactionnel avec le produit PCR recherché. Cette difficulté a déjà été reporté en PCR classique (Brownie *et al.*, 1997) mais apparaît encore plus criante en PCR quantitative (Vandesompele *et al.*, 2002). Une fois cette mise au point effectuée, les niveaux d'accumulation de transcrits de 10 gènes ont été estimés et les résultats sont présentés à la Figure 34. Neuf répétitions au minimum ont été utilisées pour chaque gène afin de s'assurer de la reproductibilité de l'expérience. Les efficacités s'échelonnent entre 77 et 99%, ce qui correspond à une activité correcte de l'enzyme *Taq* polymérase. Les niveaux d'expression varient de 0 à 108 000 % du niveau d'expression de notre gène de contrôle et, dans tous les cas, ces résultats vont dans le même sens (induction ou répression) que ceux obtenus en reverse northern même si les niveaux d'expression dans l'absolu sont parfois différents.

2.2.3. Les mécanismes moléculaires régulés durant le débourrement

Le suivi d'expression des 190 transcrits uniques différentiellement exprimés au cours du débourrement nous permet également de mettre en évidence les principaux mécanismes moléculaires mis en jeu à chaque étape du déroulement de la croissance du bourgeon. En effet, bien qu'une majorité de transcrits soient exprimés durant la totalité des stades analysés, un certain nombre sont spécifiques d'un stade particulier, ou tout du moins significativement induits ou réprimés à un stade donné.

La première catégorie fonctionnelle mise en évidence est celle des transcrits relatifs aux mécanismes de défense et protection cellulaire. En effet, cette catégorie de gènes est majoritaire durant le stade quiescent et en début de débourrement. Si l'on observe le profil d'expression du transcrit codant potentiellement pour une *LEA 5* (Figure 34.1), il apparaît évident qu'un stress hydrique affecte le bourgeon au stade quiescent. En effet, bien que la fonction biologique précise de ces protéines reste indéterminée (Wise et Tunnacliffe, 2004), les LEA ont été isolées dans un panel de tissus végétatifs de plantes sous conditions de stress, en particulier hydrique. En plus des protéines LEA, des transcrits codants pour des protéines *dehydrin-like* et *heat-shock* s'accumulent en début de débourrement, vraisemblablement afin de protéger les cellules du stress hydrique et/ou thermique. De la même manière, les niveaux de transcrits d'une *galactinol synthase* augmentent en début de débourrement (Figure 34.2). La *galactinol synthase* catalyse la première étape dans la voie de biosynthèse des Raffinose

Family Oligosaccharides (RFO), et elle a déjà été décrite comme impliquée dans la tolérance aux stress hydrique et thermique (Pukacka et Wojkiewicz, 2002; Taji *et al.*, 2002; Zhao *et al.*, 2003). Les RFO semblent jouer un rôle dans la tolérance à la dessiccation chez les graines. La conséquence de ces différents mécanismes protecteurs serait que l'intégrité des membranes et des protéines est maintenue dans les tissus du bourgeon. Par ailleurs, il a été montré (Bray *et al.*, 2000) qu'un stress hydrique activait la production d'espèces radicalaires dérivées de l'oxygène (reactive oxygen species, ROS) et augmentait la susceptibilité aux pathogènes. Dans notre expérience, on peut observer qu'un stress oxydatif prend effectivement place durant le débourrement, et est neutralisé, entre autres, par l'activation d'une *catalase* qui agit sur l'élimination des radicaux libres. De plus, la régulation du potentiel osmotique par accumulation de déhydrines, LEA ou de sucres pourrait aussi contribuer à la protection des cellules contre l'oxydation (Rhodes et Hanson, 1993). En parallèle, une *endochitinase PR4 precursor* est induite au cours du stade quiescent et en début de débourrement et pourrait contribuer à la défense contre les pathogènes fongiques.

Concernant le cycle cellulaire, les niveaux d'expression des histones *H3* et *H4*, ainsi que des facteurs de transcription *DAG2* (Figure 34.3) et *MONOPTEROS*, augmentent en début de débourrement. Alors que les niveaux d'expression élevés des histones *H3* et *H4* reflètent une augmentation de l'activité de division cellulaire en général, l'induction de certains facteurs de transcription putatifs pourrait apporter des indications quant aux processus développementaux spécifiquement induits au moment du débourrement. En effet, *DAG2* a été identifié en tant que facteur de transcription impliqué dans le contrôle maternel de la germination de la graine chez *Arabidopsis* (Gualberti *et al.*, 2002). Dans le bourgeon, il pourrait agir sur la levée de dormance, comme pour la graine, supportant l'hypothèse de l'existence de mécanismes de régulation similaires contrôlant la levée de dormance d'un bourgeon et d'une graine (Rohde *et al.*, 2000). Nous avons aussi trouvé que l'expression d'un homologue de *MONOPTEROS* (MP) est induite durant le débourrement. MP est un facteur de réponse à l'auxine (Auxin Response Factor, ARF) qui semble agir en tant qu'activateur transcriptionnel, requis pour le contrôle de la formation de l'axe dans l'embryon et dans les phénomènes d'expansion cellulaire sous la dépendance de l'auxine (Hardtke et Berleth, 1998; Hardtke *et al.*, 2004). Un rôle similaire pourrait être attribué à cet homologue MP trouvé dans les bourgeons de chêne. Par analogie avec son équivalent chez *Arabidopsis*, ce gène pourrait, en effet, permettre le transfert des signaux d'auxine, essentiels au développement du bourgeon durant le développement. L'expression de régulateurs putatifs du cycle cellulaire est également affectée par le débourrement. En effet, *SKP1* (Figure 34.4) et *pollen specific SF21*

(Figure 34.5) sont fortement induits au stade 4. Ce résultat pourrait indiquer que les phénomènes de division et de différenciation cellulaire sont décalés dans le temps au cours du débourrement, la division cellulaire débutant durant le stade quiescent tandis que la différenciation apparaîtrait principalement au stade 4 en parallèle de la division encore très présente.

Du point de vue métabolique, l'activité de glycosyl hydrolases, telles que la *glucan endo-1,3- β -glucosidase* et la *cyanogenic beta glucosidase precursor*, est induite au stade quiescent et en début de débourrement. Ces enzymes sont toutes deux induites par l'acide gibbérellique et jouent un rôle à la fois dans la mobilisation de la paroi des cellules et dans l'élongation cellulaire (Hrmova et Fincher, 2001), par l'hydrolyse des liaisons glycosidiques liant les composants des parois. En contraste, une *alpha-amylase/subtilisin inhibitor* est très fortement exprimée au stade quiescent (Figure 34.6), suggérant que l'hydrolyse des sucres stockés ou du glycogène est réprimée dans le bourgeon quiescent. La répression de l'expression de ce gène dès le stade 1 indiquerait le départ de la mobilisation des réserves de sucres à ce stade. D'une manière générale, une activation des niveaux d'expression des gènes indispensables à la fourniture d'énergie est observée durant le débourrement, tels que des transcrits codants potentiellement pour des protéines *photosystem II reaction center M*, *cytochrome B6-F complex*, *plastocyanin*, *oxygen-evolving enhancer protein 2*, *ATP synthase ϵ chain*. L'expression de la Rubis-Co est également positivement régulée, comme celle de la *glyceraldehyde 3-phosphate dehydrogenase* (Figure 34.7). L'augmentation du niveau d'expression de la G3PDH au moment de la levée de dormance a déjà été décrite par Wang *et al.* (1991b).

2.3. Sélection de GC

2.3.1. Méthodes combinées

L'emploi de la méthode SSH combinée à l'utilisation de filtres bactériens à haute densité et des filtres à ADNc nous a permis de générer puis d'isoler un certain nombre de clones d'ADNc de gènes régulés durant le débourrement. Cette démarche stratifiée permet d'identifier rapidement des gènes différentiellement exprimés comme cela a été démontré par Yang *et al.* (1999) à l'aide de la SSH et de microarray d'ADNc, ou par Nakata et McConn (2002), en combinant SSH, SH et tri différentiel en une seule méthode. Dans le cas de notre expérimentation, l'utilisation successive de la SSH, des filtres bactériens puis des filtres ADNc présente l'énorme avantage de réduire le nombre de clones différentiellement exprimés et permettre l'étude détaillée des clones sélectionnés à l'issue des différentes étapes. La SSH a d'abord permis l'isolation de transcrits rares ou spécifiques de nos différents stades de débourrement. Le tri différentiel à l'aide des filtres bactériens a ensuite réduit le nombre de clones intéressants de moitié (de 1800 clones à 900 clones) par la sélection de ceux présentant une expression différentielle significative entre les stades Q, P et F. Enfin, l'étude d'expression par filtres ADNc a été conduite pour identifier les gènes fortement régulés au cours du débourrement. En adoptant cette procédure par étapes, nous avons combiné les avantages des approches de recherche de gènes globale et ciblée. Pour en revenir à la méthode SSH, il est précisé dans la publication de Diatchenko qu'elle permet de « générer des populations d'ADNc hautement enrichies en gènes différentiellement exprimés à la fois en forte et en faible abondance ». C'est ce que nous avons pu vérifier : en effet, 6 des 10 transcrits analysés en RT-PCR sont fortement différentiellement exprimés entre le stade quiescent (#0) et le stade renflement du bourgeon (#1), à savoir *LEA5*, *galactinol synthase*, *DAG2*, *alpha-amylase/subtilisin inhibitor (RASI)*, *At2g14910* et *At4g24120* (codant pour la protéine hypothétique *T19F6.8*). De plus, les transcrits de *RASI* et *LEA5* sont respectivement 6 et 1000 fois plus abondant que le gène de contrôle au stade quiescent, alors que *DAG2* est 55 fois moins abondant que le transcrit de la protéine ribosomale 60S utilisée comme contrôle.

Les expériences de quantification par RT-PCR ont validé les résultats obtenus par Northern Blot inverse. Cette technique est une méthode puissante pour la validation des résultats issus d'études d'expression sur filtres (Chtanova *et al.*, 2001; Jones *et al.*, 2002; Pfaffl *et al.*, 2003). Son utilisation est documentée dans les domaines de la physiologie animale, l'endocrinologie, l'immunologie, la virologie, la microbiologie (Hopkins, 2002;

Pfaffl *et al.*, 2003; Pinzani *et al.*, 2004; Schams *et al.*, 2003; Stram *et al.*, 2004) ainsi que chez les végétaux (Khan et Shih, 2004; Thomas *et al.*, 2004). La PCR quantitative est considérée comme une méthode de choix pour estimer les niveaux d'expression de transcrits de gènes et elle est préférée aux méthodes plus traditionnelles des filtres (Walker, 2002). Dans notre étude, bien que nous ayons observé quelques divergences, les différences de niveaux d'expression sont d'une amplitude comparable à celles observées par Northern inverse. Ces différences d'expression peuvent être expliquées en partie par la phase de synthèse des ADNc réalisée à l'aide de deux techniques différentes, et en particulier la phase d'amplification par PCR utilisée dans le cas des sondes radioactives avec les filtres à haute-densité. En effet, bien que les réactions PCR d'amplification du second brin aient été arrêtées au cours de la phase exponentielle, quelques transcrits ont pu être artificiellement favorisés ou défavorisés. De plus, dans le cas des filtres, certaines différences peuvent provenir d'hybridations croisées entre gènes d'une même famille multigénique au contraire de la PCR quantitative où les amorces sont choisies pour être spécifiques d'un gène.

2.3.2. Des candidats expressionnels et/ou fonctionnels ?

A ce stade de la partie expérimentale et avant de passer aux étapes ultérieures de cartographie génétique et d'études d'association, nous avons réalisé la sélection d'un certain nombre de gènes sur la base de critères d'expression et/ou de fonction potentielle. C'est le cas des transcrits codants pour la *galactinol synthase*, *DAG2*, l'*alpha-amylase/subtilisin inhibitor (RASI)*, la *S-phase kinase –associated protein 1A (SKP1)*, la *pollen specific protein SF21* et trois protéines hypothétiques *At2g14910*, *At2g24120* et *Zwh21.1*. Ces derniers ont été sélectionnés parmi un certain nombre de transcrits trouvés différemment exprimés entre les stades 0 et 1. Certains n'ont été choisis que sur la base de leur profil d'expression, tels ceux codants pour les protéines hypothétiques *At2g14910* et *At2g24120* ou la *galactinol synthase*, qui présentent une forte surexpression au stade quiescent et une expression nulle aux stades 1, 2 et 3. A ce titre, ils peuvent constituer des signaux précoces menant ou résultant de la reprise d'activité du bourgeon. D'autres transcrits présentent un profil d'expression intéressant mais aussi une fonction potentiellement liée au déclenchement du débourrement.

En effet, le rôle de *DAG2* en tant que régulateur de la germination de la graine chez *Arabidopsis* (Gualberti *et al.*, 2002) en fait un gène candidat intéressant à double titre : à la fois fonctionnel et expressionnel. La protéine inhibitrice *RASI* est également intéressante d'un

point de vue fonctionnel puisque l'enzyme alpha-amylase qu'elle inhibe est impliquée dans l'hydrolyse des réserves de sucres chez la graine des céréales qui accompagne le développement de l'embryon, et pourrait, dans le bourgeon, être impliquée précocement dans la mobilisation des réserves nécessaires à la reprise de croissance. La protéine hypothétique *Zwh21.1* contient un certain nombre de domaines F-Box, qui constituent des éléments nécessaires au déroulement du cycle cellulaire. Une fonction présomptive de ces gènes dans le déroulement du débourrement reste, bien entendu, entièrement à démontrer. Certains transcrits, enfin, constituent surtout des candidats fonctionnels car leur expression ne semble pas parfaitement corrélé au déroulement du débourrement. En effet, la protéine *SKP1* est une composante du complexe SCF (SKP1-Cullin-F-Box) indispensable au déroulement de la transition entre phases G1 et S du cycle cellulaire chez la levure, et potentiellement chez les plantes (Horvath *et al.*, 2003). Ce complexe semble, par ailleurs, réguler un certain nombre d'aspects du développement floral chez *Arabidopsis* (Ni *et al.*, 2004). Ces connaissances en font un gène candidat pertinent du point de vue fonctionnel. De la même manière, la protéine *pollen specific SF21* fait partie de la famille des protéines NRDG, dont la fonction moléculaire et cellulaire exacte reste à déterminer mais qui sont impliquées dans les phénomènes de différenciation cellulaire.

L'ensemble de ces gènes forme donc un premier groupe de gènes candidats issus des études d'expression. En parallèle, un second groupe de GC est constitué des gènes reportés dans le Tableau 2, intéressants à tester dans le cas du débourrement au regard de leur implication dans la dormance ou la germination des graines. Une partie de ces GC potentiels a donc fait l'objet de la seconde partie de cette approche expérimentale qui a consisté à les cartographier sur les cartes génétiques existantes et à analyser les colocalisations résultantes avec les QTL cartographiés durant des travaux de thèse précédents (Casasoli, 2004; Scotti-Saintagne, 2003).

③ Conclusions de l'approche transcriptomique

Cette partie expérimentale, consacrée à l'étude du transcriptome, a donc confirmé le grand intérêt d'une approche combinée banques SSH, filtres bactériens puis ADNc à haute densité et, enfin, RT-PCR quantitative en temps réel pour mettre en évidence les mécanismes moléculaires fortement régulés au cours du débourrement mais aussi pour sélectionner un certain nombre de gènes candidats potentiels au regard de leur différentiel d'expression entre les stades de débourrement, et en particulier entre l'état quiescent et le début de renflement du bourgeon. L'étude est originale puisque elle constitue une des premières à suivre le niveau d'expression de transcrits en conditions naturelles durant la totalité du processus de débourrement depuis l'état quiescent jusqu'au début de croissance végétative matérialisée par l'allongement des entre-nœuds. Les phénomènes de stress hydrique, oxydatif et thermique ont été mis en évidence dès le stade quiescent et dans les premiers stades du débourrement. Le processus de division cellulaire est activé lors de la reprise de croissance visible du bourgeon mais également dès la fin du stade quiescent juste avant le début de renflement du bourgeon. Dans le même temps, les niveaux de transcrits de facteurs de transcription, liés à la reprise de croissance des graines ou à l'expansion cellulaire sous le contrôle de l'auxine, augmentent dans les tissus du bourgeon. Par exemple, l'étude nous a révélé la sur expression du facteur de transcription DAG2 dans le bourgeon quiescent. Ce résultat est original et intéressant puisqu'il supporte l'hypothèse de mécanismes similaires de levée de dormance d'un bourgeon et d'une graine, puisque DAG2 est impliqué dans le contrôle maternel de la germination de la graine chez *Arabidopsis thaliana*.

La construction de banques SSH d'EST de chêne a, par ailleurs, permis l'isolation de nombreux transcrits sans équivalent dans les bases de données protéiques ou d'EST. C'est un résultat prometteur pour l'identification et la caractérisation de gènes impliqués dans la levée de dormance et le débourrement du bourgeon, puisqu'aucun gène de fonction connue n'a encore permis de disséquer le fonctionnement de ces deux processus. La caractérisation fonctionnelle de ces transcrits codant pour des protéines inconnues pourrait permettre une meilleure compréhension de ces deux phénomènes développementaux majeurs dans la vie de l'arbre. Mais dans le cadre de cette thèse, l'objectif principal est de localiser les régions génomiques (gènes, QTL) impliquées dans la variabilité phénotypique du caractère de débourrement. Pour ce faire, les gènes candidats définis précédemment vont donc, dans un

premier temps, être cartographiés afin d'étudier leur localisation par rapport à celle des QTL impliqués dans le débourrement chez deux espèces de Fagacées, le chêne et le châtaignier. Ce travail devrait permettre d'affiner la localisation des régions génomiques contrôlant le caractère de débourrement. En parallèle, l'implication des gènes candidats dans la variabilité phénotypique observée du caractère sera analysée par le biais d'études d'association, après avoir décrit la structuration de la diversité nucléotidique au sein de ces gènes.

Dans le cadre de cette thèse, l'utilisation de la cartographie comparée doit nous permettre de préciser si les régions génomiques (QTL, GC) potentiellement impliquées dans la variation de la date du débourrement sont conservées entre espèces différentes au sein de la famille des Fagacées. En ce sens, un projet de cartographie génétique comparée avait débuté en 2000 dans le cadre du projet européen CASCADE afin d'aligner les cartes génétiques du chêne et du châtaignier européen. L'utilisation des EST issues de notre travail et des GC identifiés lors de l'approche transcriptomique est venue compléter le travail accompli précédemment.

Quatre objectifs principaux étaient poursuivis dans cette partie expérimentale. Le premier objectif était de pouvoir relier entièrement les cartes génétiques du chêne et du châtaignier en alignant l'ensemble des douze groupes de liaison existants chez les deux espèces, à l'aide de marqueurs orthologues dérivés des EST. Le projet CASCADE avait permis, dans un premier temps, d'associer 7 des douze groupes de liaison chez les deux espèces (Barreneche *et al.*, 2004) à l'aide de marqueurs microsatellites qui avaient précédemment été trouvés transférables d'une espèce à l'autre (Steinkellner *et al.*, 1997). Le second objectif était de comparer la position des QTL impliqués dans le débourrement afin de les valider chez les deux espèces. Le troisième objectif était de cartographier les GC identifiés. Enfin, le quatrième objectif était d'analyser les colocalisations entre les gènes candidats et les QTL pour le débourrement dans le but d'évaluer la pertinence du choix des GC précédemment sélectionnés. Les deux premiers objectifs, ainsi que les deux suivants dans une moindre mesure, étaient également inclus dans le travail de thèse de Manuela Casasoli (2004), qui a réalisé la majeure partie du travail expérimental de ce chapitre, avec l'aide ponctuelle de Caroline Morera-Dutrey, dans le cadre d'un stage de Licence en Sciences de la vie et de la santé.

① Matériel et Méthodes

La construction des cartes génétiques originales sur chêne (*Quercus robur*) (Barreneche *et al.*, 1998) et châtaignier (*Castanea sativa*) (Casasoli *et al.*, 2001), la cartographie comparée réalisée à l'aide de marqueurs microsatellites (Barreneche *et al.*, 2004), ainsi que les études de recherche de QTL menées sur les deux espèces (Casasoli *et al.*, 2004; Scotti-Saintagne *et al.*, 2004) ne seront pas présentées dans cette partie. Nous nous focaliserons sur la cartographie comparée réalisée à l'aide d'EST, la comparaison des QTL entre chêne et châtaignier, ainsi que la colocalisation entre GC et QTL.

1.1. Amplification et cartographie des marqueurs EST

1.1.1. L'amplification des marqueurs dérivés des EST

Au total, une centaine d'EST provenant des banques SSH d'ADNc développées au cours de l'analyse transcriptomique ont été utilisées dans cette partie du travail expérimental. Les huit GC, précédemment identifiés, ont été inclus dans l'analyse ainsi que deux des gènes candidats fonctionnels issus des connaissances disponibles sur la dormance et la germination des graines (GA 3- β -hydroxylase et GA 20-oxydase). L'objectif prioritaire étant d'aligner les douze groupes de liaison du chêne et ceux du châtaignier, le choix des amorces d'amplification a été réalisé afin de maximiser la probabilité de transfert entre les deux espèces. Dans un premier temps, une centaine de séquences, potentiellement homologues de protéines de fonctions connues, ont été utilisées pour rechercher des séquences homologues dans les bases de données, à l'aide de la procédure BLAST (Altschul *et al.*, 1997). L'ensemble des séquences homologues pour chaque EST (deux ou trois par EST en général) a été aligné avec l'EST de chêne correspondante (CLUSTALW, <http://www.ebi.ac.uk/clustalw/>) afin d'identifier les régions les plus conservées entre espèces. Des paires d'amorces ont ensuite été dessinées à l'aide du logiciel Primer3, disponible à l'URL : http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi. Trois critères principaux ont été appliqués pour le choix des paires d'amorces : i/ les amorces étaient sélectionnées dans une région du fragment d'ADN conservée entre espèces, ii/ la taille attendue du fragment était fixée entre 100 et 450 pb, iii/ la température d'hybridation était comprise entre 55°C et 60°C. Le critère de taille était dicté par deux contraintes. La première est liée aux techniques de

Tableau 9 : Liste des marqueurs STS (Site Tagged Sequences) dérivés des GC : séquences des amorces, résultats d'amplification et de cartographie.

<i>Nom de l'EST</i>	<i>Numéro d'accèsion</i>	<i>Amorce Forward (5'→3')</i>	<i>Amorce Reverse (5'→3')</i>	<i>Taille attendue (pb)</i>	<i>Taille observée (pb)</i>	<i>Groupe de liaison Q - C</i>
01B07 S-phase kinase-associated protein 1A	CR627507	CGATGAGGAGCTCAAGAACT	CCATTGGTTCTCCCTGCGAA	235	1400	nd - nd
06G05 Unknown (Zwh21.1 hypothetical protein)	CR627759	ACAACGGGGCAAGTCCATTAGA	CTATATAATCTGAGGCAGCGG	279	279	nd - nd
08A01 Galactinol synthase	CR627918	TCAAATTCAATGTTCTCTGGATGG	GAGGTACTIONTCAATGCTGGCTTCTT	210	500	3 - nd
08C11 Unknown (T19F6.8 hypothetical protein)	CR627947	TGAAGCAGGAAGCGTCCAAAGT	GCCATTCTAGGTGTTGAAGG	316	316	2 - 1
08G04 Unknown (At2g14910 hypothetical protein)	CR627986	CAGCATTACTTGGCAGGGATAGGA	CAACCGGAAAAGGGATGAAGTGT	393	1000	11 - nd
07B10 DOF zinc finger protein DAG2	CR627781	CTACAGCCTCACTCAACCAAGATA	TCCAAAGCTGAAAGATGAGTAGAG	360	360	8 - nd
09C04 Pollen specific protein SF21	CR627822	ACACATGCCATAGAAGTAGAGT	AGCCAGCTCTTGTCACCTTAT	410	2000	nd - nd
08B04 Alpha-amylase/subtilisin inhibitor (RASI)	CR627933	TTCGTGAGCCTTCGTCTTCTC	ACAGTGCCGGACGTCCTCTTCAA	327	327	9 - nd
GA 3-β-hydroxylase		TCCCTAAGCGTATGTGGTCC	CCCTCTCGTGTTGTTTTGGT	347	600	5 - nd
GA 20-oxylase		TCAGCTCAACCAGACTCATC	TTGAAAGTGCCATGAAGGTG	409	500	nd - nd

révélation du polymorphisme employées (SSCP et DGGE) qui sont à leur optimum d'efficacité pour des fragments compris entre 200 et 600 pb. La seconde est liée à la nature même des EST qui ne contiennent pas les éventuels introns présents au niveau de l'ADN génomique. Il est par conséquent plus prudent d'abaisser la limite supérieure de taille afin de tenir compte de la présence potentielle d'introns dans les fragments amplifiés. En suivant ces critères, 73 paires d'amorces ont été dessinées. Les séquences des amorces ainsi que les tailles attendues des fragments amplifiés et cartographiés avec succès sont présentées dans le Tableau 1 de l'Annexe II. La liste des marqueurs dérivés des GC, préalablement sélectionnés, est présentée dans le Tableau 9. Les amorces avaient été précédemment utilisées pour l'analyse des niveaux d'accumulation de transcrits en RT-PCR quantitative. A partir de cette étape, les marqueurs obtenus par amplification d'un fragment de gène grâce à des amorces établies sur une EST seront nommés STS (Site Tagged Sequences).

En dehors des gènes candidats pour lesquels les conditions de PCR avaient déjà été mises au point (Tableau 10), un protocole de PCR « touchdown » (Don *et al.*, 1991) a été employé pour les 73 nouvelles paires d'amorces. Le principe de la méthode est que la température d'hybridation n'est pas constante mais diminue au fur et à mesure des cycles d'amplification. Cette méthode permet d'être très spécifique durant les premiers cycles et ainsi d'éviter les mésappariements. Dans une deuxième phase, la température reste constante. Cela mène, en général, à une amplification plus efficace. Les réactions ont été faites dans un volume réactionnel total de 12,5 μ L, contenant 1,25 μ L de Tampon 10X PCR (Invitrogen), 2 mM de $MgCl_2$, 200 μ M de chaque dNTP, 0,2 μ M de chaque amorce, 100 μ g/ml de BSA, 5 à 10 ng d'ADN génomique et 0,4 U de *Taq* polymérase (Invitrogen). Les amplifications ont été conduites dans une machine GeneAmp PCR System 9600 (Perkin Elmer) avec les conditions suivantes : 10mn à 94°C, 10 cycles touchdown de 45 s à 94°C, 45 s à la température d'hybridation décroissant de 0,5°C à chaque cycle, 1mn à 72°C, puis 25 ou 30 cycles de 45 s à 94°C, 45 s à la température d'hybridation, 1mn à 72°C, et une phase d'extension finale de 10mn à 72°C. Les conditions PCR ont donc été optimisées pour obtenir une bande amplifiée unique. Les températures d'hybridation ainsi que le nombre de cycles utilisés pour chaque STS sont présentés dans le Tableau 1 de l'Annexe II. La qualité et la taille des produits d'amplification ont été vérifiés sur gel d'agarose (1,5% ; TBE 0,5X ; 0,1% BET).

Tableau 10 : Protocole d'amplification PCR des GC.

<i>Nom de l'EST</i>	<i>Protocole d'amplification*([MgCl₂])</i>
08A01 Galactinol synthase	P 94°-10' // 94°-45'' / 53°-45'' / 72°-1' (35 cycles) // 72°-10' (1,5 mM)
07B10 DAG2	P 94°-10' // 94°-45'' / 53°-45'' / 72°-1' (35 cycles) // 72°-10' (1,5 mM)
GA 3-β-hydroxylase	M 94°-10' // 94°-1' / 54°-1' / 72°-1' (35 cycles) // 72°-10' (2,5 mM)
GA 20-oxydase	M 94°-10' // 94°-1' / 50°-1' / 72°-1' (40 cycles) // 72°-10' (2 mM)
08B04 RASI	P 94°-10' // 94°-45'' / 56°-45'' / 72°-1' (35 cycles) // 72°-10' (2 mM)
08G04 At2g14910	M 94°-10' // 94°-1' / 54°-1' / 72°-1' (35 cycles) // 72°-10' (3 mM)
01B07 SKP1	P 94°-10' // 94°-45'' / 56°-45'' / 72°-1'15'' (35 cycles) // 72°-10' (1,5 mM)
09C04 Pollen specific protein SF21	P 94°-10' // 94°-45'' / 53°-45'' / 72°-1'30'' (35 cycles) // 72°-10' (1,5 mM)
06G05 Zwh21.1	P 94°-10' // 94°-45'' / 53°-45'' / 72°-1' (35 cycles) // 72°-10' (1,5 mM)
08C11 T19F6.8	P 94°-10' // 94°-45'' / 54°-45'' / 72°-1' (35 cycles) // 72°-10' (2 mM)

* **P** : machine GeneAmp PCR System 2700 ; **M** : machine RoboCycler Gradient 96

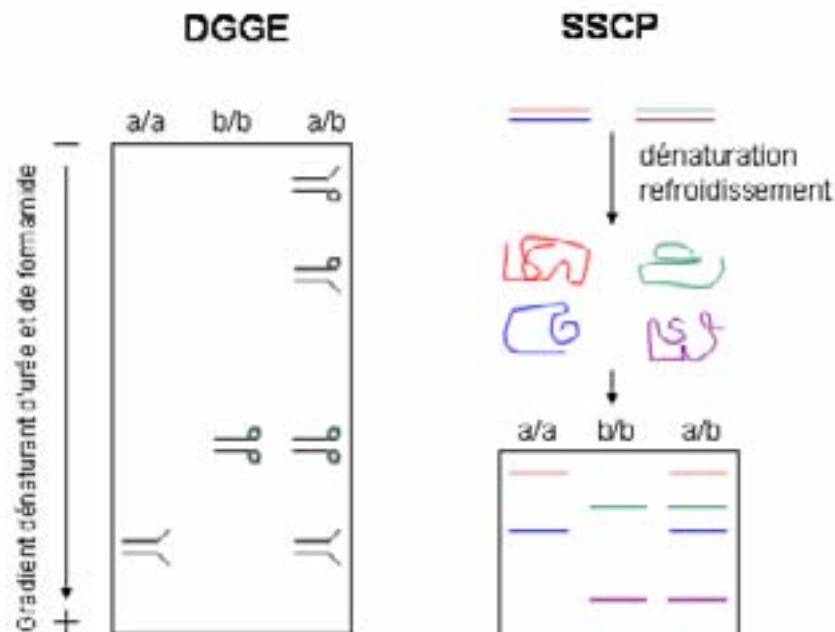


Figure 35: Principes des méthodes DGGE et SSCP. Sur le schéma de la DGGE, le point bleu sur l'ADN indique la position d'une substitution entre les fragments a et b.

1.1.2. La cartographie des STS

Les techniques DGGE (Denaturing Gradient Gel Electrophoresis, (Myers *et al.*, 1987)) et SSCP (Single Strand Conformation Polymorphism, (Orita *et al.*, 1989)) ont été utilisées pour détecter du polymorphisme au sein des marqueurs STS amplifiés (Figure 35). Les procédures expérimentales suivies sont celles décrites par Temesgen (2001) pour la méthode DGGE et Plomion (1999) pour la méthode SSCP. Dans un premier temps, les deux parents ainsi que 6 individus F₁ de chaque pedigree ont été chargés sur gel SSCP. Si un polymorphisme était révélé, la technique SSCP était appliquée pour cartographier le fragment. Sinon, les parents et les mêmes 6 individus étaient analysés par DGGE. Dans le cas où aucun polymorphisme n'était révélé par l'une ou l'autre des deux techniques, le fragment STS n'a pas été cartographié. La technique SSCP a été testée en premier lieu car elle est plus efficace à révéler des mutations et également plus facile à optimiser que la technique DGGE.

Les populations de cartographie analysées sont celles utilisées pour la construction des cartes de *Q. robur* (Barreneche *et al.*, 1998) et *C. sativa* (Casasoli *et al.*, 2001). Cinquante et un individus supplémentaires ont été génotypés pour les STS polymorphes donc 57 individus au total ont été génotypés. Cet effectif résulte d'un compromis entre la puissance de détection d'une liaison génétique et une contrainte technique. En effet, 50 individus sont suffisants pour détecter une liaison statistique à moins de 20 cM dans 90% des cas et dans 100% des cas à moins de 10 cM (Liu, 1998). Les deux cartes génétiques utilisées étant saturées et les groupes de liaison bien définis, tout nouveau marqueur sera quasiment toujours lié à un autre marqueur de la carte à moins de 20 cM de distance. Techniquement, deux gels de 28 puits (56 échantillons) sont disponibles dans une chambre d'électrophorèse ce qui convient à la migration de 51 individus, les 2 parents et 3 marqueurs de migration. Les logiciels MapMaker v.2 (Lander *et al.*, 1987) et JoinMap v.3 (Van Oijen et Voorrips, 2001) ont été utilisés pour l'analyse de liaison en considérant un seuil de LOD de liaison de 6.0 et la fonction de Kosambi (1944) pour le calcul des distances génétiques. Pour six STS, le seuil a été abaissé à 4.0 mais le résultat est confirmé par la liaison avec plusieurs marqueurs du groupe et la présence du même STS sur le groupe homologue chez l'autre espèce.

Pour finir, les marqueurs STS ont été séquencés chez le chêne et chez le châtaignier pour valider l'amplification du gène correspondant à l'EST de départ et l'homologie de séquence entre les deux espèces. Le produit amplifié de chaque espèce a été séquencé à l'aide du kit de séquençage DYEnamic™ ET Dye Terminator (Amersham Biosciences Inc.) sur

séquenceur capillaire automatique MegaBACE 1000 (Amersham Biosciences Inc.). Les séquences ont été alignées à l'aide du logiciel ALIGNn, disponible à l'URL : http://www.infobiogen.fr/services/analyseq/cgi-bin/alignn_in.pl afin d'estimer le pourcentage d'identité entre les deux séquences.

1.2. Cartes de liaison consensus et cartographie comparée des QTL

L'objectif final de cette partie est de pouvoir comparer la position des QTL impliqués dans le débourrement entre le chêne et le châtaignier. Or, les deux cartes génétiques utilisées pour la détection de QTL ne contiennent pas beaucoup de marqueurs communs entre le chêne et le châtaignier. Par ailleurs, la stratégie utilisée pour la construction des cartes (double pseudo test-cross) implique que la détection des QTL est faite séparément sur les deux cartes parentales mâle et femelle. Cependant, comme la carte utilisée pour la détection de QTL dérive de la carte contenant l'ensemble des marqueurs orthologues, il est possible d'intégrer toutes les informations (marqueurs orthologues et position des QTL) dans une seule carte chez chacune des deux espèces. Les QTL pourront alors être comparés entre espèces grâce aux marqueurs orthologues. Trois étapes sont nécessaires pour aboutir à cette comparaison : i/ construire une carte consensus de chaque espèce, ii/ projeter les QTL sur cette carte consensus grâce aux marqueurs communs, iii/ comparer la position des QTL sur les cartes des deux espèces grâce à des intervalles communs identifiés par les marqueurs orthologues.

*1.2.1. Construction des cartes de liaison consensus chez *Q. robur* et *C. sativa**

Dans un premier temps, pour chacune des deux espèces, une carte consensus a été construite afin de regrouper les marqueurs et les liaisons entre ces marqueurs, issus des différentes études de cartographie, dans une seule et même carte pour chaque espèce. Dans ces deux nouvelles cartes consensus, ont été regroupés certains des marqueurs contenus dans chacune des cartes parentales utilisées pour la détection de QTL (Casasoli *et al.*, 2004; Scotti-Saintagne *et al.*, 2004), tous les microsatellites cartographiés précédemment chez les deux espèces (Barreneche *et al.*, 2004) ainsi que la totalité des STS cartographiés dans cette étude. Pour chaque espèce, deux étapes ont été nécessaires afin de construire une carte génétique consensus unique, à l'aide des logiciels MapMaker et JoinMap:

- MapMaker a d'abord été utilisé pour construire une carte jalon de chaque parent dans chacune des deux espèces. Quarante quatre individus ont été utilisés chez le chêne et 96 chez le châtaignier, dont 57 avaient été génotypés avec les STS. L'ordre des marqueurs a été optimisé grâce à l'option « ripple », qui permet d'obtenir l'ordre le plus probable des marqueurs. Les marqueurs positionnés avec un LOD de liaison ≤ 2.0 ont été retirés.

L'ordre des marqueurs dans ces deux cartes jalons coïncide avec celui de la carte QTL, à quelques exceptions près pour des marqueurs étroitement liés ;

- le logiciel JoinMap a ensuite permis d'intégrer les données de ségrégation des deux cartes parentales dans une carte consensus. La liaison est rendue possible par l'utilisation des marqueurs informatifs chez les deux parents et ségrégant 3:1, 1:2:1 ou 1:1:1:1. Comme la densité de marqueurs est augmentée par rapport aux cartes parentales et que les couples de marqueurs 1:1/3:1 et 1:1/1:2:1 sont peu informatifs pour estimer le taux de recombinaison (Ritter *et al.*, 1990), l'optimisation de l'ordre des marqueurs est rendue plus délicate et plusieurs combinaisons peuvent apparaître avec la même probabilité d'occurrence. C'est pourquoi l'option « fixed order » proposée par le logiciel JoinMap a été utilisée lorsque l'introduction des marqueurs 3:1 et 1:2:1 entraînait de fortes incohérences (marqueurs inversés à une distance supérieure à 5 cM) par rapport à l'ordre optimisé de MapMaker et à l'ordre sur la carte de détection des QTL. L'ordre de certains marqueurs a alors été fixé pour correspondre à ces derniers puis les autres marqueurs ont été intégrés. L'option « fixed order » n'a, en revanche, pas été utilisée si l'ordre proposé par JoinMap correspondait à celui donné par MapMaker pour la construction de la carte jalon de détection des QTL. Par ailleurs, tous les marqueurs orthologues ont été inclus dans les deux cartes consensus.

Grâce à cette stratégie, la carte consensus contient un certain nombre de marqueurs en commun avec la carte jalon de détection des QTL, qui vont permettre d'inclure les informations relatives aux QTL, mais aussi la totalité des marqueurs orthologues entre les deux espèces ; deux prérequis indispensables à la comparaison des positions des QTL contrôlant le débourrement entre le chêne et le châtaignier.

1.2.2. Projection et cartographie comparée des QTL

Grâce aux marqueurs communs aux cartes consensus et de détection des QTL, les QTL ont pu être positionnés sur la carte consensus de chaque espèce. A cette fin, le logiciel BioMercator (Arcade *et al.*, 2004) a été utilisé. Il permet, en effet, par un processus de projection homothétique⁶, de synthétiser des données de marqueurs et de QTL issus de différentes expériences dans une seule carte génétique. Une carte consensus contenant tous les

⁶ Projection homothétique : projection qui conserve les rapports des distances entre les QTL ou les marqueurs projetés (cf. homothétie).

gènes, les marqueurs génétiques et la position des QTL a donc été construite. Concernant les QTL, la position, mais aussi l'intervalle de confiance sur cette position ont été projetés. De cette manière, les 27 et 21 QTL détectés année par année pour le débourrement, respectivement sur le chêne et le châtaignier, ont été projetés sur la carte consensus de chaque espèce. Afin de pouvoir comparer les positions respectives des QTL entre *Q. robur* et *C. sativa*, des intervalles communs aux deux cartes ont été identifiés grâce aux marqueurs orthologues. Un intervalle commun est défini, sur un groupe de liaison, comme la région qui partage les mêmes marqueurs orthologues à ses extrémités. Ces intervalles sont comparables s'ils représentent au minimum 5 cM de longueur. En dessous de cette distance, l'ordre des marqueurs est peu fiable pour des marqueurs génotypés sur 57 individus. Lorsque plus d'un marqueur était cartographié à l'extrémité d'un intervalle commun, le marqueur le plus informatif ou génotypé sur le plus grand nombre d'individus a été choisi comme marqueur limite de l'intervalle. A titre d'exemple, le marqueur EMCs11 génotypé sur 96 individus a été préféré à l'EST Cons126 génotypée sur 57 individus F₁ sur le groupe C5. Sur les groupes de liaison Q12-C12, l'EST 06E10, ségrégant 1 :1 a été préférée à l'EST Cons105, ségrégant 1 :3. Une fois les intervalles identifiés entre les cartes consensus du chêne et du châtaignier, des QTL uniques ont été définis pour chaque espèce. En effet, plusieurs QTL peuvent être positionnés dans la même région génomique, comme des QTL détectés sur deux années différentes ou des QTL ayant la même position sur la carte femelle et la carte mâle. Comme il peut s'agir du même QTL détecté plusieurs fois, nous avons considéré qu'il s'agissait d'un QTL unique lorsque la position la plus probable d'un ensemble de QTL se trouve dans le même intervalle identifié par les marqueurs orthologues,. Si la position la plus probable est contenue entre deux intervalles, le QTL est alors positionné dans l'intervalle contenant les autres QTL pour le caractère ou la plus grande longueur de l'intervalle de confiance. Cette définition du QTL unique est évidemment une définition pratique et subjective qui permet la comparaison de la position des QTL entre les deux espèces. Cette stratégie nous permet de conclure que lorsque la position la plus probable est contenue dans le même intervalle, les QTL sont colocalisés entre *Q. robur* et *C. sativa*. Enfin, la probabilité que les colocalisations apparaissent au hasard a été calculée selon la formule donnée par Lin *et al.* (1995), avec de légères modifications décrites par Feltus (2003). Ainsi, sous l'hypothèse de distribution aléatoire des QTL, la probabilité que la coïncidence entre les QTL dans une région donnée du génome soit due au hasard est égale à :

$$p = \frac{\binom{1}{m} \binom{n-1}{s-m}}{\binom{n}{s}}$$

où n = nombre d'intervalles comparés ; l = nombre de QTL détectés dans l'expérience avec le plus grand nombre de QTL ; s = nombre de QTL détectés dans la seconde expérience ; m = nombre de QTL « colocalisés » entre les deux expériences. Les expériences correspondent, dans notre cas, aux deux espèces.

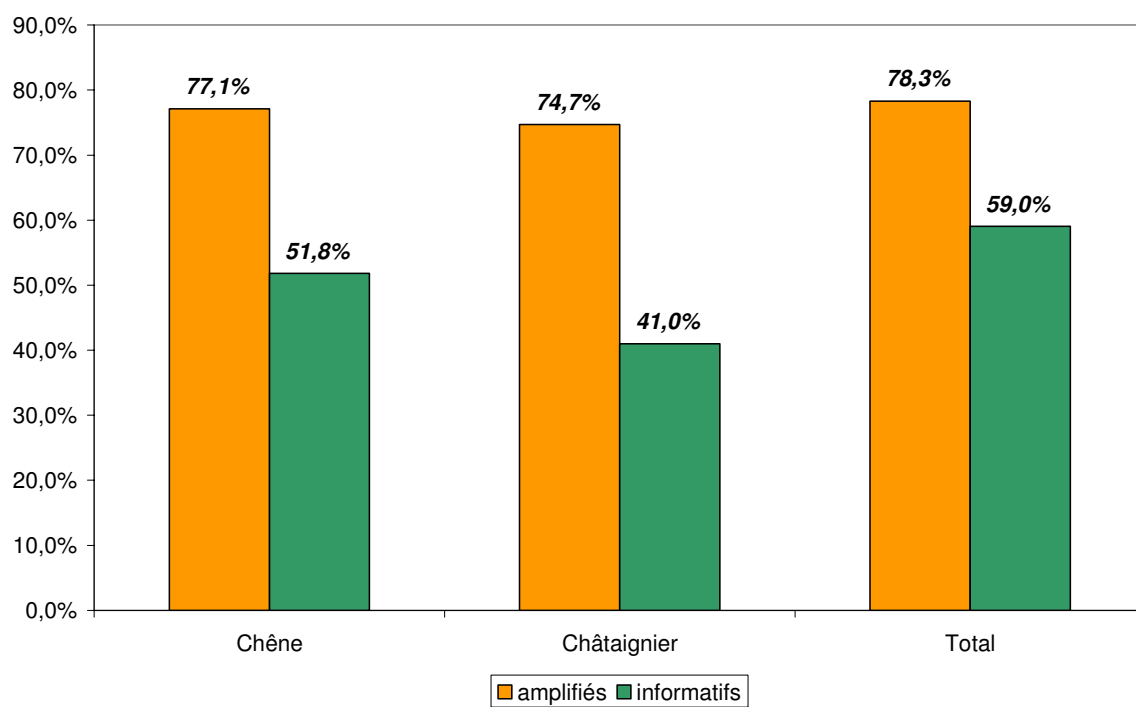


Figure 36 : Pourcentage de STS amplifiés et informatifs chez le chêne et le châtaignier par rapport aux 82 paires d'amorces initialement dessinées.

② Résultats et Discussion

2.1. L'alignement des douze groupes de liaison grâce aux STS

2.1.1. L'amplification et la cartographie des STS

Les résultats de l'amplification des STS chez les deux espèces sont très similaires. En effet, parmi les 83 paires d'amorce dessinées, respectivement 64/83 (77%) et 62/83 (75%) ont amplifié une bande unique et bien visible chez le chêne et le châtaignier (Figure 36). Seuls trois STS amplifiés chez le chêne n'ont pas été amplifiés correctement chez le châtaignier et un seul STS en sens inverse, c'est-à-dire amplifié correctement chez le châtaignier mais pas chez le chêne. Les amorces écartées de l'étude pour la cartographie sont celles dont le résultat est un produit d'amplification présentant plusieurs bandes, même à température d'hybridation élevée (65°C), ou celles qui ne fournissent aucune amplification. Seuls les STS amplifiés et cartographiés avec succès ont été reportés dans le tableau 1 de l'Annexe II. Sur la base de la séquence EST, les tailles attendues des fragments amplifiés varient entre 115 pb et 454 pb, alors que les tailles obtenues varient entre 120 pb et 1600 pb (Tableau 9 et Table 1 Annexe II). Soit la taille correspond à celle attendue pour les deux espèces, soit elle correspond à celle attendue chez une seule des deux espèces, soit la taille ne correspond pas à la taille attendue chez les deux espèces. Parmi les 65 STS correctement amplifiés chez l'une ou l'autre des deux espèces, 42% (27/65) résultent en un fragment de taille équivalente à l'EST de départ chez les deux espèces tandis que 37% (24/65) présentent une bande de taille supérieure à celle attendue chez les deux espèces. Ces produits d'amplification, de longueur supérieure à celle prédite sur la base de la séquence EST, contiennent vraisemblablement des introns. Enfin, 14 STS (21%) amplifient un fragment de taille différente entre le chêne et le châtaignier. A ce stade, on peut déjà observer que la structure des gènes est bien conservée entre les deux espèces puisqu'il y a concordance, dans la plupart des cas (79%), pour la présence éventuelle d'introns dans le fragment amplifié.

La technique SSCP a été privilégiée pour la cartographie des fragments STS amplifiés du fait de sa plus grande efficacité à révéler des mutations mais aussi sa plus grande simplicité d'optimisation des conditions de migration. Seuls six fragments ont été cartographiés par la technique DGGE et 15 fragments correctement amplifiés n'ont pas pu être cartographiés par absence de polymorphisme ou parce que les conditions de migration en SSCP ou DGGE n'ont pas pu être optimisées. Au final donc, 50 fragments issus du séquençage des banques SSH,

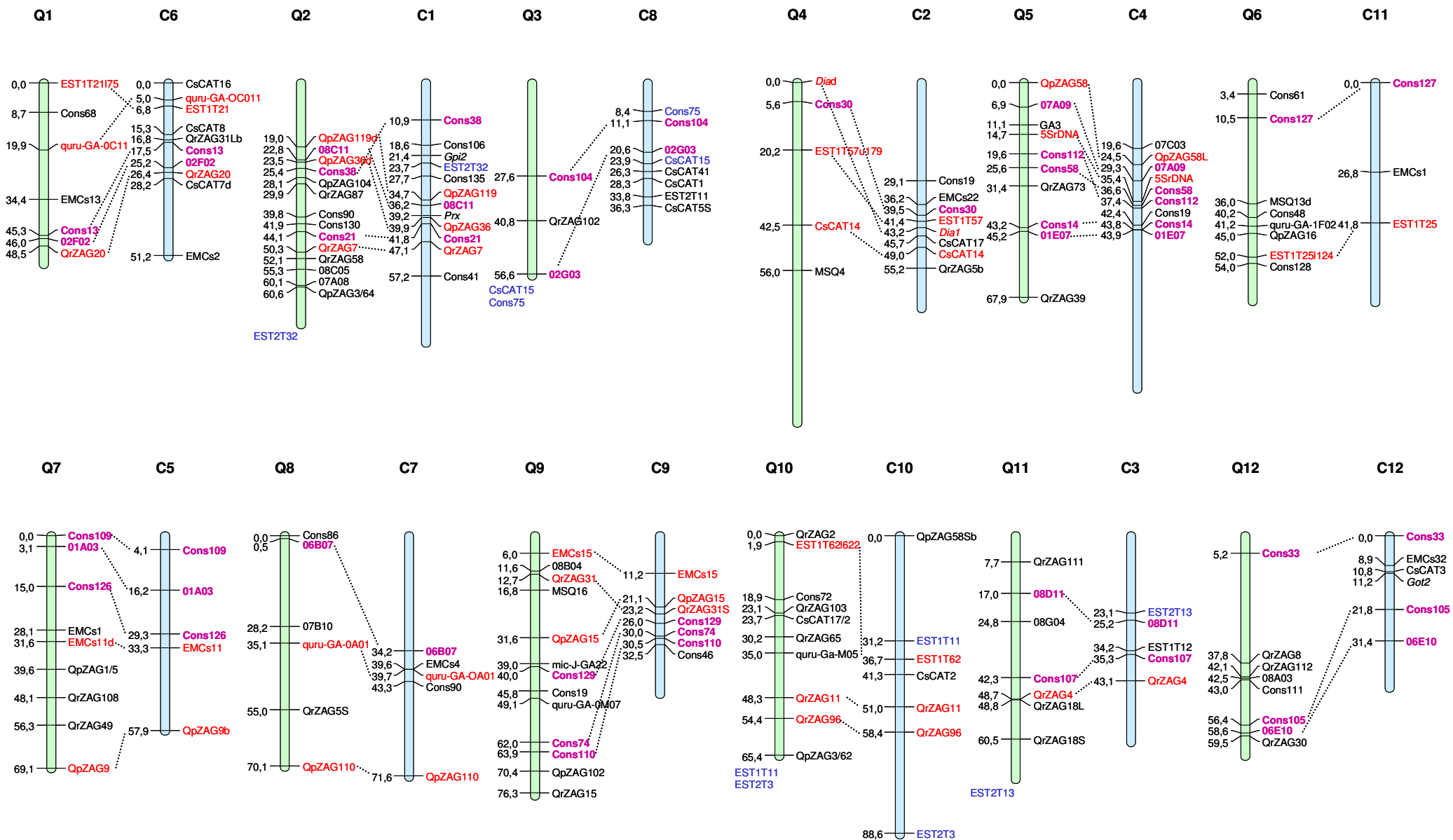


Figure 37: Les douze couples de groupes homologues entre le châtaigne (vert) et le châtaignier (bleu). En rouge et violet apparaissent les marqueurs orthologues utilisés dans ce travail; en violet les STS dérivés des EST de châtaigne. Un sous échantillon de marqueurs de la carte consensus de chaque espèce est représenté ici. En bleu, en dessous des groupes, les marqueurs orthologues cartographiés sur le pedigree interspécifique du châtaigne.

dont 6 sélectionnés en tant que gènes candidats pour le débourrement, ont été cartographiés chez l'une ou l'autre des deux espèces. Dans le détail, 43 des 64 STS (67%) amplifiés chez le chêne sont informatifs dans la famille de pleins frères de chêne utilisée pour la cartographie, et 34 des 62 fragments (55%) amplifiés chez le châtaignier sont informatifs dans la famille de pleins frères de châtaignier. Les deux techniques de cartographie employées dans cette étude se montrent donc parfaitement efficaces à révéler des mutations de séquence. L'utilisation de marqueurs STS n'est certainement pas étrangère à ce résultat puisque l'amplification de séquences d'introns ou d'extrémités 3'-UTR, comme c'est le cas de 58% des STS chez au moins une des deux espèces, augmente la probabilité de détecter des polymorphismes. Les exons sont, en effet, moins variables que les deux types de séquences précédemment cités. Par ailleurs, six de ces marqueurs STS ont été cartographiés avec des valeurs de LOD comprises entre 4.0 et 6.0 (Table 1 Annexe II) mais la localisation est confirmée par la liaison avec plus d'un marqueur au sein du même groupe de liaison. Pour tous les autres marqueurs, la valeur de LOD était supérieure à 6.0, confirmant par là même qu'un effectif de 57 individus est suffisant pour pouvoir cartographier de nouveaux marqueurs sur les deux cartes génétiques saturées de *Q. robur* et *C. sativa*.

2.1.2. L'alignement des cartes génétiques du chêne et du châtaignier

La figure 37 représente les douze couples de groupes de liaison homologues entre le chêne et le châtaignier. Au total, 55 marqueurs orthologues ont permis de relier les deux cartes dont 18 microsatellites, précédemment utilisés pour la comparaison des cartes génétiques de *Q. robur* et *C. sativa* (Barreneche *et al.*, 2004), 1 locus isozyme, 1 locus 5SrdNA et 35 STS, dont 27 développés grâce aux EST de chêne issues des banques SSH de notre étude. Les STS dérivés des EST de chêne représentent donc la moitié des marqueurs orthologues et sont parfois les seuls au sein d'un couple de groupes homologues (Q3-C8 ; Q12-C12). Dans la plupart des cas, l'analyse des séquences des fragments amplifiés chez les deux espèces a montré que le produit d'amplification correspondait à la séquence de l'EST et que l'homologie était respectée entre le chêne et le châtaignier (Table 1 Annexe II). Il y a deux exceptions à cela : les STS 01E07 et Cons129 pour lesquelles moins de 100 paires de base s'alignent avec l'EST de départ. On peut penser que le reste de la séquence correspond à un intron ce qui expliquerait ce résultat. Pour cinq fragments, la mauvaise qualité des séquences n'a pas permis de conclure sur l'homologie. Par ailleurs, 4 STS et 1 locus

microsatellite ont été cartographiés chez une famille de chêne interspécifique qui n'est pas celle utilisée dans ce travail. Mais cette carte a été reliée en partie à la carte du croisement 3PxA4 utilisée dans notre étude (Jauffrès, 2003) et les quatre marqueurs ont ainsi pu être positionnés indirectement sur la carte 3PxA4. De plus, deux locus microsatellites supplémentaires ont été cartographiés chez le chêne dans cette version de la carte. Il s'agit des locus CsCAT15 (Jauffrès, 2003) et EMCs11 (Barreneche, communication personnelle). Une différence importante avec la première étude comparative (Barreneche *et al.*, 2004) a été observée dans notre étude. En effet, le groupe de liaison Q7 avait été initialement associé au groupe C11 grâce au microsatellite EMCs1. Grâce aux nouveaux marqueurs orthologues cartographiés dans notre étude, les groupes Q7 et C11 ne semblent pas être homologues. Les nouveaux couples de groupes homologues sont Q6-C11 et Q7-C5 (Figure 37).

Le microsatellite EMCs1 amplifiait plusieurs fragments de différente taille chez le châtaignier (Barreneche *et al.*, 2004). Dans le croisement interspécifique, ce marqueur a été cartographié sur un groupe de liaison différent du croisement intraspécifique 3PxA4 (Jauffrès, 2003). Il semblerait ici que le locus cartographié chez le chêne et le châtaignier soit un locus paralogue, comme cela a déjà été montré dans des cas similaires d'amplification de fragments de différentes tailles (Barreneche *et al.*, 2004). Par ailleurs, deux STS ne semblent pas positionnés sur des groupes homologues entre les deux espèces : les STS Cons90 et Cons19. Dans les deux cas, les annotations fonctionnelles correspondent à des familles multigéniques. En effet, le STS Cons19 correspond au gène codant pour une protéine ribosomale L13 pour laquelle 8 locus ont été cartographiés sur 3 chromosomes différents chez *Arabidopsis* (Barakat *et al.*, 2001), où les gènes codant pour les protéines ribosomales semblent organisés en familles avec plusieurs copies distribuées sur les cinq chromosomes. Chez le chêne et le châtaignier, une seule bande a été amplifiée mais deux locus indépendants ont été cartographiés chez le châtaignier (groupes 2 et 4) et un troisième locus, indépendant des deux premiers, a été identifié chez le chêne sur le groupe 9. Nous pourrions être en présence de plusieurs membres de la même famille multigénique. Le STS Cons90 présente deux fragments d'amplification de taille très différente chez les deux espèces et localisés sur deux groupes non homologues (Q2 et C7). L'homologie de séquence n'a malheureusement pas pu être vérifiée puisque nous n'avons pas obtenu les séquences pour ce fragment. Néanmoins, nous sommes là aussi vraisemblablement en présence d'une famille multigénique puisque l'EST de départ code pour une metallothionein-like protein de type 3, qui compte 7 gènes cartographiés à ce jour chez *Arabidopsis* (Guo *et al.*, 2003).

Trois résultats principaux ressortent de cet alignement des cartes génétiques du chêne pédonculé et du châtaignier européen. Tout d'abord, les douze groupes de liaison apparaissent homologues et ne présentent pas de réarrangements chromosomiques majeurs malgré les 60 millions d'années écoulées depuis la divergence des deux espèces. Ensuite, à l'exception de deux marqueurs (Cons19 et Cons90), la macro synténie⁷ semble conservée. Enfin, la macro colinéarité⁸ semble aussi globalement respectée sauf pour cinq marqueurs. Sur ce dernier point, dans les cinq cas, les marqueurs sont à moins de 5 cM de distance et, à cette distance, la fiabilité dans l'ordre des marqueurs n'est pas très bonne avec 57 individus, comme nous l'avons déjà signalé. Ce résultat provient donc certainement d'erreurs de cartographie. Cependant, ces résultats demeurent préliminaires d'autant que la densité des marqueurs orthologues entre les deux espèces est encore faible. Si l'on prend l'exemple du riz et du maïs, ce n'est que très récemment que des auteurs ont mis en évidence des nouvelles duplications et des micro-réarrangements (Paterson *et al.*, 2004; Salse *et al.*, 2004) grâce à l'exploitation de données de séquence.

2.1.3. Les marqueurs dérivés des EST : un outil de choix pour la cartographie comparée

La disponibilité de séquences de fragments de gènes, tels que les EST, s'est récemment accrue de manière fulgurante grâce aux développements des programmes de séquençage et d'études d'expression sur de nombreuses espèces végétales. Les analyses transcriptomiques représentent un énorme potentiel pour l'étude de la physiologie et des processus développementaux chez les plantes, comme nous l'avons montré chez le chêne pour le caractère de débourrement dans le chapitre B. Par ailleurs, ces études réalisées sur le transcriptome fournissent également de nombreuses séquences de gènes disponibles à travers les bases de données EST. Dans le cas des espèces non modèles, comme le sont la plupart des essences forestières, et en particulier le chêne et le châtaignier, ces études sont d'une importance particulière pour les études génétiques et moléculaires. Etant donné la forte conservation de séquence entre les espèces, l'intérêt des EST dans le cadre des études de cartographie comparée est grand et a déjà été mis en évidence chez les conifères (Brown *et al.*, 2001). De plus, certaines EST colocalisant avec des QTL contrôlant la qualité du bois ont

⁷ Synténie : conservation du contenu des gènes dans certaines portions du génome d'espèces différentes.

⁸ Colinéarité : la colinéarité est respectée lorsqu'au moins trois marqueurs orthologues sont dans le même ordre chez des espèces différentes.

été considérées en tant que gènes candidats pour ce caractère (Brown *et al.*, 2003; Chagne *et al.*, 2003). Les résultats obtenus dans le cadre de notre étude viennent confirmer à la fois l'efficacité des marqueurs dérivés des EST pour des études de cartographie comparée mais aussi, nous le verrons par la suite, leur intérêt pour l'identification de gènes candidats.

Les résultats de l'amplification des STS montrent clairement qu'une forte proportion ($\approx 75\%$) des marqueurs dérivés des EST a amplifié chez le chêne et a été facilement transféré chez le châtaignier. Il s'agit en effet de séquences codantes beaucoup plus conservées entre espèces que les séquences flanquantes de marqueurs microsatellites par exemple. De plus, les amorces de ces marqueurs ont été préférentiellement choisies dans les régions les plus conservées de ces EST. Par ailleurs, près de 60% des STS amplifient un fragment de taille supérieure à celle attendue sur la base de la séquence de l'EST de départ. Ce résultat peut s'expliquer par la présence de séquences introniques. Cette présence augmente les chances de détecter du polymorphisme mais augmente aussi le risque d'amplifier difficilement les STS du fait de l'efficacité moindre de la réaction PCR sur des fragments de grande taille. C'est pour ces raisons qu'il apparaît judicieux de dessiner des amorces amplifiant des fragments de taille relativement petite (100-400 pb). Par ailleurs, parmi les différentes techniques de détection de mutations disponibles, nous avons choisi les méthodes SSCP et DGGE pour plusieurs raisons. D'une part, elles ne nécessitent pas de caractériser au préalable la séquence du marqueur STS, d'autre part le coût est plus faible que les techniques de génotypage de SNP et l'équipement pour les mettre en œuvre est relativement simple et peu cher. De plus, plusieurs polymorphismes peuvent être révélés à la fois ce qui permet d'obtenir des marqueurs co-dominants. A l'aide des méthodes SSCP et DGGE, un polymorphisme informatif pour la cartographie a été détecté pour 67% (43/64) et 55% (34/62) des EST correctement amplifiés, respectivement sur le chêne et le châtaignier. Ces résultats montrent clairement que les EST sont une source importante de marqueurs pour les études de cartographie comparée. A titre de comparaison, la transférabilité des marqueurs microsatellites était beaucoup plus faible entre les deux espèces. En effet, 31% des microsatellites de chêne avaient pu être cartographiés chez le châtaignier, et 17% des microsatellites de châtaignier avaient pu être cartographiés chez le chêne (Barreneche *et al.*, 2004). Bien que les microsatellites offrent de nombreuses applications ultérieures dans le domaine de la génétique des populations et aient représenté les premiers points d'ancrage pour la cartographie comparée chez les Fagacées, les marqueurs dérivés des EST ont été plus efficaces pour l'alignement des cartes génétiques des différentes espèces.

Tableau 11 : Nombre et valeurs de variance phénotypique expliquée des QTL impliqués dans le débourrement détectés chez *Q. robur* (Q) et *C. sativa* (C).

<i>Débourrement</i>	Q1-C6	Q2-C1	Q3-C8	Q4-C2	Q5-C4	Q6-C11
Nb QTL	1-3	4-3	1-0	4-1	1-1	2-2
PEV (%)	10,1-7,3	7,1-9,1	5,7-0	5,9-8,5	4,8-10,3	8,0-7,9
<i>Débourrement</i>	Q7-C5	Q8-C7	Q9-C9	Q10-C10	Q11-C3	Q12-C12
Nb QTL	3-1	0-1	4-4	4-0	2-2	1-3
PEV (%)	8,5-6,7	0-4,5	5,9-8,2	8,0-0	8,2-7,6	5,3-5,8

2.2. Le déterminisme génétique du débourrement chez *C. sativa* et *Q. robur*

Avant de présenter et de discuter les résultats de comparaison des QTL entre le chêne pédonculé et le châtaignier, nous pouvons rappeler quelques limites de l'approche qui seront utiles afin d'avoir une lecture et une interprétation éclairées des résultats obtenus. Trois faiblesses de l'étude limitent en effet les conclusions qui peuvent en être tirées : i/ les QTL détectés ne forment pas la totalité des QTL contrôlant le caractère de débourrement chez ces espèces, et les QTL à effet faible ne sont pas détectés ; ii/ la position estimée des QTL est relativement imprécise, et la plupart des intervalles de confiance recouvrent plus de la moitié du groupe de liaison ; iii/ la carte consensus sur laquelle les QTL sont projetés contient un certain nombre de marqueurs qui n'ont pas servi à la détection des QTL. A l'inverse, les données phénotypiques utilisées pour la détection des QTL ont été mesurées durant trois années chez les deux espèces et dans différents dispositifs chez le chêne. Les QTL détectés sont donc des QTL stables qui ne dépendent pas des conditions environnementales.

2.2.1. Comparaison des positions des QTL chez les deux espèces

Afin de pouvoir comparer la position des QTL impliqués dans le débourrement entre le chêne pédonculé et le châtaignier européen, la première étape a consisté à transférer les QTL détectés sur la carte consensus de chaque espèce. Les données de QTL avaient été obtenus par Casasoli (2004) et Scotti-Saintagne (2004). Les résultats de recherche de QTL contrôlant le débourrement sont synthétisés dans le tableau 11. Au sein de chaque couple de groupes homologues, le nombre total de QTL détectés et la valeur moyenne de variance phénotypique expliquée (PEV⁹) sont reportés pour chaque espèce. Le niveau de signification (α) choisi, qui rend compte du risque de première espèce (faux positif = risque de détecter un QTL alors qu'il n'est pas présent), était de $\alpha = 0,20$ au niveau du génome chez le châtaignier. C'est pourquoi, seuls les QTL détectés avec un risque de première espèce inférieur ou égal à cette valeur ont été considérés chez le chêne. Par ailleurs, pour les deux espèces nous n'avons considéré que les QTL détectés avec les données phénotypiques mesurées année par année. Les QTL détectés durant trois expériences indépendantes (trois années sur le châtaignier, trois années et différents environnements sur le chêne) ont donc été comparés dans cette partie

⁹ PEV : Phenotypic Explained Variance.

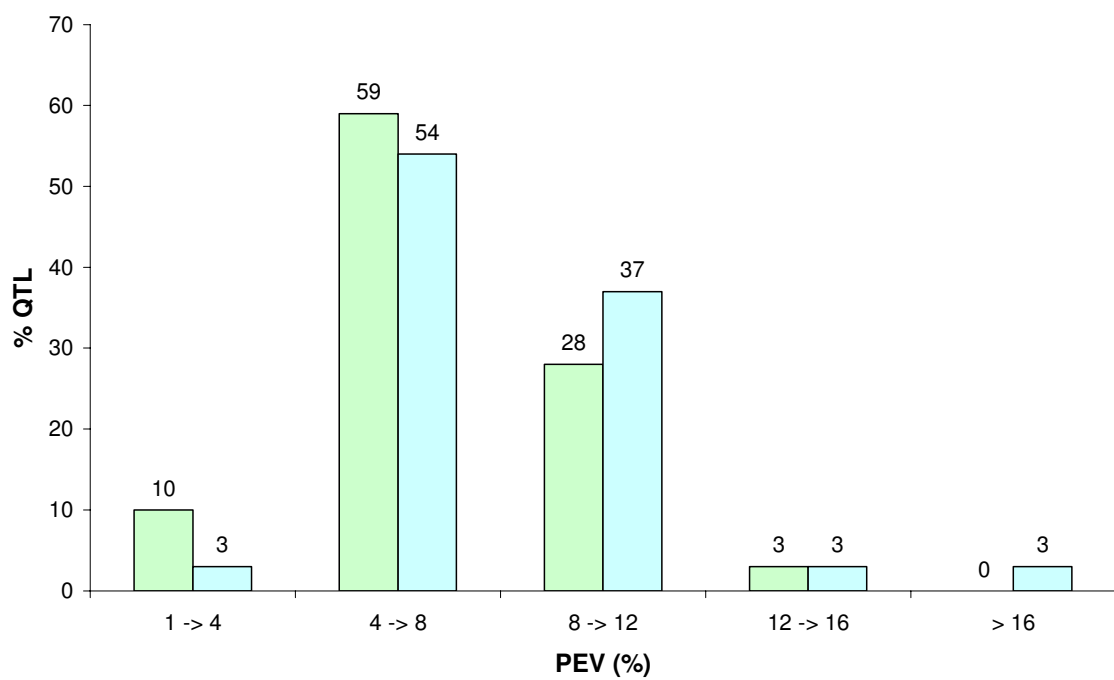


Figure 38 : Distribution des valeurs de variance phénotypique expliquée par les QTL détectés chez le chêne (bleu) et le châtaignier (vert).

expérimentale. Vingt sept et vingt et un QTL individuels ont été détectés, respectivement chez le chêne pédonculé et le châtaignier européen. Les valeurs moyennes de pourcentage de variance phénotypique expliquée (PEV) varient entre 4,8% et 10,1% chez le chêne, et entre 4,5 et 10,3% chez le châtaignier. Il faut noter que les valeurs de PEV ne sont comparables qu'à la condition que les deux expériences de détection de QTL aient été réalisées avec le même effectif (Beavis, 1995). Malgré le nombre supérieur d'individus dans la descendance de chêne (278) par rapport à celle de châtaignier (186), le nombre moyen d'individus utilisés pour la cartographie des QTL impliqués dans le débourrement était de 174 et 153, respectivement pour le chêne et le châtaignier. Il est possible, dans ces conditions, de comparer les valeurs de PEV estimées (Figure 38). Chez les deux espèces, une majorité de QTL à effet faible ou modéré¹⁰ ont été détectés plutôt que des QTL à effet fort. Le tableau 11 montre que les groupes Q2-C1 et Q9-C9 contiennent des QTL stables pour les deux espèces. A l'inverse, sur certains groupes, on observe de grandes différences comme sur le groupe Q10 qui contient 4 QTL alors que le groupe homologue du châtaignier (C10) n'en contient aucun.

La figure 39 présente les deux cartes consensus contenant tous les marqueurs orthologues cartographiés chez les deux espèces et certains marqueurs de la carte de détection des QTL ayant permis la projection de ces QTL sur ces nouvelles cartes. Au bas de la figure, se trouvent les deux cartes consensus représentées avec les QTL projetés et peu de marqueurs, ce qui permet d'identifier aisément les intervalles communs entre le chêne pédonculé et le châtaignier européen. Trente quatre intervalles communs ont été identifiés sur les cartes consensus de *Q. robur* et *C. sativa* grâce aux marqueurs orthologues. Selon la définition de QTL unique décrite dans le paragraphe 1.2.2., nous avons identifié 15 et 12 QTL uniques impliqués dans le débourrement, respectivement chez le châtaignier et chez le chêne. A titre d'exemple, sur le groupe Q2, 4 QTL ont été identifiés qui correspondent à 2 QTL uniques (cf. figure 39), l'un correspond aux trois QTL du haut et est situé dans l'intervalle colorié en jaune et l'autre correspond au QTL du bas et est situé dans l'intervalle colorié en rose. Les trois QTL identifiés chez le châtaignier correspondent à un QTL unique situé dans l'intervalle jaune. Parmi ces QTL uniques, onze ont été colocalisés entre les deux espèces. En suivant la formule proposée par Lin (1995), la probabilité d'obtenir ces colocalisations par hasard est $p = 0,0001$. Les QTL conservés entre les deux espèces ont été positionnés dans les mêmes régions génomiques.

¹⁰ Les effets des QTL sont considérés faibles ou modérés si la proportion de la variance phénotypique expliquée est comprise entre 4 et 12%.

Intervalles comparés : 3

QTL Q : 1 débourement

QTL C : 1 débourement

Colocalisations : 1

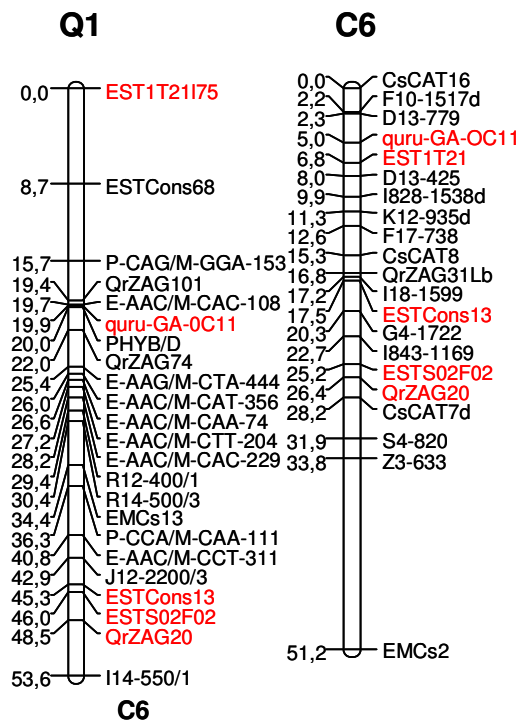
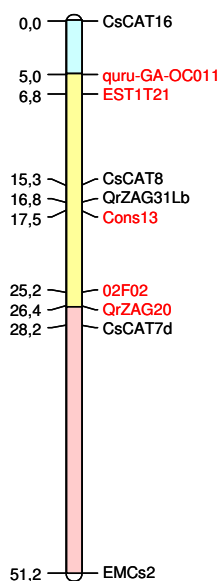
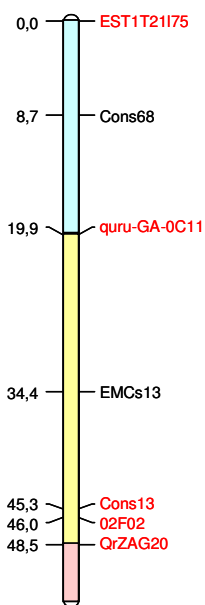


Figure 39

Q1

C6

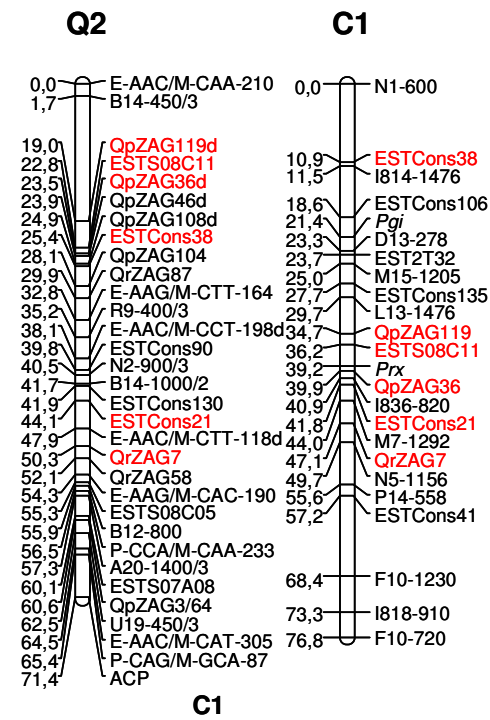


Intervalles comparés : 4

QTL Q : 2 débourement

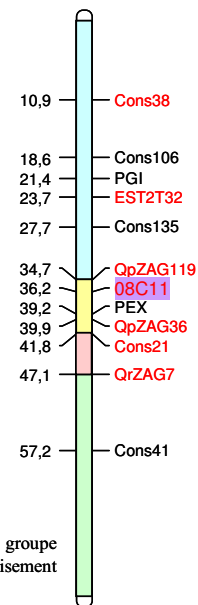
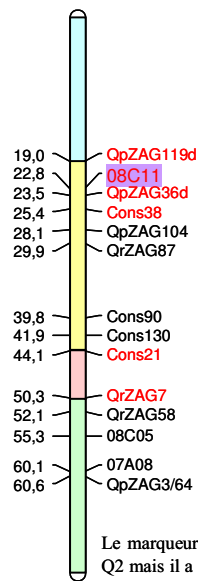
QTL C : 1 débourement

Colocalisations : 1



Q2

C1



Le marqueur EST2T32 se trouve sur le groupe Q2 mais il a été cartographié dans le croisement interspécifique (voir texte).

Intervalles comparés : 2

QTL Q : 1 débourement

QTL C : aucun

Colocalisations : aucune

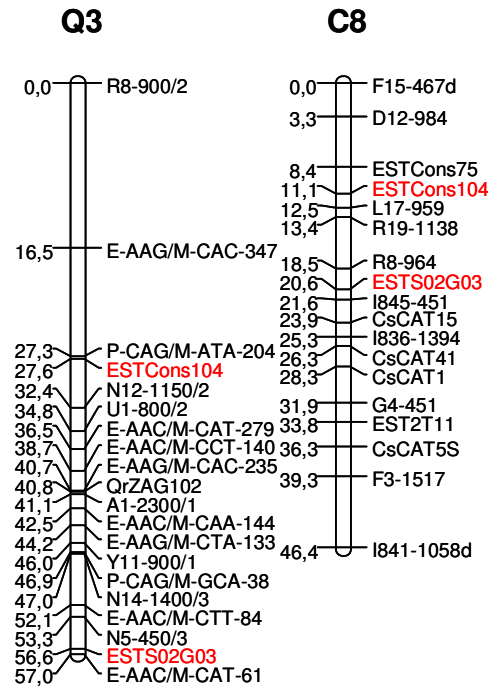
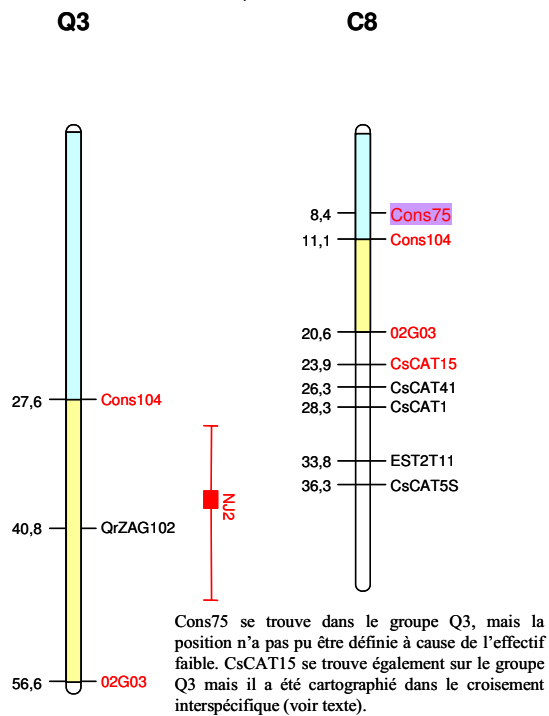


Figure 39

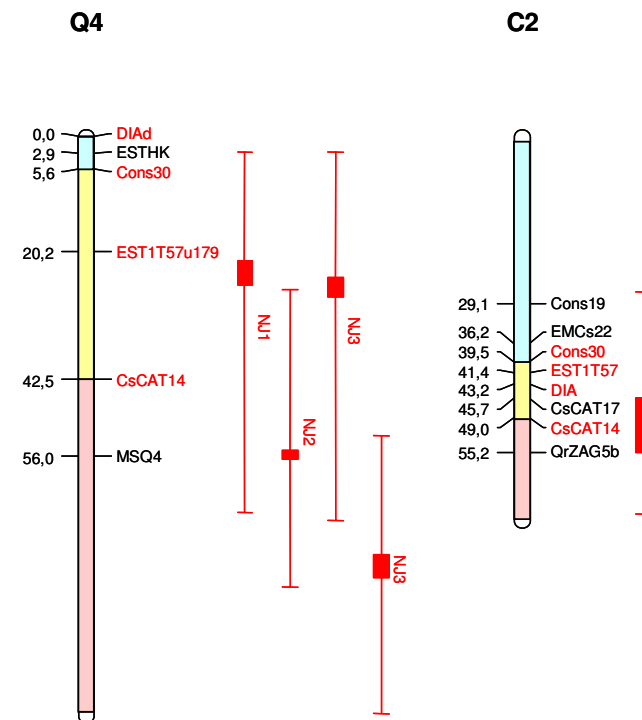
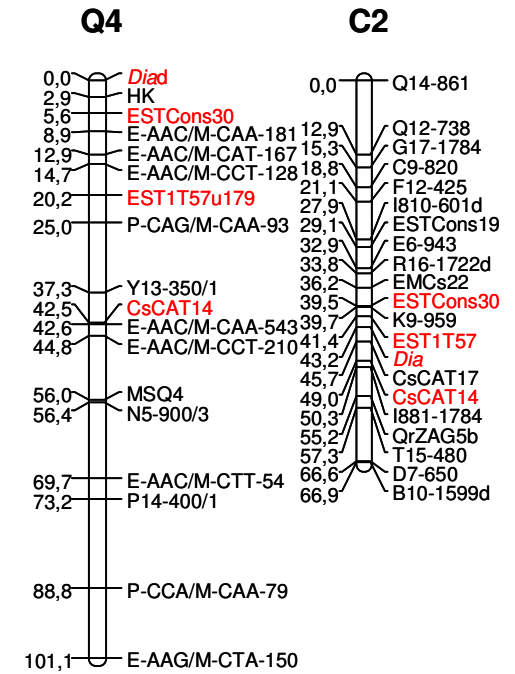


Intervalles comparés : 3

QTL Q : 2 débourement

QTL C : 1 débourement

Colocalisations : 1

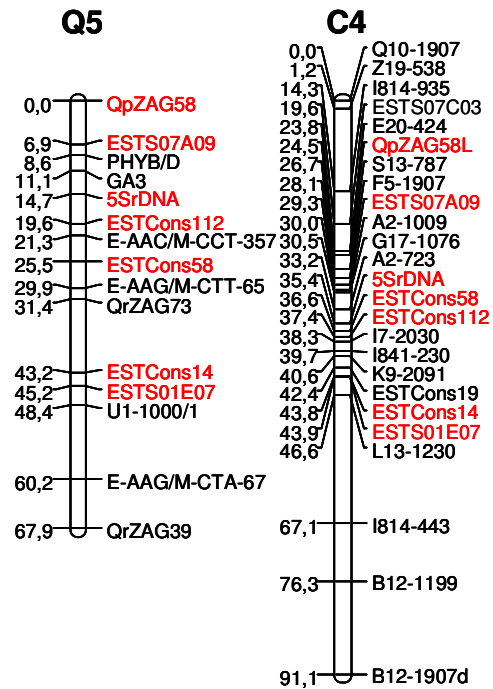


Intervalles comparés : 3

QTL Q : 1 débourement

QTL C : 1 débourement

Colocalisations : 1



Intervalles comparés : 2

QTL Q : 1 débourement

QTL C : 1 débourement

Colocalisations : 1

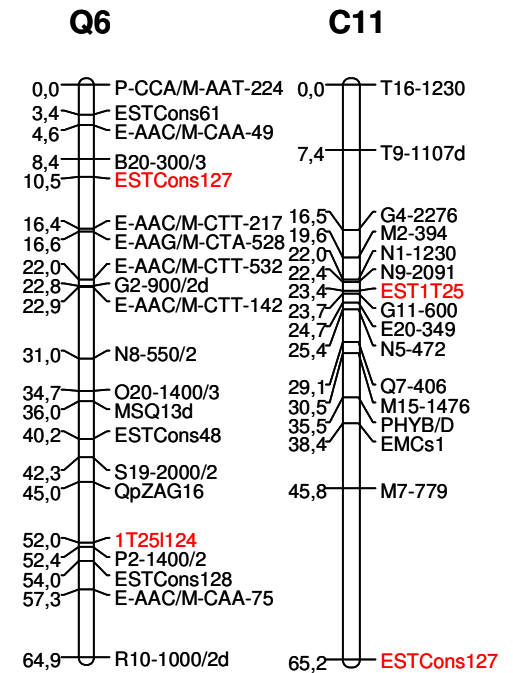
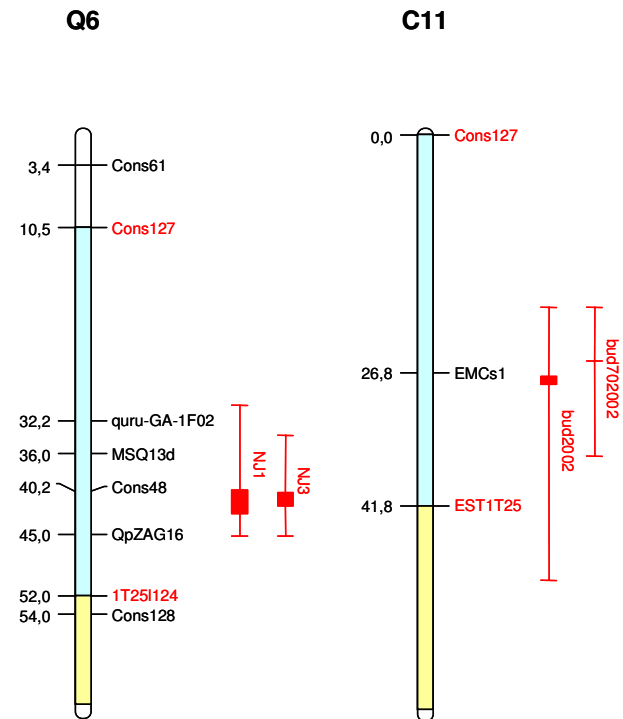
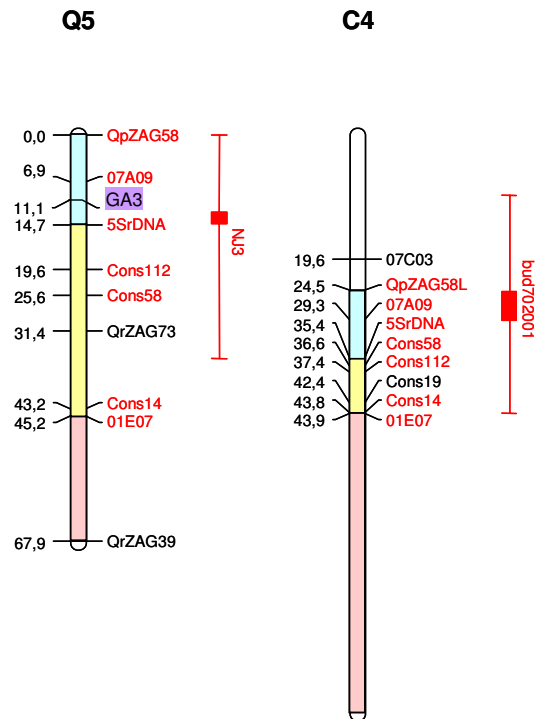


Figure 39

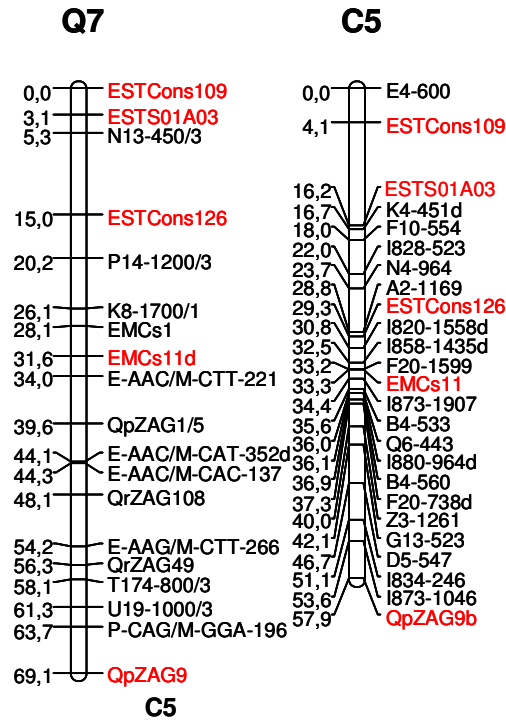


Intervalles comparés : 2

QTL Q : 1 débourement

QTL C : 1 débourement

Colocalisations : 1



Intervalles comparés : 2

QTL Q : 0 débourement

QTL C : 1 débourement

Colocalisations : 1

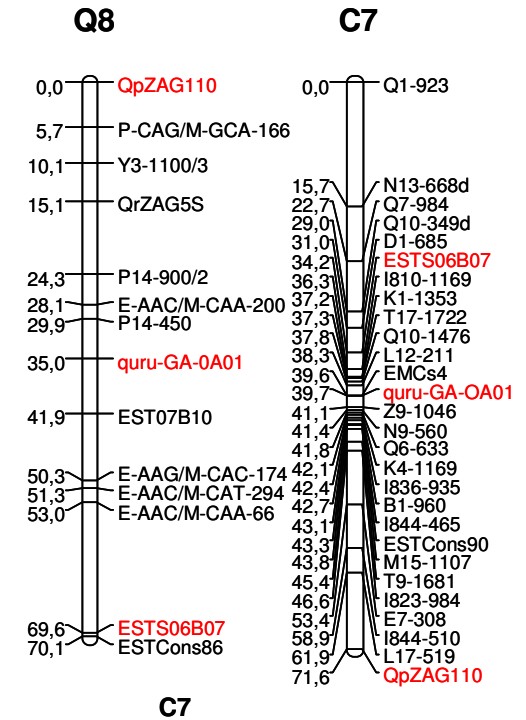
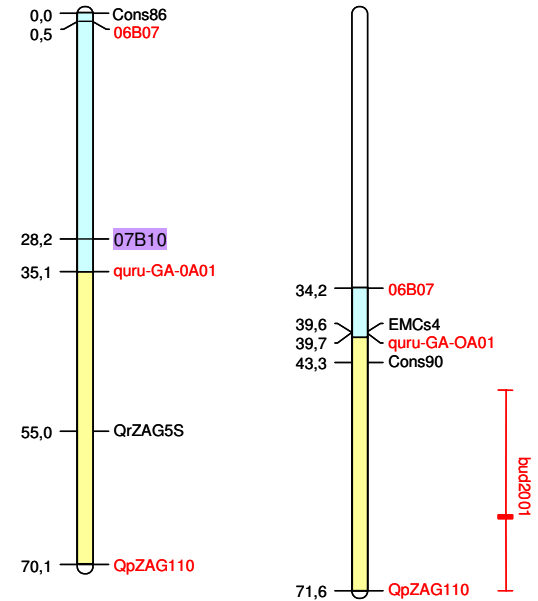
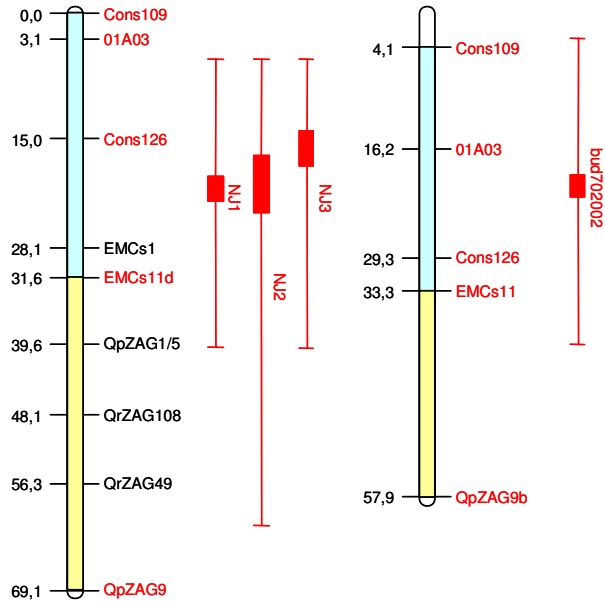


Figure 39 Q7



Intervalles comparés : 4
 QTL Q : 2 débourement
 QTL C : 2 débourement
 Colocalisations : 2

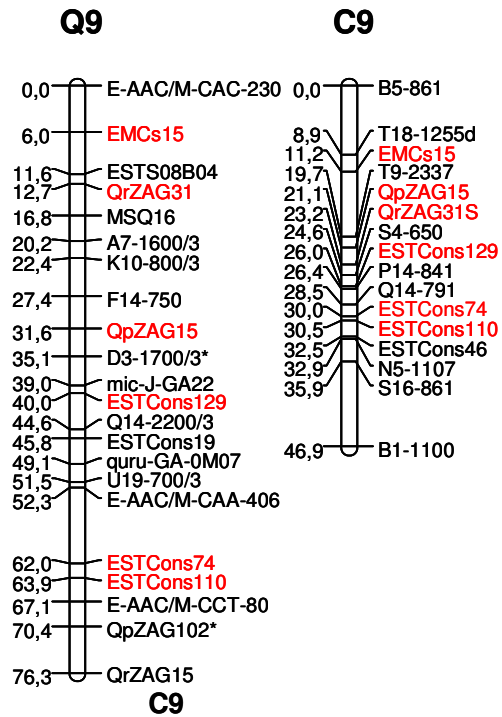
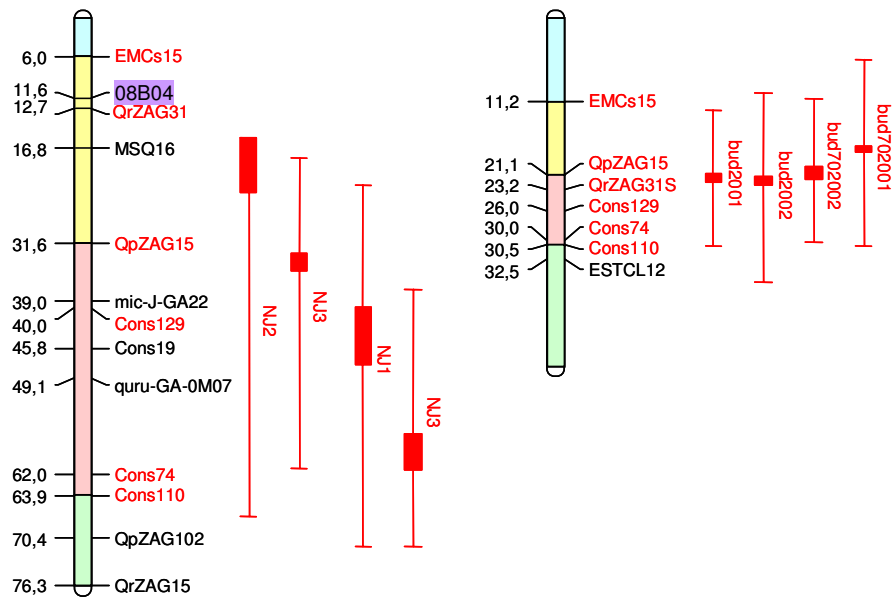
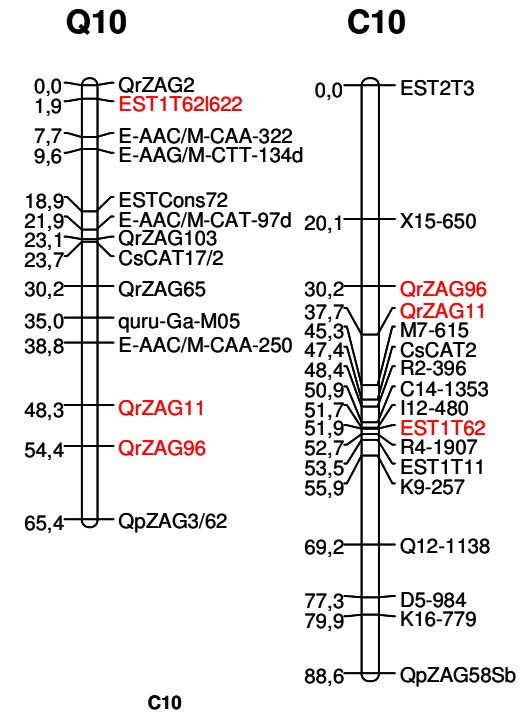


Figure 39
 Q9

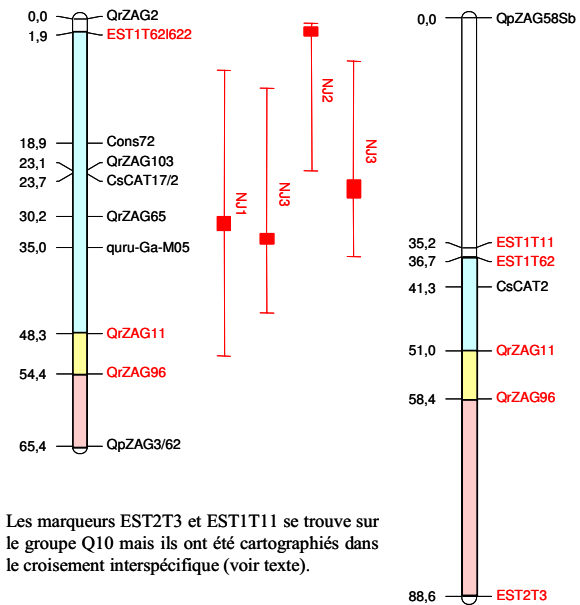


Intervalles comparés : 3
 QTL Q : 1 débourement
 QTL C : aucun
 Colocalisations : aucune



Q10

C10



Les marqueurs EST2T3 et EST1T11 se trouve sur le groupe Q10 mais ils ont été cartographiés dans le croisement interspécifique (voir texte).

Intervalles comparés : 4
 QTL Q : 2 débourement
 QTL C : 2 débourement
 Colocalisations : 2

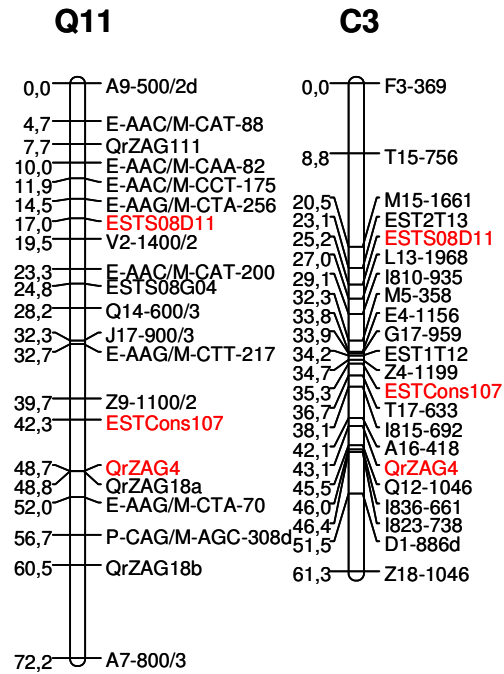
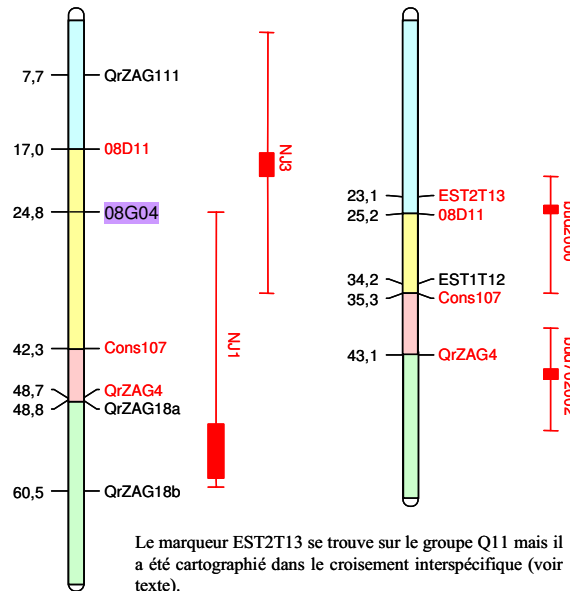


Figure 39

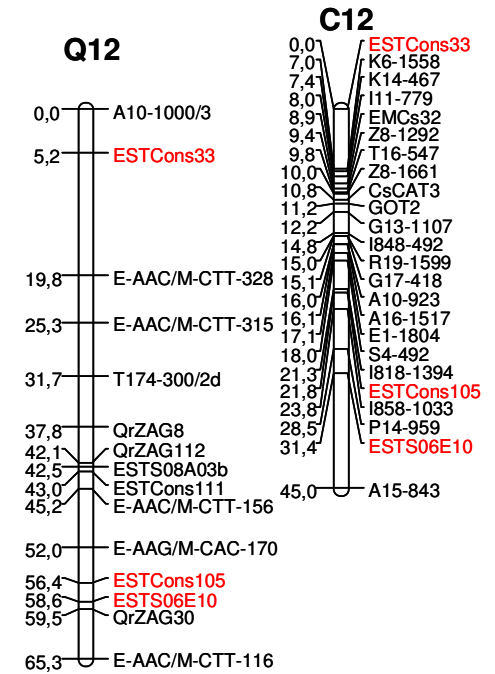
Q11

C3



Le marqueur EST2T13 se trouve sur le groupe Q11 mais il a été cartographié dans le croisement interspécifique (voir texte).

Intervalles comparés : 2
 QTL Q : 1 débourement
 QTL C : 1 débourement
 Colocalisations : 1



Q12

C12

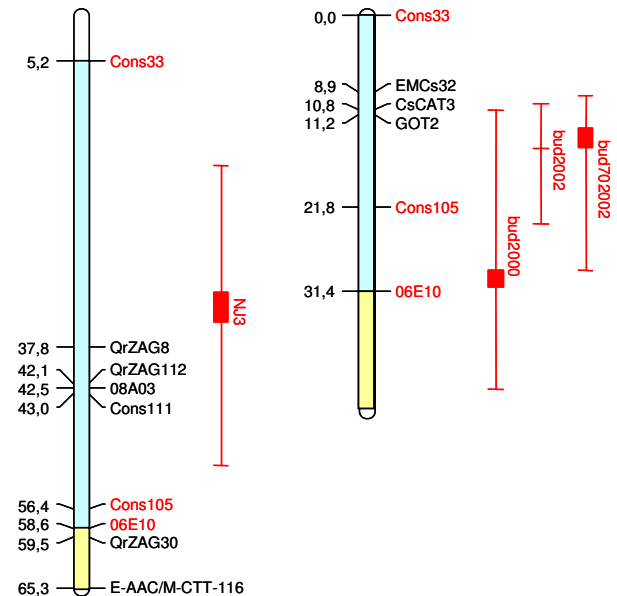


Figure 39 : les deux cartes consensus de *Q. robur* (Q) et *C. sativa* (C) sont montrées en haut avec tous les marqueurs utilisés pour la construction de ces cartes. Les noms des groupes se réfèrent à ceux définis par Barreneche *et al.* (1998) et Casasoli *et al.* (2001). Les groupes de liaison du chêne ont été pris comme référence et ordonnés de Q1 à Q12.

En bas, les mêmes couples sont représentés avec peu de marqueurs pour mieux visualiser les intervalles communs, matérialisés par des couleurs différentes. Les marqueurs EST2T32, EST2T3, EST1T11, EST2T13 et CsCAT15 n'ont pas été représentés sur les groupes de liaison du chêne parce qu'ils ont été cartographiés à l'aide d'une population de cartographie différente, comme c'est expliqué dans le texte. Le marqueur cons75, a été cartographié dans le groupe Q3 mais n'a pas été représenté parce que son introduction perturbe l'ordre des marqueurs. Les marqueurs orthologues sont représentés en rouge. Les gènes candidats pour le débourrement sont identifiés par un fond coloré violet. Les QTL projetés sont représentés sur la droite des groupes de liaison.

2.2.2. Un déterminisme génétique similaire

Les études de comparaison de QTL ont souvent été utilisées pour rechercher les régions génomiques, contrôlant des caractères quantitatifs, conservées tout au long de l'évolution (Paterson *et al.*, 1995). Cependant, les échantillonnages hétérogènes, comme les faibles tailles de populations, qui sont la plupart du temps utilisées dans les expériences de détection de QTL, peuvent introduire des biais dans les résultats (Barton et Keightley, 2002). Dans le cas de notre étude comparative, nous avons déjà cité les limites importantes à considérer en introduction du chapitre 2.2. Malgré cela, et parce que des stratégies similaires d'échantillonnage ont été utilisées chez les deux espèces, nous pouvons tirer certaines conclusions de cette analyse comparative. Tout d'abord, le déterminisme génétique du caractère de débourrement est comparable entre les deux espèces en termes de nombre de QTL et de contribution à la variance phénotypique. Les QTL à effet faible ou modéré sont plus fréquents que ceux à effet fort, et ce résultat paraît être un caractère commun à la génétique des caractères adaptatifs. On peut émettre l'hypothèse que les QTL stables correspondent à des gènes impliqués dans les processus de base contrôlant le caractère de débourrement et qui auraient été conservés au cours de l'évolution. Ces QTL seraient ainsi presque toujours détectés chez les deux espèces. La stabilité des QTL liés au débourrement est en accord avec les résultats obtenus par Jermstad (2003; 2001) et semble confirmer l'héritabilité élevée de ce caractère. Par ailleurs, les différences observées peuvent provenir des dispositifs de détection mais aussi de la biologie des espèces. En effet, seul un sous ensemble du jeu complet des QTL contrôlant le débourrement a été détecté. De plus, les conditions expérimentales des deux essais, conditions écologiques des tests notamment, étaient différentes et indépendantes, ce qui implique que les mécanismes de déclenchement du processus de débourrement peuvent être différents dans ces deux cas. En dernier lieu, ces différences peuvent s'expliquer par le fait que le débourrement soit contrôlé, en partie, par des mécanismes différents chez le chêne et le châtaignier.

La position de ces QTL est remarquablement conservée puisque, parmi les QTL uniques détectés chez le chêne (15 QTL) ou chez le châtaignier (12 QTL), 11 sont colocalisés avec une forte probabilité ($p = 0,001\%$) d'occurrence. Ce résultat est d'autant plus remarquable si l'on considère le temps de divergence estimé à 60 millions d'années entre le chêne et le châtaignier mais aussi le fait que d'autres caractères adaptatifs tels que la croissance en hauteur et l'efficacité d'utilisation de l'eau (Annexe II) n'apparaissent pas

conservés. La stabilité plus forte des QTL impliqués dans le débourrement peut, en partie, expliquer ce résultat. La date de débourrement est probablement un caractère moins complexe que la discrimination isotopique du carbone ou la croissance du fait d'un nombre plus restreint de processus physiologiques contribuant à son expression. Comme nous l'avons vu en détail dans la synthèse bibliographique, la date de débourrement est principalement influencée par la température, et la photopériode dans une moindre mesure. En contraste, de nombreux facteurs environnementaux sont susceptibles d'influencer la discrimination isotopique du carbone et la croissance et donc contribuer à ce que ces caractères soient sous la dépendance de nombreux processus physiologiques. Cela implique que les QTL détectés pour ces caractères peuvent fortement varier en fonction du facteur environnemental prépondérant agissant dans tel ou tel expérimentation. La mise en place d'expériences de détection de QTL en conditions contrôlées pourrait aider à valider cette hypothèse.

Les résultats obtenus soulèvent également la question du maintien du polymorphisme aux mêmes locus chez le chêne et le châtaignier. En effet, les locus contrôlant le débourrement étaient hétérozygotes chez les individus parentaux du croisement contrôlé utilisé pour la détection des QTL. Bien que le débourrement soit fortement influencé par la sélection naturelle chez les arbres forestiers, la variabilité intra population est maintenue à des niveaux élevés (Howe *et al.*, 2003). Nos résultats suggèrent que les pressions de sélection agissent sur les mêmes régions génomiques par des mécanismes capables de maintenir la diversité durant de longues périodes d'évolution. C'est le cas de la sélection balancée¹¹ ou de la sélection disruptive¹². L'une ou l'autre de ces deux types de sélection pourraient provenir, dans le cas du chêne au moins, d'un phénomène de co-évolution entre l'arbre et les populations d'insectes défoliateurs, qui peuvent être responsables du maintien de la variabilité intra population (Tikkanen et Tiitto, 2003). Si l'on considère les conditions climatiques, la sélection favorisant les génotypes tardifs peu coexister avec des pressions sélectives favorisant des génotypes précoces qui auront un avantage sélectif pour la croissance en conditions favorables. Il en résulterait une élimination préférentiel des individus à débourrement « intermédiaire ». Un phénomène de sélection disruptive apparaîtrait alors sous l'action des conditions climatiques. La conservation des positions des QTL n'implique pas que les gènes responsables du QTL soient les mêmes. En effet, chez les Poacées, le gène *tb1*, responsable du contrôle du tallage chez le maïs, n'a qu'un effet mineur chez le millet (Doust

¹¹ Sélection balancée : dans ce cas, la fitness des hétérozygotes est plus grande que celle des homozygotes.

¹² Sélection disruptive : dans ce cas, les phénotypes extrêmes sont favorisés par rapport aux phénotypes intermédiaires.

et al., 2004). D'autres gènes, situés dans la même région génomique, représentent des meilleurs candidats pour le contrôle du caractère. Donc, malgré la conservation exceptionnelle des QTL contrôlant le débourrement chez *Q. robur* et *C. sativa*, ce travail ne permet pas encore de tirer des conclusions quant à la correspondance des locus impliqués dans le caractère chez les deux espèces.

2.3. Identification de gènes candidats pour le débourrement

L'utilisation des marqueurs dérivés des EST dans le cadre de ce travail n'a pas seulement permis d'obtenir des marqueurs orthologues chez le chêne pédonculé et le châtaignier européen, mais aussi d'identifier, ou de préciser l'intérêt d'un certain nombre de gènes candidats pour le débourrement, au regard de leur localisation par rapport aux QTL contrôlant le débourrement. D'une part, les gènes candidats potentiels préalablement identifiés (Chapitre B) ont été inclus dans le travail de cartographie ; d'autre part, les autres EST cartographiées peuvent également constituer des gènes candidats intéressants. En effet, des EST différentiellement exprimées qui colocalisent avec les QTL impliqués dans le débourrement et conservés au cours de l'évolution, représentent potentiellement des gènes candidats contrôlant une partie de la variation quantitative du caractère. Cependant, comme nous l'avons déjà mentionné précédemment, le faible pouvoir résolutif de l'étude en termes de position à la fois des gènes et des QTL, fait que la colocalisation d'un marqueur STS avec un QTL contrôlant le débourrement est loin d'être une preuve suffisante de l'implication du gène dans le processus physiologique. Inversement, seule une partie des QTL sont détectés dans l'étude et l'absence de colocalisation n'exclut pas que l'EST considérée soit impliquée dans le déterminisme du caractère.

2.3.1. Cartographie des gènes candidats expressionnels

Le premier couple à contenir un des gènes candidats potentiels est le couple Q2-C1. Pour rappel, le chêne et le châtaignier contiennent, respectivement, 4 et 3 QTL contrôlant le débourrement et 1 colocalisation a été observée (Figure 39). Le STS dérivant du GC 08C11 a été cartographié dans le même intervalle que le QTL conservé entre les deux espèces. Les autres STS d'intérêt sont Cons38 (petite sous-unité de la RubisCo), Cons90 (metallothionein type 3), Cons130 (sous-unité de la glyceraldehyde 3-phosphate déshydrogénase chloroplastique) et Cons21 (protéine ribosomale S27). Chez le châtaignier, deux isozymes : la peroxydase et la 6-phospho gluconate déshydrogénase ont également été cartographiées sur ce groupe. La plupart de ces enzymes et protéines voient leur niveau d'expression fortement varier au moment du débourrement (Rohde *et al.*, 2000) et cela explique peut être la localisation de nombreux QTL impliqués dans le débourrement sur ces deux groupes de liaison. Pour ce qui concerne le STS 08C11, sa localisation au niveau de ce QTL conservé

renforce son intérêt en tant que gène candidat pour le débourrement puisqu'il constituait déjà un GC expressionnel. On peut également revenir à ce stade sur son annotation fonctionnelle puisque sa fonction potentielle au moment de la phase de sélection de GC expressionnel était une protéine hypothétique contenant un domaine actif codant pour un transporteur de tétrapeptides. Cette annotation s'est affinée par la suite et ce transcrite code potentiellement pour le transporteur YSL1, de la famille des transporteurs d'oligopeptides (OPT). Au même titre que les familles de transporteurs ABC ou PTR, cette fonction pourrait lui conférer un rôle dans la perception d'un signal menant au débourrement. A ce titre, il peut également être considéré comme un GC fonctionnel et donc être pertinent à titre triple.

Le marqueur dérivé du GC GA3 est localisé sur le couple Q5-C4, qui contient également, dans l'intervalle de confiance du QTL conservé, le STS 07A09 (transporteur ABC), Cons58 (dormancy-associated protein), Cons112 (Histone H3), le STS 01E07 (auxin response factor 3), Cons19 (protéine ribosomale L13) et Cons14 (protéine ribosomale L18). La présence dans cette région de deux gènes codant pour des protéines de réponse à l'auxine (Cons58 et 01E07) est très intéressante puisque l'auxine et les gibbérellines (GA3 code en effet pour une enzyme de la voie de biosynthèse des gibbérellines) sont des hormones clés du débourrement et de la croissance cellulaire. Par ailleurs, les transporteurs ABC sont impliqués dans la perception du signal lumineux, qui joue un rôle dans le déclenchement du débourrement chez certaines espèces mais aussi dans le bourgeonnement automnal, y compris chez le chêne. De plus, un homologue du phytochrome B ou D a été localisé entre GA3 et l'EST codant pour le transporteur ABC (Figure 39). Nous ne l'avons pas décrit dans notre étude puisque les recherches doivent se poursuivre afin d'isoler et cartographier les membres distincts de la famille chez le chêne et le châtaignier. Quoi qu'il en soit, la localisation dans le même intervalle de ces trois gènes à des positions très proches est très intéressante puisque le phytochrome B, les hormones gibbérelliques et les transporteurs ABC interagissent fortement dans de nombreux processus. La présence d'une histone H3 est également à remarquer puisque ces dernières pourraient jouer un rôle dans le contrôle épigénétique de la dormance (cf. § 2.2.3. de la Synthèse Bibliographique). GA3 constitue donc un gène candidat à la fois fonctionnel et positionnel.

Le STS dérivé du GC 08B04, homologue potentiel du gène codant pour une protéine alpha-amylase/subtilisin inhibitor est localisé sur le groupe Q9 chez le chêne. Le couple d'homologues Q9-C9 est fortement impliqué dans le débourrement chez les deux espèces avec 2 QTL uniques conservés. En plus de l'EST 08B04, on trouve également, dans les intervalles de confiance des 2 QTL, Cons 129 (cytokinin-specific binding protein), Cons19

(protéine ribosomale L13), Cons74 (glutathione S-transferase) et Cons110 (PVR3-like protein). Cons129 et Cons74 sont potentiellement impliquées dans la dormance par leur fonction (Horvath *et al.*, 2003; Rohde *et al.*, 2000). Ces colocalisations font de l'EST 08B04 un gène candidat pertinent du point de vue de sa position, en plus de son profil d'expression et de sa fonction.

Le dernier GC cartographié, 08G04, se trouve sur le groupe Q11, homologue du groupe C3. Deux QTL conservés ont été détectés sur ces groupes. Dans le même intervalle, on trouve en plus Cons107 (hemoglobine II) et le STS 08D11 (beta-1,3-glucanase), dont les fonctions n'amènent pas d'information pertinente vis à vis du débourrement. Cependant, le GC 08G04 constitue lui aussi un candidat positionnel, en plus de son expression différentielle entre le bourgeon quiescent et le début du débourrement.

Le Cons75, dont fait partie le GC 08A01 est localisé sur le groupe C8 chez le châtaignier. Ce groupe ne contient aucun QTL pour le débourrement chez le châtaignier et, par conséquent, aucune colocalisation n'a été observée mais le chêne contient un QTL pour le débourrement sur le groupe homologue Q3. De façon similaire, le couple Q8-C7 contient le GC 07B10 codant potentiellement pour le facteur de transcription DAG2, et ne contient pas de QTL chez le chêne. Ces deux résultats ne remettent pas en cause l'intérêt de ces GC d'un point de vue expressionnel et/ou fonctionnel. Enfin, les STS dérivés des EST codant pour les quatre autres gènes candidats, précédemment identifiés, n'ont pu être cartographiés mais ce travail est amené à se poursuivre et ces gènes seront prochainement cartographiés.

2.3.2. Des clusters de gènes informatifs ?

En dehors de l'analyse des couples homologues contenant les gènes candidats identifiés lors de l'étape précédente (Chapitre B), les autres EST cartographiées peuvent également constituer des gènes candidats pertinents pour le caractère de débourrement. Au delà de l'objectif d'identification de GC, des clusters de gènes localisés dans la même région génomique et impliqués dans des processus physiologiques ou développementaux similaires peuvent apparaître, comme l'exemple précédent des groupes Q5 et C4.

En effet, dans l'intervalle d'intérêt des groupes Q1 et C6, ont été cartographiés les STS Cons13 (Histone H4), 02F02 (Histone H3) et un autre locus issu de l'amplification du phytochrome PHY B/D (Figure 39). Les gènes du phytochrome, impliqué chez les plantes dans la perception de la lumière, ont déjà été localisés avec des QTL impliqués dans le

débourrement et l'arrêt de croissance chez le peuplier (Frewen *et al.*, 2000), ce qui n'apparaît pas étonnant puisque la perception de la diminution de la longueur du jour est essentiel dans le processus d'arrêt de croissance et de début de période de dormance. Dans notre étude, même si l'identité des différents locus amplifiés a été vérifiée avec la séquence de départ (85-95%), issu de *Quercus turbinella*, l'identification des différents locus amplifiés est maintenant nécessaire pour pouvoir interpréter plus en avant ces résultats. Concernant les histones, des évidences d'ordre à la fois expressionnel et fonctionnel supportent le rôle de ces protéines comme gènes candidats du débourrement. Les histones H4 et H3, qui reflètent une augmentation du niveau de division cellulaire en général, sont différentiellement exprimées au cours du débourrement (cf. Chapitre B). Chez le pois, les niveaux de transcrits de l'histone H4, faibles durant le débourrement, augmentent durant la croissance du bourgeon et décroissent à nouveau ensuite quand les bourgeons stoppent leur croissance (Devitt et Stafstrom, 1995). Par ailleurs, chez le châtaignier, dans cette même région chromosomique, en plus des QTL contrôlant le débourrement, des QTL impliqués dans le bourgeonnement ont été localisés. Un résultat intéressant peut être tiré de l'étude faite par Casasoli (2004). L'effet de substitution allélique de ces deux QTL, identifiés sur la carte mâle (groupe LG6M), contribue à un débourrement plus tardif et un bourgeonnement plus précoce ou vice-versa. On peut émettre l'hypothèse que ces deux phénomènes soient sous la dépendance des gènes codant pour les histones, ou du moins que ces gènes de régulation de l'activité mitotique soient responsables de la présence de ces deux QTL. Ces différents résultats font des histones des gènes candidats pertinents à tester.

Sur les groupes Q4 et C2, un QTL conservé entre les deux espèces a été trouvé. L'intervalle de confiance de ce dernier contient le STS Cons30 (protéine inconnue). Sur les groupes Q10 et C10, aucun QTL conservé n'a été observé.

Les groupes Q6 et C11 contiennent les STS Cons 127 (early light induced protein) et Cons 48 (protéine déhydrine de réponse à l'ABA) mais aussi un locus du gène PHY B/D cartographié sur le châtaignier, dans l'intervalle de confiance du QTL conservé entre les deux espèces. La présence d'une déhydrine à proximité de la position la plus probable du QTL conservé est intéressante puisque ces protéines sont impliquées dans les phénomènes de protection contre le stress hydrique. Nous avons vu, dans le chapitre B, que le débourrement était associé à un fort stress hydrique des tissus du bourgeon et que les différentes déhydrines isolées étaient toutes différentiellement exprimées.

L'intervalle d'intérêt des groupes Q7 et C5 contient les STS Cons109 (metallothionein-like protein), Cons 126 (protéine ribosomale L24) et 01A03 (protéine

ribosomale S23). Une remarque intéressante est que les protéines ribosomales L24 et S23, cartographiés sur ces groupes font partie du groupe d'expression III, identifiés dans l'article en Annexe I, alors que les protéine ribosomales L13 et L18, cartographiés à proximité sur les groupes Q5 et C4, font partie du groupe d'expression II. Dans les deux cas, la position proche sur le génome est associée à un profil d'expression similaire. Les gènes de protéines ribosomales sont organisés en régions dupliquées en tandem dans le génome et certains de ces gènes proches sont certainement exprimés de façon coordonnée.

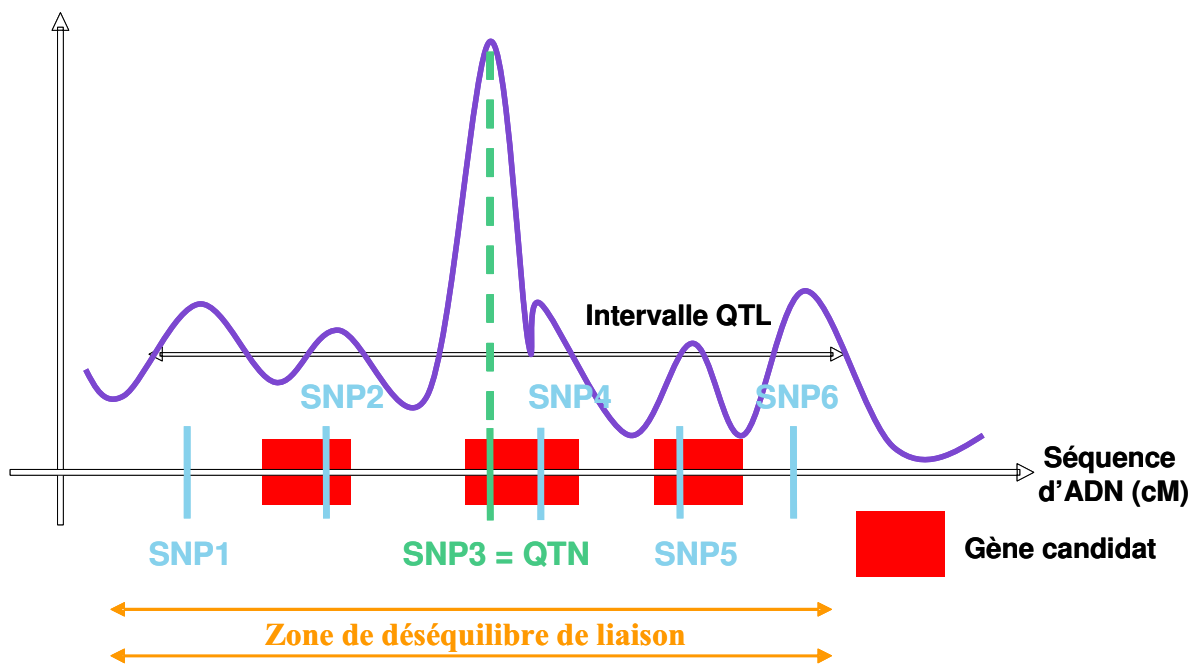
Le dernier couple d'homologues (Q12-C12) contient, dans l'intervalle d'intérêt du QTL conservé, Cons33 (Defender against cell death), 08A03 (squamosa promoter binding protein), Cons111 (oxigen-evolving complex related), Cons105 (phosphoenolpyruvate carboxykinase, PEPCK) et 06E10 (malate dehydrogenase). Il est intéressant de noter la présence d'un facteur de transcription impliqué dans le développement floral précoce chez *Arabidopsis*, ainsi que d'enzymes impliquées dans le cycle de la glycolyse ou le métabolisme énergétique. Le facteur de transcription hypothétiquement codé par l'EST 08A03 constitue un gène candidat pertinent pour le débourrement.

③ Conclusions de l'approche de cartographie comparée

L'utilisation des marqueurs orthologues, dont une majorité d'EST isolées dans le cadre de l'approche transcriptomique, cartographiés à la fois chez le chêne et chez le châtaignier a permis d'aligner les douze groupes de liaison de ces deux espèces. Les EST se sont révélées être un puissant outil pour les études de cartographie comparée, comme cela avait été montré chez les conifères auparavant (Brown *et al.*, 2001). Cet alignement des cartes consensus a donc rendu possible la comparaison de la position des QTL impliqués dans le débourrement, mais aussi la croissance et la discrimination isotopique du carbone 13 (Annexe II). En termes de nombre de QTL détectés et de variance phénotypique expliquée, le déterminisme génétique du caractère de débourrement semble bien conservé entre les deux espèces, et ce malgré les limites de l'approche discutées tout au long du chapitre. La stabilité de ces QTL impliqués dans le débourrement avait été démontrée au niveau intraspécifique, aussi bien chez *Q. robur* (Scotti-Saintagne *et al.*, 2004) que chez *C. sativa* (Casasoli *et al.*, 2004), et semble respectée au niveau interspécifique. Malgré les 60 millions d'années de divergence entre les deux espèces, il semblerait que les mêmes régions génomiques sont impliquées dans le contrôle du caractère de débourrement. Bien que les mêmes gènes ne soient pas nécessairement impliqués dans le caractère, cette stabilité des QTL suggère une stabilité du génome des Fagacées, suggestion renforcée par la similarité du caryotype, de la taille génétique et physique des deux génomes ainsi que la conservation de la macro synténie et de la macro colinéarité. La diversification interspécifique au sein des Fagacées entre le chêne et le châtaignier ne se serait donc pas produite par le biais de réarrangements chromosomiques majeurs. Par ailleurs, la cartographie de gènes candidats pour le caractère de débourrement semble être une stratégie pertinente même si l'identification des GC doit s'appuyer sur d'autres critères que la simple position sur les cartes génétiques, étant donné les intervalles de confiance très larges des QTL contenant des centaines, voire des milliers de gènes potentiels. L'utilisation des EST a également permis la mise en évidence de clusters de gènes proches qui pourraient être régulés de manière coordonnée. Cependant, ces clusters ne sont pas significatifs vu le faible nombre de gènes cartographiés dans notre étude.

D. Diversité nucléotidique et études d'association de gènes candidats pour le débourrement

Effet du SNP sur le caractère quantitatif



Dans le cadre de l'approche gène candidat employée dans notre étude, les deux précédents chapitres ont donc permis la mise en évidence de gènes candidats sur la base de critères d'expression au cours du débourrement et/ou de fonction potentielle et/ou de position sur les cartes génétiques du chêne et du châtaignier. L'étape suivante de cette approche GC consiste en l'étude de la structuration de la diversité, la recherche de signatures moléculaires de la sélection au sein de ces GC, ainsi que l'étude des associations entre la variabilité nucléotidique et la valeur phénotypique de la date de débourrement. Plus généralement, l'ensemble des données de séquençage et de génotypage sera analysé avec les outils de la génétique des populations à notre disposition.

L'approche QTL nous a renseigné sur les régions du génome impliquées dans la variation du caractère de débourrement mais avec une grande imprécision sur la localisation. En effet, la fenêtre de déséquilibre de liaison est très large au sein des pedigrees de cartographie et la région contrôlant une partie du caractère peut contenir jusqu'à des milliers de gènes (cf. Chapitre C). A l'inverse, la fenêtre de déséquilibre de liaison autour des sites fonctionnels est très réduite dans le cas des populations naturelles du fait des nombreuses générations de recombinaison qui se sont accumulées au cours du temps. L'approche sous forme d'étude d'association doit donc nous permettre de détecter directement le gène, voire le polymorphisme impliqué. Ce chapitre présente la partie expérimentale mise en place, dans le cadre de cette thèse, pour répondre à cet objectif. Le travail expérimental relatif à l'étude d'association sur les trois locus SNP a été effectué par Alexandre Jauffrès (Jauffrès, 2004).

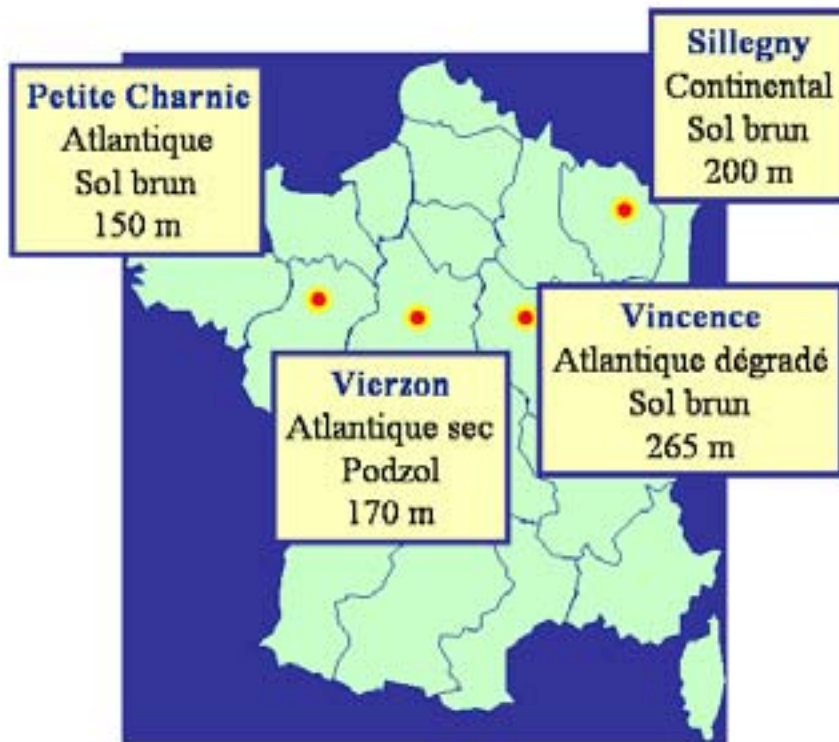


Figure 40 : Emplacement et caractéristiques (régime climatique, type de sol, altitude) des différents tests de provenances de chêne.



Figure 41 : Les populations analysées en études d'association. En violet, le sous échantillon utilisé pour la recherche des SNP. La note de débourement moyenne de la population apparaît sous forme d'un rond dont l'échelle de notation est indiquée en bas à droite de la figure. Les valeurs en rouge indiquent des populations dont la notation est supérieure à la moyenne (plutôt précoce) ; les populations en bleue sont celles dont la notation est inférieure à la moyenne (plutôt tardive).

① Matériel et Méthodes

La partie expérimentale qui suit présente les résultats issus de l'analyse de la diversité nucléotidique d'un certain nombre de gènes candidats préalablement identifiés (Chapitres B et C), parmi lesquels deux ont fait l'objet d'une étude d'association. La liste des principaux gènes candidats pour le débourrement ne se limite en aucun cas à ceux présentés dans cette partie. Les recherches sont en cours concernant les autres fragments identifiés qui constituent des GC tout aussi intéressants.

1.1. Du matériel végétal au génotypage des marqueurs SNP

1.1.1. Les populations échantillonnées au sein de l'aire naturelle

L'échantillonnage des populations utilisées dans le cadre des études d'association a été réalisé au sein de tests de provenances contenant plus d'une centaine de populations de chêne sessile originaires de l'ensemble de l'aire de répartition. Ces tests de provenances ont été installés sur quatre sites géographiquement et écologiquement distincts (Figure 40) afin de caractériser les différentes populations pour un certain nombre de paramètres relatifs à la croissance, la phénologie mais aussi l'architecture des arbres. Ces tests ont été mis en place entre 1989 et 1991 après récolte des glands sur place, germination en pépinière puis plantation. Chaque provenance est représentée en moyenne par 240 arbres plantés au sein de parcelles unitaires de 24 arbres, elles mêmes réparties au sein de blocs écologiquement homogènes. La récolte du matériel a été effectuée sur le site de la Petite Charnie, à l'exception de la population provenant de Lapwald, récoltée à Vincence. Nous avons récolté 21 populations originaires de l'ensemble de l'aire de répartition et contrastées pour la notation phénotypique du débourrement (Figure 41). Ces différentes provenances peuvent être réunies en trois groupes distincts de 7 populations (Tableau 12) : un groupe de provenances allemandes (G1), un groupe de provenances françaises (G2) et un groupe de provenances contrastées au maximum pour le débourrement. Deux critères principaux ont orienté l'échantillonnage : i/ chaque groupe présente une variabilité maximale pour la date de débourrement, à savoir des populations précoces, intermédiaires et tardives, ii/ les deux premiers groupes réunissent des populations issues d'une même zone géographique ce qui limite énormément l'impact de l'histoire (appartenance à des populations sources différentes,

Tableau 12 : Caractéristiques des 21 populations échantillonnées. La note stade de débourrement correspond à la moyenne observée sur tous les arbres d'une même provenance, au sein de chaque test de provenances.

<i>Groupe</i>	<i>Forêt</i>	<i>Pays</i>	<i>Long.</i>	<i>Lat.</i>	<i>Alt.</i>	<i>Notation de stade de débourrement</i>			
			°cg	°cg	m	Charnie	Vierzon	Vincence	Sillegny
1	Sprakensehl	D	10,6	52,8	115	0,713	1,112	1,015	1,643
1	Gohrde	D	10,9	53,1	90	0,962	1,211	1,016	1,947
1	Luss	D	10,3	52,8	110	1,094	1,419	∅	2,004
1	Johanneskreuz	D	7,8	49,4	460	1,153	∅	1,127	2,137
1	Lapwald	D	10,9	52,2	180	∅	1,536	1,561	2,243
1	Wolfgangz	D	9,1	50,2	160	1,677	2,101	1,528	2,622
1	Cochem	D	7,1	50,1	400	1,862	2,172	2,019	3,22
2	Saint Jean	F	6,7	48,8	227	1,016	1,229	1,001	1,509
2	Romersberg	F	6,7	48,8	220	1,258	1,747	1,383	2,262
2	Still	F	7,3	48,6	688	1,586	1,954	1,606	2,956
2	Haslach	F	7,4	48,6	265	1,997	1,812	1,864	2,88
2	Lembach	F	7,8	48,9	260	2,243	2,278	2,011	3,206
2	Westhoffen	F	7,5	48,6	400	2,767	2,697	2,632	3,213
2	Longchamps	F	5,3	47,3	235	2,763	2,829	2,75	3,546
3	Mölln	D	10,8	53,6	36	0,576	1,184	0,906	1,48
3	Fontainebleau	F	2,7	48,4	80	0,839	0,787	0,709	1,16
3	Blakeney	GB	-2,5	51,8	76	1,227	∅	1,264	1,297
3	Adé	F	0	43,1	425	1,741	1,967	1,569	2,594
3	Neuville en Hez	F	2,3	49,4	70	2,152	1,868	1,931	3,048
3	Obora	CZ	19	48,6	∅	3,169	∅	2,623	3,712
3	Klostermarienber	A	16,6	47,4	310	3,278	3,802	4,154	4,079

divergences dues à la colonisation) sur la différenciation moléculaire de ces populations. Le troisième groupe, quant à lui, inclus un maximum de variabilité pour le caractère de débourrement, mais de ce fait inclus aussi des populations provenant de régions géographiques très éloignées. Cet échantillonnage devrait permettre de structurer les 21 populations, soit par groupes géographiques, soit par classes de précocité du débourrement lors des analyses des niveaux de différenciation. Un sous ensemble de 9 populations a été utilisé pour l'étape de recherche des SNP. Ces populations sont signalées par la couleur violette sur la figure 41. Ce sous échantillon respecte également les deux critères édictés précédemment, à savoir que trois populations sont originaires du Nord Est de la France (Westhoffen, Haslach, Saint Jean), trois autres sont originaires d'Allemagne (Cochem, Johanneskreuz, Sprakensehl) et les trois dernières sont très contrastées pour le débourrement (Klostermarienbergr (A), La Neuville en Hez (F), Mölln (D)). Dans la pratique, quelques dizaines de bourgeons pour 30 arbres de chaque provenance ont été récoltés sur le site de la Petite Charnie et conservés à -80°C pour une extraction d'ADN ultérieure. L'étude phénotypique du stade de débourrement a été réalisée à l'aide d'une échelle de notation exactement équivalente à celle utilisée pour la récolte du matériel lors de l'approche transcriptomique (cf. Figure 20). Une note moyenne prenant en compte l'ensemble des bourgeons de l'arbre est attribuée à chaque individu, ainsi évalué pour le débourrement entre 1993 et 1995 sur chacun des sites de tests. Les notes moyennes obtenues pour chaque population sont présentées dans le tableau 12.

1.1.2. La détection de polymorphisme

L'étape préliminaire aux études d'association est la recherche des SNP au sein des individus issus de populations naturelles. Pour ce faire, 3 individus originaires de chacune des 9 provenances précédemment décrites, ont été utilisés. L'ADN de ces individus a d'abord été extraits en tube Eppendorf 1,5 ml. Les échantillons ont tout d'abord été broyés à l'aide d'azote liquide puis transférés en tube 1,5 ml. Un ml de tampon (NaCl : 1,4M, TrisHCl : 100 mM, EDTA : 20 mM, ATMAAB : 2%, DTT : 1,25%, PVP soluble : 1,67%), préalablement chauffé à 65°C, a été ajouté dans chaque échantillon mis à incuber à 55°C pendant 45mn. Après ajout de 500 µL de dichlorométhane puis centrifugation 10mn à 13 000 g, le surnageant a été transféré dans un nouveau tube. Six cent µL d'isopropanol ont été ajoutés et les échantillons placés 1 heure à -20°C. Le surnageant a été retiré puis les culots séchés et repris

Tableau 13 : Caractéristiques des GC étudiés au niveau nucléotidique.

<i>Nom du gène</i>	<i>Fonction hypothétique</i>	<i>Taille attendue Taille obtenue</i>	<i>Origine</i>	<i>Groupes de liaison</i>	<i>Co-localisation*</i>
<i>CI08A01 GALA</i>	Galactinol synthase	<i>210 bp 505 bp</i>	<i>Banques SSH</i>	Q 3 & C 8	Non
<i>CI07B10 DAG2</i>	Zinc finger protein DAG2	<i>360 bp 360 bp</i>	<i>Banques SSH</i>	Q 8	Non
<i>CI08B04 RASI</i>	Alpha- amylase/subtilisin inhibitor (RASI)	<i>327 bp 327 bp</i>	<i>Banques SSH</i>	Q 9	Oui
<i>GA 3-β- hydroxylase GA3</i>	Synthesis of active Gibberellin	<i>347 bp 652 bp</i>	<i>Cucurbita maxima</i>	Q 5	Oui
<i>CI08C11 YSL1</i>	Metal-nicotianamine transporter YSL1	<i>316 bp 316 bp</i>	<i>Banques SSH</i>	Q 2 & C 1	Oui
<i>GA 20- oxydase GA20</i>	Synthesis of active Gibberellin	<i>409 bp 554 bp</i>	<i>Fagus sylvatica</i>		
<i>Cons58 AUX-REP</i>	Auxin-repressed 12.5 kDa protein	<i>255 bp 447 bp</i>	<i>Banques SSH</i>	Q 5	Oui

* co-localisation avec un QTL de débourrement

dans 80 µL de tampon Tris HCl (10 mM, pH 8). Le lendemain, les échantillons ont été traités à la RNase (2µg) et incubés 30mn à 37°C. L'ADN a été reprécipité par ajout de 250µL d'éthanol 95% et 1/20 de volume de NaCl 5M, puis centrifugé 5mn à 13 000 g. Après avoir éliminé le surnageant, 1 ml d'éthanol 76% a été ajouté et les échantillons centrifugés 5mn à 13 000 g. Enfin, les culots ont été séchés puis repris dans 100 à 200 µL de tampon Tris HCl 10mM. La qualité ainsi que la quantité de chaque échantillon ont été estimées par mesure de la densité optique à 260 et 280nm ainsi que par électrophorèse sur gel d'agarose 1,2%. Les échantillons de qualité médiocre (bandes de dégradation visibles, concentration trop faible, $DO_{260}/DO_{280}<1,6$) ont été extraits de nouveau pour assurer une qualité homogène des ADN des différents individus.

Dans un premier temps, chaque gène candidat sélectionné a été amplifié sur les 27 individus constituant le panel représentatif de l'ensemble de l'aire naturelle, selon les conditions PCR décrites dans le tableau 10. Parmi ces gènes candidats, six étaient issus des analyses des niveaux d'expression au cours du débourrement et le septième a été identifié lors de l'approche de cartographie, à savoir Cons58 qui code potentiellement pour une protéine auxin-repressed. Les informations relatives à ces 7 locus étudiés sont présentées dans le tableau 13. Pour les six premiers GC, les réactions PCR ont été réalisées dans un volume réactionnel de 20 µL contenant 2 µL de Tampon 10X PCR (Invitrogen), x mM de MgCl₂ (cf. tableau 10), 200 µM de chaque dNTP, 0,2 µM de chaque amorce, 4 à 8 ng d'ADN génomique et 0,4 U de *Taq* polymérase (Invitrogen). Les amplifications ont été conduites dans une machine GeneAmp PCR System 2700 (Perkin Elmer) avec les conditions suivantes : 10mn à 94°C, 35 ou 40 cycles de 45s ou 1mn à 94°C, 45s ou 1mn à la température d'hybridation, 1mn à 72°C, et une phase d'extension finale de 10mn à 72°C. Les températures d'hybridation ainsi que la durée des différentes phases pour chaque fragment sont présentées au tableau 10. Pour ce qui concerne Cons58, les conditions ont été les suivantes : 10mn à 94°C, 10 cycles de touchdown de 45s à 94°C, 45s à 58°C avec une température décroissante de 0,5°C à chaque cycle, 1mn à 72°C, suivis de 25 cycles de 45s à 94°C, 45s à 53°C et 1mn à 72°C. Une phase d'élongation finale de 10mn à 72°C a été appliquée. Deux mM de MgCl₂ ont été utilisés dans le mélange réactionnel pour Cons58.

Chaque produit PCR a ensuite été cloné à l'aide du kit TOPO TA cloning (Invitrogen) puis les inserts amplifiés à l'aide des amorces M13 forward et reverse (cf. protocole Chapitre B § 1.2.3. et 1.3.3.). Les clones ont été séquencés à l'aide du kit DYEnamic™ ET Dye Terminator (Amersham Biosciences Inc.) et révélés sur un séquenceur capillaire Megabace 1000 automated DNA sequencer (Amersham Biosciences Inc.). Un à trois clones issus de

Tableau 14 : Règles de priorités pour la sélection des SNP (Tiré de Tabor *et al.*, 2002).

<i>Type de variant</i>	<i>Position</i>	<i>Effet fonctionnel</i>	<i>Fréquence dans le génome</i>	<i>Prédiction du risque relatif sur le phénotype</i>
Non sens	Séquence codante	Arrêt prématuré de la séquence d'acides aminés	Très faible	Très élevé
Antisens / Non synonyme (Non conservatif)	Séquence codante	Changement de la protéine en une autre aux propriétés différentes	Faible	Modéré à très fort, dépend de la position
Antisens / Non synonyme (Conservatif)	Séquence codante	Changement de la protéine en une autre aux propriétés similaires	Faible	Faible à très élevé, dépend de la position
Insertions/délétions (hors cadre)	Séquence codante	Modifie le cadre de lecture dans la région codant pour la protéine, avec des conséquences très négatives sur la protéine	Faible	Très élevé, dépend de la position
Insertion/délétions (dans le cadre)	Codant ou non codant	Change la séquence d'acides aminés	Faible	Faible à très élevé
Sens/Synonyme	Séquence codante	Ne change pas la séquence des acides aminés dans la protéine mais peut altérer l'épissage	Modéré	Faible à élevé
Promoteur/région régulatrice	Promoteur, 5' UTR, 3' UTR	N'affecte pas la séquence protéique mais peut modifier le niveau, la position ou le timing dans l'expression du gène	Faible à modéré	Faible à élevé
Site d'épissage/liaison intron-exon	Extrémité de l'exon	Peut modifier le motif d'épissage ou l'efficacité des introns	Faible	Faible à élevé
Intronique	Dans l'intron	Pas de fonction connue, mais peut affecter l'expression ou la stabilité de l'ARNm	Modéré	Très faible
Intergénique	Régions intergéniques	Pas de fonction connue, mais peut affecter l'expression	Elevé	Très faible

chaque fragment ont été séquencés et la séquence de meilleure qualité a été conservée. Les 27 séquences ont été alignées puis les sites polymorphes potentiels identifiés à l'aide d'une suite de programmes d'analyses de séquence d'ADN décrite dans l'article de Le Dantec *et al.* (2004). Brièvement, le programme *Phred* (Ewing et Green, 1998; Ewing *et al.*, 1998) est d'abord utilisé pour identifier chacune des bases de la séquence et lui assigner un score de qualité. Ensuite, *Phrap* (<http://www.phrap.org>) permet d'assembler l'ensemble des séquences en un contig unique. Enfin, *PolyBayes* (Marth *et al.*, 1999), à l'aide de méthodes d'inférence bayésienne, détermine pour chacun des sites au sein de l'alignement, une probabilité pour le site d'être polymorphe. Chacun des sites polymorphes, quelle que soit la probabilité associée, a été inspecté visuellement afin de déterminer la présence ou non d'un polymorphisme de type SNP ou insertions/délétions. Les alignements ont été visualisés à l'aide du logiciel *Consed* (Gordon *et al.*, 1998) sous environnement Unix. Les 7 locus ont ainsi été traités et l'ensemble des polymorphismes de type SNP ou insertions/délétions présents dans la séquence a été identifié.

1.1.3. Le génotypage des SNP

Deux des sept locus précédemment décrits ont été l'objet d'une étude de diversité sur certains sites SNP dans le cadre du travail de Master d'Alexandre Jauffrès (2004), réalisé au sein de l'UMR Biogeco. Il s'agit du GC C108A01 (GALA) codant potentiellement pour l'enzyme Galactinol synthase et du GC C107B10 (DAG2) codant potentiellement pour le facteur de transcription DAG2. Après le choix des locus, l'étape suivante a consisté en la sélection des sites SNP à étudier. A cet effet, les informations concernant la position ainsi que le type de variants alléliques rencontrés peuvent servir à établir une liste des SNP par ordre de priorité (Tabor *et al.*, 2002). Le tableau 14 reprend les critères de priorité établis par Tabor *et al.* (2002), qui nous serviront dans le choix des sites SNP étudiés. Les niveaux de différenciation entre populations mesurés au niveau de ces sites polymorphes sont également une information à prendre en considération dans ce choix.

La technique de génotypage des sites SNP utilisée est basée sur la méthode d'extension d'amorce. Une amorce spécifique, dont l'extrémité 3' se termine juste avant le site polymorphe, est utilisée dans une réaction d'amplification. La base ajoutée est un didéoxynucléotide marqué à l'aide d'un fluorochrome, chaque base étant marquée par un fluorochrome différent. Une fois l'incorporation des bases fluorescentes réalisée, la lecture

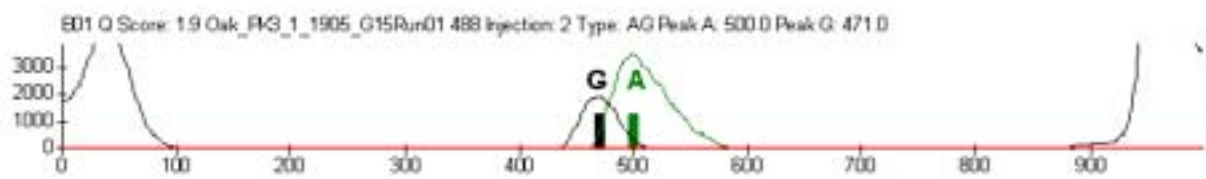


Figure 42 : Lecture d'un polymorphisme de type A/G au site SNP GalaLoc455, à l'aide du logiciel SNP Profiler.

peut se faire soit à l'aide d'un séquenceur capillaire soit à l'aide d'un appareil mettant en jeu la fluorescence polarisée. Une douzaine d'amorces a ainsi été testée mais seulement trois ont pu être mises au point, à savoir une amplifiant un site SNP de DAG2 en position 25 (5' CAG CCT CAC TCA ACC AAG ATA 3') et deux amplifiant des sites SNP, en positions 395 et 455, au sein du GC codant pour la Galactinol synthase (GALA395 : 5' GAA CAT RTT CAA AAA GTC CTG AAA 3' ; GALA455 : 5' TAA CAT GGC AAG CAC AAG ATT 3'). Ces trois sites SNP ont donc été génotypés chez un peu plus de 500 individus issus de 21 populations naturelles.

Sur le plan expérimental, les ADN génomiques des 30 individus récoltés par provenance ont été extraits en suivant le protocole décrit par Shagaï-Marooof *et al.* (1984) légèrement modifié. Après évaluation de la qualité et de la quantité des ADN extraits, 24 individus par provenances ont été retenus en prenant soin d'éliminer les échantillons de mauvaise qualité (cf. § 1.1.2.). Un total de 504 individus a ainsi été arrangé en plaques 96 puits. Chaque individu a, dans un premier temps, été amplifié avec les amorces correspondant aux fragments GALA et DAG2, et conformément au protocole décrit dans le tableau 10. Les produits PCR obtenus ont été vérifiés sur gel d'agarose 2% puis purifiés, en suivant les instructions données par le fournisseur du kit de génotypage MegaBACE SnuPe Genotyping (Amersham Biosciences Inc.). L'amplification des sites SNP a ensuite été réalisée à l'aide du même kit selon les instructions fournies. Après purification des produits issus de la réaction de génotypage, la révélation des SNP a été réalisée sur séquenceur MegaBACE (Amersham Biosciences Inc.). Dans ce cas, la séparation des produits amplifiés est effectuée par la technique d'électrophorèse capillaire. Chacune des quatre bases marquées est lue à l'aide d'un laser de longueur d'onde spécifique du dNTP considéré. Les profils obtenus sont révélés à l'aide du logiciel SNP Profiler, fournis par le fabricant (Amersham Biosciences Inc.). Les allèles sont ainsi visualisés grâce à des pics colorés, localisés et encadrés par un marqueur d'injection multiple (Figure 42).

1.2. Outils d'analyse de la diversité et des études d'association

Les analyses de diversité ont porté d'une part sur les données issues du séquençage des fragments et, d'autre part, sur les données de génotypage aux sites SNP. Dans toute la suite, les GC seront nommés relativement à leur fonction hypothétique, tels que décrits dans le tableau 13.

1.2.1. Analyse des séquences

Après inspection visuelle de chacune des séquences issues de l'alignement sous le logiciel *Consed* et validation de l'ensemble des sites polymorphes, les données pour chaque locus ont été transférées et mises en forme dans un format compatible avec les logiciels utilisés par la suite, grâce au logiciel ProSeq v 2.91 (Filatov, 2002). Au final, 26 ou 27 séquences d'individus issus de 9 provenances européennes étaient ainsi disponibles pour chaque fragment analysé, sauf le fragment GA20 pour lequel 21 séquences ont été obtenues. L'utilisation de produit de clonage a ainsi permis la définition directe des haplotypes présents (combinaisons multi locus des polymorphismes) sans avoir recours à un algorithme EM (Long *et al.*, 1995) pour déterminer, entre autres paramètres, les fréquences haplotypiques.

Les paramètres descriptifs tels que le nombre de polymorphismes SNP, d'insertions délétions (INDELS), de mutations synonymes (S) ou non synonymes (NS) ont été déterminés à l'aide du logiciel DnaSP v 4.00.5 (Rozas *et al.*, 2003, <http://www.ub.es/dnasp>). La diversité nucléotidique a été estimée par les paramètres θ_w (Watterson, 1975) et π (Nei, 1987). Le paramètre π représente le nombre moyen de différences nucléotidiques observées, par site, entre deux séquences, et son estimateur est calculé comme suit : $\hat{\pi} = \frac{n}{n-1} \sum_{ij} \hat{x}_i \hat{x}_j \pi_{ij}$, où n est le nombre de séquences analysées, $\hat{x}_{i(j)}$ est la fréquence dans la population du $i^{\text{ème}}$ ($j^{\text{ème}}$) variant d'ADN et π_{ij} la proportion des différents nucléotides entre le $i^{\text{ème}}$ et le $j^{\text{ème}}$ variant d'ADN. θ_w représente le nombre moyen de sites polymorphes par nucléotide, et son estimateur est calculé comme suit : $\hat{\theta}_w = \frac{S}{n} \left/ \sum_{i=1}^{m-1} \frac{1}{i} \right.$, où S et n sont, respectivement, le nombre de sites polymorphes et le nombre de nucléotides dans la séquence, et m le nombre d'haplotypes échantillonnés.

Ces deux paramètres ont également été calculés dans DnaSP v 4.00.5 (Rozas *et al.*, 2003), sans considérer les INDELS dans les calculs de diversité. Le paramètre π a été estimé à trois niveaux au sein de la séquence : i/ sur l'ensemble de la séquence, ii/ dans les régions non codantes (introns, 3' et 5' UTR), iii/ dans les régions codantes subdivisées en régions synonymes (S) et non synonymes (NS). Le nombre d'haplotypes ainsi que la diversité haplotypique ont également été calculés à l'aide du logiciel DnaSP (Rozas *et al.*, 2003). Enfin, les indices de différenciation, au niveau de la séquence et pour chaque site SNP, ont été estimés à l'aide du paramètre F_{ST} selon la méthode de Weir et Cockerham (1984) au sein du logiciel GENEPOP (Raymond et Rousset, 1995). Deux niveaux hiérarchiques ont été considérés : i/ un premier niveau en comparant les populations par provenances géographiques, ii/ un second niveau en comparant les populations par groupes de précocité de débourrement.

Pour estimer si les fragments considérés suivaient ou non le modèle neutre de l'évolution moléculaire (Kimura, 1983), nous avons appliqué le test statistique d'écart à la neutralité du D de Tajima (1989). La significativité de la valeur de D a été testée à l'aide de simulations de coalescence, disponibles dans DnaSP (Rozas *et al.*, 2003). Les simulations ont été réalisées avec un nombre donné de sites en ségrégation pour chaque locus, égal au nombre de sites effectivement observés sur l'échantillon, et sans recombinaison, ce qui rend le test conservatif. Malgré le nombre d'individus limités mais sachant que les niveaux de différenciation entre ces derniers sont relativement faibles (cf. Tableau 18), le déséquilibre de liaison a été estimé au sein des locus entre les sites polymorphes (singletons exclus). Les locus DAG2, GA3 et GA20 ont été retirés de l'analyse comme ils ne présentaient respectivement que deux, trois et quatre sites SNP. L'ensemble des séquences disponibles a été utilisé et le calcul du coefficient de corrélation des fréquences alléliques, entre paires de sites polymorphes, noté R (Hill et Robertson, 1968), a été effectué à l'aide de DnaSP. La significativité des associations détectées a été estimée grâce à des tests de Fisher exact en utilisant la correction de Bonferroni.

1.2.2. Diversité nucléotidique et différenciation entre populations à certains sites SNP

Les analyses décrites en suivant ont porté sur un ensemble des 504 individus récoltés parmi 21 provenances (cf. § 1.1.1.). Nous utiliserons dans cette partie des indices de la

diversité basés sur les fréquences alléliques. L'autre type d'indice est en effet basé sur la richesse allélique ce qui ne convient pas aux marqueurs SNP qui sont bi alléliques (richesse allélique attendue = 1 ou 2). L'une des mesures les plus communes de diversité est l'index de diversité écologique de Simpson (1949) dont dérivent un certain nombre de mesures. La plus largement utilisée est la diversité génétique de Nei (1973). Pour une population donnée, elle est définie comme la probabilité pour que deux allèles tirés au hasard avec remise dans la population sont différents. Si à un locus donné, on trouve m allèles, alors : $H_i = 1 - \sum_{i=1}^m p_i^2$,

où p_i est la fréquence de l'allèle i au locus étudié. Comme on s'intéresse dans notre étude à plusieurs populations, on calculera la diversité génétique moyenne à l'intérieur des sous populations (H_S est la moyenne des H_i sur toutes les sous populations) et la diversité génétique totale (H_T est la valeur de H_i pour la population fictive englobant toutes les sous populations). Nous estimerons également l'hétérozygotie observée, notée H_0 , qui se définit comme étant le nombre d'individus hétérozygotes dans la population divisé par le nombre total d'individus. La structuration de cette diversité intra population peut être estimée par le paramètre F_{is} , plus communément appelé coefficient de consanguinité et définit comme suit : $F_{is} = 1 - H_0 / H_S$. Wright (1969) l'a initialement défini sous le nom de coefficient de consanguinité mais la consanguinité n'est pas le seul facteur induisant un écart à l'équilibre de Hardy-Weinberg (1908). En effet, l'homogamie, la dérive, la sélection et la différenciation génétique locale y participent également.

L'analyse de la diversité est généralement faite dans plusieurs populations, comme dans notre étude. Deux niveaux hiérarchiques peuvent alors être considérés : la diversité entre populations, généralement appelée différenciation, et la diversité intra population, que nous avons évoqué précédemment. On peut montrer (Nei, 1973) que la différenciation entre population s'exprime aussi comme la différence des diversités entre deux niveaux hiérarchiques : $d_{ST} = H_T - H_S$. La différenciation est alors exprimée sous forme relative (G_{ST}), en la rapportant à la diversité totale : $G_{ST} = d_{ST} / H_T$, d'où $G_{ST} = 1 - H_S / H_T$. Il est intéressant de noter que la différenciation, contrairement à la diversité, est indépendante de l'unité de mesure et que les comparaisons entre marqueurs sont donc possibles. Il est à noter que l'ensemble de ces paramètres est calculé par le biais d'estimateurs non biaisés, c'est-à-dire dont la variance est minimale. Les paramètres H_S , H_T , H_0 et G_{ST} ont ainsi été estimés à l'aide du logiciel Fstat v 2.9.3 (Goudet, 2001). Le paramètre F_{is} a été estimé selon la méthode de Weir et Cockerham (1984) à l'aide du logiciel GENEPOP (Raymond et Rousset, 1995).

1.2.3. Etudes d'association entre variants alléliques et notes de débourement

Il faut rappeler ici que l'objectif de ces études est de valider l'importance des gènes candidats sélectionnés par l'étude des corrélations entre les variants alléliques et la variation phénotypique observée en populations naturelles. Pour évaluer cette corrélation à chaque site SNP étudié, nous avons procédé à une analyse à deux niveaux distincts :

- d'une part, la corrélation a été testée au niveau individuel, par comparaison du génotype de chaque individu au site nucléotidique considéré avec la note de débourement mesurée sur le site de test de la Petite Charnie. Deux approches ont été employées pour cela. La première a consisté à réaliser une comparaison de moyennes des stades de débourement pour les trois classes génotypiques (11, 12, 22) en testant la significativité des différences observées (test t de Student) entre chaque couple de classes génotypiques. La seconde approche a consisté à analyser les interactions possibles (effets d'association) entre les trois locus. Pour ce faire, nous avons réalisé une analyse de variance en modèle linéaire en testant l'effet individuel de chaque locus, les interactions deux à deux et l'effet d'interaction des trois locus. Le modèle utilisé peut s'écrire de la manière suivante :

$$Y_{ijkl} = \mu + \text{GALA455}_i + \text{GALA395}_j + \text{DAG25}_k + \text{GALA455} * \text{GALA395}_{ij} + \text{GALA455} * \text{DAG25}_{ik} + \text{GALA395} * \text{DAG25}_{jk} + \text{GALA455} * \text{GALA395} * \text{DAG25}_{ijk} + e_{ijkl}$$

avec Y_{ijkl} = note de débourement du l^{ème} individu,

μ = note moyenne de débourement sur l'ensemble des individus,

GALA455_i = effet du locus GALA455 (i=1-3),

GALA395_j = effet du locus GALA395 (j=1-3),

DAG25_k = effet du locus DAG25 (k=1-3),

$\text{GALA455} * \text{GALA395}_{ij}$ = effet d'interaction entre les locus GALA455 et GALA395,

$\text{GALA455} * \text{DAG25}_{ik}$ = effet d'interaction entre les locus GALA455 et DAG25,

$\text{GALA395} * \text{DAG25}_{jk}$ = effet d'interaction entre les locus GALA395 et DAG25,

$\text{GALA455} * \text{GALA395} * \text{DAG25}_{ijk}$ = effet d'interaction entre les trois locus,

e_{ijkl} = effet résiduel non pris en compte par les effets précédents ;

- d'autre part, la corrélation a été testée au niveau populationnel. Dans chaque plantation comparative, nous avons calculé la corrélation entre la note moyenne de débourrement de tous les arbres d'une provenance (240 individus) et la fréquence allélique de cette même provenance au niveau du site SNP.

Ces différentes analyses ont été réalisées à l'aide du logiciel SAS (Statistical Analysis System version 6.12, SAS Institute, Cary, NC).

Tableau 15 : Longueur des régions analysées et nombre de sites polymorphes.

<i>Locus</i>	Codant	Non-codant	SNP	Singletons	INDELS
<i>GALA</i>	210	295	17	8	28 pb 1 pb
<i>RASI</i>	327	0	10	13	3 pb
<i>AUX-REP</i>	259	188	7	12	1 pb (x2) 2 pb
<i>YSL1</i>	316	0	8	4	1 pb (x2)
<i>DAG2</i>	360	0	2	4	1 pb (x2)
<i>GA3</i>	347	305	3	9	6 pb
<i>GA20</i>	425	129	4	11	27 pb
<i>Total</i>	2243	917	51	61	12 INDELS

Tableau 16 : Fréquences et nombre de singletons potentiels résultant d'erreur de l'enzyme *Taq* polymérase.

<i>Locus</i>	F ₁ *	Nb Singl.	F ₂ *	Nb Singl.
<i>GALA</i>	0,152	1,2	0,303	2,4
<i>RASI</i>	0,104	1,4	0,196	2,6
<i>AUX-REP</i>	0,141	1,7	0,268	3,2
<i>YSL1</i>	0,101	0,4	0,189	0,8
<i>DAG2</i>	0,115	0,5	0,216	0,9
<i>GA3</i>	0,111	1,0	0,391	3,5
<i>GA20</i>	0,136	1,5	0,332	3,7

*F₁ : fréquence d'erreurs de *Taq*, sur les singletons détectés, lors de la PCR sur colonies après clonage ; F₂ : fréquence d'erreurs de *Taq*, sur les singletons détectés, cumulée lors de l'amplification du fragment et de la PCR sur colonies.

② Résultats et discussion

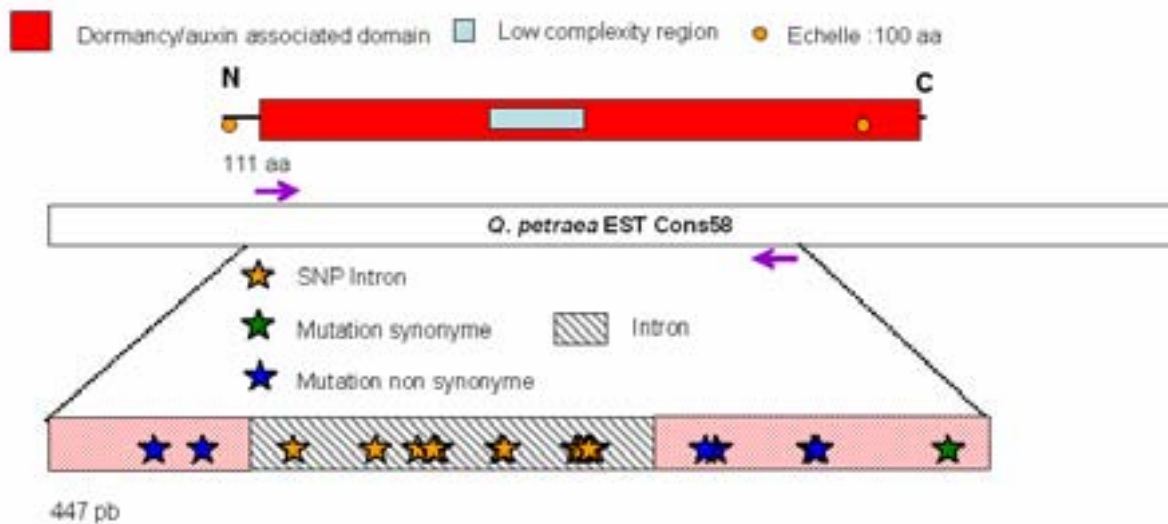
2.1. Le polymorphisme révélé au sein des gènes candidats

2.1.1. De la difficulté de mettre en évidence des polymorphismes au sein des séquences

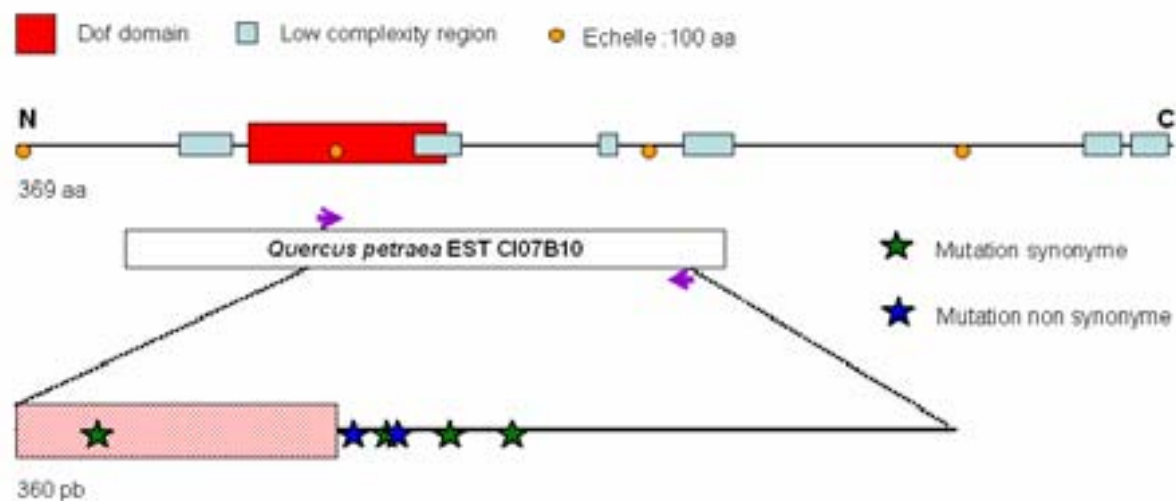
Dans un premier temps, il est intéressant d'estimer la capacité de notre échantillon à détecter du polymorphisme au sein des gènes candidats étudiés, autrement dit la puissance de détection des sites polymorphes. La probabilité P de détecter deux allèles à un locus SNP dépend de trois paramètres que sont : i/ le nombre de gamètes échantillonnés (N), ii/ la fréquence d'un allèle rare dans la population (p), et iii/ l'organisation de la diversité génétique. Si l'on suppose l'absence de différenciation entre les populations, alors $P = 1 - (1 - p)^N$. Dans le cas de notre étude, 26 gamètes en moyenne ont été séquencés, la probabilité de détection d'un allèle présent à la fréquence de 10% était donc de 94%. Inversement, il est possible de calculer la fréquence minimale qu'un allèle doit présenter pour être détecté dans 90% des cas. La fréquence doit être au minimum de 8,4% pour que la mutation soit détectée dans 90% des cas, soit 2,2 individus. La mutation doit donc être portée par trois individus au moins. L'échantillon apparaît donc satisfaisant pour la détection de sites polymorphes, y compris ceux résultant d'un événement de mutation rare.

La seconde difficulté à révéler des sites polymorphes « vrais » réside dans des aspects liés aux techniques employées, et en particulier l'étape de clonage. En effet, une tendance à l'excès de singletons apparaît dans la majorité des locus étudiés (Tableau 15). Les séquences ont été obtenues à partir de produits PCR clonés et ce résultat peut résulter d'erreurs de la *Taq* polymérase pendant la phase d'amplification du fragment ou ensuite lors de la PCR sur les fragments clonés. Cependant, connaissant la fréquence d'erreur de l'enzyme *Taq* polymérase de 8×10^{-6} /pb/duplication (Cline *et al.*, 1996), les erreurs de PCR seraient à l'origine de 1 à 4 singletons par fragment (Tableau 16). Ces erreurs de l'enzyme polymérase ne semblent donc pas expliquer l'excès de singletons observé. Par ailleurs, chez le peuplier (*Populus tremula* L.), Ingvarsson (2005) a également trouvé un excès de singletons dans les 5 locus de son étude, alors même que 5 à 10 clones ont été séquencés pour chaque individu afin de passer outre les erreurs dues à l'enzyme polymérase. Il est également indispensable de travailler sur des produits de séquençage de très bonne qualité afin de ne pas « polluer » la lecture des différentes bases par des erreurs de séquençage. La suite de programmes précédemment

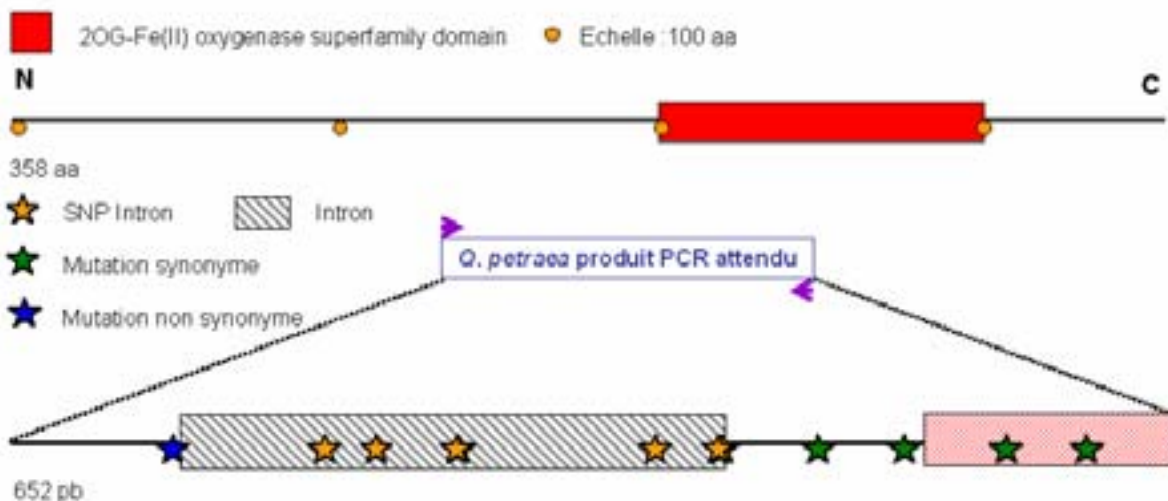
Auxin-repressed 12.5 kDa protein *Fragaria ananassa* (Q05349)
 e-value : 9×10^{-26}



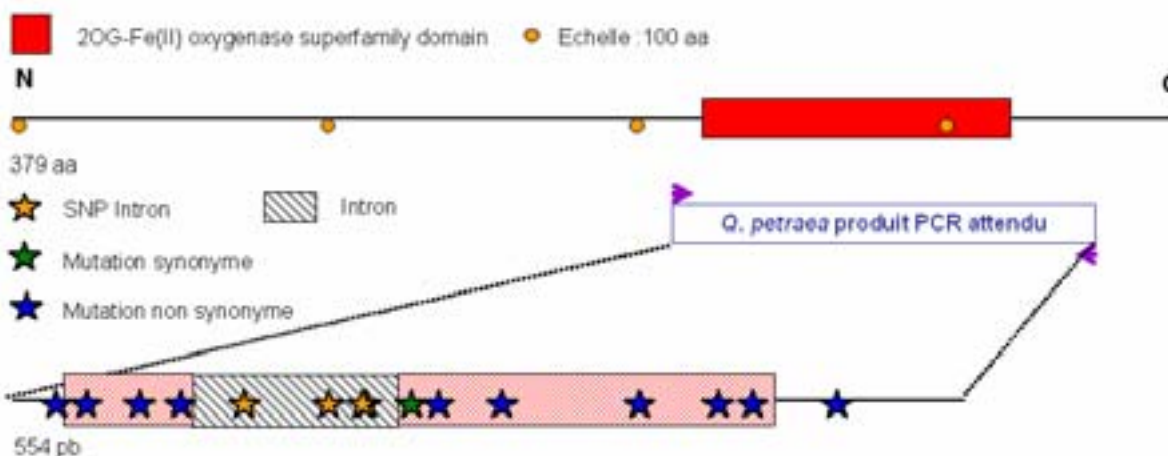
Dof zinc finger protein DOF2.5 (DAG2) *A. thaliana* (Q9ZPY0)
 e-value : 4×10^{-43}



Gibberellin 3- β -hydroxylase *Cucurbita maxima* (Q9M4P2)
 e-value : 2×10^{-68}

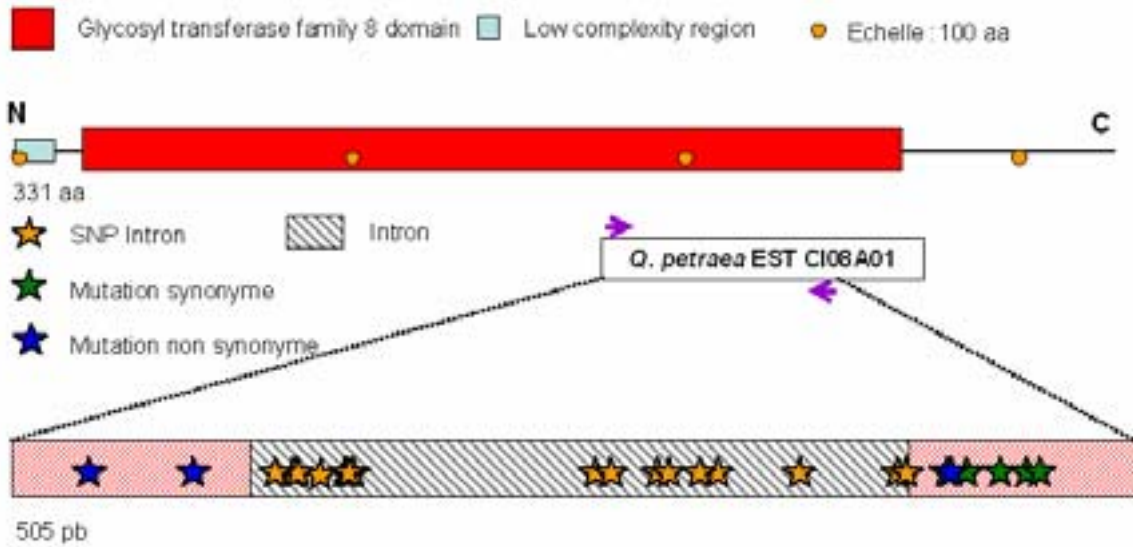


Gibberellin 20-oxylase 1 *Fagus sylvatica* (Q8VXL2)
 e-value : 6×10^{-65}



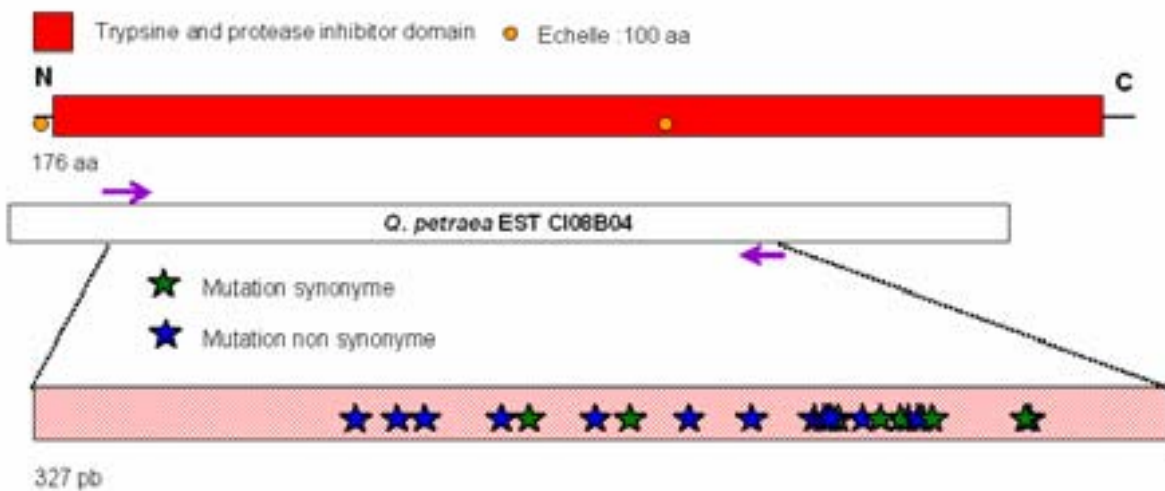
Galactinol synthase *Cucumis melo* (Q8S9H1)

e-value : 2×10^{-39}



Alpha-amylase/subtilisin inhibitor *Oryza sativa* (P29421)

e-value : 5×10^{-14}



Metal-nicotianamine transporter YSL1 *A. thaliana* (Q6R3L0)
e-value : 1×10^{-58}

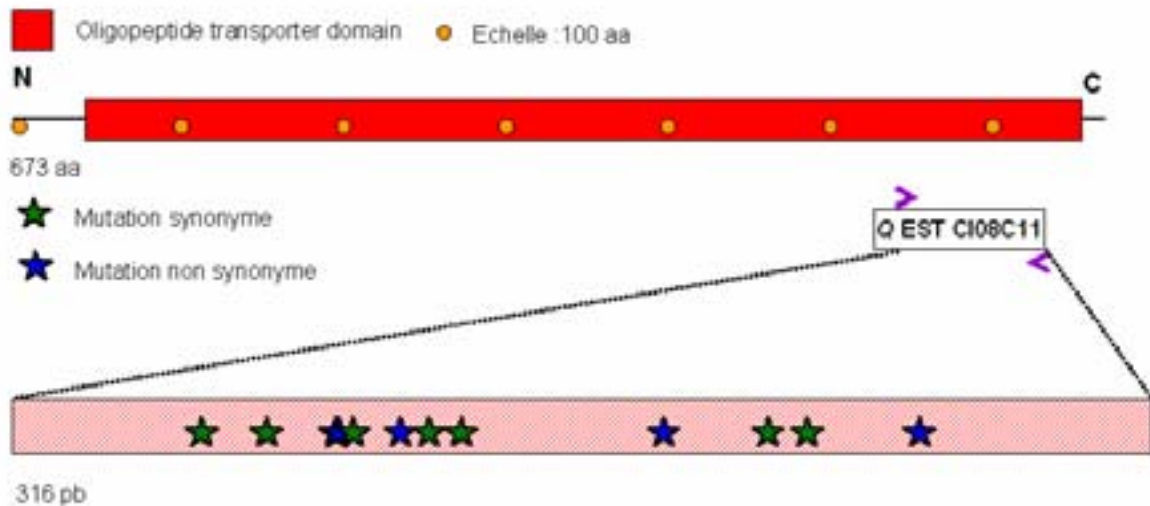


Figure 43 : Représentation schématique des sites SNP identifiés pour chaque fragment d'ADN génomique étudié (bas de la figure). L'EST de départ (ou le produit PCR attendu sur la base des amorces dessinés) est également représentée (milieu de la figure) à la même échelle que la protéine potentielle correspondante, issue de l'annotation bioinformatique ou sur laquelle les amorces ont été dessinées. Les différents sites actifs ou de faible complexité identifiés sur la protéine sont projetés sur la séquence d'ADN génomique afin d'apprécier l'effet possible des mutations sur la protéine. La valeur de e-value reportée en haut de la figure correspond à celle du BLAST effectué avec la séquence d'ADN génomique obtenue.

décrite pour aligner puis identifier les sites polymorphes permet de remédier à cela puisqu'un score de qualité est attribué à chaque base. Les bases de mauvaise qualité ont ainsi été retirées de l'analyse.

2.1.2. Les sites polymorphes identifiés au sein des GC

Pour chacun des locus, nous avons obtenu et analysé les séquences de 26 ou 27 individus du panel représentatif de l'aire de distribution naturelle, à l'exception du locus GA20, pour lequel 21 séquences ont été obtenues, ce qui représente en moyenne 26 individus par locus. La longueur des régions analysées variait entre 316 et 652 pb (INDELS inclus, cf. tableau 13). Quatre des 7 locus analysés contenaient des régions introniques non codantes, représentant entre 129 pb (GA20) et 305 pb (GA3). Au total, 3,16 kb ont donc été étudiés dont 2,24 kb dans des régions codantes et un peu plus de 0,9 kb dans des régions non codantes (introns) (cf. tableau 15). Un total de 112 sites SNP, dont 61 singletons¹³ a été détecté, ainsi que 12 INDELS couvrant de 1 à 29 bases. Deux des SNP détectés sont tri nucléotidiques dans notre échantillon. Sur l'ensemble des 7 locus, un site SNP est donc détecté en moyenne toutes les 28 paires de base. Ce résultat est variable selon les locus, avec un maximum mesuré de 1 SNP/14 pb pour le locus RASI et un minimum de 1 SNP/60 pb pour le locus DAG2. Nous pouvons étudier plus en détail, locus par locus, la position des différents sites SNP. La figure 43 propose une représentation schématique de chacun des locus amplifié indiquant les régions codantes (domaine actif ou non) et non codantes déduites de la séquence de la protéine hypothétique correspondante.

Nous pouvons tout d'abord remarquer que les régions amplifiées se situent préférentiellement du côté de la région 3' du gène. C'est le cas des locus amplifiés sur GA3, GA20, GALA et YSL1. A l'inverse, le locus amplifié sur DAG2 se situe du côté 5'. Les locus amplifiés sur RASI et AUX-REP couvrent quasiment la totalité du gène hypothétique. Dans ces deux derniers cas, les valeurs de probabilités associées à la procédure BLAST sont les deux plus faibles observées, respectivement 5×10^{-14} et 9×10^{-26} , et la séquence codant pour la protéine hypothétique est entièrement contenue dans l'EST. Il est donc fort probable que ces deux EST codent pour une autre protéine que celle obtenue par recherche de similarités de séquences. Cependant, nous discuterons des SNP associés à ces deux locus dans ce qui suit en suivant ce cadre de lecture. Il n'est pas surprenant de voir que la majorité des EST se situe, ou

¹³ Singleton : mutation présente chez un seul individu de l'échantillon.

Tableau 17 : Patron de variation nucléotidique des 7 locus étudiés.

Locus	Provenances	N	Polymorphisme SNP													
			Total				Non codant		Codant				Nb haplotypes	Hd		
			S	θ_w	π	D (Tajima) ¹	S	π	Total		Synonyme				N.synonyme	
						S	π	S	π	S	π	S	π	S	π	
GALA	Total	26	25	0,01382	0,01502	0,31757	17	0,01755	8	0,01184	5	0,03905	3	0,00420	18	0,945
	France	8	19	0,01543	0,01647	-	14	0,02022	5	0,01173	4	0,04162	1	0,00333	7	0,964
	Allemagne	9	16	0,01242	0,01184	-	11	0,01452	5	0,00847	4	0,02998	1	0,00242	7	0,917
	Contrasté	8	17	0,01380	0,01624	-	10	0,01792	7	0,01412	4	0,04143	3	0,00647	8	1,000
RASI	Total	26	23	0,01860	0,01024	-1,63371**	-	-	23	0,01024	9	0,01583	14	0,00853	15	0,908
	France	9	8	0,00908	0,00660	-	-	-	8	0,00660	3	0,01036	5	0,00546	5	0,806
	Allemagne	8	17	0,02005	0,01606	-	-	-	17	0,01606	6	0,02047	11	0,01480	6	0,893
	Contrasté	8	10	0,01179	0,00961	-	-	-	10	0,00961	5	0,01994	5	0,00630	7	0,964
AUX-REP	Total	27	19	0,01115	0,00569	-1,83479**	12	0,01061	7	0,00204	1	0,00124	6	0,00235	16	0,863
	France	9	8	0,00663	0,00438	-	4	0,00556	4	0,00350	0	-	4	0,00470	7	0,917
	Allemagne	9	6	0,00497	0,00526	-	5	0,01111	1	0,00087	1	0,00371	0	-	5	0,806
	Contrasté	8	10	0,00873	0,00606	-	8	0,01159	2	0,00197	0	-	2	0,00264	6	0,893
YSL1	Total	27	12	0,00995	0,01039	-0,11937	-	-	12	0,01039	7	0,02903	5	0,00419	13	0,937
	France	8	8	0,00986	0,01312	-	-	-	8	0,01312	6	0,03869	2	0,00459	5	0,857
	Allemagne	9	8	0,00940	0,00976	-	-	-	8	0,00976	4	0,02555	4	0,00452	7	0,944
	Contrasté	9	8	0,00940	0,00958	-	-	-	8	0,00958	5	0,02558	3	0,00428	6	0,917
DAG2	Total	27	6	0,00451	0,00251	-1,20968	-	-	6	0,00251	4	0,00814	2	0,00078	7	0,664
	France	9	3	0,00308	0,00279	-	-	-	3	0,00279	2	0,00917	1	0,00083	4	0,750
	Allemagne	9	2	0,00205	0,00217	-	-	-	2	0,00217	2	0,00906	0	-	3	0,639
	Contrasté	9	3	0,00308	0,00233	-	-	-	3	0,00233	2	0,00524	1	0,00146	4	0,694
GA3	Total	27	12	0,00482	0,00228	-1,75622**	7	0,00326	5	0,00145	4	0,00565	1	0,00027	13	0,860
	France	9	7	0,00399	0,00292	-	4	0,00409	3	0,00192	3	0,00881	0	-	7	0,917
	Allemagne	9	3	0,00171	0,00155	-	2	0,00204	1	0,00112	1	0,00514	0	-	4	0,778
	Contrasté	8	5	0,00299	0,00221	-	4	0,00394	1	0,00072	0	-	1	0,00093	6	0,929
GA20	Total	21	15	0,00793	0,00438	-1,63056**	4	0,01001	11	0,00277	1	0,00102	10	0,00302	14	0,895
	France	8	7	0,00513	0,00482	-	3	0,01160	4	0,00288	1	0,00268	3	0,00298	6	0,893
	Allemagne	6	6	0,00499	0,00455	-	2	0,00904	4	0,00326	0	-	4	0,00428	6	1,000
	Contrasté	7	6	0,00465	0,00361	-	4	0,00888	2	0,00210	0	-	2	0,00183	5	0,857
Total		181	112	0,01011	0,00722		40	0,01036	72	0,00589	31	0,01428	41	0,00333	13,7	0,867

¹ : ** test significatif à 5%.

couvre, la partie 3' du gène correspondant dans la mesure où les banques d'ADNc ont été construites après isolation des ARNm par la queue polyA, située en 3'. La deuxième observation est que toute ou partie de la séquence codant hypothétiquement pour les sites actifs des protéines est présente dans les régions amplifiées. Ce résultat est prometteur si l'on se réfère aux règles de priorité établies par Tabor *et al.* (2002) résumées dans le Tableau 14. En effet, les mutations non synonymes apparaissant au sein de la séquence codant pour le site actif de la protéine sont très intéressantes à considérer puisqu'elles ont un impact très fort sur la séquence et donc potentiellement sur l'activité de la protéine. Or, les locus AUX-REP, GA20, GALA, RASI et YSL1 contiennent des mutations non synonymes dans la partie de la séquence codant potentiellement pour le site actif de la protéine associée.

2.1.3. Diversité nucléotidique

Les calculs de diversité qui suivent ont donc été effectués en considérant la totalité des sites SNP détectés, singletons inclus, les INDELS ayant été retirés de l'analyse. Le nombre de sites polymorphes (S), les niveaux de diversité (π et θ_w), le nombre d'haplotypes, ainsi que la diversité haplotypique (Hd) estimés sur chacun des 7 locus sont présentés dans le tableau 17. S et π sont calculés sur la totalité de la séquence, au sein des régions codantes et non codantes, en distinguant dans ce dernier cas, les régions synonymes et les régions non synonymes. Par ailleurs, les résultats sont présentés au niveau de l'ensemble des séquences mais également par groupes géographiques précédemment définis, à savoir le groupe allemand (G1), le groupe français (G2) et le groupe contrasté (G3). Les niveaux de polymorphisme sont en règle générale assez élevés puisque la diversité moyenne sur les 7 locus est de $\pi = 7,22 \times 10^{-3}$, bien que cette valeur varie jusqu'à 7 fois entre locus, de $2,28 \times 10^{-3}$ pour le locus GA3 à $15,02 \times 10^{-3}$ pour le locus GALA. Cette valeur moyenne de diversité est proche de celles reportée sur maïs ($8,71 \times 10^{-3}$) (pour une revue, voir White et Doebley, 1999) et sur sapin de Douglas ($8,53 \times 10^{-3}$) (Neale et Savolainen, 2004). La plus forte valeur reportée à ce jour chez une espèce forestière concerne le peuplier (Ingvarsson, 2005), avec $\pi = 11,1 \times 10^{-3}$ sur un ensemble de 5 locus étudiés parmi quatre populations (2 suédoises, 1 française, 1 autrichienne). *A contrario*, des espèces telles que les pins apparaissent beaucoup moins diversifiées, avec des valeurs de $\pi = 3,98 \times 10^{-3}$ chez *Pinus taeda* (Brown *et al.*, 2004) et $\pi = 1,4 \times 10^{-3}$ chez *Pinus sylvestris* (Dvornyk *et al.*, 2002). Globalement, la diversité dans les

régions non codantes ($10,36 \times 10^{-3}$) est 2 fois plus élevée que dans les régions codantes ($5,89 \times 10^{-3}$) et ce résultat se vérifie sur les 4 gènes contenant des régions non codantes, le ratio π -codant/ π -noncodant variant de 1,5 sur le gène GALA à 5,2 sur le locus AUX-REP. Dans les régions codantes, la diversité des zones synonymes est en moyenne 4 fois plus élevée que la diversité des zones non synonymes ($14,28 \times 10^{-3} > 3,33 \times 10^{-3}$). Cependant, la tendance est inverse sur les locus AUX-REP et GA20 puisque la diversité des régions non synonymes est plus forte que celle des régions synonymes, respectivement $2,35 \times 10^{-3} > 1,24 \times 10^{-3}$ et $3,02 \times 10^{-3} > 1,02 \times 10^{-3}$. La diversité haplotypique moyenne a été estimée à 0,867 ce qui correspond à 13,7 haplotypes en moyenne sur chacun des locus. Pour chaque locus, la diversité totale estimée sur l'ensemble des séquences est à peu près équivalente à la diversité estimée dans chacun des groupes de provenances. Les différences reportées entre diversité des zones codantes et des zones non codantes, tout comme entre régions synonymes et non synonymes, sont également respectées au sein des groupes de provenances.

Tableau 18 : Niveaux de différenciation (F_{ST}) entre populations.

<i>Gène</i>	<i>F_{ST} global</i>	<i>Groupe Géo.¹</i>	<i>F_{ST} Géo.²</i>	<i>Groupe Deb.¹</i>	<i>F_{ST} Deb.²</i>
DAG2	-0,0147	FRA <-> DE	-0,0338	Prec <-> Inter	0,0100
		FRA <-> NDE	0,0637	Prec <-> Tard	-0,0239
		DE <-> NDE	0,0462	Inter <-> Tard	0,0233
GALA	0,0434	FRA <-> DE	-0,0384	Prec <-> Inter	-0,0318
		FRA <-> NDE	0,1138	Prec <-> Tard	0,0589
		DE <-> NDE	0,0630	Inter <-> Tard	0,0324
YSL1	0,0437	FRA <-> DE	-0,0044	Prec <-> Inter	0,0065
		FRA <-> NDE	0,0030	Prec <-> Tard	0,1028
		DE <-> NDE	0,0097	Inter <-> Tard	-0,0250
AUX-REP	0,0257	FRA <-> DE	0,0519	Prec <-> Inter	0,0500
		FRA <-> NDE	0,1969	Prec <-> Tard	0,0688
		DE <-> NDE	0,0115	Inter <-> Tard	-0,0378
RASI	0,0037	FRA <-> DE	0,0081	Prec <-> Inter	-0,0506
		FRA <-> NDE	-0,0057	Prec <-> Tard	0,0013
		DE <-> NDE	-0,0292	Inter <-> Tard	-0,0259
GA3	-0,0092	FRA <-> DE	0,0175	Prec <-> Inter	0,0030
		FRA <-> NDE	-0,0314	Prec <-> Tard	-0,0206
		DE <-> NDE	0,0123	Inter <-> Tard	0,0421
GA20	0,0325	FRA <-> DE	0,0186	Prec <-> Inter	-0,0190
		FRA <-> NDE	-0,0277	Prec <-> Tard	0,0480
		DE <-> NDE	-0,0407	Inter <-> Tard	0,0105

¹ Groupes FRA : populations françaises, DE : populations allemandes, NDE : populations d'Allemagne du Nord, Prec : populations précoces, Inter : populations intermédiaires, Tard : populations tardives (cf. § 2.2.1 « Différenciation entre populations »).

² En gras : valeur de F_{ST} la plus élevée pour chaque gène.

2.2. Des gènes candidats différenciés et soumis à l'action de la sélection ?

2.2.1. Différenciation entre les populations

Une fois les niveaux de diversité estimés au sein de chaque GC, la question qui se pose alors est de connaître la structuration de cette diversité entre les populations échantillonnées. Pour ce faire, nous avons calculé les indices de différenciation (F_{ST}) au niveau de la séquence en considérant : i/ la population totale incluant tous les individus séquencés, ii/ trois groupes géographiques : France (Westhoffen, Haslach, Saint Jean), Allemagne (Cochem, Johanneskreuz), Allemagne du Nord (Sprakensehl, Mölln), iii/ trois groupes de précocité : Tardif (Saint Jean, Sprakensehl, Mölln), Intermédiaire (Haslach, Johanneskreuz), Précoce (Westhoffen, Cochem). La population totale inclue les séquences issues des neuf populations échantillonnées pour la recherche de SNP. En revanche, l'analyse par groupes géographiques n'a pas permis d'inclure les populations de Klostermarienbergr et La Neuville en Hez (issues du groupe contrasté), trop éloignées géographiquement des trois groupes utilisés (cf. Figure 41). De ce fait, ces deux populations ont également été retirées de l'analyse par groupes de précocité afin de comparer le même échantillon de populations. Le tableau 18 présente les résultats des analyses des indices de diversité aux trois niveaux précédemment définis.

Si l'on considère la population totale, les niveaux de différenciation au niveau de chaque locus sont relativement faibles puisqu'ils s'échelonnent entre 0,0037 pour le locus RASI et 0,0437 pour le locus YSL1. La différenciation au niveau de ces régions codantes du génome est du même ordre de grandeur que les valeurs mesurées à l'aide des marqueurs neutres. En effet, Mariette *et al.* (2002) ont reporté des valeurs de différenciation intraspécifique pour le chêne sessile de 0,023 pour des marqueurs microsatellites et 0,044 pour des marqueurs AFLP. L'analyse au niveau de la population totale ne semble donc pas mettre en évidence de différenciation forte dans la séquence des gènes candidats entre des provenances pourtant très contrastées à la fois géographiquement et du point de vue du débourement. L'analyse par groupes géographiques ou de précocité peut donc se révéler plus informative. En effet, les provenances sont parfois très différenciées entre elles selon leur localisation géographique. C'est le cas des gènes GALA et AUX-REP, pour lesquels l'indice de différenciation entre les groupes de populations françaises et d'Allemagne du Nord est respectivement égal à 0,1138 et 0,1969. Les valeurs de F_{ST} entre les groupes de populations tardives et précoces sont moindres, respectivement 0,0589 et 0,0688 pour GALA et AUX-

REP. Le gène DAG2 présente également des valeurs de différenciation plus forte entre groupes géographiques qu'entre groupes de débourrement avec un $F_{ST} = 0,0637$ entre les populations françaises et celles du Nord de l'Allemagne, contre 0,0233 entre le groupe tardif et le groupe intermédiaire. Néanmoins, les valeurs de différenciation élevées, observées entre groupes géographiques pour ces trois gènes, ne résultent pas nécessairement d'événements démographiques puisque dans ce cas, les valeurs de F_{ST} auraient également été fortes au niveau de la population totale. Ces valeurs peuvent donc provenir d'un processus de sélection, agissant sur des caractères autres que le débourrement, et ayant mené à la différenciation de ces trois groupes géographiques. Le locus RASI, quant à lui, n'est absolument pas différencié, quelle que soit la structuration des populations.

En revanche, pour les gènes YSL1, GA3 et GA20, les valeurs de différenciation sont plus élevées en considérant la précocité de débourrement qu'en considérant la localisation géographique. Le gène YSL1, en particulier, présente un F_{ST} égal à 0,1028 entre les groupes précoce et tardif alors que ce gène n'est absolument pas différencié d'un point de vue géographique ($-0,0044 < F_{ST} < 0,0097$). Dans ce cas, on peut donc émettre l'hypothèse que des pressions de sélection, liées à la précocité de débourrement, ont façonné la diversité au sein de ce gène et donc que ce gène est en partie lié au caractère de débourrement. Cette hypothèse peut également être avancée, dans une moindre mesure, pour les deux gènes GA3 et GA20. En effet, les gènes GA3 et GA20 présentent respectivement des valeurs maximales de F_{ST} de 0,0421 et 0,0480 entre groupes de débourrement alors que les valeurs maximales, entre groupes géographiques, sont respectivement de 0,0175 et 0,0186. On peut noter également que pour 5 gènes sur 7, en ne considérant que la différenciation entre groupes de débourrement, la différenciation entre groupes extrêmes (tardif et précoce) est supérieure à la différenciation entre groupes contigus (tardif et intermédiaire ; précoce et intermédiaire).

2.2.2. Des gènes qui s'écartent du modèle évolutif neutre ?

Dans le but de tester l'écart à la neutralité des 7 locus étudiés, nous avons calculé les valeurs du D de Tajima sur l'ensemble des individus séquencés. Les valeurs de D pour chaque locus sont reportées dans le tableau 17 et la significativité des valeurs obtenues a été testée par simulation de coalescence (cf. § 1.2.1.). Les locus RASI, AUX-REP, GA3 et GA20 s'écartent significativement ($p < 0,05$) de l'hypothèse nulle de neutralité avec des valeurs négatives de D dans les quatre cas. Le problème est que l'échantillon regroupe des populations contrastées à

Tableau 19 : Valeurs et significativité du test D de Tajima.

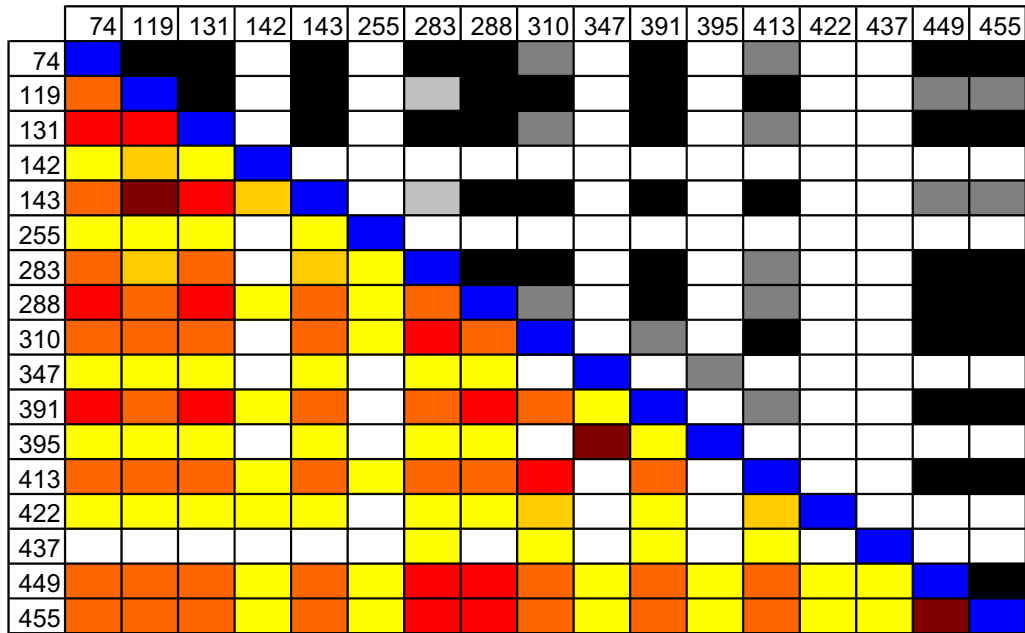
Gène	RASI		AUX-REP		GA3		GA20		
	Pop.	Totale	France	Totale	France	Totale	France	Totale	France
N		26	9	27	9	27	9	21	8
S		23	8	19	8	12	3	15	7
D¹		-1,63371*	-1,62966*	-1,83479*	-1,5405*	-1,75622*	-0,35929	-1,63056*	-0,2891

N = nombre de séquences ; S = nombre de sites polymorphes informatifs ; D = valeur du test D.

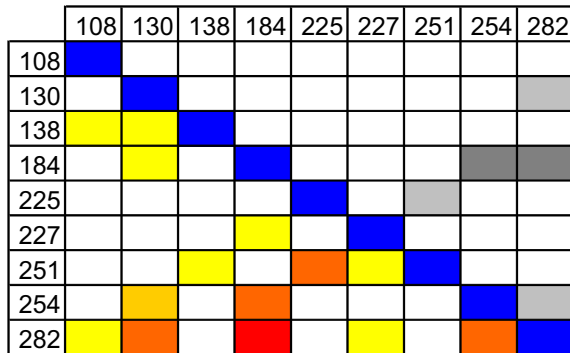
* : significatif à 5%.

la fois pour leur localisation géographique et pour leur comportement vis-à-vis du débourrement. La situation idéale aurait été de pouvoir tester, en suivant, la valeur du D de Tajima au sein d'une même population. Mais l'échantillonnage que nous avons réalisé avait pour objectif principal de pouvoir bien estimer les paramètres de différenciation entre populations et n'est pas du tout adapté à la mesure de paramètres génétiques intra population. Cependant, les trois populations originaires du Nord Est de la France, incluses dans cet échantillon peuvent être considérées comme une seule et même provenance d'un point de vue géographique (cf. tableau 12 et figure 41). Nous avons donc testé la valeur du D de Tajima au sein des trois populations Saint Jean, Haslach et Westhoffen dont les individus ont été regroupés en une seule population. Les résultats de ces analyses ainsi que ceux des simulations effectuées sur la population totale sont présentés au tableau 19. La significativité des valeurs négatives du test D de Tajima est donc conservée pour les gènes RASI et AUX-REP au sein du groupe des trois populations françaises. Il semblerait donc bien que les gènes RASI et AUX-REP s'écartent du modèle d'évolution neutre et qu'ils aient été la cible d'effets de la sélection au cours de leur histoire évolutive. Par ailleurs, ces deux locus présentent des valeurs de diversité moyennes à fortes, avec $\pi = 10,24 \times 10^{-3}$ et $5,69 \times 10^{-3}$, respectivement pour RASI et AUX-REP, ainsi qu'une diversité haplotypique élevée, estimée respectivement à 0,86 et 0,91. Les allèles en fréquence faible sont, par ailleurs, très présents au sein de ces deux gènes. En revanche, ils ne se comportent pas de la même manière en terme de différenciation puisque AUX-REP présente une valeur de F_{ST} de 0,1969 entre le groupe français et celui d'Allemagne du Nord alors que RASI n'est pas différencié ($F_{ST} \text{ max} = 0,0013$). Etant donné ces valeurs et un rapport de substitutions non synonymes par rapport aux synonymes (dN/dS) égal à 6 dans le cas du locus AUX-REP, on peut faire l'hypothèse d'une sélection positive diversifiante ayant agi sur ce gène vis-à-vis du caractère de débourrement. Pour ce qui concerne RASI, les valeurs observées indiqueraient une sélection positive directionnelle ayant subi un balayage sélectif récent, ce qui expliquerait l'excès d'allèles rares. En revanche, pour les gènes GA3 et GA20, la valeur du test reste négative mais non significative ce qui semblerait indiquer que la valeur sur l'ensembles des 9 populations résulte plus probablement d'une structuration génétique sous jacente de ces populations contrastées que d'un écart réel à la neutralité de ces locus. Ces résultats doivent cependant être interprétés avec une extrême précaution compte tenu de l'échantillonnage effectué, et de l'aggrégation d'individus issus de populations différentes. Enfin des valeurs négatives de D peuvent ne pas

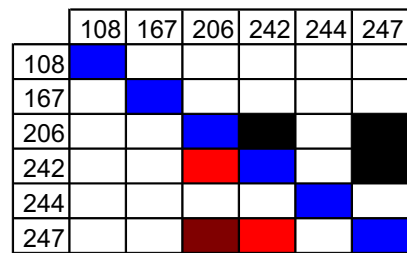
Locus GALA



Locus RASI



Locus AUX-REP



Locus YSL1

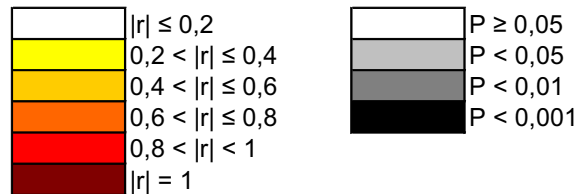
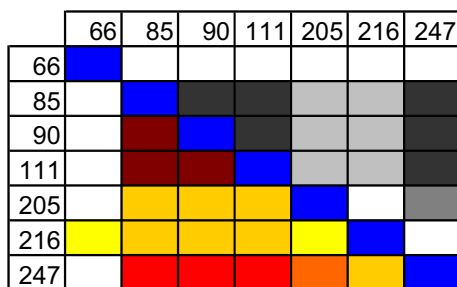


Figure 44 : Déséquilibre de liaison entre les sites SNP au sein des GC étudiés au niveau nucléotidique. Les valeurs absolues de R sont indiquées en dessous de la diagonale et les valeurs des probabilités associées (test de Fisher exact) au dessus de la diagonale. Les sites SNP en moucheté sont situés dans les régions introniques.

être exclusivement générées par des effets de sélection directionnelle, mais également par une expansion démographique des populations.

2.2.3. Le déséquilibre de liaison au sein des GC

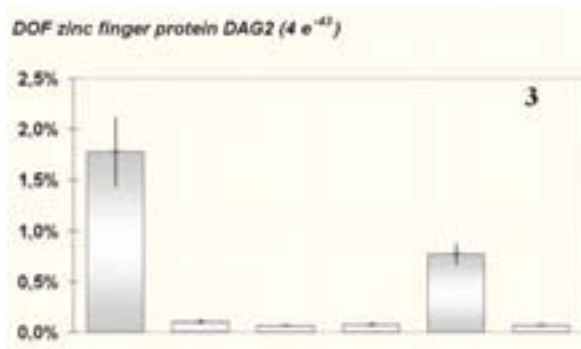
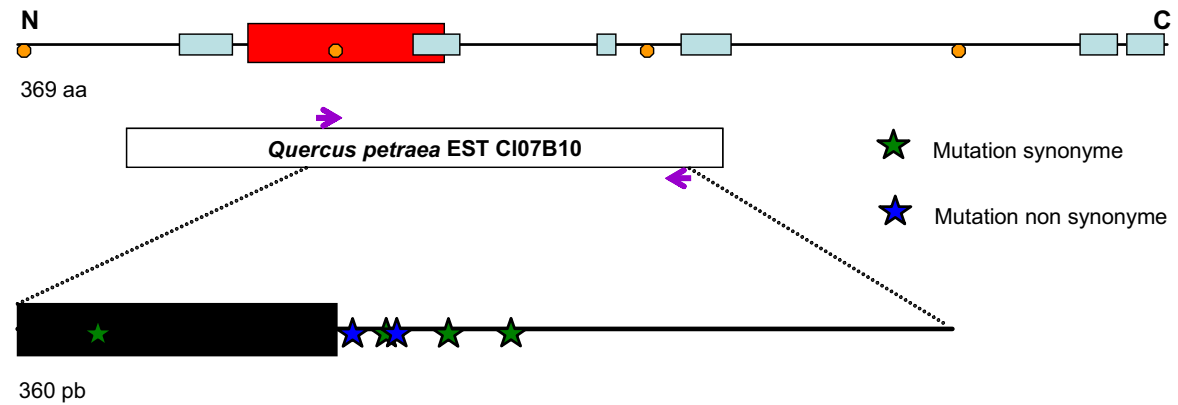
Nous avons estimé le déséquilibre de liaison entre les sites SNP (singletons exclus) au sein de 4 locus. Les gènes DAG2, GA3 et GA20, qui ne présentent respectivement que 2, 3 et 4 sites SNP informatifs, n'ont pas été considérés dans cette analyse. Les résultats sont présentés locus par locus, la longueur des fragments amplifiés ne permettant pas de réaliser une analyse globale du déséquilibre en fonction de la distance nucléotidique. La figure 44 présente les résultats de ces analyses.

Au sein du locus GALA, certaines valeurs de R sont fortes et significatives tout au long des 400 pb étudiées. Globalement, 10 sites SNP sont plus ou moins liés avec des valeurs de $|R|$ significatives comprises entre 0,522 et 1. Il s'agit des locus SNP en positions 74, 119, 131, 143, 283, 288, 310, 391, 413, 449, 455. Parmi ces derniers, les couples 119-143 et 449-455 sont entièrement liés avec une probabilité inférieure à 0,1%. On peut noter qu'entre les sites SNP en position 74 et en position 455, la valeur de R est encore de -0,791 ($p < 0,1\%$) donc le déséquilibre au sein de ce GC est présent au moins jusqu'à 380 paires de bases. Un second groupe de sites apparaît, faiblement liés entre eux ($|R|_{\max} = 0,348$), à l'exception des sites 347 et 395, et faiblement liés au premier groupe ($|R|_{\max} = 0,476$). Il s'agit des locus 142, 255, 347, 395, 422, 437. En moyenne, la valeur de $|R|$ d'un site pris au sein de ce groupe avec un site pris au sein du premier groupe est comprise entre 0,2 et 0,4. On peut enfin remarquer que les sites peuvent être liés au sein des régions introniques ou exoniques comme entre sites présents dans l'intron ou dans l'exon. A titre d'exemple, le site 74 (intron) est lié avec le site 288 (exon) avec une valeur de $|R| = 0,926$ ($p < 0,1\%$). Sur le locus YSL1, hormis le SNP en position 66, les 6 autres SNP sont liés entre eux. Les sites 85, 90, 111 et 247 sont fortement liés entre eux ($0,9 < |R| < 1$) et plus faiblement aux sites 205 et 216. Le déséquilibre est fort au sein de ce locus mais estimé seulement sur 200 paires de base.

Il n'apparaît quasiment pas de déséquilibre de liaison dans les régions amplifiées de RASI et AUX-REP à l'exception de quelques sites SNP. En effet, au sein du locus RASI, seuls deux couples de sites sont liés avec une probabilité associée à la valeur de R inférieure à 1 %. Il s'agit des locus 184 et 254 ($R = 0,799$) et 184 et 282 ($R = 0,847$). Trois autres liaisons sont significatives à 5%. D'une manière générale, le déséquilibre est quasi inexistant au sein

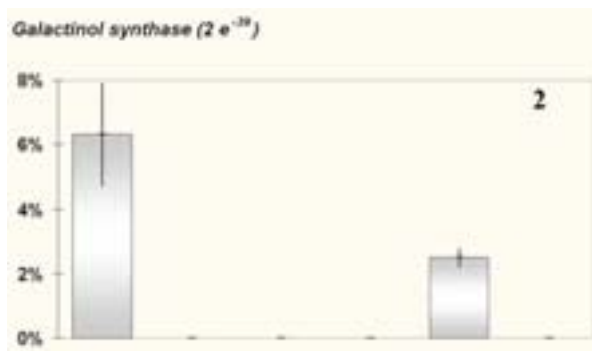
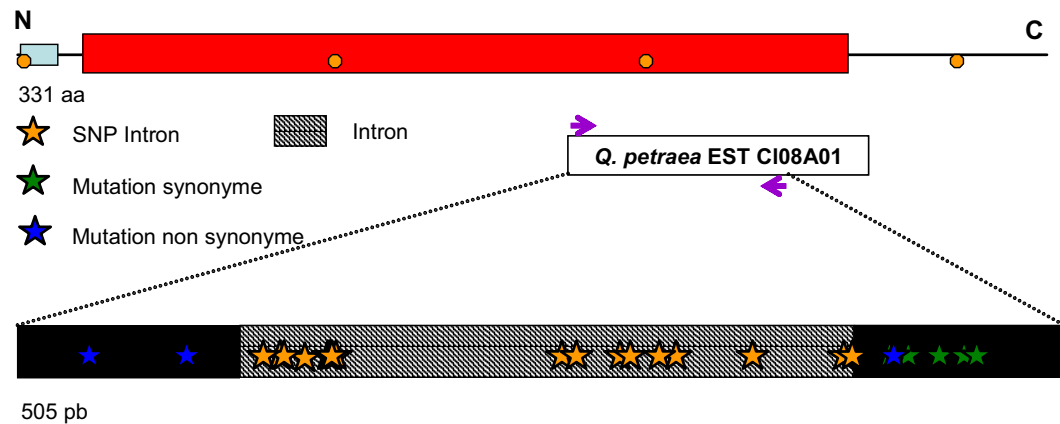
de ce gène ce qui serait en adéquation avec l'hypothèse d'un balayage sélectif récent ou d'une sélection directionnelle en cours de fixation, précédemment évoquée. Pour le locus AUX-REP, trois couples de sites sont liés significativement. Le site SNP en position 242 est lié aux sites 206 et 247 avec $|R| = 0,892$ ($p < 0,1\%$), ces deux derniers étant entièrement liés ($p < 0,1\%$).

Dof zinc finger protein DOF2.5 (DAG2) *A. thaliana* (Q9ZPY0)



$\pi_{tot} = 2,51 \times 10^{-3}$
 7 haplotypes / 27 individus
 $F_{ST} = -0,0147$
 Groupe de liaison Q8
 Pas de co-localisation QTL

Galactinol synthase *Cucumis melo* (Q8S9H1)



$\pi_{tot} = 15,02 \times 10^{-3}$
 18 haplotypes / 26 individus
 $F_{ST} = 0,0434$
 Groupes de liaison Q3 et C8
 Pas de co-localisation QTL

Figure 45: Caractéristiques principales (sites polymorphes, niveaux d'accumulation de transcrits, diversité, différenciation, cartographie génétique) de DAG2 et GALA.

2.3. Etudes d'association

Les résultats présentés ici ne constituent qu'une étude préliminaire sur deux locus, la totalité de l'échantillon des GC devant faire l'objet d'une étude d'association par la suite.

2.3.1. Choix des locus et des sites SNP

Dans un premier temps, nous avons choisi l'homologue potentiel du facteur de transcription DAG2, qui nous semblait très intéressant puisque DAG2 est impliqué dans le contrôle de la germination de la graine chez *Arabidopsis* (Gualberti *et al.*, 2002) et que, par analogie, un rôle similaire dans la transcription du signal pouvait être attribué à ce locus chez le chêne. Le second locus analysé code potentiellement pour l'enzyme Galactinol synthase, qui participe à la protection des cellules contre la déshydratation (Taji *et al.*, 2002), par son implication précoce dans la synthèse de RFO (Raffinose Family Oligosaccharides). Nous sommes ainsi en présence d'un gène de régulation et d'un gène de structure dont les principales caractéristiques sont présentées à la figure 45. Pour savoir sur quels sites SNP travailler, nous nous sommes appuyés sur les résultats issus de l'analyse fonctionnelle de chacun des sites (position, statut fonctionnel, ...) ainsi que ceux issus de l'analyse des niveaux de différenciation entre les populations.

Tout d'abord, nous nous sommes intéressés aux sites SNP informatifs, c'est-à-dire singletons exclus. Dans le cas de l'homologue potentiel de DAG2, les possibilités étaient donc restreintes aux deux sites nucléotidiques en position 25 et 123. La mutation au locus 123 est non synonyme et donc, *a priori*, plus susceptible d'avoir un impact sur la structure de la séquence protéique et donc sur son activité (cf. tableau 14). Nous avons donc testé les deux amorces susceptibles d'amplifier ce site en amont et en aval de la position 123. Malheureusement, les conditions d'amplification n'ont pu être mises au point sur ces deux amorces. La technique utilisée pour le génotypage des sites SNP (extension d'amorces) s'est révélée très sensible et délicate à mettre en œuvre. D'une part, la région située autour du site SNP considéré doit être exempte de mutations afin de ne pas inclure de bases dégénérées non spécifiques dans la séquence de l'amorce. D'autre part, l'incorporation de la base supplémentaire au niveau du site SNP considéré à la suite de l'amorce doit être très spécifique et il faut s'assurer de l'incorporation correcte à l'aide d'individus témoins, préalablement séquencés au niveau de ce site. Ces critères n'étaient pas respectés pour les amorces

Tableau 20 : Fréquences alléliques des différentes populations sur les trois sites SNP étudiés en association.

Population	GALA455*			GALA395*			DAG25*		
	A	G	Nombre d'allèles	A	G	Nombre d'allèles	C	T	Nombre d'allèles
Fontainebleau	0,526	0,474	38	0,262	0,738	42	0,833	0,167	18
Blakeney	0,579	0,421	38	0,214	0,786	42	0,875	0,125	32
Adé	0,583	0,417	36	0,143	0,857	42	1	0	18
Longchamps	0,533	0,467	30	0,368	0,632	38	0,775	0,225	40
Sprakensehl	0,625	0,375	40	0,333	0,667	48	0,813	0,188	48
Lembach	0,553	0,447	38	0,136	0,864	44	0,773	0,227	44
Wolfgangz	0,386	0,614	44	0,438	0,563	48	0,896	0,104	48
Saint-Jean	0,633	0,367	30	0,368	0,632	38	0,842	0,158	38
Still	0,548	0,452	42	0,261	0,739	46	0,739	0,261	46
Westhoffen	0,588	0,412	34	0,455	0,545	44	0,818	0,182	44
Mölln	0,563	0,438	32	0,175	0,825	40	0,909	0,091	22
Klostermarienbergr	0,563	0,438	32	0,395	0,605	38	0,658	0,342	38
Gorhde	0,676	0,324	34	0,413	0,587	46	0,891	0,109	46
Johanneskreuz	0,396	0,604	48	0,313	0,688	48	0,813	0,188	48
Luss	0,368	0,632	38	0,239	0,761	46	0,857	0,143	42
Romersberg	0,706	0,294	34	0,381	0,619	42	0,891	0,109	46
Haslach	0,6	0,4	30	0,3	0,7	30	0,842	0,158	38
Cochem	0,341	0,659	44	0,409	0,591	44	0,913	0,087	46
Neuville en Hez	0,575	0,425	40	0,405	0,595	42	0,786	0,214	42
Obora	0,553	0,447	38	0,325	0,675	40	0,975	0,025	40
Lapwald	0,295	0,705	44	0,318	0,682	44	0,976	0,024	42
21 populations	0,523	0,477	784	0,317	0,683	892	0,846	0,154	826

* En gras : F_{IS} significativement différent de 0 ($p < 5\%$).

amplifiant le site en position 123. Nous nous sommes donc intéressés au site en position 25 qui porte une mutation synonyme mais néanmoins située dans la région codant pour le site actif de la protéine (cf. figure 45) ce qui peut mener à des variations du niveau d'expression ou de la stabilité de cette dernière. Concernant le locus codant potentiellement pour l'enzyme Galactinol synthase, 17 sites SNP informatifs étaient à notre disposition, dont un seul site non synonyme en position 74 et 12 sites situés dans la partie intronique non codante. Les sites en position 74 (NS), 288, 310, 391, 395 (N), 449 et 455 (S) ont été testés. L'amplification du site en position 74 n'a pas pu être optimisée et nous avons conservé pour l'analyse le site en position 395 (intron) et le site en position 455 (codant synonyme). Il nous paraissait intéressant de pouvoir comparer l'information portée par une mutation en position intronique et une autre dans la partie codante. Par ailleurs, le site en position 455 est significativement et assez fortement lié ($R = -0,791$; $p < 0,1\%$) avec le site en position 74. Parmi les sites testés dans la partie intronique, le site en position 395 présente un F_{ST} entre groupes de débourrement tardif et précoce égal à 0,1342, ce qui représente la plus forte valeur observée entre groupes de débourrement. Ces caractéristiques en font deux sites intéressants à tester dans le cadre de notre étude d'association.

Les deux exemples précédents illustrent la manière dont le choix des sites pertinents peut être effectué avec, par ordre de priorité, le statut de la mutation (synonyme, non synonyme) puis sa position dans la séquence (domaine actif, régulateur ...). En plus de ces critères, les indices de différenciation, vis-à-vis du caractère étudié, estimés en populations naturelles au niveau de chaque site SNP, constituent également des critères de choix. D'une manière générale, toute information relative aux conséquences connues ou potentielles de la mutation sur le niveau d'expression de la protéine, en dernier ressort, est à prendre en compte dans le choix des sites les plus pertinents à étudier.

2.3.2. Diversité et différenciation des sites SNP étudiés

Dans un premier temps, nous nous sommes intéressés aux fréquences alléliques associées aux trois sites SNP étudiés. Le tableau 20 présente les fréquences alléliques estimées au sein des 21 populations. Le locus DAG25 possède un allèle très majoritaire (C), *a priori* fixé sur la population Adé. Cela se traduit par une majorité d'homozygotes CC (289) et un faible taux d'hétérozygotie (121). Seuls trois individus homozygotes TT sont présents. Les fréquences de cet allèle sont comprises entre 0,658 (Klostermarienbergr) et 1 (Adé). Le locus

Tableau 21 : Paramètres de diversité génétique au sein des 21 populations.

Nom du locus SNP	H ₀	H _S	H _T	G _{ST}	Fis
GALA455	0,754	0,48	0,498	0,036	-0,575
GALA395	0,543	0,423	0,433	0,024	-0,289
DAG25	0,283	0,246	0,254	0,03	-0,156

H₀ : hétérozygotie observée ; H_S : diversité génétique moyenne intra population ; H_T : diversité totale ; G_{ST} : différenciation inter population ; F_{IS} : coefficient de consanguinité.

Tableau 22 : Paramètres de diversité au sein des trois groupes géographiques.

Groupe géographique	Nom du locus SNP	H ₀	H _S	H _T	G _{ST}	Fis
France	DAG25	0,735	0,483	0,483	0,001	-0,548
	GALA395	0,532	0,428	0,439	0,026	-0,251
	GALA455	0,361	0,307	0,306	-0,004	-0,182
Allemagne	DAG25	0,698	0,461	0,494	0,066	-0,513
	GALA395	0,619	0,454	0,457	0,007	-0,36
	GALA455	0,24	0,21	0,212	0,012	-0,148
Contraste	DAG25	0,829	0,496	0,493	-0,007	-0,667
	GALA395	0,479	0,387	0,399	0,031	-0,234
	GALA455	0,245	0,221	0,239	0,072	-0,113

Tableau 23 : Paramètres de diversité des groupes de débourement.

Nom du locus SNP	H ₀	H _S	H _T	G _{ST}	Fis
GALA455	0,757	0,497	0,499	0,004	-0,519
GALA395	0,546	0,434	0,434	0	-0,256
DAG25	0,293	0,258	0,26	0,007	-0,133

GALA395 possède également un allèle majoritaire (G) dont les fréquences s'échelonnent entre 0,545 (Westhoffen) et 0,864 (Lembach). Les hétérozygotes restent cependant majoritaires avec 243 génotypes AG, 183 génotypes GG et 20 génotypes AA. Sur le locus GALA455, les deux allèles A et G sont en équilibre sur l'ensemble des populations, la fréquence de l'allèle G variant entre 0,294 (Romersberg) et 0,705 (Lapwald). Les hétérozygotes restent majoritaires sur ce locus avec 296 génotypes AG, 57 génotypes AA et 39 génotypes GG. Globalement, les fréquences alléliques varient peu entre populations sur les trois locus SNP.

Les indices de diversité et de différenciation ont été calculés sur l'ensemble des provenances pour chacun des sites polymorphes étudiés. Les paramètres estimés sont présentés au tableau 21. Une manière d'apprécier les valeurs de H_S et H_T est de calculer la diversité minimale possible, dans le cas où un seul allèle est présent ($H = 1 - 1^2$), qui est de 0, et la diversité maximale possible, dans le cas où les deux allèles sont en fréquence égale ($H = 1 - (\frac{1}{2}^2 + \frac{1}{2}^2)$), qui est de 0,5. Les valeurs de diversité obtenues sont donc moyennes (dag25) à fortes (gala395 et gala455). Mais le résultat le plus remarquable est qu'il existe une forte structuration de la diversité intra population, traduite par les valeurs négatives du paramètre F_{IS} , qui indiquent un excès d'hétérozygotes parmi les trois locus SNP. Une analyse au sein de chaque population a permis de voir que les valeurs sont toutes négatives quelle que soit la population considérée mais que seules les valeurs associées aux deux locus GALA sont significatives. Le déficit significatif en homozygotes est équitablement réparti chez gala455 contrairement à gala395 où l'homozygote GG est plus présent que l'homozygote AA. L'analyse effectuée par groupes géographiques de populations (France, Allemagne, Contrasté) ne semble pas présenter d'effets dus à la localisation géographique (Tableau 22). En effet, les niveaux de diversité et de différenciation au sein de chaque groupe sont équivalents aux niveaux estimés sur les 21 populations dans leur ensemble. Il en est de même pour la structuration de la diversité en fonction de la précocité de débourrement. En effet, les niveaux de différenciation entre les groupes de précocité, ainsi que la valeur moyenne de F_{IS} au sein de chaque groupe, est équivalente à la valeur moyenne estimée sur les 21 populations considérées individuellement (Tableau 23). A ce niveau de l'analyse, il est donc impossible de conclure sur un effet des allèles, à ces trois sites SNP, sur la note de débourrement. C'est pourquoi les études d'association sont nécessaires.

Tableau 24 : Marqueurs utilisés dans le cadre du test de Beaumont et Nichols (1996).

Marqueurs	Nb de locus	Nb de pops	Nb arbres/pop	Références
AFLPs (<i>PstI/MseI</i>)	155	7	50	Mariette <i>et al.</i> (2002)
RAPDs	41	21	23	Le Corre <i>et al.</i> (1997)
Microsatellites	6	7	200	Mariette <i>et al.</i> (2002)
SNPs stress osmotique	24	10	10	Porth <i>et al.</i> (2004)
SNPs débourrement	3	21	24	Notre étude
SCARs	13	8	45	Bodenes <i>et al.</i> (1997)
Isozymes	8	81	150	Zanetto et Kremer (1995)

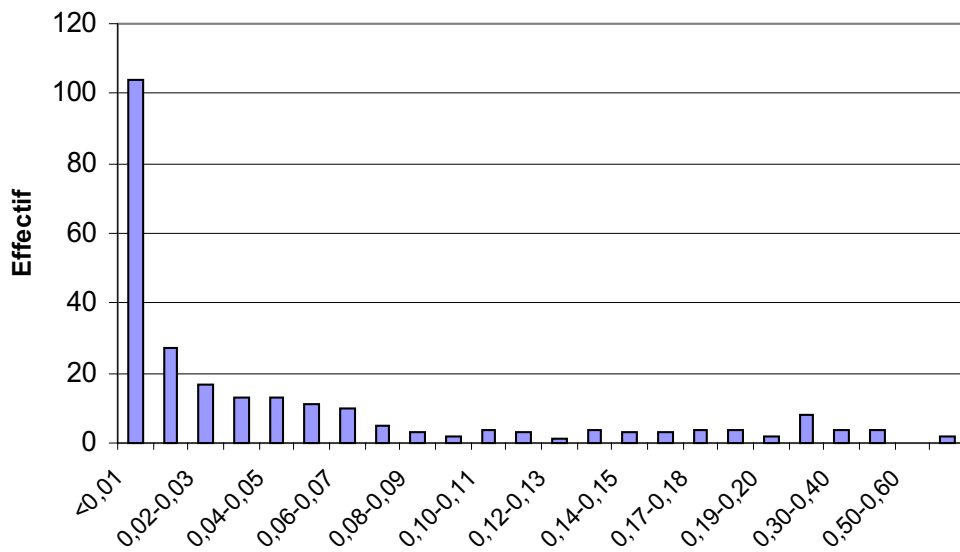


Figure 46 : Distribution des valeurs de F_{ST} estimées sur plus de 250 marqueurs (cf. tableau 24). La valeur moyenne est égale à 0,058.

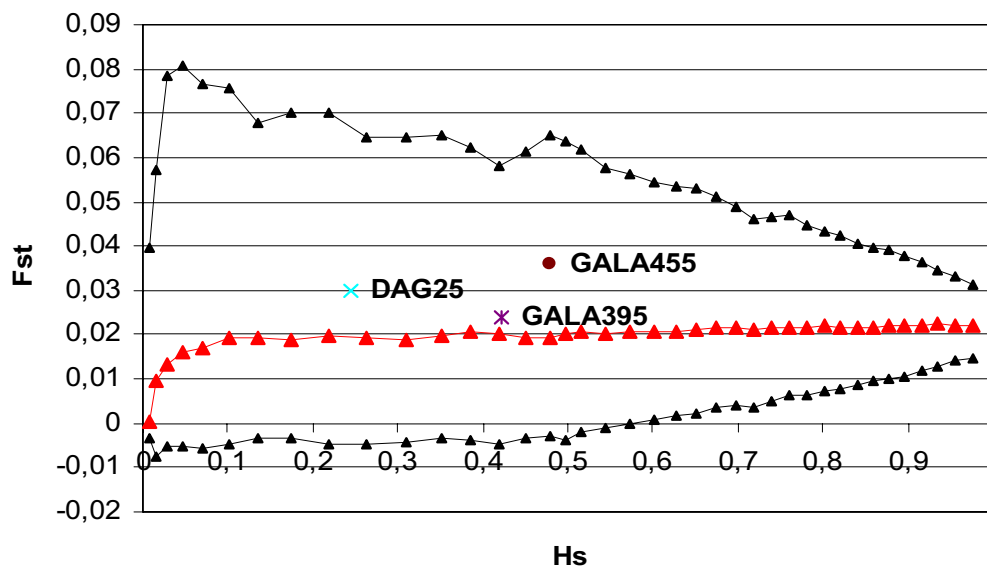


Figure 47 : Enveloppe de distribution des valeurs de différenciation associées aux marqueurs chez le chêne sessile et position des trois locus SNP de notre étude.

2.3.3. Des locus SNP en déséquilibre et soumis à l'action de la sélection naturelle ?

Une analyse intéressante et complémentaire des calculs effectués dans le paragraphe précédent est de rechercher si les trois locus ont été soumis à l'action de la sélection naturelle à l'aide du test décrit par Beaumont et Nichols (1996), basé sur les indices de différenciation. Pour ce faire, la première étape consiste en la construction d'une enveloppe de distribution des valeurs de différenciation (F_{ST}) en fonction de l'hétérozygotie observée (H_S). A cet effet, l'ensemble des données moléculaires (valeurs de F_{ST} et H_S obtenues avec divers marqueurs) disponibles sur le chêne sessile a été utilisé et la synthèse est présentée au tableau 24. Les marqueurs utilisés peuvent être des AFLP, des microsatellites, des RAPD, des SNP, des SCAR ou des Isozymes. La distribution des valeurs de F_{ST} suit une courbe en L typique avec une valeur moyenne estimée à 0,058 (Figure 46). Cette valeur moyenne est ensuite utilisée pour déterminer la courbe moyenne, ainsi que l'enveloppe, avec un intervalle de confiance de 95%, de distribution des valeurs de F_{ST} en fonction de H_S . Les valeurs sont simulées par coalescence à l'aide d'un modèle d'évolution en îles. Le processus est itératif et les valeurs individuelles de chaque marqueur sont, en suivant, reportées sur le graphe. Une nouvelle enveloppe de distribution est alors calculée en prenant soin de retirer les marqueurs à l'extérieur de l'enveloppe précédente. La nouvelle valeur de différenciation est alors utilisée pour calculer de nouveau l'enveloppe de distribution des valeurs de différenciation en fonction de l'hétérozygotie observée. La figure 47 présente les résultats de cette analyse avec les trois locus SNP de notre étude reportés sur le graphe. Les résultats indiquent clairement que, d'après ce test, les trois locus étudiés ne manifestent pas de valeurs de différenciation situées en dehors de l'enveloppe de l'hypothèse neutre. Ce résultat est en accord avec la valeur non significative du D de Tajima estimée au niveau de la séquence sur les 27 individus constituant le panel de recherche de SNP (cf. § 2.2.2.).

Il est également intéressant de mesurer le déséquilibre entre ces trois sites SNP afin de préjuger de possibles effets de covariation sur le caractère de débourement. Sur l'ensemble des populations, seuls les locus GALA455 et GALA395 sont globalement en déséquilibre avec une valeur du test de χ^2 infinie et une probabilité associée très significative et non déterminée. En ce qui concerne les deux autres associations, elles ne sont globalement pas significatives. En revanche, à l'intérieur de chacune des 21 populations on trouve deux associations significatives à 10%, l'une entre les locus GALA455 et DAG25 au sein de la

Tableau 25 : Classes génotypiques et moyennes de débourement.

GALA455			GALA395			DAG25		
Génotype	Note	Proba.*	Génotype	Note	Proba.*	Génotype	Note	Proba.*
AA	1,596	0,2	AA	1,4	0,136	CC	1,735	0,184
AG	1,763	0,418	AG	1,819	0,25	CT	1,882	0,26
GG	1,543	0,148	GG	1,655	0,123	TT	2,667	0,29

* Probabilité associée au test t : la première valeur correspond à la comparaison entre le premier homozygote et l'hétérozygote, la seconde correspond à la comparaison entre les deux homozygotes et la dernière correspond à la comparaison entre l'hétérozygote et le second homozygote.

Tableau 26 : Tableau récapitulatif de l'Anova incluant les 3 locus et leurs interactions.

<i>Facteur*</i>	<i>d.l.</i>	<i>SCE (type III)</i>	<i>CM</i>	<i>Valeur de F</i>	<i>Pr > F</i>
GALA455	2	2,020	1,010	0,49	0,613
GALA395	2	7,527	3,763	1,83	0,163
DAG25	2	3,667	1,834	0,89	0,412
GALA455*GALA395	3	4,422	1,474	0,72	0,544
GALA455*DAG25	2	6,128	3,064	1,49	0,228
GALA395*DAG25	3	10,049	3,350	1,63	0,184
GALA455*GALA395*DAG25	2	2,232	1,116	0,54	0,582

* Le r² du modèle est de 6,70%.

Tableau 27 : Coefficients de corrélation entre fréquences alléliques et notes moyennes de débourement estimées au niveau de chaque population.

Locus SNP	Charnie	Vierzon	Vincence	Sillegny
GALA455	0,01	0,06	0,05	0,08
GALA395	0,29	0,28	0,31	0,31
DAG25	0,27	0,51*	0,41	0,30

* Probabilité significative à 5%.

population d'Obora ($p=0,069$) et l'autre entre les locus GALA395 et DAG25 au sein de la population Haslach ($p=0,070$). Il n'y a donc pas d'association significative entre les nucléotides aux différents locus ce qui laisse supposer que des effets de covariation entre ces locus ne peuvent se mettre en place du fait de l'indépendance des allèles, du moins entre GC différents. Les locus GALA455 et GALA3958 semblent en effet liés mais ils appartiennent au même gène codant potentiellement pour l'enzyme Galactinol synthase et leur association significative reflète donc un effet de liaison physique.

2.3.4. Corrélation entre variants alléliques et notes de débourrement

Nous avons commencé par regrouper l'ensemble des individus génotypés en une seule population pour estimer la corrélation entre les variants alléliques aux sites SNP et les notes individuelles de débourrement. La significativité des écarts observés entre les notes moyennes de débourrement a été testée à l'aide du test t de Student. La population dans son ensemble présente une note moyenne de débourrement égale à 1,776. Les effectifs, les notes moyennes par classes génotypiques et les probabilités associées pour chaque site pris individuellement sont présentés dans le tableau 25. Aucune des comparaisons entre les notes moyennes de débourrement pour chaque classe génotypique n'est significative. Il n'y a donc pas d'effet statistique des allèles pour chaque locus pris individuellement. En revanche, des effets d'interaction peuvent exister. Le tableau 26 présente les résultats de l'analyse de variance effectuée en considérant chaque locus individuellement et toutes les interactions possibles entre les locus SNP. La première remarque est que le modèle explique globalement 6,70% de la variation observée des notes de débourrement, ce qui est assez faible. Les effets des classes génotypiques à chaque locus pris individuellement ne sont pas significatifs ($16\% < \text{Proba.} < 62\%$), ce qui confirme les résultats de l'analyse précédente de comparaison des moyennes à l'aide du test t de Student. Les associations entre locus deux à deux ne semblent pas significatives non plus ($18\% < \text{Proba.} < 55\%$), tout comme l'effet d'association entre les trois locus ($\text{Proba.} = 58\%$). Il semblerait donc, aux vues des résultats de ces différentes analyses, que les trois sites SNP étudiés, ne présentent pas d'effet significatif sur la valeur phénotypique des individus vis-à-vis de la date de débourrement puisqu'aucun effet n'a pu être détecté.

Nous avons testé en suivant les corrélations entre les fréquences alléliques et la note moyenne de débourrement estimées au niveau de chaque population. Le tableau 27 présente

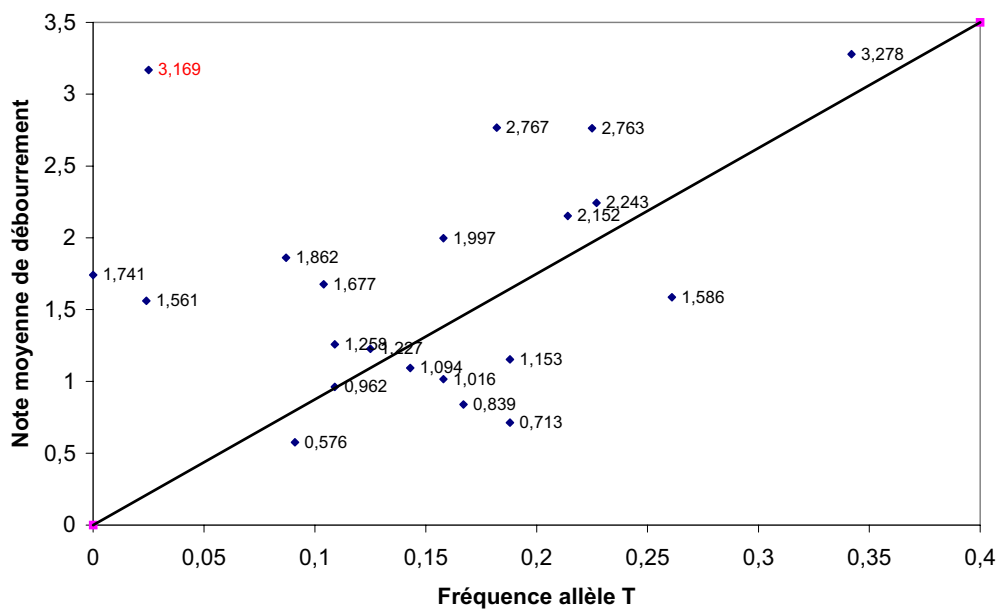


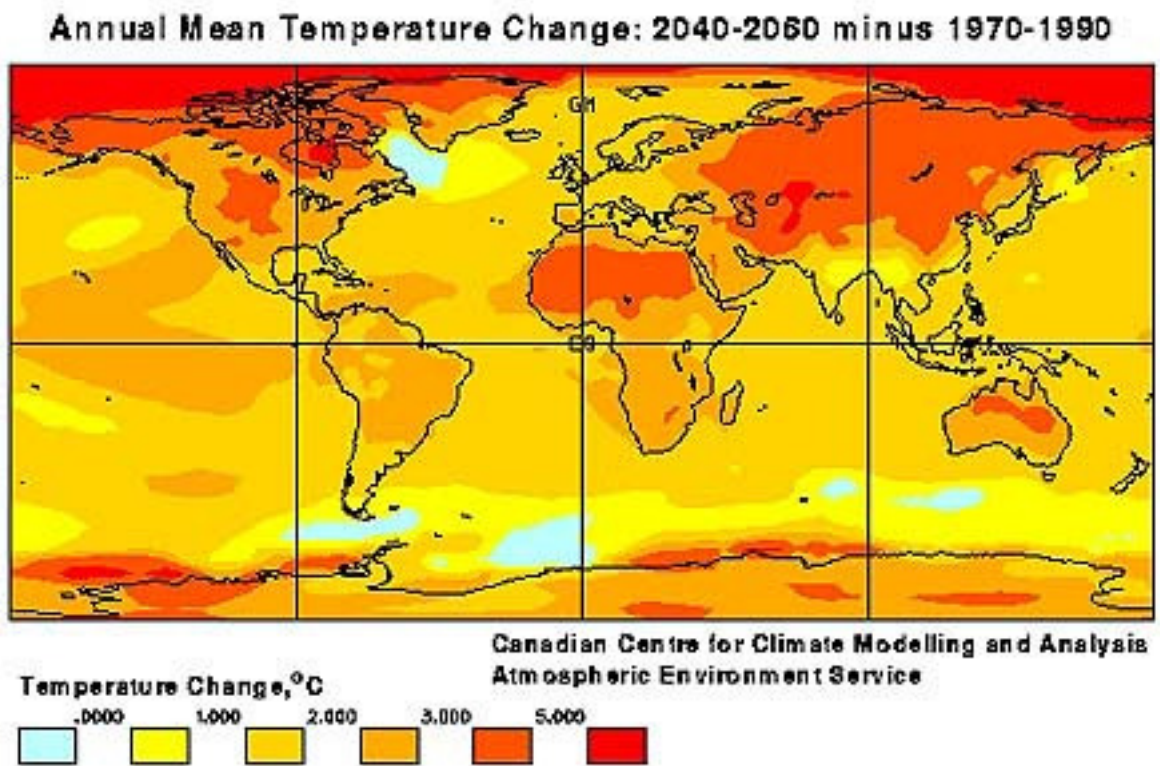
Figure 48 : Distribution des notes moyennes de débourement estimées sur le site de la Petite Charrie en fonction de la fréquence de l'allèle T au locus DAG25 ($r^2 = 0,27$). En rouge, la population d'Obora.

les coefficients de corrélation estimés au sein de chaque test de provenances (Petite Charnie, Vierzon, Vincence, Sillegny). Pour le locus DAG25, c'est la fréquence de l'allèle T qui est utilisée, celle de l'allèle A pour le locus GALA395 et celle de l'allèle G pour le locus GALA455. Les notes moyennes de débourrement par provenance ont été estimées sur l'ensemble des individus de chaque population disponibles par site, à savoir 240 arbres. Pour ce test, l'hypothèse H_0 est : « le stade de débourrement moyen d'une population est indépendant des fréquences alléliques ». Avec 20 degrés de liberté (= Nb populations - 1), H_0 est rejeté, avec un risque de 5%, si le coefficient est supérieur à 0,43, et H_0 est rejeté, avec un risque de 1%, si le coefficient est supérieur à 0,55. Une seule valeur de coefficient de corrélation est significative à 5%. Il s'agit de la corrélation entre la fréquence de l'allèle T au locus DAG25 et la note moyenne de débourrement estimée sur le site de Vierzon. Cette corrélation n'est pas significative sur les trois autres sites. L'explication de ce résultat est à rechercher dans l'absence de la population Obora sur le site du test de provenances de Vierzon. En effet, le comportement de cette population est particulièrement atypique. Si l'on observe le nuage de points représentant chaque population avec la fréquence allélique de l'allèle T au locus DAG25 en abscisse et la note moyenne de débourrement estimée sur chacune des populations en ordonnée, on remarque que la population Obora présente une faible fréquence de l'allèle T et une très forte précocité de débourrement alors que la tendance générale est à l'inverse (Figure 48). C'est ce qui explique que la valeur du coefficient de corrélation atteigne une valeur particulièrement élevée sur le site de Vierzon du fait de l'absence de la population Obora. L'ensemble des autres coefficients de corrélation est non significatif même si on peut remarquer que les valeurs sont beaucoup plus élevées pour GALA395 et DAG25, comprises entre 0,27 et 0,41, que pour GALA455, pour lequel les valeurs sont proches de 0.

③ Conclusions des études de diversité et d'association

A ce stade de l'analyse, nous avons donc analysé les patrons de diversité nucléotidique de sept locus candidats pour le débourrement. La diversité au sein du locus codant potentiellement pour YSL1, et dans une moindre mesure pour GA3 et GA20, semble en partie façonnée par les pressions de sélection liées à la précocité de débourrement. Nous avons également recherché d'éventuelles signatures moléculaires de la sélection au sein de ces locus. Les résultats permettent de faire l'hypothèse d'effets de la sélection naturelle sur les locus AUX-REP et RASI, sélection positive respectivement diversifiante et directionnelle sur ces deux locus. Une étude d'association a ensuite été menée sur trois sites SNP, deux appartenant au locus GALA et le troisième au locus DAG2. Les deux locus GALA et DAG2 apparaissent peu diversifiés et différenciés (géographiquement et vis-à-vis du débourrement), ils ne semblent pas avoir été soumis à l'action de la sélection, du moins avec notre échantillon, et ils ne présentent quasiment pas de mutations non synonymes. Aucune association significative n'a été détectée entre les variants alléliques des trois sites nucléotidiques étudiés et la variabilité phénotypique du débourrement. La pertinence du choix de sites nucléotidiques au sein des deux gènes candidats GALA et DAG2 pour des études d'association se pose clairement et sera discutée dans le chapitre de synthèse.

E. Conclusions et perspectives



① La stratégie mise en place est-elle pertinente pour l'étude du déterminisme génétique du débourrement ?

- L'utilisation de marqueurs orthologues dérivés des EST, identifiés à l'aide d'expériences de génomique menées sur le chêne, a permis l'alignement des douze groupes de liaison du chêne et du châtaignier. Ce travail a fourni les résultats permettant la comparaison des QTL contrôlant le débourrement chez les deux espèces, et par là même l'étude de la conservation du caractère entre ces deux mêmes espèces. Deux résultats principaux sont à rappeler :

- le nombre et les effets (variance phénotypique expliquée) des QTL impliqués dans le contrôle du caractère semblent conservés entre les deux espèces malgré les limites de l'approche discutées dans le chapitre C ;
- la stabilité de ces QTL semble respectée au niveau inter spécifique, après avoir été vérifiée au niveau intra spécifique.

Le résultat le plus remarquable est qu'il semblerait que les mêmes régions génomiques soient impliquées dans le contrôle du caractère de débourrement, malgré les 60 millions d'années de divergence entre les deux espèces. Les mêmes gènes ne sont pas nécessairement impliqués dans le contrôle du caractère mais cette stabilité des QTL suggère très fortement que les mêmes gènes soient en cause. Cette stabilité vient renforcer l'hypothèse d'un groupe de gènes uniques contrôlant le caractère au sein de chaque espèce, et en particulier du chêne, pour lequel les QTL sont stables pour différents environnements. En revanche, l'existence d'un grand nombre de locus impliqués dans le débourrement n'est pas favorable à la recherche de gènes candidats pour le caractère. En effet, l'implication d'un grand nombre de gènes tend à « diluer » la contribution de chacun d'entre eux, mais par contre à favoriser les effets de covariation (dus à des associations d'allèles) des gènes différents. Dans le cadre de cette thèse, les études d'association étaient malheureusement limitées à l'analyse de trois sites SNP, ce qui semble insuffisant pour mettre en évidence l'effet combiné de plusieurs locus sur le débourrement.

- La caractérisation de gènes candidats sur la base de critères d'expression mesurée durant le processus de débourrement a répondu, en partie, à l'objectif d'identification de gènes contrôlant potentiellement le caractère :

- trois des cinq gènes candidats « expressionnels », cartographiés chez au moins une des deux espèces, sont colocalisés avec un QTL impliqué dans le débourrement et conservé entre le chêne et le châtaignier ;
- les 5 GC, identifiés par approche transcriptomique, se révèlent, en moyenne, plus diversifiés que les GC identifiés dans la littérature, à l'exception du transcrit codant potentiellement pour DAG2 ;
- deux des GC ont révélé une structure de la diversité nucléotidique compatible avec l'action de la sélection naturelle, à savoir RASI et AUX-REP.

Cependant, la colocalisation entre un GC et un QTL contrôlant le caractère n'implique pas forcément que le gène considéré soit responsable de l'effet observé. Inversement, l'absence de colocalisation ne permet pas de conclure sur l'absence d'effet du gène sur le débourrement puisque seul un sous ensemble des QTL contrôlant le caractère a été détecté. Par ailleurs, les niveaux de diversité nucléotidique présents au sein des GC ne préfigurent en rien le comportement de ces gènes vis-à-vis de l'action de la sélection naturelle. C'est la structuration de cette diversité qui va nous apporter des éléments de réponse sur l'action potentielle de la sélection naturelle. Dans notre étude, l'échantillonnage était optimisé pour l'estimation de la différenciation entre populations contrastées, à savoir beaucoup de populations échantillonnées (9), avec peu d'individus par population (3). A l'inverse, les études liées à la recherche de signatures moléculaires de la sélection sont efficaces lorsque l'échantillonnage intra population est conséquent. Il apparaît nécessaire d'augmenter cet échantillonnage au sein d'un nombre restreint de populations afin d'estimer avec plus de précision l'écart à la neutralité des gènes considérés, tout en conservant un nombre élevé de populations pour une bonne estimation des niveaux de différenciation. De plus, dans le cas d'un gène fortement différencié pour le caractère de débourrement, cet échantillon supplémentaire permettrait également de valider ces niveaux de différenciation au niveau intra population, entre groupes d'individus contrastés pour le débourrement.

- La stratégie employée consistait à étudier simultanément la diversité nucléotidique au sein de fragments restreints de gènes. Une autre possibilité eut été de focaliser l'étude sur la totalité de la séquence d'un gène. Cette seconde stratégie aurait permis d'estimer avec plus de précision l'écart à la neutralité du gène étudié ainsi que l'étendue du déséquilibre de liaison au sein d'un même locus, deux questions auxquelles notre étude n'a pu répondre que très partiellement. Cependant, cette stratégie alternative aurait, dans le même temps, limité la

portée de l'étude. En effet, comme indiqué précédemment, le caractère semble être sous le contrôle d'un grand nombre de gènes contribuant chacun faiblement au déterminisme du caractère. Il apparaît donc indispensable de s'intéresser à un grand nombre de gènes candidats afin de tester le maximum de combinaisons de locus. A ce titre, il n'est donc pas exclu que les deux gènes DAG2 et GALA agissent sur le caractère en interaction avec d'autres gènes même si l'étude n'a pas permis la mise en évidence d'associations significatives entre le polymorphisme au sein de ces deux locus et la variation phénotypique du débourrement. Enfin, pour estimer au mieux l'étendue du déséquilibre de liaison, il aurait également été intéressant de pouvoir étudier des GC situés à proximité sur le même groupe de liaison.

En conclusion, il apparaît que la démarche menant à l'identification des gènes candidats les plus pertinents à tester en études d'association doit nécessairement être intégrative. Chaque élément considéré individuellement ne peut en effet permettre de conclure avec certitude sur l'implication probable du gène dans le contrôle du caractère. Les trois critères principaux qu'il nous semble indispensable de prendre en compte sont : i/ la colocalisation entre le GC et un QTL impliqué dans le débourrement, ii/ l'existence de niveau de différenciation élevé pour le caractère de débourrement au sein du GC, iii/ la compatibilité de la structure nucléotidique du GC avec l'action de la sélection naturelle. On peut également y ajouter un quatrième qui est la variation clinale des fréquences alléliques avec la notation de débourrement.

② Evaluation des techniques employées

- L'utilisation combinée de banques SSH, filtres bactériens puis ADNc à haute densité et, enfin, RT-PCR quantitative en temps réel dans le cadre de l'approche transcriptomique s'est révélée efficace, à plusieurs titres :
 - de nombreux transcrits sans équivalent dans les bases de données protéiques ou d'EST ont été isolés ;
 - les mécanismes moléculaires fortement régulés au cours du débourrement (stress hydrique, oxydatif et thermique ; division cellulaire ; métabolisme énergétique...) ont été mis en évidence ;
 - un certain nombre de gènes candidats différentiellement exprimés entre les stades de débourrement a pu être identifié.

L'étude est originale puisqu'elle nous a permis de suivre le niveau d'expression d'un grand nombre de transcrits en conditions naturelles durant la totalité du processus de débourrement. En revanche, le suivi de l'expression des transcrits pourrait être sensiblement amélioré par un échantillonnage plus précis au cours de la quiescence du bourgeon. Cet objectif était, en partie, poursuivi dans notre étude avec un prélèvement de bourgeons quiescents plus d'un mois avant le premier signe visible d'activité et quelques jours avant le débourrement. Mais les quantités très limitées de bourgeons nous ont amené à considérer ce matériel comme correspondant à un seul stade. Un échantillonnage plus approprié consisterait à réaliser deux à trois prélèvements dans les quelques semaines précédant le début du débourrement afin d'identifier avec plus de précision les signaux moléculaires précoces de reprise d'activité du méristème.

- Au-delà de la localisation des GC vis-à-vis des QTL impliqués dans le débourrement, le choix des EST comme source principale de marqueurs orthologues s'est révélé très pertinent :
 - les EST ont, en effet, confirmé leur puissance comme outil privilégié des études de cartographie comparée ;

- des clusters de gènes localisés dans la même région génomique et impliqués dans des processus physiologiques ou développementaux similaires apparaissent et pourraient être régulés de manière coordonnée.

Cependant, ces clusters ne sont pas significatifs vu le faible nombre de gènes cartographiés dans notre étude. La cartographie des gènes candidats pour le caractère de débourrement semble pertinente mais l'identification des GC doit s'appuyer sur d'autres critères connaissant les intervalles de confiance très larges des QTL (centaines, voire milliers de gènes). Il est toutefois possible de réduire ces intervalles par l'utilisation d'un plus grand nombre d'individus dans la population de cartographie. Une autre approche consiste à synthétiser, par méta analyse, l'information provenant de la recherche de QTL effectuée dans des pedigrees indépendants. Les différents QTL détectés dans une région donnée du génome peuvent représenter plusieurs estimations de la position d'un même et unique QTL. Dans la plupart des cas, la longueur de l'intervalle de confiance de la position consensus est réduite si l'on compare au plus petit intervalle de confiance du QTL initial (Arcade *et al.*, 2004). Par ailleurs, la procédure employée pour l'identification des sites polymorphes au sein des gènes candidats pourrait être améliorée. En effet, bien que l'excès de singletons observé sur la plupart des GC ne résulte pas directement de l'étape de clonage, il est possible d'améliorer cette dernière afin de s'assurer de l'absence d'artefacts de sites polymorphes simplement dus aux erreurs de l'enzyme *Taq* polymérase. Pour cela, plusieurs clones par individu peuvent être séquencés afin de s'assurer de la présence d'un polymorphisme et/ou le taux d'erreurs lors de l'amplification des inserts peut être diminué par l'utilisation d'une *Taq* polymérase « proofreading ».

③ Perspectives de cette étude

La suite immédiate de cette étude est la poursuite des études d'association sur l'ensemble des gènes candidats identifiés durant ce travail de thèse afin de rechercher des associations significatives entre les variations alléliques au sein de ces GC, et entre ces GC, et la variation phénotypique de la date de débourrement. De plus, un certain nombre d'améliorations d'ordre technique ou stratégique peut être apporté à ce travail, comme la caractérisation de la séquence complète de certains gènes. A cette fin, deux techniques peuvent être envisager : i/ le criblage de banques d'ADNc, qui sera rendu possible d'ici peu grâce à la disponibilité d'une banque d'ADNc de bourgeons dont la construction est en cours au sein du Laboratoire (Salin, communication personnelle), ii/ l'utilisation de la technique RACE (Amplification Rapide des Extrémités des ADNc), susceptible cependant de détruire l'intégrité des ADNc. La caractérisation complète au niveau génomique pourra se faire par des procédés de « gene walking » permettant l'accès aux régions introniques et flanquantes du gène. Ce travail ne pourra s'effectuer que sur un nombre restreint de locus et la sélection de ces derniers doit se faire en prenant en compte les trois critères principaux précédemment évoqués, à savoir la localisation de ces GC vis-à-vis des QTL contrôlant le débourrement, la présence de signatures moléculaires de la sélection naturelle sur la diversité nucléotidique et l'existence de niveaux de différenciation élevés entre populations contrastées pour le débourrement. Afin d'estimer au mieux le second critère (neutralité sélective), l'échantillonnage doit également être renforcé au sein d'un nombre restreint de populations, ce qui permettra également de confirmer ou d'infirmer la présence de niveaux de différenciation élevés entre individus contrastés appartenant à la même population.

Au-delà des améliorations à apporter à la stratégie adoptée dans cette thèse, lorsque des associations auront été mises en évidence entre les gènes et le caractère de débourrement, il sera nécessaire de rechercher les bases moléculaires de ces dernières. Pour ce faire, un certain nombre d'expériences complémentaires sont possibles :

- localisation *in situ* de l'expression du gène par des expériences d'hybridation *in situ* ;
- estimation de la fonction potentielle des gènes par la recherche de mutants dans les bases de données d'*Arabidopsis* ;
- modification de la structure des protéines ;

- estimation de la variation individuelle du niveau d'expression par cartographie des régions impliquées dans la variation d'expression (Transcript Quantitative Loci : TQL, Protein Quantitative Loci : PQL), ou étude des niveaux de différenciation au niveau expressionnel.

Références bibliographiques

- Aalen, R.B., Opsahl-Ferstad, H.G., Linnestad, C. et Olsen, O.A. 1994. *Transcripts encoding an oleosin and a dormancy-related protein are present in both the aleurone layer and the embryo of developing barley (*Hordeum vulgare* L.) seeds.* **Plant Journal** 5 (3): 385-396.
- Aas, G. 1993. *Taxonomical impact of morphological variation in *Quercus robur* and *Q. petraea*: a contribution to the hybrid controversy.* **Annals of forest science** 50 (Suppl. 1): 107-113.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. et Lipman, D.J. 1997. *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* **Nucleic Acids Research** 25 (17): 3389-3402.
- Arcade, A., Labourdette, A., Falque, M., Mangin, B., Chardon, F., Charcosset, A. et Joets, J. 2004. *BioMercator: integrating genetic maps and QTL towards discovery of candidate genes.* **Bioinformatics** 20 (14): 2324-2326.
- Bacilieri, R., Ducouso, A. et Kremer, A. 1995. *Genetic, morphological, ecological and phenological differentiation between *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. in a mixed stand in the northwest of France.* **Silvae Genetica** 44 (1): 1-10.
- Bacilieri, R., Ducouso, A., Petit, R.J. et Kremer, A. 1996. *Mating system and asymmetric hybridization in a mixed stand of European oaks.* **Evolution** 50 (2): 900-908.
- Badeck, F.-W., Bondeau, A., Bottcher, K., Doktor, D., Lucht, W., Schaber, J. et Sitch, S. 2004. *Responses of spring phenology to climate change.* **New Phytologist** 162 (2): 295-309.
- Bailey, P.C., Lycett, G.W. et Roberts, J.A. 1996. *A molecular study of dormancy breaking and germination in seeds of *Trollius ledebouri*.* **Plant Molecular Biology** 32 (3): 559-564.
- Barakat, A., Szick-Miranda, K., Chang, I.F., Guyot, R., Blanc, G., Cooke, R., Delseny, M. et Bailey-Serres, J. 2001. *The organization of cytoplasmic ribosomal protein genes in the *Arabidopsis* genome.* **Plant Physiology** 127 (2): 398-415.
- Barreneche, T., Bodenes, C., Lexer, C., Trontin, J.F., Fluch, S., Streiff, R., Plomion, C., Roussel, G., Steinkellner, H., Burg, K., Favre, J.M., Glossl, J. et Kremer, A. 1998. *A genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR,*

- microsatellite, minisatellite, isozyme and 55 rDNA markers*. **Theoretical and Applied Genetics** 97 (7): 1090-1103.
- Barreneche, T., Casasoli, M., Russell, K., Akkac, A., Meddour, H., Plomion, C., Villani, F. et Kremer, A. 2004. *Comparative mapping between Quercus and Castanea using simple-sequence repeats (SSRs)*. **Theoretical and Applied Genetics** 108 (3): 558-566.
- Barton, N.H. et Keightley, P.D. 2002. *Understanding quantitative genetic variation*. **Nature Reviews Genetics** 3: 11-21.
- Bastow, R., Mylne, J.S., Lister, C., Lippman, Z., Martienssen, R.A. et Dean, C. 2004. *Vernalization requires epigenetic silencing of FLC by histone methylation*. **Nature** 427 (6970): 164-167.
- Beaubien, E.G. et Freeland, H.J. 2000. *Spring phenology trends in Alberta, Canada: links to ocean temperature*. **International Journal of Biometeorology** 44: 53-59.
- Beaumont, M.A. et Nichols, R.A. 1996. *Evaluating loci for use in the genetic analysis of population structure*. **Proceedings of the Royal Society of London Series B-Biological Sciences** 263 (1377): 1619-1626.
- Beavis, W.D. 1995. *The power and deceit of QTL experiments: lessons from comparative QTL studies*. **49th Annual Corn and Sorghum Industry Research Conference**, Washington DC.
- Bechtold, N., Ellis, J. et Pelletier, G. 1993. *In planta Agrobacterium mediated gene transfer by infiltration of adult Arabidopsis thaliana plants*. **Comptes Rendus de l'Academie des Sciences. Serie III, Sciences de la Vie** 316 (10): 1194-119.
- Becker, M. et Levy, G. 1990. *Le point sur l'écologie comparée du chêne sessile et du chêne pédonculé*. **Revue Forestiere Francaise** 42: 148-153.
- Birks, H.J.B. 1986. Late-Quaternary biotic changes in terrestrial and lacustrine environments, with particular reference to north-west Europe. In: Berglund E. (ed), *Handbook of Holocene Palaeoecology and Palaeohydrology*, **Wiley and Sons Ltd**, pp. 3-65.
- Bodénès, C., Joandet, S., Laigret, F. et Kremer, A. 1997. *Detection of genomic regions differentiating two closely related oak species Quercus petraea (Matt.) Liebl. and Quercus robur L.* **Heredity** 78 (4): 433-444.
- Bohncke, S.J.P. 1993. *Lateglacial environmental changes in the Netherlands: Spatial and temporal patterns : A contribution to the 'North Atlantic seaboard programme' of IGCP-253, 'Termination of the Pleistocene'*. **Quaternary Science Reviews** 12 (8): 707-717.

- Bouchez, D. et Höfte, H. 1998. *Functional Genomics in Plants*. **Plant Physiology** 118 (3): 725-732.
- Boury, C. 2003. *Contribution à la construction d'un micro-réseau à ADNc pour l'étude du transcriptome du pin maritime en condition de stress hydrique*. Université Denis Diderot Paris VII, pp. 53.
- Bradshaw, H.D., Jr. et Stettler, R.F. 1995. *Molecular genetics of growth and development in Populus. IV. Mapping QTLs with large effects on growth, form, and phenology traits in a forest tree*. **Genetics** 139 (2): 963-973.
- Bray, E.A., Bailey-Serres, J. et Weretilnyk, E. 2000. Responses to abiotic stresses. In: Buchanan B., Gruissem W., Jones R. (eds), *Biochemistry & molecular biology of plants*, **American Society of Plant Physiologists**, Rockville, pp. 1158-1203.
- Brendel, O., Pot, D., Plomion, C., Rozenberg, P. et Guehl, J.M. 2002. *Genetic parameters and QTL analysis of 13C and ring width in maritime pine*. **Plant, Cell and Environment** 25: 945-953.
- Brewer, S., Cheddadi, R., Beaulieu, J.L.d. et Reille, M. 2002. *The spread of deciduous Quercus throughout Europe since the last glacial period*. **Forest Ecology and Management** 156 (1/3): 27-48.
- Brookes, A.J. 1999. *The essence of SNPs*. **Gene** 234 (2): 177-186.
- Brown, C.L. 1971. Primary growth. In: Zimmermann M.H., Brown C.L. (eds), *Trees: Structure and Function*, **Springer Verlag**, New York, pp. 1-66.
- Brown, G.R., Bassoni, D.L., Gill, G.P., Fontana, J.R., Wheeler, N.C., Megraw, R.A., Davis, M.F., Sewell, M.M., Tuskan, G.A. et Neale, D.B. 2003. *Identification of quantitative trait loci influencing wood property traits in loblolly pine (Pinus taeda L.). III. QTL verification and candidate gene mapping*. **Genetics** 164 (4): 1537-1546.
- Brown, G.R., Gill, G.P., Kuntz, R.J., Langley, C.H. et Neale, D.B. 2004. *Nucleotide diversity and linkage disequilibrium in loblolly pine*. **Proceedings of the National Academy of Sciences of the United States of America** 101 (42): 15255-15260.
- Brown, G.R., Kadel, E.E., III, Bassoni, D.L., Kiehne, K.L., Temesgen, B., Buijtenen, J.P.v., Sewell, M.M., Marshall, K.A. et Neale, D.B. 2001. *Anchored reference loci in loblolly pine (Pinus taeda L.) for integrating pine genomics*. **Genetics** 159 (2): 799-809.
- Brownie, J., Shawcross, S., Theaker, J., Whitcombe, D., Ferrie, R., Newton, C. et Little, S. 1997. *The elimination of primer-dimer accumulation in PCR*. **Nucleic Acids Research** 25 (16): 3235-3241.

- Brumfield, R.T., Beerli, P., Nickerson, D.A. et Edwards, S.V. 2003. *The utility of single nucleotide polymorphisms in inferences of population history*. **Trends in Ecology & Evolution** 18 (5): 249-256.
- Bundock, P.C. et Henry, R.J. 2004. *Single nucleotide polymorphism, haplotype diversity and recombination in the *Isa* gene of barley*. **Theoretical and Applied Genetics** 109 (3): 543-551.
- Byrne, P.F. et McMullen, M.D. 1996. *Defining genes for agricultural traits: QTL analysis and the candidate gene approach*. **Probe** 7: 24-27.
- Caicedo, A.L., Stinchcombe, J.R., Olsen, K.M., Schmitt, J. et Purugganan, M.D. 2004. *Epistatic interaction between *Arabidopsis* *FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait*. **Proceedings of the National Academy of Sciences of the United States of America** 101 (44): 15670-15675.
- Calvo, A.P., Nicolas, C., Nicolas, G. et 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 (4): 623-630.
- Campbell, R.K. et Sugano, A.I. 1979. *Genecology of Bud-Burst Phenology in Douglas-Fir: Response to Flushing Temperature and Chilling*. **Botanical Gazette** 140 (2): 223-231.
- Cannell, M.G.R. et Smith, R.I. 1983. *Thermal time, chill days and prediction of budburst in *Picea sitchensis**. **Journal of Applied Ecology** 20 (3): 951-963.
- Cannell, M.G.R. et Smith, R.I. 1986. *Climatic warming, spring budburst and frost damage on trees*. **Journal of Applied Ecology** 23 (1): 177-191.
- Cardon, L.R. et Bell, J.I. 2001. *Association study designs for complex diseases*. **Nature Reviews Genetics** 2 (2): 91-99.
- Casoli, M. 2004. *Cartographie génétique comparée chez les Fagacées*. **Doctorat de Biologie Cellulaire et Moléculaire**. Université Bordeaux I, pp. 190.
- Casoli, M., Mattioni, C., Cherubini, M. et Villani, F. 2001. *A genetic linkage map of European chestnut (*Castanea sativa* Mill.) based on RAPD, ISSR and isozyme markers*. **Theoretical and Applied Genetics** 102 (8): 1190-1199.
- Casoli, M., Pot, D., Plomion, C., Monteverdi, M.C., Barreneche, T., Lauteri, M. et Villani, F. 2004. *Identification of QTLs affecting adaptive traits in *Castanea sativa* Mill.* **Plant, Cell and Environment** 27 (9): 1088-1101.

- Cervera, M.T., Remington, D., Frigerio, J.M., Storme, V., Ivens, B., Boerjan, W. et Plomion, C. 2000. *Improved AFLP analysis of tree species*. **Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere** 30 (10): 1608-1616.
- Chagne, D., Brown, G., Lalanne, C., Madur, D., Pot, D., Neale, D. et Plomion, C. 2003. *Comparative genome and QTL mapping between maritime and loblolly pines*. **Molecular Breeding** 12 (3): 185-195.
- Chang, S., Puryear, J. et Cairney, J. 1993. *A simple and efficient method for isolating RNA from pine trees*. **Plant Molecular Biology Reporter** 11 (2): 113-116.
- Chen, Z., Kolb, T.E. et Clancy, K.M. 2001. *Mechanisms of Douglas-fir resistance to western spruce budworm defoliation: bud burst phenology, photosynthetic compensation and growth rate*. **Tree Physiology** 21 (16): 1159-1169.
- Chiang, H.H., Hwang, I. et Goodman, H.M. 1995. *Isolation of the Arabidopsis GA4 Locus*. **Plant Cell** 7 (2): 195-201.
- Chtanova, T., Kemp, R.A., Sutherland, A.P.R., Ronchese, F. et Mackay, C.R. 2001. *Gene microarrays reveal extensive differential gene expression in both CD4+ and CD8+ type 1 and type 2 T cells*. **Journal of Immunology** 167 (6): 3057-3063.
- Chuine, I. et Beaubien, E.G. 2001. *Phenology is a major determinant of tree species range*. **Ecology Letters** 4 (5): 500-510.
- Chuine, I. et Cour, P. 1999. *Climatic determinants of budburst seasonality in four temperate-zone tree species*. **New Phytologist** 143 (2): 339-349.
- Cline, J., Braman, J. et Hogrefe, H. 1996. *PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases*. **Nucl. Acids Res.** 24 (18): 3546-3551.
- Coart, E., Lamote, V., Loose, M.d., Bockstaele, E.v., Lootens, P. et Roldan-Ruiz, I. 2002. *AFLP markers demonstrate local genetic differentiation between two indigenous oak species Quercus robur L. and Quercus petraea (Matt.) Liebl. in Flemish populations*. **Theoretical and Applied Genetics** 105 (2/3): 431-439.
- Collins, F.S., Brooks, L.D. et Chakravarti, A. 1998. *A DNA Polymorphism Discovery Resource for Research on Human Genetic Variation*. **Genome Research** 8 (12): 1229-1231.
- Crawley, M.J. et Akhteruzzaman, M. 1988. *Individual variation in the phenology of oak trees and its consequences for herbivorous insects*. **Functional Ecology** 2 (3): 409-415.
- Debeaujon, I. et Koornneef, M. 2000. *Gibberellin requirement for Arabidopsis seed germination is determined both by testa characteristics and embryonic abscisic acid*. **Plant Physiology** 122 (2): 415-424.

- Devitt, M.L. et Stafstrom, J.P. 1995. *Cell cycle regulation during growth-dormancy cycles in pea axillary buds*. **Plant Molecular Biology** 29 (2): 255-265.
- Diatchenko, L., Lau, Y.-F.C., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D. et Siebert, P.D. 1996. *Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries*. **Proceedings of the National Academy of Sciences of the United States of America** 93 (12): 6025-6030.
- Doerge, R.W. 2002. *Mapping and analysis of quantitative trait loci in experimental populations*. **Nature Reviews Genetics** 3: 43-52.
- Don, R., Cox, P., Wainwright, B., Baker, K. et Mattick, J. 1991. *'Touchdown' PCR to circumvent spurious priming during gene amplification*. **Nucl. Acids Res.** 19 (14): 4008-.
- Doust, A.N., Devos, K.M., Gadberry, M.D., Gale, M.D. et Kellogg, E.A. 2004. *Genetic control of branching in foxtail millet*. **Proceedings of the National Academy of Sciences of the United States of America** 101 (24): 9045-9050.
- Dubos, C., Provost, G.I., Pot, D., Salin, F., Lalane, C., Madur, D., Frigerio, J.M. et Plomion, C. 2003. *Identification and characterization of water-stress-responsive genes in hydroponically grown maritime pine (*Pinus pinaster*) seedlings*. **Tree Physiology** 23 (3): 169-179.
- Ducouso, A., Bacilieri, R., Demesure, B., Dumolin-Lapegue, S., Kremer, A., Petit, R. et Zanetto, A. 1997. *Geographic structure of genetic diversity for European deciduous oaks*. **Bulletin Technique - Office National des Forêts** 33.
- Ducouso, A. et Bordacs, S. 2004. *EUFORGEN, Technical Guidelines for genetic conservation and use for pedunculate and sessile oaks (*Quercus robur* and *Q. petraea*)*. **International Plant Genetic Resources Institute**, Rome, Italy, pp. 6.
- Ducouso, A., Guyon, J.P. et 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 (2/3): 775-782.
- Ducouso, A. et Petit, R. 1996. *Le geai des chênes, premier reboiseur européen*. **Forêt-Entreprise** 97: 60-64.
- Dvornyk, V., Sirvio, A., Mikkonen, M. et Savolainen, O. 2002. *Low nucleotide diversity at the *pall* locus in the widely distributed *Pinus sylvestris**. **Molecular Biology and Evolution** 19 (2): 179-188.

- Eriksson, G., Ekberg, I., Dormling, I., Matern, B. et Wettstein, D.v. 1978. *Inheritance of bud-set and bud-flushing in Picea abies (L.) Karst.* **Theoretical and Applied Genetics** 52 (1): 3-19.
- Ewing, B. et Green, P. 1998. *Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities.* **Genome Research** 8 (3): 186-194.
- Ewing, B., Hillier, L., Wendl, M.C. et Green, P. 1998. *Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment.* **Genome Research** 8 (3): 175-185.
- Falusi, M. et Calamassi, R. 1990. *Bud dormancy in beech (Fagus sylvatica L.). Effect of chilling and photoperiod on dormancy release of beech seedlings.* **Tree Physiology** 6 (4): 429-438.
- Falusi, M. et Calamassi, R. 2003. *Dormancy of Fagus sylvatica L. buds. III. Temperature and hormones in the evolution of dormancy in one-node cuttings.* **Plant Biosystems** 137 (2): 185.
- Farmer, R.E. 1968. *Sweetgum dormancy release: effects of chilling, photoperiod and genotype.* **Physiologia Plantarum** 21: 1241-1248.
- Feltus, F.A., Hart, G.E. et Paterson, A.H. 2003. *QTL correspondence in diverse sorghum populations: statistically significant overlap of tiller height and testa factor QTLs.* **International Plant and Animal Genome Conference**, Town & Country Convention Center San Diego, CA.
- Feurtado, J.A., Ambrose, S.J., Cutler, A.J., Ross, A.R.S., Abrams, S.R. et Kermode, A.R. 2004. *Dormancy termination of western white pine (Pinus monticola Dougl. Ex D. Don) seeds is associated with changes in abscisic acid metabolism.* **Planta** 218 (4): 630-639.
- Filatov, D.A. 2002. *ProSeq: a software for preparation and evolutionary analysis of DNA sequence data sets.* **Molecular Ecology Notes** 2: 621-624.
- Frewen, B.E., Chen, T.H.H., Howe, G.T., Davis, J., Rohde, A., Boerjan, W. et Bradshaw, H.D., Jr. 2000. *Quantitative trait loci and candidate gene mapping of bud set and bud flush in Populus.* **Genetics** 154 (2): 837-845.
- Gaut, B.S. et Long, A.D. 2003. *The Lowdown on Linkage Disequilibrium.* **Plant Cell** 15 (7): 1502-1506.
- Geldermann, H. 1975. *Investigations on inheritance of quantitative characters in animals by gene markers. I. methods.* **Theoretical and Applied Genetics** 46: 319-330.

- Germain, E. 1992. Le noyer. In: Gallais A., Bannerot H. (eds), *Amélioration des espèces végétales cultivées, objectifs et critères de sélection*, **INRA Editions**, Paris, pp. 620-632.
- Giraudat, J., Hauge, B.M., Valon, C., Smalle, J., Parcy, F. et Goodman, H.M. 1992. *Isolation of the Arabidopsis ABI3 Gene by Positional Cloning*. **Plant Cell** 4 (10): 1251-1261.
- Glazier, A.M., Nadeau, J.H. et Aitman, T.J. 2002. *Finding Genes That Underlie Complex Traits*. **Science** 298 (5602): 2345-2349.
- Gordon, D., Abajian, C. et Green, P. 1998. *Consed: a graphical tool for sequence finishing*. **Genome Research** 8: 195-202.
- Goudet, J. 2001. *FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3)*.
- Grandjean, G. et Sigaud, P. 1987. *Contribution to the taxonomy and ecology of oaks in the Berry region, France*. **Annales des Sciences Forestières** 44 (1): 35-65.
- Grattapaglia, D., Bertolucci, F.L. et Sederoff, R.R. 1995. *Genetic mapping of QTLs controlling vegetative propagation in Eucalyptus grandis and E. urophylla using a pseudo-testcross strategy and RAPD markers*. **Theoretical and Applied Genetics** 90: 933-947.
- Gualberti, G., Papi, M., Bellucci, L., Ricci, I., Bouchez, D., Camilleri, C., Costantino, P. et Vittorioso, P. 2002. *Mutations in the Dof zinc finger genes DAG2 and DAG1 influence with opposite effects the germination of Arabidopsis seeds*. **Plant Cell** 14 (6): 1253-1263.
- Guillet-Claude, C., Birolleau-Touchard, C., Manicacci, D., Rogowsky, P.M., Rigau, J., Murigneux, A., Martinant, J.P. et Barriere, Y. 2004. *Nucleotide diversity of the ZmPox3 maize peroxidase gene: relationships between a MITE insertion in exon 2 and variation in forage maize digestibility*. **BMC Genetics** 5: Art. No. 19.
- Guo, W.J., Bundithya, W. et Goldsbrough, P.B. 2003. *Characterization of the Arabidopsis metallothionein gene family: tissue-specific expression and induction during senescence and in response to copper*. **New Phytologist** 159 (2): 369-381.
- Gutierrez, C., Ramirez-Parra, E., Castellano, M.M. et Pozo, J.C.d. 2002. *G1 to S transition: more than a cell cycle engine switch*. **Current Opinion in Plant Biology** 5 (6): 480-486.
- Hardtke, C.S. et Berleth, T. 1998. *The Arabidopsis gene MONOPTEROS encodes a transcription factor mediating embryo axis formation and vascular development*. **EMBO Journal** 17 (5): 1405-1411.

- Hardtke, C.S., Ckurshumova, W., Vidaurre, D.P., Singh, S.A., Stamatiou, G., Tiwari, S.B., Hagen, G., Guilfoyle, T.J. et Berleth, T. 2004. *Overlapping and non-redundant functions of the Arabidopsis auxin response factors MONOPTEROS and NONPHOTOTROPIC HYPOCOTYL 4*. **Development** 131 (5): 1089-1100.
- He, Y. et Gan, S. 2004. *A novel zinc-finger protein with a proline-rich domain mediates ABA-regulated seed dormancy in Arabidopsis*. **Plant Molecular Biology** 54 (1): 1-9.
- Heide, O.M. 1993a. *Daylength and thermal time responses of budburst during dormancy release in some northern deciduous trees*. **Physiologia Plantarum** 88 (4): 531-540.
- Heide, O.M. 1993b. *Dormancy release in beech buds (Fagus sylvatica) requires both chilling and long days*. **Physiologia Plantarum** 89 (1): 187-191.
- Hill, W.G. et Robertson, A. 1968. *Linkage disequilibrium in finite populations*. **Theoretical and Applied Genetics** 38: 226-231.
- Hinderhofer, K. et Zentgraf, U. 2001. *Identification of a transcription factor specifically expressed at the onset of leaf senescence*. **Planta** 213 (3): 469-473.
- Hopkins, J. 2002. *Molecular immunology - gene regulation and signal transduction*. **Veterinary Immunology and Immunopathology** 87 (3/4): 245-249.
- Horvath, D.P., Anderson, J.V., Chao, W.S. et Foley, M.E. 2003. *Knowing when to grow: signals regulating bud dormancy*. **Trends in Plant Science** 8 (11): 534-540.
- Howe, G.T., Aitken, S.N., Neale, D.B., Jermstad, K.D., Wheeler, N.C. et Chen, T.H.H. 2003. *From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees*. **Canadian Journal of Botany** 81 (12): 1247-1266.
- Hrmova, M. et Fincher, G.B. 2001. *Structure-function relationships of beta -D-glucan endo- and exohydrolases from higher plants*. **Plant Molecular Biology** 47 (1/2): 73-91.
- Hunter, A.F. et Lechowicz, M.J. 1992. *Predicting the time of budburst in temperate trees*. **Journal of Applied Ecology** 29 (3): 597-604.
- Huntley, B. 1988. Europe. In: Huntley B., Webb I.T. (eds), *Vegetation History*, **Kluwer Academic Publishers**, London, pp. 341-384.
- Huntley, B. et Prentice, I.C. 1993. Europe. In: al. H.E.W.e. (ed), *Global Climates Since the Last Glacial Maximum*, **University of Minnesota Press**, USA, pp. 136-168.
- Ingvarsson, P.K. 2005. *Nucleotide Polymorphism and Linkage Disequilibrium Within and Among Natural Populations of European Aspen (Populus tremula L., Salicaceae)*. **Genetics** 169 (2): 945-953.

- IPCC. 2001. *IPCC Third Assessment Report - Climate Change 2001. Intergovernmental Panel on Climate Change.*
- Ismail, A.M., Hall, A.E. et Close, T.J. 1999. *Allelic variation of a dehydrin gene cosegregates with chilling tolerance during seedling emergence. Proceedings of the National Academy of Sciences of the United States of America* 96 (23): 13566-13570.
- Israelsson, M., Mellerowicz, E., Chono, M., Gullberg, J. et Moritz, T. 2004. *Cloning and Overproduction of Gibberellin 3-Oxidase in Hybrid Aspen Trees. Effects on Gibberellin Homeostasis and Development. Plant Physiology* 135 (1): 221-230.
- Jablanczy, A. 1971. Changes due to age in apical development in spruce and fir, *Bi-monthly research notes, Canadian Forest Service.*
- Jacamon, M. 1984. *Guide de dendrologie. Tome II, Feuillus. ENGREF, Nancy.*
- Jacobsen, J.V., Pearce, D.W., Poole, A.T., Pharis, R.P. et Mander, L.N. 2002. *Abscisic acid, phaseic acid and gibberellin contents associated with dormancy and germination in barley. Physiologia Plantarum* 115 (3): 428.
- Jarvis, S.B., Taylor, M.A., MacLeod, M.R. et Davies, H.V. 1996. *Cloning and characterisation of the cDNA clones of three genes that are differentially expressed during dormancy-breakage in the seeds of Douglas fir (Pseudotsuga menziesii). Journal of Plant Physiology* 147 (5): 559-566.
- Jauffrès, A. 2003. *Cartographie comparée de Fagacées à l'aide de microsatellites. UFR sciences et techniques côte Basque.*
- Jauffrès, A. 2004. *Diversité nucléotidique de gènes candidats du débourrement chez le chêne sessile (Quercus petraea (Matt.) Liebl.). Universités Bordeaux I - Bordeaux II, pp. 27.*
- Jensen, J.S. 1993. *Variation of growth in Danish provenance trials with oak (Quercus robur L and Quercus petraea Mattuschka Liebl). Annales des Sciences Forestieres* 50: 203-207.
- Jermstad, K.D., Bassoni, D.L., Jech, K.S., Ritchie, G.A., Wheeler, N.C. et Neale, D.B. 2003. *Mapping of Quantitative Trait Loci Controlling Adaptive Traits in Coastal Douglas Fir. III. Quantitative Trait Loci-by-Environment Interactions. Genetics* 165 (3): 1489-1506.
- Jermstad, K.D., Bassoni, D.L., Jech, K.S., Wheeler, N.C. et Neale, D.B. 2001. *Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. I. Timing of vegetative bud flush. Theoretical and Applied Genetics* 102 (8): 1142-1151.

- Johnson, R.R., Cranston, H.J., Chaverra, M.E. et Dyer, W.E. 1995. *Characterization of cDNA clones for differentially expressed genes in embryos of dormant and nondormant Avena fatua L. caryopses*. **Plant Molecular Biology** 28 (1): 113-122.
- Jones, E.W. 1959. *Biological flora of the British Isles: Quercus L.* **Journal of Ecology (Oxford)** 47 (1): 160-222.
- Jones, P.G., Allaway, D., Gilmour, D.M., Harris, C., Rankin, D., Retzel, E.R. et Jones, C.A. 2002. *Gene discovery and microarray analysis of cacao (Theobroma cacao L.) varieties*. **Planta** 216 (2): 255-264.
- Karsai, A., Müller, S., Platz, S. et Hauser, M.T. 2002. *Evaluation of a home-made SYBR Green I reaction mixture for real-time PCR quantification of gene expression*. **BioTechniques** 32 (4): 1-5.
- Khan, A.A. et Shih, D.S. 2004. *Molecular cloning, characterization, and expression analysis of two class II chitinase genes from the strawberry plant*. **Plant Science** 166 (3): 753-762.
- Kim, J., Chung, Y., Paek, K., Park, Y., Kim, J., Yu, S., Oh, B. et Shin, J. 1999. *Isolation and characterization of a cDNA encoding the cysteine proteinase inhibitor, induced upon flower maturation in carnation using suppression subtractive hybridization*. **Molecules and Cells** 9 (4): 392-392.
- Kimura, M. 1983. *The neutral theory of Molecular Evolution*. **Cambridge University Press**, Cambridge, Massachusetts.
- Kleinschmit, J., Kremer, A. et Roloff, A. 1995a. *Sind stieleiche und traubeneiche zwei getrennte arten?* **AFZ** 26: 1453-1456.
- Kleinschmit, J.R.G., Bacilieri, R., Kremer, A. et Roloff, A. 1995b. *Comparison of morphological and genetic traits of pedunculate oak (Q. robur L.) and sessile oak (Q. petraea (Matt.) Liebl.)*. **Silvae Genetica** 44 (5/6): 256-269.
- Koornneef, M., Bentsink, L. et Hilhorst, H. 2002. *Seed dormancy and germination*. **Current Opinion in Plant Biology** 5 (1): 33-36.
- Kosambi, D.D. 1944. *The estimation of map distances from recombination values*. **Annu. Eugenics** 12: 172-175.
- Kramer, K. 1994. *Selecting a model to predict the onset of growth of Fagus sylvatica*. **Journal of Applied Ecology** 31 (1): 172-181.
- Kremer, A., Dupouey, J.L., Deans, J.D., Cottrell, J., Csaikl, U., Finkeldey, R., Espinel, S., Jensen, J., Kleinschmit, J., Dam, B.v., Ducouso, A., Forrest, I., Lopez de Heredia, U., Lowe, A.J., Tutkova, M., Munro, R.C., Steinhoff, S. et Badeau, V. 2002a. *Leaf*

- morphological differentiation between Quercus robur and Quercus petraea is stable across western European mixed oak stands. Annals of forest science* 59 (7): 777-787.
- Kremer, A., Petit, R.J. et Ducouso, A. 2002b. *Developmental biology and genetic diversity of sessile and pedunculate oak. Revue Forestiere Francaise* 54 (2): 111-130.
- Kuittinen, H. et Aguade, M. 2000. *Nucleotide Variation at the CHALCONE ISOMERASE Locus in Arabidopsis thaliana. Genetics* 155 (2): 863-872.
- Kulikova, T., Aldebert, P., Althorpe, N., Baker, W., Bates, K., Browne, P., van den Broek, A., Cochrane, G., Duggan, K., Eberhardt, R., Faruque, N., Garcia-Pastor, M., Harte, N., Kanz, C., Leinonen, R., Lin, Q., Lombard, V., Lopez, R., Mancuso, R., McHale, M., Nardone, F., Silventoinen, V., Stoehr, P., Stoesser, G., Tuli, M.A., Tzouvara, K., Vaughan, R., Wu, D., Zhu, W. et Apweiler, R. 2004. *The EMBL Nucleotide Sequence Database. Nucleic Acids Research* 32: 27-30.
- Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S.E. et Newburg, L. 1987. *MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics* 1 (2): 174-181.
- Lang, G.A. 1987. *Dormancy: a new universal terminology. HortScience* 22 (5): 817-820.
- Lanier, L. 1994. *Précis de sylviculture. ENGREF, Nancy.*
- Laval, H., Medus, J. et Roux, M. 1991. *Palynological and sedimentological records of Holocene human impact from the Etang de Berre, southeastern France. The Holocene* 1: 269-272.
- Law, R.D. et Suttle, J.C. 2003. *Transient decreases in methylation at 5'-CCGG-3' sequences in potato (Solanum tuberosum L.) meristem DNA during progression of tubers through dormancy precede the resumption of sprout growth. Plant Molecular Biology* 51 (3): 437-447.
- Le Corre, V., Machon, N., Petit, R.J. et Kremer, A. 1997. *Colonization with long-distance seed dispersal and genetic structure of maternally inherited genes in forest trees: a simulation study. Genetical Research* 69 (2): 117-125.
- Le Dantec, L., Chagne, D., Pot, D., Cantin, O., Garnier-Gere, P., Bedon, F., Frigerio, J.-M., Chaumeil, P., Leger, P., Garcia, V., Laigret, F., de Daruvar, A. et Plomion, C. 2004. *Automated SNP Detection in Expressed Sequence Tags: Statistical Considerations and Application to Maritime Pine Sequences. Plant Molecular Biology* 54 (3): 461-470.

- Leubner-Metzger, G. 2001. *Brassinosteroids and gibberellins promote tobacco seed germination by distinct pathways*. **Planta** 213 (5): 758-763.
- Levins, R. 1969. Dormancy as an adaptive strategy. In: Woolhouse H.W. (ed), *Dormancy and survival*, **Cambridge University Press**, pp. 1-10.
- Li, J. et Chory, J. 1997. *A Putative Leucine-Rich Repeat Receptor Kinase Involved in Brassinosteroid Signal Transduction*. **Cell** 90 (5): 929-938.
- Li, J., Nagpal, P., Vitart, V., McMorris, T.C. et Chory, J. 1996. *A Role for Brassinosteroids in Light-Dependent Development of Arabidopsis*. **Science** 272 (5260): 398-401.
- Liepe, K. 1993. *Growth-chamber trial on frost hardiness and field trial on flushing of sessile oak (Quercus petraea Liebl)*. **Annals of forest science** 50 (Suppl 1): 208-214.
- Lieth, H. 1974. Phenology and seasonality modeling, *Phenology and seasonality modeling.*, **Springer-Verlag Inc.**, New York USA, pp. 444.
- Lin, Y.R., Schertz, K.F. et Paterson, A.H. 1995. *Comparative Analysis of QTLs Affecting Plant Height and Maturity Across the Poaceae, in Reference to an Interspecific Sorghum Population*. **Genetics** 141 (1): 391-411.
- Liu, B.-H. 1998. *Statistical Genomics. Linkage, mapping and QTL analysis*. Boca Raton, New York.
- Long, A.D. et Langley, C.H. 1999. *The Power of Association Studies to Detect the Contribution of Candidate Genetic Loci to Variation in Complex Traits*. **Genome Research** 9 (8): 720-731.
- Long, J.C., Williams, R.C. et Urbanek, M. 1995. *An E-M Algorithm and Testing Strategy for Multiple-Locus Haplotypes*. **American Journal of Human Genetics** 56 (3): 799-810.
- Madoka, Y. et Mori, H. 2000a. *Acropetal disappearance of PsAD1 protein in pea axillary buds after the release of apical dominance*. **Plant and Cell Physiology** 41 (5): 556-564.
- Madoka, Y. et Mori, H. 2000b. *Two novel transcripts expressed in pea dormant axillary buds*. **Plant and Cell Physiology** 41 (3): 274-281.
- Mariette, S. 2001. *Mesure de la diversité génétique intra et inter-populations: échantillonnage intra-génome et choix des marqueurs. Applications expérimentales sur Quercus petraea, Quercus robur et Pinus pinaster*. **Doctorat de Sciences Forestières**. Ecole Nationale du Génie Rural, des Eaux et Forêts, pp. 90.
- Mariette, S., Cottrell, J., Csaikl, U.M., Goikoechea, P., König, A., Lowe, A.J., Van Dam, B.C., Barreneche, T., Bodenes, C., Streiff, R., Burg, K., Groppe, K., Munro, R.C., Tabbener, H. et Kremer, A. 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 (2-3): 72-79.
- Marth, G.T., Korf, I., Yandell, M.D., Yeh, R.T., Gu, Z.J., Zakeri, H., Stitzel, N.O., Hillier, L., Kwok, P.Y. et Gish, W.R. 1999. *A general approach to single-nucleotide polymorphism discovery. Nature Genetics* 23 (4): 452-456.
- Matziris, D.I. 1994. *Genetic variation in the phenology of flowering in black pine. Silvae Genetica* 43 (5/6): 321-328.
- Menzel, A. et Fabian, P. 1999. *Growing season extended in Europe. Nature* 397 (6721): 659.
- Muir, G., Fleming, C.C. et Schlotterer, C. 2000. *Species status of hybridizing oaks. Nature* 405 (6790): 1016.
- Murray, M.B., Cannell, M.G.R. et Smith, R.I. 1989. *Date of budburst of fifteen tree species in Britain following climatic warming. Journal of Applied Ecology* 26 (2): 693-700.
- Myers, R.M., Maniatis, T. et Lerman, S. 1987. *Detection and localization of single base changes by denaturing gradient gel-electrophoresis. Methods in Enzymology* 155: 501-527.
- Myneni, R.B., Keeling, C.D., Tucker, C.J., Asrar, G. et Nemani, R.R. 1997. *Increased plant growth in the northern high latitudes from 1981 to 1991. Nature (London)* 386 (6626): 698-702.
- Nakata, P.A. et McConn, M.M. 2002. *Sequential subtractive approach facilitates identification of differentially expressed genes. Plant Physiology and Biochemistry* 40 (4): 307-312.
- Neale, D.B. et Savolainen, O. 2004. *Association genetics of complex traits in conifers. Trends in Plant Science* 9 (7): 325-330.
- Nei, M. 1973. *Analysis of Gene Diversity in Subdivided Populations. Proceedings of the National Academy of Sciences of the United States of America* 70 (12, Part I): 3321-3323.
- Nei, M. 1987. *Molecular Evolutionary Genetics. Columbia Univ. Press, New York.*
- Ni, W., Xie, D., Hobbie, L., Feng, B., Zhao, D., Akkara, J. et Ma, H. 2004. *Regulation of Flower Development in Arabidopsis by SCF Complexes. Plant Physiology* 134 (4): 1574-1585.
- Nicolas, C., Nicolas, G. et Rodriguez, D. 1996. *Antagonistic effects of abscisic acid and gibberellic acid on the breaking of dormancy of Fagus sylvatica seeds. Physiologia Plantarum* 96 (2): 244-250.

- Ok, S., Park, H., Kim, J., Bahn, S., Bae, J., Suh, M., Jeung, J., Kim, K. et Shin, J. 2003. *Identification of differentially expressed genes during flower development in carnation (*Dianthus caryophyllus*)*. **Plant Science** 165 (2): 291-297.
- Or, E., Belausov, E., Popilevsky, I. et 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 & Biotechnology** 75 (2): 190-194.
- Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. et Sekiya, T. 1989. *Detection of Polymorphisms of Human DNA by Gel Electrophoresis as Single-Strand Conformation Polymorphisms*. **PNAS** 86 (8): 2766-2770.
- Paran, I. et Zamir, D. 2003. *Quantitative traits in plants: beyond the QTL*. **Trends in Genetics** 19: 303-306.
- Paterson, A.H., Bowers, J.E. et Chapman, B.A. 2004. *Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics*. **Proceedings of the National Academy of Sciences of the United States of America** 101 (26): 9903-9908.
- Paterson, A.H., Lin, Y., Li, Z., Schertz, K.F., Doebley, J.F., Pinson, S.R.M., Liu, S., Stansel, J.W. et Irvine, J.E. 1995. *Convergent domestication of cereal crops by independent mutations at corresponding genetic loci*. **Science (Washington)** 269 (5231): 1714-1718.
- Pawlowski, T.A., Bergervoet, J.H.W., Bino, R.J. et Groot, S.P.C. 2004. *Cell cycle activity and beta -tubulin accumulation during dormancy breaking of *Acer platanoides* L. seeds*. **Biologia Plantarum** 48 (2): 211-218.
- Penuelas, J. et Filella, I. 2001. *Phenology: Responses to a Warming World*. **Science** 294 (5543): 793-795.
- Perez-Flores, L., Carrari, F., Osuna-Fernandez, R., Rodriguez, M.V., Enciso, S., Stanelloni, R., Sanchez, R.A., Bottini, R., Iusem, N.D. et Benech-Arnold, R.L. 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 (390): 2071-2079.
- Pfaffl, M.W., Gerstmayer, B., Bosio, A. et Windisch, W. 2003. *Effect of zinc deficiency on the mRNA expression pattern in liver and jejunum of adult rats: monitoring gene expression using cDNA microarrays combined with real-time RT-PCR*. **Journal of Nutritional Biochemistry** 14 (12): 691-702.

- Pflieger, S., Lefebvre, V. et Causse, M. 2001. *The candidate gene approach in plant genetics: a review*. **Molecular Breeding** 7 (4): 275-291.
- Pinzani, P., Bonciani, L., Pazzagli, M., Orlando, C., Guerrini, S. et Granchi, L. 2004. *Rapid detection of *Oenococcus oeni* in wine by real-time quantitative PCR*. **Letters in Applied Microbiology** 38 (2): 118-124.
- Plomion, C., Durel, C.E. et O'Malley, D.M. 1996. *Genetic dissection of height in maritime pine seedlings raised under accelerated growth conditions*. **Theoretical and Applied Genetics** 93: 849-858.
- Plomion, C., Hurme, P., Frigerio, J.M., Ridolfi, M., Pot, D., Pionneau, C., Avila, C., Gallardo, F., David, H., Neutelings, G., Campbell, M., Canovas, F.M., Savolainen, O., Bodenes, C. et Kremer, A. 1999. *Developing SSCP markers in two *Pinus* species*. **Molecular Breeding** 5 (1): 21-31.
- Porth, I., Koch, M., Berenyi, M., Burg, A. et Burg, K. 2004. *Identification of adaptation specific differences in the mRNA expression of sessile and pedunculate based on osmotic stress induced genes*. **Tree Physiology** sous presse.
- Pukacka, S. et Wojkiewicz, E. 2002. *Carbohydrate metabolism in Norway maple and sycamore seeds in relation to desiccation tolerance*. **Journal of Plant Physiology** 159 (3): 273-279.
- Rameau, J.C. 1990. *Comportement dynamique du chêne pédonculé et du chêne sessile dans les successions forestières*. **Revue Forestiere Francaise** 42: 155-164.
- Raymond, M. et Rousset, F. 1995. *GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism*. **Journal of Heredity** 86: 248-249.
- Réaumur, R.A.F.d. 1735. *Observations du thermomètre, faites à Paris pendant l'année 1735, comparées avec celles qui ont été faites sous la ligne, à l'Isle de France, à Alger et en quelques unes de nos isles d'Amérique.*, *Mémoires de l'Académie Royale des Sciences*, Paris, pp. 545-576.
- Rhodes, D. et Hanson, A.D. 1993. *Quaternary Ammonium and Tertiary Sulfonium Compounds in Higher Plants*. **Annual Review of Plant Physiology and Plant Molecular Biology** 44 (1): 357-384.
- Richards, J.H. et Larson, P.R. 1981. *Morphology and development of *Populus deltoides* branches in different environments*. **Botanical Gazette** 142 (3): 382-393.
- Rinne, P., Tuominen, H. et Junttila, O. 1994. *Seasonal changes in bud dormancy in relation to bud morphology, water and starch content, and abscisic acid concentration in adult trees of *Betula pubescens**. **Tree Physiology** 14 (6): 549-561.

- Ritter, E., Gebhardt, C. et Salamini, F. 1990. *Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents*. **Genetics** 125: 645-654.
- Rohde, A., Howe, G.T., Olsen, J.E., Moritz, T., Van Montagu, M., Junttila, O. et Boerjan, W. 2000. Molecular Aspects of bud dormancy in trees. In: Minocha S.M.J.a.S.C. (ed), *Molecular Biology of Woody Plants*, **Kluwer Academic Publishers**, Netherlands, pp. 89-134.
- Rohde, A., van Montagu, M. et Boerjan, W. 1999. *The ABSCISIC ACID-INSENSITIVE 3 (ABI3) gene is expressed during vegetative quiescence processes in Arabidopsis*. **Plant, Cell and Environment** 22 (3): 261-270.
- Rothschild, M.F. et Soller, M. 1997. *Candidate gene analysis to detect genes controlling traits of economic importance in domestic livestock*. **Probe** 8: 13-20.
- Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X. et Rozas, R. 2003. *DnaSP, DNA polymorphism analyses by the coalescent and other methods*. **Bioinformatics** 19 (18): 2496-2497.
- Saghai-Marooif, M.A., Soliman, K.M., Jorgensen, R.A. et Allard, R.W. 1984. *Ribosomal DNA Spacer-Length Polymorphisms in Barley: Mendelian Inheritance, Chromosomal Location, and Population Dynamics*. **PNAS** 81 (24): 8014-8018.
- Saintagne, C., Bodenes, C., Barreneche, T., Pot, D., Plomion, C. et Kremer, A. 2004. *Distribution of genomic regions differentiating oak species assessed by QTL detection*. **Heredity** 92 (1): 20-30.
- Salse, J., Piegu, B., Cooke, R. et Delseny, M. 2004. *New in silico insight into the synteny between rice (Oryza sativa L.) and maize (Zea mays L.) highlights reshuffling and identifies new duplications in the rice genome*. **Plant Journal** 38 (3): 396-409.
- Sarvas, R. 1974. Investigations on the annual cycle of development of forest trees. II. Autumn dormancy and winter dormancy, *Metsantutkimuslaitoksen Julkaisuja*, Helsinki.
- Savenstrand, H., Brosche, M., Aneghagen, M. et Strid, A. 2000. *Molecular markers for ozone stress isolated by suppression subtractive hybridization: specificity of gene expression and identification of a novel stress-regulated gene*. **Plant, Cell and Environment** 23 (7): 689-700.
- Schams, D., Kohlenberg, S., Amselgruber, W., Berisha, B., Pfaffl, M.W. et Sinowatz, F. 2003. *Expression and localisation of oestrogen and progesterone receptors in the bovine mammary gland during development, function and involution*. **Journal of Endocrinology** 177 (2): 305-317.

- Schlarbaum, S.E. et Bagley, W.T. 1981. *Intraspecific genetic variation of Quercus rubra L., northern red oak*. **Silvae Genetica** 30 (2/3): 50-56.
- Schmitz, N., Xia, J.H. et Kermode, A.R. 2001. *Dormancy of yellow cedar seeds is terminated by gibberellic acid in combination with fluridone or with osmotic priming and moist chilling*. **Seed Science and Technology** 29 (2): 331-346.
- Scotti-Saintagne, C. 2003. *Distribution des régions génomiques différenciant deux espèces proches: le chêne sessile (Quercus petraea) et le chêne pédonculé (Quercus robur)*. **Doctorat de Biologie Forestière**. Université Henri Poincaré, Nancy I, pp. 169.
- Scotti-Saintagne, C., Bodénès, C., Barreneche, T., Bertocchi, E., Plomion, C. et Kremer, A. 2004. *Detection of quantitative trait loci controlling bud burst and height growth in Quercus robur L.* **Theoretical and Applied Genetics** 109: 1648-1659.
- Sewell, M.M., Bassoni, D.L., Megraw, R.A., Wheeler, N.C. et Neale, D.B. 2000. *Identification of QTLs influencing wood property traits in loblolly pine (Pinus taeda L.). I. Physical wood properties*. **Theoretical and Applied Genetics** 101: 1273-1281.
- Sewell, M.M. et Neale, D.B. 2000. Mapping quantitative traits in forest trees. In: Minocha S.M.J.a.S.C. (ed), *Molecular Biology of Woody Plants*, **Kluwer Academic Publishers**, Netherlands, pp. 407-423.
- Shimizu, S. et Mori, H. 1998. *Analysis of cycles of dormancy and growth in pea axillary buds based on mRNA accumulation patterns of cell cycle-related genes*. **Plant and Cell Physiology** 39 (3): 255-262.
- Simpson, E.H. 1949. *Measurement of diversity*. **Nature** 163: 688.
- Stafstrom, J.P., Ripley, B.D., Devitt, M.L. et Drake, B. 1998. *Dormancy-associated gene expression in pea axillary buds. Cloning and expression of PsDRM1 and PsDRM2*. **Planta** 205 (4): 547-552.
- Starkel, L. 1991. *Environmental changes at the Younger Dryas - Preboreal Transition and during the early Holocene: some distinctive aspects in central Europe*. **The Holocene** 1: 234-242.
- Steber, C.M. et McCourt, P. 2001. *A role for brassinosteroids in germination in Arabidopsis*. **Plant Physiology** 125 (2): 763-769.
- Steiner, K.C. 1979. *Patterns of variation in bud-burst timing among populations in several Pinus species*. **Silvae Genetica** 28 (5/6): 185-194.
- Steinhoff, S. 1993. *Results of species hybridization with Quercus robur L. and Q. petraea (Matt.) Liebl.* **Annals of forest science** 50 (Suppl. 1): 137-143.

- Steinkellner, H., Lexer, C., Turetschek, E. et Glossl, J. 1997. *Conservation of (GA)n microsatellite loci between Quercus species*. **Molecular Ecology** 6 (12): 1189-1194.
- Stram, Y., Kuznetzova, L., Guini, M., Rogel, A., Meiom, R., Chai, D., Yadin, H. et Brenner, J. 2004. *Detection and quantitation of Akabane and Aino viruses by multiplex real-time reverse-transcriptase PCR*. **Journal of Virological Methods** 116 (2): 147-154.
- Streiff, R. 1998. *Structure génétique spatiale et flux polliniques dans une chênaie mixte: Quercus petraea (Matt.) Liebl., Quercus robur L.* **Doctorat de Biologie des Populations**. Université Montpellier II, pp. 84.
- Sung, S. et Amasino, R.M. 2004. *Vernalization in Arabidopsis thaliana is mediated by the PHD finger protein VIN3*. **Nature** 427 (6970): 159-164.
- Syvänen, A.-C. 2001. *Assessing genetic variation: genotyping single nucleotide polymorphisms*. **Nature Reviews Genetics** 2 (12): 930-942.
- Tabor, H.K., Risch, N.J. et Myers, R.M. 2002. *Candidate-gene approaches for studying complex genetic traits: practical considerations*. **Nature Reviews Genetics** 3 (5): 391-A396.
- Taji, T., Ohsumi, C., Iuchi, S., Seki, M., Kasuga, M., Kobayashi, M., Yamaguchi-Shinozaki, K. et Shinozaki, K. 2002. *Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in Arabidopsis thaliana*. **Plant Journal** 29 (4): 417-426.
- Tajima, F. 1989. *Statistical method for testing the neutral mutation hypothesis by DNA polymorphism*. **Genetics** 123: 585-595.
- Tamura, F., Tanabe, K. et Itai, A. 2002. *Regulation of endodormancy in Japanese pear*. **Acta Horticulturae** 2 (No.587): 325-336.
- Tanksley, S.D. 1993. *Mapping polygenes*. **Annual Review of Genetics** 27: 205-233.
- Tanksley, S.D., Ganal, M.W. et Martin, G.B. 1995. *Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes*. **Trends in Genetics** 11 (2): 63-68.
- Temesgen, B., Brown, G.R., Harry, D.E., Kinlaw, C.S., Sewell, M.M. et Neale, D.B. 2001. *Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (Pinus taeda L.)*. **Theoretical and Applied Genetics** 102 (5): 664-675.
- Thomas, C., Meyer, D., Humber, C. et Steinmetz, A. 2004. *Spatial expression of a sunflower SERK gene during induction of somatic embryogenesis and shoot organogenesis*. **Plant Physiology and Biochemistry** 42 (1): 35-42.

- Thornsberry, J.M., Goodman, M.M., Doebley, J., Kresovich, S., Nielsen, D. et Buckler, E.S.I. 2001. *Dwarf8 polymorphisms associate with variation in flowering time*. **Nature Genetics** 28 (3): 286-289.
- Tikkanen, O.P. et Tiitto, R.J. 2003. *Phenological variation as protection against defoliating insects: the case of *Quercus robur* and *Operophtera brumata**. **Oecologia** 136 (2): 244-251.
- Tsarouhas, V., Gullberg, U. et Lagercrantz, U. 2002. *An AFLP and RFLP linkage map and quantitative trait locus (QTL) analysis of growth traits in *Salix**. **Theoretical and Applied Genetics** 105: 277-288.
- Tsarouhas, V., Gullberg, U. et Lagercrantz, U. 2003. *Mapping of quantitative trait loci controlling timing of bud flush in *Salix**. **Hereditas** 138 (3): 172-178.
- Tuomi, J., Niemelä, P., Jussila, I., Vuorisalo, T. et Jormalainen, V. 1989. *Delayed budbreak: a defensive response of mountain birch to early-season defoliation?* **Oikos** 54 (87-91).
- van Buijtenen, J.P. 2001. *Genomics and quantitative genetics*. **Canadian Journal of Forest Research** 31: 617-622.
- Van Oijen, J.W. et Voorrips, R.E. 2001. *Joinmap 3.0, software for the calculation of genetic linkage maps*.
- Vandesompele, J., De Paepe, A. et Speleman, F. 2002. *Elimination of Primer-Dimer Artifacts and Genomic Coamplification Using a Two-Step SYBR Green I Real-Time RT-PCR*. **Analytical Biochemistry** 303 (1): 95-98.
- Vegis, A. 1964. *Dormancy in Higher Plants*. **Annual Review of Plant Physiology** 15 (1): 185-224.
- Velichko, A.A. 1993. *Evolution of Landscapes and Climates of Northern Eurasia. Late Pleistocene-Holocene elements of prognosis*. **Moscow 'Nauke'** 2.
- Vignal, A., Milan, D., SanCristobal, M. et Eggen, A. 2002. *A review on SNP and other types of molecular markers and their use in animal genetics*. **Genetics, Selection, Evolution** 34 (3): 275-305.
- Vilaine, F., Palauqui, J.C., Amselem, J., Kusiak, C., Lemoine, R. et Dinant, S. 2003. *Towards deciphering phloem: a transcriptome analysis of the phloem of *Apium graveolens**. **Plant Journal** 36 (1): 67-81.
- Visser, M.E. et Holleman, L.J.M. 2001. *Warmer springs disrupt the synchrony of oak and winter moth phenology*. **Proceedings of the Royal Society of London. Series B, Biological Sciences** 268 (1464): 289-294.

- Voiblet, C., Duplessis, S., Encelot, N. et Martin, F. 2001. *Identification of symbiosis-regulated genes in Eucalyptus globulus-Pisolithus tinctorius ectomycorrhiza by differential hybridization of arrayed cDNAs*. **Plant Journal** 25 (2): 181-191.
- Walker, N.J. 2002. *A technique whose time has come*. **Science** 296 (5567): 557-559.
- Walther, G.R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J.C., Fromentin, J.M., Hoegh-Guldberg, O. et Bairlein, F. 2002. *Ecological responses to recent climate change*. **Nature** 416 (6879): 389-395.
- Wang, S.Y. et Faust, M. 1994. *Changes in the antioxidant system associated with budbreak in 'Anna' apple (Malus domestica Borkh.) buds*. **Journal of the American Society for Horticultural Science** 119 (4): 735-741.
- Wang, S.Y., Jiao, H.J. et Faust, M. 1991a. *Changes in ascorbate, glutathione, and related enzyme activities during thidiazuron-induced bud break of apple*. **Physiologia Plantarum** 82 (2): 231-136.
- Wang, S.Y., Jiao, H.J. et Faust, M. 1991b. *Changes in metabolic enzyme activities during thidiazuron-induced lateral budbreak of apple*. **HortScience** 26 (2): 171-173.
- Wang, S.Y., Jiao, H.J. et Faust, M. 1991c. *Changes in superoxide dismutase activity during thidiazuron-induced lateral budbreak of apple*. **HortScience** 26 (9): 1202-1204.
- Wang, S.Y., Jiao, H.J. et Faust, M. 1991d. *Changes in the activities of catalase, peroxidase, and polyphenol oxidase in apple buds during bud break induced by thidiazuron*. **Journal of Plant Growth Regulation** 10 (1): 33-39.
- Wareing, P.F. 1956. *Photoperiodism in Woody Plants*. **Annual Review of Plant Physiology** 7 (1): 191-214.
- Watterson, G.A. 1975. *On the number of segregating sites in genetical models without recombination*. **Theoretical Population Biology** 7: 256-276.
- Wayne, M.L. et McIntyre, L.M. 2002. *Combining mapping and arraying: an approach to candidate gene identification*. **Proceedings of the National Academy of Sciences of the United States of America** 99 (23): 14903-14906.
- Weber, J.A. et Nooden, L.D. 1973. *Induction of dormant bud formation in Myriophyllum verticillatum*. **Plant Physiology** 51 (Supplement): 57.
- Weir, B.S. et Cockerham, C.C. 1984. *Estimating F-statistics for the analysis of population structure*. **Evolution** 38: 1358-1370.
- Wellmer, F., Riechmann, J.L., Alves-Ferreira, M. et Meyerowitz, E.M. 2004. *Genome-wide analysis of spatial gene expression in Arabidopsis flowers*. **Plant Cell** 16 (5): 1314-1326.

- White, M.A., Running, S.W. et Thornton, P.E. 1999. *The impact of growing-season length variability on carbon assimilation and evapotranspiration over 88 years in the eastern US deciduous forest*. **International Journal of Biometeorology** 42 (3): 139-145.
- White, S.E. et Doebley, J.F. 1999. *The Molecular Evolution of terminal ear1, a Regulatory Gene in the Genus Zea*. **Genetics** 153 (3): 1455-1462.
- Wise, M.J. et Tunnacliffe, A. 2004. *POPP the question: what do LEA proteins do?* **Trends in Plant Science** 9 (1): 13-17.
- Worrall, J. 1983. *Temperature - bud-burst relationships in amabilis and subalpine fir provenance tests replicated at different elevations*. **Silvae Genetica** 32 (5/6): 203-209.
- Worrall, J. et Mergen, F. 1967. *Environmental and genetic control of dormancy in Picea abies*. **Physiologia Plantarum** 20: 733-745.
- Wright, J.W. 1976. *Introduction to forest genetics*. New York.
- Wright, S. 1969. *The Theory of Gene Frequencies, Evolution and the Genetics of Populations*, **University of Chicago Press**, Chicago.
- Yamaguchi, S., Smith, M.W., Brown, R.G.S., Kamiya, Y. et Sun, T. 1998. *Phytochrome regulation and differential expression of gibberellin 3 beta -hydroxylase genes in germinating Arabidopsis seeds*. **Plant Cell** 10 (12): 2115-2126.
- Yamauchi, Y., Ogawa, M., Kuwahara, A., Hanada, A., Kamiya, Y. et Yamaguchi, S. 2004. *Activation of gibberellin biosynthesis and response pathways by low temperature during imbibition of Arabidopsis thaliana seeds*. **Plant Cell** 16 (2): 367-378.
- Yanagisawa, S. 2002. *The Dof family of plant transcription factors*. **Trends in Plant Science** 7 (12): 555-560.
- Yang, G.P., Ross, D.T., Kuang, W.W., Brown, P.O. et Weigel, R.J. 1999. *Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes*. **Nucleic Acids Research** 27 (6): 1517-1523.
- Zanetto, A. et Kremer, A. 1995. *Geographical Structure of Gene Diversity in Quercus-Petraea (Matt) Liebl .1. Monolocus Patterns of Variation*. **Heredity** 75: 506-517.
- Zeng, Y., Raimondi, N. et Kermode, A.R. 2003. *Role of an ABI3 homologue in dormancy maintenance of yellow-cedar seeds and in the activation of storage protein and Em gene promoters*. **Plant Molecular Biology** 51 (1): 39-49.
- Zhang, X., Friedl, M.A., Schaaf, C.B. et Strahler, A.H. 2004. *Climate controls on vegetation phenological patterns in northern mid- and high latitudes inferred from MODIS data*. **Global Change Biology** 10 (7): 1133-1145.

- Zhao, T.-Y., Meeley, R.B. et Downie, B. 2003. *Aberrant processing of a Maize Galactinol Synthase transcript is caused by heat stress*. **Plant Science** 165 (1): 245-256.
- Zhu, Y.L., Song, Q.J., Hyten, D.L., Van Tassell, C.P., Matukumalli, L.K., Grimm, D.R., Hyatt, S.M., Fickus, E.W., Young, N.D. et Cregan, P.B. 2003. *Single-Nucleotide Polymorphisms in Soybean*. **Genetics** 163 (3): 1123-1134.
- Zoldos, V., Siljak-Yakovlev, S., Papes, D., Sarr, A. et Panaud, O. 2001. *Representational difference analysis reveals genomic differences between *Q. robur* and *Q. suber*: implications for the study of genome evolution in the genus *Quercus**. **Molecular Genetics and Genomics** 265 (2): 234-241.

Résumé – Mots clés : changements globaux, débourrement, étude d'association, gène candidat.

La productivité et la distribution des espèces longévives telles que les arbres forestiers, sont intimement liées à la phénologie, et en particulier à la date de débourrement. Sous l'effet du réchauffement climatique, le processus de débourrement pourrait être fortement affecté, et la question de l'adaptation des essences forestières à ces changements se pose clairement. Leur capacité d'adaptation dépend, pour une part, de la diversité génétique présente dans les populations naturelles. Cette thèse a pour objectif de déterminer les régions génomiques contrôlant le débourrement chez les chênes afin de mettre au point des outils d'évaluation de la diversité adaptative de ce caractère.

Sur le plan des méthodes, nous avons procédé en trois étapes, associant des approches génétiques et moléculaires. Nous avons tout d'abord suivi une approche transcriptomique pour identifier des gènes candidats (GC) pertinents au regard de leur profil d'expression au cours du débourrement chez le chêne sessile (*Quercus petraea*). Cette étape a été menée à l'aide de banques SSH, d'expériences de macroarrays et de RT-PCR quantitative. Dans un deuxième temps, ce lot de gènes candidats et certains GC identifiés dans la littérature, ont été cartographiés chez *Quercus robur* (chêne pédonculé) et *Castanea sativa* (châtaignier européen), et leur position comparée à celle des QTL impliqués dans le débourrement et conservés chez les deux espèces. Une étude de diversité au niveau nucléotidique a été menée sur sept de ces GC, dont deux ont fait l'objet d'une étude d'association entre les variants nucléotidiques et la date de débourrement. L'étude a permis la mise en évidence des processus fonctionnels régulés au cours du débourrement, ainsi que l'identification d'un certain nombre de transcrits sans équivalent dans les bases de données. Le caractère de débourrement apparaît remarquablement conservé entre le chêne et le châtaignier et les EST se sont révélés un formidable outil pour les études de cartographie comparée. Le niveau de diversité moyen est élevé au sein des locus analysés, dont certains présentent une structure compatible (niveaux de différenciation, écart à la neutralité) avec l'action de pressions de sélection liées au débourrement. Cependant, aucune association significative n'a été détectée entre variants nucléotidiques et note de débourrement.

Ce travail a permis d'aboutir à un certain nombre de résultats intéressants et encourageants concernant le déterminisme génétique du caractère de débourrement et ouvrent des perspectives pour l'étude de la diversité adaptative dans les populations naturelles.

Abstract – Keywords: association study, bud burst, candidate gene, global change.

Productivity and distribution of long-lived species, such as forest trees, are closely related to phenological events, especially bud burst. Forest trees may be greatly affected by global warming in the near future and their adaptation is challenged by the rapid climatic changes. Their capacity to respond will depend on the levels of genetic diversity in natural populations. The main objective of this thesis is to identify genes contributing to the variability of bud burst in oaks in order to develop tools for assessing diversity of these genes in natural populations.

We used a three steps strategy, by using genetic and molecular approaches to identify genes of interest. First, we followed a transcriptomic approach in sessile oak (*Quercus petraea*) to identify relevant expressional candidate genes (CG), by using SSH libraries, macroarrays experiments and quantitative RT-PCR. These candidate genes, complemented by genes already published, were mapped for *Quercus robur* (pedunculate oak) and *Castanea sativa* (european chestnut) on existing genetic maps and their locations compared to those of QTLs of bud burst. Then, nucleotide diversity was assessed for seven CG, and we studied associations between nucleotide diversity and the date of bud burst for two of them. During this study, an inventory was made of transcripts that are differentially expressed during bud burst, some of these transcripts were reported for the first time. We also used EST as markers to align linkage groups between oak and chestnut, and we showed that QTLs of bud burst are remarkably conserved between the two species. When compared to other species and genes, average nucleotide diversity is high in oaks across the seven studied loci. Some genes exhibited diversity patterns (differentiation levels, departure from neutrality) suggesting that selection might be involved as an evolutionary force. However, no significant association between nucleotide variants and bud burst score has been found.

This work has produced several interesting and promising results that can be implemented for future research on adaptive diversity in oaks.