

Numéro d'ordre : 2903

Thèse

présentée à

L'UNIVERSITÉ DE BORDEAUX I

École Doctorale : Sciences du Vivant, Géosciences, Sciences de l'Environnement

par Manuela CASASOLI

pour obtenir le grade de

DOCTEUR

Spécialité : Biologie Cellulaire et Moléculaire

Cartographie génétique comparée chez les Fagacées

Soutenance publique le 14 décembre 2004

Membres du Jury :

Président :	Jean-Pierre RENAUDIN	Professeur, Université de Bordeaux I
Rapporteurs :	Jean-Marc GUEHL	Directeur de Recherche, INRA, Nancy
	Brigitte MANGIN	Directeur de Recherche, INRA, Toulouse
Examineurs :	Elisabeth DIRLEWANGER	Chargé de Recherche, INRA, Bordeaux
	Nathalie FRASCARIA	Maître de Conférences, Université d'Orsay, Paris XI
	Antoine KREMER	Directeur de Recherche, INRA, Pierroton (Directeur de thèse)

INRA-UMR BIOGECO, 69 Route d'Arcachon, BP 45, 33612 Cestas, France

Allo zio Alfredo

Perché il ricordo di te, *tranquillo e beato*,
resterà per sempre vivo dentro i miei occhi.

*Nel mezzo del cammin di nostra vita
mi ritrovai per una selva oscura
ché la diritta via era smarrita.*

(Dante Alighieri, Divina Commedia, Inferno)

GRAZIE!

Voici donc la fin de cette thèse franco-italienne et de toute une expérience humaine et scientifique que je porterai toujours dans mon cœur.

Je tiens tout d'abord à remercier les membres du jury : Jean-Marc Guehl et Brigitte Mangin pour avoir accepté de relire ce manuscrit, Elisabeth Dirlewanger, Nathalie Frascaria et Jean-Pierre Renaudin pour avoir accepté de faire partie du jury. Mais avant de commencer votre lecture je vous invite à découvrir tous les gens qui ont participé à ce travail, qui m'ont appris une infinité de choses et qui m'ont aidée, conseillée soutenue à tout moment.

Je voudrais remercier trois personnes tout particulièrement car sans eux je n'aurais jamais eu la possibilité de vivre une expérience aussi belle et enrichissante. Et pour vous expliquer le rôle de ces trois « grands » je vous raconte comment je suis arrivée dans ce laboratoire extraordinaire qu'est le laboratoire de génétique de Pierroton et comment je n'ai plus eu la force de m'en détacher ! C'était le mois de juillet 1998 et en faisant la biblio je suis tombée sur Plomion et al...intéressant...il fait de la cartographie génétique, c'est la France...j'y vais ! En janvier 1999 Christophe Plomion a accepté de me recevoir pendant trois semaines à Pierroton. Je suis arrivée en lui disant que je venais de commencer la cartographie génétique et que je voulais donc voir de quoi il s'agissait...en trois semaines j'ai compris tout de suite à qui j'avais à faire ! Le dynamisme scientifique personnifié, Christophe m'a appris comment construire une carte, il m'a donné la biblio qu'il fallait, il m'a montré les logiciels à utiliser (« Enjoy yourself !!! », il m'a dit), il m'a pris rendez-vous avec tous les chercheurs et les thésards de la station pour me permettre d'avoir une idée de l'ensemble des recherches qui se déroulaient à Pierroton. Ce court séjour à côté de Christophe m'a donné une motivation qui m'a énormément aidée par la suite. Quand je suis rentrée en Italie, il a continué à m'aider et à me conseiller par mail. Pour les articles il a été toujours là, prêt à corriger tout ce qui fallait, pour transformer mes « premiers jets » en vrais articles, grâce à son expérience, ses connaissances et son esprit critique, si précieux pour moi. Je garde encore deux de ses messages d'encouragement (en septembre 1999 et surtout en octobre 2002) qui m'ont vraiment aidée à ne pas jeter l'éponge ! Merci Christophe mille fois ! Et toujours Christophe pendant le séjour de 1999 m'a fait travailler avec quelqu'un qui en 1998 avait publié la carte génétique du chêne pédonculé...article affiché dans mon ancien bureau de Porano en Italie avec dessus écrit « target ! »...

Ce quelqu'un s'appelle Teresa Barreneche, croyez-moi c'est quelqu'un de très bien et je me sens très chanceuse de l'avoir rencontrée. Teresa et moi, ça a été tout de suite un binôme indissociable, un « coup de foudre » entre deux esprits latins qui restera toujours une des choses les plus belles de ma vie. Et c'est grâce à Teresa, à son admirable sens de la justice, son inclination à aider les autres et son action diplomatique que j'ai pu valoriser mon travail par une thèse, m'inscrire à Bordeaux I et continuer le travail de cartographie comparée. Coté scientifique Teresa a participé directement à cette thèse car elle a réalisé le transfert de marqueurs microsatellites entre les deux espèces et la première comparaison des deux cartes génétiques du chêne et du châtaignier. Je la remercie pour tous les bons moments passés au laboratoire, pour avoir rendue si agréable cette collaboration, mais surtout pour m'avoir confié tous les secrets de la cartographie, pour m'avoir appris plein de choses que j'ignorais sur la génétique des plantes et pour le temps passé à corriger et améliorer ce manuscrit. On a bien ri ensemble, n'est pas Teresa ? Et ce n'est pas fini...tu verras qu'on « s'éclatera » ensemble en Italie ! Coté humain, je te remercie Teresa pour m'avoir permis de redécouvrir l'amitié, pour m'avoir aidée et soutenue dans mes moments difficiles, pour m'avoir accueillie

dans ta maison où j'ai passé des moments inoubliables avec toi, Imanol et Erwan. Merci aussi aux plus beaux et sympathiques adolescents du monde, le futur pilote d'avions, Imanol, et le futur avocat (réalisateur dans le temps libre !), Erwan, qui m'ont tellement fait rire ! Je n'oublierai jamais la première fois qu'on a regardé ensemble Rabbi Jacob !!! La France, la France...clip, clop, clip, clop...merci Teresa de m'avoir fait vivre tout cela et pour cette thèse!

Il manque un des trois « grands »...et je pense que vous avez tous compris de qui s'agit-il. Quand, lors de mon deuxième séjour à Pierroton en février et mars 2001 j'ai eu l'occasion de mieux connaître le grand chef, Antoine Kremer, j'ai compris pourquoi Pierroton est tel qu'il est ! Je dis un très grand merci à Antoine pour m'avoir permis de faire cette thèse, pour m'avoir intégrée dans son équipe et ses projets de recherche, et surtout pour m'avoir donné un coup de main inoubliable en décembre 2002 qui m'a permis de terminer ce travail. A partir de mars 2003 je me suis installée à Pierroton et tout au long de ce séjour Antoine m'a permis d'avoir les conditions idéales pour travailler à ma thèse et profiter de ce laboratoire extraordinaire. Il m'a toujours encouragé avec son optimisme que j'apprécie beaucoup, il m'a fait confiance pour la cartographie des EST, il m'a permis d'avoir mon comité de thèse (je remercie Véronique Jorge, Charles Eric Durel, Brigitte Mangin et Christophe Plomion d'y avoir participé et de m'avoir donné des conseils précieux), et il a été toujours là quand j'ai eu besoin d'un avis, de son énorme connaissance et vision intégrée de la génétique des arbres forestiers, de sa vision critique ou tout simplement d'une signature. Merci pour le champagne pour les douze homologues !, tout cet ensemble de choses m'a redonné la motivation et l'enthousiasme pour ce travail. Cela m'a fait énormément plaisir d'apprendre qu'au sommet de la recherche il y a de gens comme Antoine. Merci Antoine pour tout cela, et si j'avais trouvé un « petit accélérateur de particules » pas loin de Bordeaux, j'aurais tout fait pour continuer à avoir un chef comme toi !

Puis, il y a tous les gens que j'ai connu à Pierroton, qui contribuent à faire de cet endroit le laboratoire idéal pour faire une thèse, mais aussi pour avoir toute une série « d'à côtés » qui permettent même à une italienne de se sentir bien hors d'Italie ! Pendant tous mes séjours à Pierroton, j'ai toujours rencontré des gens extraordinaires qui n'ont jamais hésité à me donner un coup de main...comme vous verrez la liste est longue...

Voici tous ceux que j'ai connu lors des mes premiers séjours en 1999 et 2001 jusqu'à ceux qui m'ont assistée dans ces tous derniers mois de rédaction...

Merci Stéphanie pour ta rigueur scientifique (que j'adore !), ton sens aigu de la justice, ton français parfait, les discussions AFLP, SSR, coloration nitrate d'argent...RFLP-Scan (!!!!), mais surtout pour ton amitié, je suis sûre que toi, moi et Teresa on aurait pu faire le labo « utopia » ! Je te souhaite de vivre une superbe expérience dans ton nouveau labo et j'espère bientôt venir te voir à Madrid.

Merci Christian pour ta passion pour mes pâtes, les discussions sur Galileo Galilei, Leonardo da Vinci et les campagnes d'Italie des français quand il faisait beau. Je garde des très beaux souvenirs de la co-location à la cantine (y compris les assiettes dans le frigo !). Merci pour tous les mails d'encouragement, je croise les doigts pour ton concours...mais j'ai confiance en toi ! Courage Christian !

Merci à Delphine, pour sa gentillesse et son message d'encouragement en août 2001 quand mon inscription en thèse avait l'air plutôt compliqué, un mot que j'ai bien apprécié et qui m'a fait très plaisir.

Merci à Laetitia pour ne m'avoir jamais oubliée et pour tous ses messages auxquels je réponds souvent en vitesse, mais j'aurais plus de temps maintenant. A bientôt j'espère.

Merci Marie-Hélène, pour toutes les discussions sur la vraie vie, tes mots d'encouragement, toujours là quand il fallait.

Merci à David (Chagné), que j'ai connu en 1999 (jeune rebelle !) et j'ai trouvé transformé après son expérience à Davis...merci pour toutes les discussions cartographie comparée, la biblio, la relecture de l'article transfert des microsattellites, mais aussi pour les repas à la maison avec Karine et désolée de n'avoir pas pu participer à votre mariage. Je vous souhaite plein de bonheur de l'autre coté du monde et merci pour les superbes photos.

Un gros merci à David (Pot). Quand je suis arrivée en 2001 j'ai donc fait la connaissance du grand David, grand en hauteur, mais surtout pour ces qualités scientifiques et humaines. Sa contribution à l'analyse des données et à l'article QTL a été fondamentale. Merci David de trouver toujours le temps pour donner un coup de main aux autres, le monde entier et certainement celui de la recherche ont besoin de gens comme toi. Merci aussi pour l'amusante expérience : planche à voile ! "In bocca al lupo" à toi et Manu pour votre nouvelle vie à Montpellier et autour du monde.

Merci à Caroline (Scotti-Saintagne) pour m'avoir donné les fichiers QTL du chêne, alors as-tu commencé à parler en Italien ? Grazie anche a te Ivan per la qualità dei tuoi lavori, so che diventerai un grande ricercatore ed è la stessa cosa che ho pensato quando hai fatto il tuo seminario a Porano. In bocca al lupo per il concorso!

Merci à Philippe, pour m'avoir aidée plusieurs fois avec mon portable et sa disponibilité pour tout le monde. Merci à toi et à Gwenaëlle pour le dîner avec Mar.

Merci à Céline Latouche pour un coup de main avec le micro et les discussions qu'on a eu au cours de mes premiers séjours à Pierroton.

Merci à Jorge pour les nombreuses soirées passées tous ensembles dans son appartement, pour son gâteau délicieux et sa passion pour la musique classique.

Merci à Alex pour la cartographie de CsCAT15, pour sa gentillesse avec moi et une soirée passée à discuter l'été dernière, très sympathique. Courage pour la suite !

Merci à Fabiano pour ses discussions sur Barrichello et les microsattellites.

Merci à Akim, pour sa joie de vivre et pour m'avoir surnommée Speedy Gonzalez, bip, bip...

Merci à Manu pour avoir apprécié mes conseils, courage pour la suite de cette thèse.

Merci al gran profesor Carlos Magni, pour tous les repas, les beaux moments passés avec Loreto, Antonia et Amelia et pour toutes les questions « microsattellites » qui m'ont souvent fait beaucoup réfléchir, je viendrai un jour au Chili rendre de nouveau visite à ta famille...peut-être quand il y aura un GP de formule 1...

Merci à Henri pour les informations concernant Berlusconi, les discussions du matin au labo acry, sa sympathie, ses mots prononcés toujours au bon moment et le repas à la maison...derrière cette barbe il y a un grand homme.

Merci à Caroline (Morera), stagiaire exemplaire et super organisée!!!! Merci d'avoir supporté mes programmes chargés, 8 gènes cartographiés en une semaine ! cela restera un record. Merci beaucoup pour le livre sur la cuisine française et le tennis !

J'ai laissé exprès pour la fin deux thésards de Pierroton que je dois remercier particulièrement car ils ont fait beaucoup d'efforts pendant ces deux années pour rendre mon séjour à Pierroton très agréable et ils ont toujours compris quand il fallait me remonter le moral. Merci donc à JD & JD...Merci à Jérémy, le seul homme qui m'a fait tomber dans les pommes ! Merci pour avoir accepté que j'utilise tes EST, merci pour toutes les discussions qu'on a eu autour de ces fameux gènes de Jérémy, merci pour ta vision fonctionnelle et tout ce que tu m'as appris sur le débourement. Merci pour le grand boulot de correction de ce manuscrit. Travailler avec toi a été vraiment un plaisir. Mais, merci surtout pour les repas avec Alice (ils sont magnifiques tes albums de photos ! merci Alice), la compétition France Italie pendant les jeux olympiques...comme tu le sais bien en pratique l'Italie a gagné, merci d'avoir acheté des chaussures italiennes pour jouer au foot !!!!!...merci Jérémy pour ton amitié et pour m'avoir fait beaucoup rire surtout pendant les mois de rédaction! Et merci à l'autre JD, merci Jérôme d'avoir tout essayé pour faire que je me sente bien en France...mais tu sais bien maintenant que l' « Italie manquera toujours aux italiens » (publicité française Buitoni). En tout cas tu as bien réussi. Tu m'as impressionnée avec ta cuisine et je me suis dit, ce garçon n'aurait aucun problème à se marier en Italie...et en effet...j'avais raison. Merci de m'avoir fait goûter les délices de la cuisine française. Mais merci surtout de m'avoir fait comprendre que je pouvais toujours compter sur toi, merci pour l'assistance pendant la rédaction de la thèse, pour le temps passé à corriger mon franco-italien et merci surtout pour toutes les discussions sur les relations humaines qui m'ont beaucoup appris. Et s'il y a une chose que j'adore recevoir comme cadeau c'est des livres et tu m'as donné un des plus beaux livres que je n'ai jamais reçu, merci Jérôme à toi aussi pour ton amitié précieuse.

Et merci à tous le gens de Pierroton qui ont été avec moi toujours très gentils. Merci à Christophe d'avoir organisé les sorties à la plage (moi j'irai en décembre après la thèse !), à Olivier, Laurent, Erwan, Iria, Muriel pour les discussions du coin café. Merci à Thibaut pour son travail sur le châtaignier et son calme. Merci à Céline Lalanne et Agnes pour les coups de main à la paillasse. Merci à Jean Marc d'avoir partagé son bureau avec moi et pour tous ces mots d'encouragement, merci également pour sa gentillesse à Marie France, merci à Dominique pour m'avoir souvent accompagnée à l'aéroport et pour les discussions du matin, merci à Sabine et Pierre-Yves pour les discussions dans l'ancien bureau et à Cyril dans le nouveau. Merci à Valérie et Patrick pour certains des meubles que j'ai à la maison et leur gentillesse lors de mon déménagement à Cestas. Merci à Guy pour les châtaignes cuites et ses mots gentils. Merci à Jeannot pour sa passion pour ma belle région en Italie. Merci à Mimi pour son sourire. Merci à Marie-Pierre pour avoir bien compris mon expérience difficile dans le labo italien. Merci à Annie, Florence et Corinne pour simplifier beaucoup de choses et leur sympathie. Merci à Frédéric pour l'aide dans l'organisation de l'impression de cette thèse. Merci à Christian pour sa gentillesse et un très beau concert ! Merci à Annie, Sophie, Jean Marc, Remy, Pauline, Alexis, Frank pour les discussions que de temps en temps on a pu avoir. Merci à Céline et à Frank pour leur gentillesse quand j'étais à la cantine et les bons repas de midi.

Thanks to Mar. I will never forget the “part of your body” falling down after having cut a small carrot...I enjoyed a lot with you, I am sure you will succeed in biology, take care...hoping to meet you soon. Thanks Teresa for your kind messages!

Grazie Marta dei piacevoli momenti passati in terra straniera insieme e della mattinata “all’ospedale”!!!, in bocca al lupo per la scrittura della tua tesi e a presto in Italia.

Grazie Michela, della tua gentilezza e dolcezza. Spero che a tesi finita e quando sarò in Italia potremo incontrarci e passare un po’ di tempo insieme a discutere in pace senza la fretta e lo stress. Grazie di tutti i tuoi messaggi estivi e prometto che in futuro sarò una migliore amica di penna. E in bocca al lupo anche a te per la fine della tua tesi.

I would like thank all the CASCADE team, especially Fina to have kindly provided to me very nice pictures of chestnut buds and precious advises for bud set measurements. Cécile for the nice collaboration, Emily, Daniela, Karen and Roberto for SSR primers, Phil for several interesting discussions and for the course at the MAICH in Crete. Thanks a lot to all teachers and students I met in Crete during two very nice weeks at the MAICH in September 2002!

Thanks a lot Paul (Sisco) for the interesting collaboration with you, for your encouraging messages. I hope to meet you soon and to see again chestnut forests in USA.

E finalmente posso sfogarmi nella mia lingua!!!!!! Durante i cinque anni passati a Porano molte persone mi hanno aiutata direttamente o indirettamente e approfitto quindi per dire a tutti loro grazie dal profondo del cuore. Grazie Claudia dei bei cinque anni passati insieme, di tutto il tuo contributo in questo lavoro, grazie per tutte le volte che mi hai sostenuta e aiutata, grazie per i pomeriggi passati a ridere immaginando Crick che spediva il suo CV a Porano....avremo altre occasioni per ridere insieme! Grazie Enrico e Roberto per quei mesi incredibili passati a scherzare nel laboratorio di genetica, tra Enrico che cantava nella “mitica camera oscura” e Roberto che lo dirigeva dall’esterno. Grazie della vostra preziosa amicizia. Una cena tra amici è il minimo che posso promettervi al mio ritorno. E poi grazie a Marcello per l’incrocio controllato e tutto il lavoro su questi benedetti castagni (siamo stati bene però con la Mattioni ad Antrodoco, eh!?), grazie a Bruno per tutte le sue consulenze di statistica e le innumerevoli interessanti discussioni avute insieme, grazie a Cristina che mi ha aiutata nel momento più difficile e questo non lo dimenticherò mai, grazie a Stefano per alcuni chiarimenti “carboidratici” e la sua passione per la Rossa e tutti i piacevoli messaggi scambiati dopo ogni GP. Grazie a Andrea (Scartazza) per le sue spiegazioni fisiologiche. Grazie a Piero per i suoi “non lavorare troppo, Manuela”, ma se non avessi lavorato un po’ troppo non avrei mai avuto l’occasione di vivere questa straordinaria esperienza, grazie a Andrea (Pisanelli) per aver un giorno pronunciato le parole “Manuela non c’entra niente”, per ricordarti si trattava dell’affare Antrodoco, mai nessuno così apertamente aveva preso le mie difese. Grazie a Alessandro che ha scucito lo scudetto dalla maglia della Lazio per metterlo su quella della Roma. Grazie alla grande Daniela per le innumerevoli discussioni avute in tarda serata sulla vita e la gente. Grazie a Marco per il suo aiuto nell’interpretazione fisiologica del lavoro QTL. Grazie a Alberto per la stima che mi ha dimostrato. Grazie a Enrico per avermi sostenuta a inizio 2000, mi dispiace che sia mancato il tempo di discutere con qualcuno che stimo scientificamente. Grazie a Caterina per le correzioni del mio inglese. Grazie a Peppe, Quinto e Silde per essere sempre stati gentili. Grazie a Massimo per le fotocopie, la posta e la sua passione per i viaggi. Grazie a Angela e Silvia, la padana, per la loro simpatia. Grazie a Simona e Paola per avermi seguita nell’avventura del journal club insieme ovviamente a tutti gli altri che hanno partecipato. Grazie a Mario per aver assemblato un ottimo PC per me e

Simone. Grazie a Giovanni per l'aiuto in serra e a Luciano per l'aiuto nella triturazione di alcuni campioni. Grazie a Lucia per essersi a volte confidata con me. Thanks Dong to have opened my mind to China, thanks to be my friend.

Un grazie speciale alla mia professoressa di francese del liceo, Anna Maria Angeli, perché grazie al suo amore per questa lingua e ai suoi insegnamenti, la scrittura di questa tesi è stata meno difficile!

E adesso un grazie ai miei più cari amici e a tutti coloro che rendono la mia vita semplicemente felice.

Grazie Marcella, sei la persona più forte e coraggiosa che conosca e il solo pensare a te mi dà la forza di affrontare le mie piccole difficoltà. Spero potremo stare più insieme al mio ritorno e incrocio le dita perché tu possa trovare il lavoro che desideri.

Grazie a Stefano e Consuelo per il loro bel matrimonio e le cene insieme, allora io ritorno il campo da calcetto è pronto o il progetto è cambiato?

Grazie a Enrico (Ladi), per tutti i GP sofferti insieme durante le stagioni '98, '99, NOI ferraristi e schumacheriani nell'anima lo sapevamo che prima o poi il dominio sarebbe cominciato.

Grazie a Francesca e Paolo per tutti i pranzi, le cene, i torroni, i crostoni, i cocomeri...e così via mangiati a casa vostra!

Grazie a PierLuigi "il commenda milanese", ormai ai vertici della SanPellegrino e alla sua Claudia per il bellissimo weekend a Bominaco. Mi dicono che il milanese coltiva ormai la passione della pipa, le gite in yacht...bella vita la tua Ripe!

Grazie a Silvio per essere sempre così giovane, dinamico, splendido e soprattutto per il suo incantevole modo di parlare il perfetto italiano. Grazie a Paola per la sua dolcezza e per i tanti bei momenti (cene di fine anno, compleanni e scampagnate al fiume) passati insieme.

Grazie a Francesca per tutto il suo affetto, per i suoi bellissimi regali (un esempio l'anellino e il braccialetto fatti a mano! una vera opera d'arte), per i suoi dolci sempre perfetti e buonissimi, per la sua personalità forte che a me piace tanto! Da ora in poi spero avrò più tempo per venirti a trovare a Torino. E grazie ad Antonio per il buonissimo vino d'Avellino e la sua dedizione a un lavoro importante.

Grazie a Gabriella e Saverio per il bene che mi volete e per i tanti momenti passati insieme che da ora in poi saranno spero più frequenti.

Grazie allo zio Tonino, per la sua passione per la storia e i viaggi, per essermi stato sempre di buon consiglio, per il suo costante interesse alle mie varie vicende e per essere sempre così vicino a tutta la famiglia.

Grazie a Fabrizio, per essere prima di tutto un caro amico, per avermi fatta divertire durante la costruzione della casa, per le mille cene e avventure trascorse insieme e tutte quelle che ancora verranno, per la tua passione per Schumy e soprattutto perché rendi felice la mia sorellina. E grazie a Roberta e Sandro per tutti i pranzi e le cene!

Un grazie tutto speciale, e sicuramente non arriverò a ben esprimere i miei sentimenti, va a mia sorella Margherita, ai miei genitori e all'uomo che mi rende la persona più felice di tutto l'universo (conosciuto e sconosciuto!).

Margherita ti ringrazio perché so che sei sempre pronta ad aiutarmi a sacrificarti per me, a facilitarmi le cose e tutto questo mi fa sentire il bene che mi vuoi. Sei sempre stata e sempre sarai la mia migliore amica e in ogni momento brutto o bello mi hai dimostrato che sorella speciale mi sia capitata. Sappi che il bene che ti voglio è altrettanto grande e che farò sempre tutto quello che mi è possibile perché tu sia felice.

Cari Mamma Antonia e Babbo Adolfo siete semplicemente i migliori genitori che un figlio possa desiderare. Vi ringrazio per tutto quello che avete fatto, che fate e che farete per me, per come mi avete cresciuto, perché mi avete sempre permesso di fare tutto quello che ho desiderato, perché mi fate sentire amata e protetta, sicura di avere sempre qualcuno su cui contare. Sono orgogliosa di voi e felice grazie a voi.

E ora un grazie che non avrà mai fine all'uomo che ho sempre amato, che amo e che amerò per tutta la mia vita. Grazie Simone per questi primi $4,88328 \times 10^{20}$ ps di felicità inebriante durante i quali mi hai fatto sempre sentire la donna più amata e desiderata di questo universo dandomi la forza per affrontare tutte le difficoltà. Mai più mi separerò da te e dedicherò ogni istante futuro della mia vita a renderti altrettanto felice...*percorreremo assieme le vie che portano all'essenza. I profumi d'amore inebrieranno i nostri corpi, la bonaccia d'agosto non calmerà i nostri sensi. Tesserò i tuoi capelli come trame di un canto. Conosco le leggi del mondo, e te ne farò dono. Supererò le correnti gravitazionali, lo spazio e la luce per non farti invecchiare. Ti salverò da ogni malinconia, perché sei un essere speciale ed io avrò cura di te...io sì, che avrò cura di te.* (e grazie a Franco Battiato per queste parole).

E infine grazie a qualcuno che non conosco ma è come se fosse un mio grande amico, grazie Michael (Schumacher, ovviamente). Ricordo ancora quando Simone, una domenica d'agosto del 1991, mi disse: "è arrivato un fenomeno". Grazie per aver preso alla lettera le parole di Enzo Ferrari: "Il secondo è il primo degli ultimi", e lui sì che sarebbe stato fiero di te e di come guidi la Ferrari. Grazie perché mi fai divertire, perché dopo ogni GP questa estate ho scritto due o tre pagine in più del previsto di questa tesi. E grazie perché quando sono depressa mi rivedo quel capolavoro del GP di Barcellona 1996!

Table des matières

Table des matières.....	10
Préambule	14
A. Introduction.....	16
<i>A.1. LA CARTOGRAPHIE GÉNÉTIQUE COMPARÉE.....</i>	<i>17</i>
A.1.1. La comparaison de génomes entiers.....	17
A.1.2. Cartographie génétique comparée : définition et intérêt.	19
A.1.3. Les études de cartographie génétique et génomique comparées chez les plantes : une perspective phylogénétique.	22
<i>A1.3.1. Les gymnospermes : l'exemple de la famille des Pinaceae.</i>	<i>24</i>
<i>A1.3.2. Les monocotylédones : le « système génétique unique » des Poaceae.....</i>	<i>26</i>
<i>A1.3.3. Les eudicotylédones : Asteraceae, Solanaceae, Brassicaceae, Malvaceae et Fabaceae, cinq exemples pour comprendre la structure du génome des Angiospermes eudicotylédones.</i>	<i>36</i>
<i>A1.3.4. Comparaison entre taxa éloignés phylogénétiquement.</i>	<i>46</i>
A.1.4. Les principaux résultats et les questions issues des études de comparaison des génomes chez les plantes.....	49
<i>A.2. LES FAGACÉES.....</i>	<i>54</i>
A.2.1. Taxonomie.....	54
A.2.2. Origine et phylogénie.	54
A.2.3. Distribution géographique.....	56
A.2.4. Intérêt écologique et économique.....	57
A.2.5. Connaissances sur le génome des Fagacées et contexte de la thèse.....	57
<i>A.3. OBJECTIFS DE LA THÈSE.</i>	<i>59</i>
B. Cartographie génétique du châtaignier	61
<i>B.1. LA CARTOGRAPHIE GÉNÉTIQUE CHEZ LES ARBRES FORESTIERS.....</i>	<i>62</i>
B.1.1. Principes de la cartographie génétique.	62
B.1.2. Ségrégation mendélienne, estimation de la liaison génétique et de la distance entre marqueurs.	64
B.1.3. Fonctions de cartographie.....	66
B.1.4. Détermination de l'ordre des marqueurs.	68
B.1.5. Influence du type de marqueurs et de la taille de la population.	68
B.1.6. Saturation de la carte et taille du génome.....	69
B.1.7. La stratégie de double pseudo test-cross.	69
B.1.8. Les logiciels « MapMaker » et « JoinMap ».	71
<i>B.2. LES POPULATIONS BURSA ET HOPA ET LE CROISEMENT CONTROLÉ.....</i>	<i>73</i>
B.2.1. Les études de diversité génétique chez le châtaignier européen.	73
B.2.2. Identification de deux populations de châtaignier adaptées à deux environnements contrastés.....	74
B.2.3. Le croisement contrôlé.	76
<i>B.3. LES MARQUEURS MOLÉCULAIRES UTILISÉS.....</i>	<i>79</i>

B.3.1. Le choix des marqueurs.....	79
B.3.2. L'optimisation du protocole.....	80
B.3.3. La fiabilité des marqueurs.....	80
<i>B.4. LA CARTE GÉNÉTIQUE DU CHÂTAIGNIER.....</i>	<i>81</i>
B.4.1. Le criblage des marqueurs informatifs et le génotypage.....	81
B.4.2. La construction de la carte.....	83
B.4.3. L'évolution de la carte du châtaignier.....	87
C. Détection de QTL impliqués dans les caractères d'adaptation chez le châtaignier	92
<i>C.1. LES CARACTÈRES ADAPTATIFS.....</i>	<i>93</i>
C.1.1. L'adaptation et les caractères adaptatifs.....	93
C.1.2. La phénologie du bourgeon (débourrement et arrêt de croissance).....	95
C.1.3. La discrimination isotopique du carbone 13.....	96
C.1.4. La croissance.....	98
C.1.5. L'étude des caractères adaptatifs chez les arbres forestiers dans le cadre des changements climatiques globaux.....	98
<i>C.2. LA DÉTECTION DES QTL.....</i>	<i>100</i>
C.2.1. La détection des QTL : l'étude du contrôle génétique des caractères complexes. Comment peut-elle aider à la compréhension des bases génétiques de l'adaptation ? ..	100
C.2.2. Les approches statistiques pour la détection des QTL et le choix des niveaux de signification des QTL.....	103
C.2.3. Limites de la détection des QTL.....	108
C.2.4. Les études de détection des QTL chez les arbres forestiers.....	110
<i>C.3. DÉTECTION DES QTL IMPLIQUÉS DANS LA PHÉNOLOGIE DES BOURGEONS, LA DISCRIMINATION ISOTOPIQUE DU CARBONE 13 ET LA CROISSANCE CHEZ LE CHÂTAIGNIER.....</i>	<i>113</i>
C.3.1. Les mesures phénotypiques : 2000, 2001 et 2002.....	113
C.3.2. Le logiciel MultiQTL et la stratégie d'analyse des données.....	122
C.3.3. Résultats de la détection des QTL.....	125
C.3.3.1. <i>Phénologie des bourgeons.....</i>	<i>127</i>
C.3.3.2. <i>Croissance.....</i>	<i>128</i>
C.3.3.3. <i>Discrimination isotopique du carbone 13.....</i>	<i>128</i>
C.3.3.4. <i>Distribution sur le génome et stabilité des QTL.....</i>	<i>131</i>
C.3.3.5. <i>Variance phénotypique expliquée par les QTL.....</i>	<i>132</i>
C.3.3.6. <i>La plupart des QTL ont été détectés chez le parent femelle : adaptation de cet écotype à la sécheresse ?</i>	<i>134</i>
C.3.3.7. <i>« Colocalisations » des QTL : pour le même caractère et entre caractères... ..</i>	<i>135</i>
C.3.3.8. <i>Conclusions : les résultats semblent indiquer une adaptation locale des deux parents à leur environnement d'origine.....</i>	<i>138</i>
D. Cartographie génétique comparée chez les Fagacées.....	141
<i>D.1. LA CARTOGRAPHIE GÉNÉTIQUE COMPARÉE CHEZ LES FAGACÉES : OBJECTIFS ET ACQUIS.....</i>	<i>142</i>
D.1.1. La carte génétique du chêne pédonculé.....	142
D.1.2. Les marqueurs orthologues.....	143
<i>D.2. LE TRANSFERT DE MICROSATELLITES ENTRE CHÊNE ET CHÂTAIGNIER... ..</i>	<i>147</i>

D.2.1. Le microsatellites disponibles.....	147
D.2.2. Le transfert entre les deux genres.....	147
D.2.3. La cartographie génétique des microsatellites et la première comparaison des cartes du chêne et du châtaignier.....	152
D.2.4. La comparaison des locus transférés au niveau de séquence.....	156
D.3. LA CARTOGRAPHIE DES EST CHEZ LE CHÊNE ET LE CHÂTAIGNIER.....	160
D.3.1. Les ressources en EST.....	160
D.3.2. La cartographie des EST : le géotypage SNP et les techniques SSCP et DGGE.....	161
D.3.2.1. Les techniques de géotypage SNP : Victor et MegaBACE.....	163
D.3.2.2. Les techniques SSCP et DGGE.....	165
D.3.3. Les critères de choix des amorces pour l'amplification des EST.....	170
D.3.4. L'amplification des EST chez les deux espèces.....	172
D.3.5. La cartographie des EST et l'alignement des cartes génétiques du chêne et du châtaignier.....	175
D.3.6. Quelques conclusions sur la structure du génome des Fagacées.....	182
E. Cartographie comparée des QTL chez le chêne pédonculé et le châtaignier européen.....	184
E.1. LA DÉTECTION DE QTL CHEZ LE CHÊNE PÉDONCULÉ.....	185
E.1.1. Les caractères étudiés chez le chêne pédonculé.....	185
E.1.2. Les résultats obtenus chez le chêne sont-ils comparables à ceux du châtaignier ?	186
E.2. LA CONSTRUCTION D'UNE CARTE CONSENSUS CHEZ LE CHÊNE ET CHEZ LE CHÂTAIGNIER ET LA COMPARAISON DES QTL.....	190
E.2.1. La stratégie de construction de la carte consensus.....	190
E.2.2. Le logiciel BioMercator et la projection des QTL sur la carte consensus de chaque espèce.....	192
E.2.3. La comparaison des QTL.....	193
E.2.3.1. Comparaison du nombre et des effets des QTL.....	194
E.2.3.2. Comparaison de la position des QTL.....	209
E.2.3.3. Quelles conclusions peuvent être tirées de cette comparaison ?.....	212
E.3. IDENTIFICATION DE GÈNES CANDIDATS POUR LE CARACTÈRE DE DÉBOURREMENT.....	220
E.3.1. Le concept de gène candidat.....	220
E.3.2. Quelques gènes candidats positionnels : « colocalisation » entre QTL et EST... ..	221
F. Conclusion et perspectives.....	224
F.1. LA DÉTECTION DE QTL CONTRÔLANT LES CARACTÈRES ADAPTATIFS : EST-ELLE UNE STRATÉGIE INTÉRESSANTE POUR L'ÉTUDE DES BASES GÉNÉTIQUES DE L'ADAPTATION ?	225
F.2. LA COMPARAISON DES QTL ENTRE LE CHÊNE ET LE CHÂTAIGNIER : POUVONS-NOUS TIRER QUELQUES CONCLUSIONS CONCERNANT L'ÉVOLUTION DES CARACTÈRES ADAPTATIFS ?	226
F.3. LA CARTOGRAPHIE GÉNÉTIQUE COMPARÉE CHEZ LE CHÊNE ET LE CHÂTAIGNIER : VERS LE « SYSTÈME GÉNÉTIQUE UNIQUE » DES FAGACÉES....	227

Références bibliographiques	230
Annexe I : A genetic linkage map of European chestnut (<i>Castanea sativa</i> Mill.) based on RAPD, ISSR and isozyme markers.....	262
Annexe II : Identification of QTLs affecting adaptive traits in <i>Castanea sativa</i> Mill.	263
Annexe III : Comparative mapping between <i>Quercus</i> and <i>Castanea</i> using simple-sequence repeats (SSRs).....	264
Annexe IV : Comparison of QTLs for adaptive traits between oak and chestnut based on an EST consensus map.	265
Annexe V : Protocole de génotypage SNP (VICTOR).....	266

Préambule

La construction d'une carte génétique s'appuie sur le concept de liaison génétique élaboré en 1913 par Sturtevant, thésard du généticien Morgan, qui a établi pour la première fois une carte génétique à partir de données de ségrégation (Sturtevant, 1913). Dix ans plus tard, Sax (1923), dans une étude menée sur un croisement contrôlé entre deux variétés de haricot, trouva une association statistique entre un caractère qualitatif et un caractère quantitatif, qu'il expliqua grâce à la liaison génétique entre le locus responsable du caractère qualitatif et l'un des locus contrôlant le caractère quantitatif. Il s'agit évidemment du premier QTL (« Quantitative Trait Locus ») détecté. Mais ce n'est que grâce à l'essor des techniques de marquage moléculaire et, plus en général, de la biologie moléculaire, que ces deux concepts, élaborés au début du siècle passé, ont pu être largement appliqués.

La disponibilité d'un très grand nombre de marqueurs moléculaires permet facilement de construire des cartes génétiques qui couvrent tout le génome d'une espèce donnée. Ces cartes peuvent être utilisées d'une part pour l'identification des régions génomiques qui contrôlent les caractères quantitatifs (détection des QTL), d'autre part pour la comparaison avec d'autres cartes génétiques développées dans des espèces phylogénétiquement proches (cartographie génétique comparée). Par conséquent, la construction d'une carte génétique permet, comme le suggère Liu (1998), d'avoir une base pour l'étude de la structure, de l'évolution et du fonctionnement du génome d'une espèce donnée. L'étude menée dans le cadre de cette thèse s'inscrit parfaitement dans ce contexte. Comme nous le verrons tout au long de ce manuscrit, la construction de la carte génétique du châtaignier européen (*Castanea sativa* Mill.), son utilisation pour la détection des QTL contrôlant des caractères adaptatifs et la comparaison avec la carte génétique du chêne pédonculé (*Quercus robur* L.) ont apporté une contribution importante à la compréhension de certaines caractéristiques du génome des Fagacées.

Dans l'introduction de cette thèse (Partie A) nous présenterons tout d'abord une synthèse bibliographique sur l'état des connaissances en cartographie génétique comparée chez plusieurs espèces de plantes. Cela nous permettra de faire le point sur les différents mécanismes impliqués dans l'évolution et la différenciation des génomes des angiospermes afin de mieux interpréter les résultats que nous avons obtenus chez les Fagacées. Cette famille d'espèces forestières ainsi que tout particulièrement les deux espèces, *Q. robur* et *C. sativa*,

étudiées dans le cadre de cette thèse, seront également présentées dans la dernière partie de ce chapitre.

La partie B sera dédiée à la cartographie génétique du châtaignier. Après une présentation des principes de la cartographie génétique, nous décrirons la famille de plein-frères utilisée pour la construction de la carte génétique du châtaignier européen et l'évolution de cette carte due à l'intégration de différents types de marqueurs moléculaires.

Dans la partie C, nous décrirons d'abord certains aspects de la génétique de l'adaptation et les caractères adaptatifs étudiés dans le cadre de cette thèse. Les méthodes statistiques utilisées lors de la détection des QTL et les limites de ce type d'analyse seront présentées dans une deuxième partie de ce chapitre. Enfin, nous discuterons les résultats de la détection des QTL, contrôlant le débourrement, l'arrêt de croissance, la discrimination isotopique du carbone 13 et la croissance, obtenus chez le châtaignier.

La partie D traitera de la cartographie génétique comparée entre le châtaignier européen et le chêne pédonculé. Deux types de marqueurs moléculaires seront utilisés afin de relier les cartes génétiques de ces deux espèces : les microsatellites et les marqueurs dérivés de séquences EST. Nous présenterons dans ce chapitre l'alignement des 12 groupes de liaison du chêne et du châtaignier, ce qui nous permettra de tirer quelques conclusions sur la structure du génome des Fagacées.

Dans la partie E, les QTL contrôlant les mêmes caractères adaptatifs détectés chez le chêne et le châtaignier seront comparés sur la base d'une carte consensus construite dans chacune des deux espèces. Quelques hypothèses sur le contrôle génétique de ces caractères complexes seront formulées dans la dernière partie de ce chapitre.

La conclusion générale de ce travail de thèse et les perspectives seront enfin présentées dans la partie F.

A. Introduction

It has been said that nothing makes sense except in the light of evolution, but it is also true that very little in evolution makes much sense without a phylogenetic context.

(Doyle J.J. et Gaut B.S. (2000) «Evolution of genes and taxa: a primer » Plant. Mol. Biol. 42 : 1-23)

A.1. LA CARTOGRAPHIE GÉNÉTIQUE COMPARÉE.

A.1.1. La comparaison de génomes entiers.

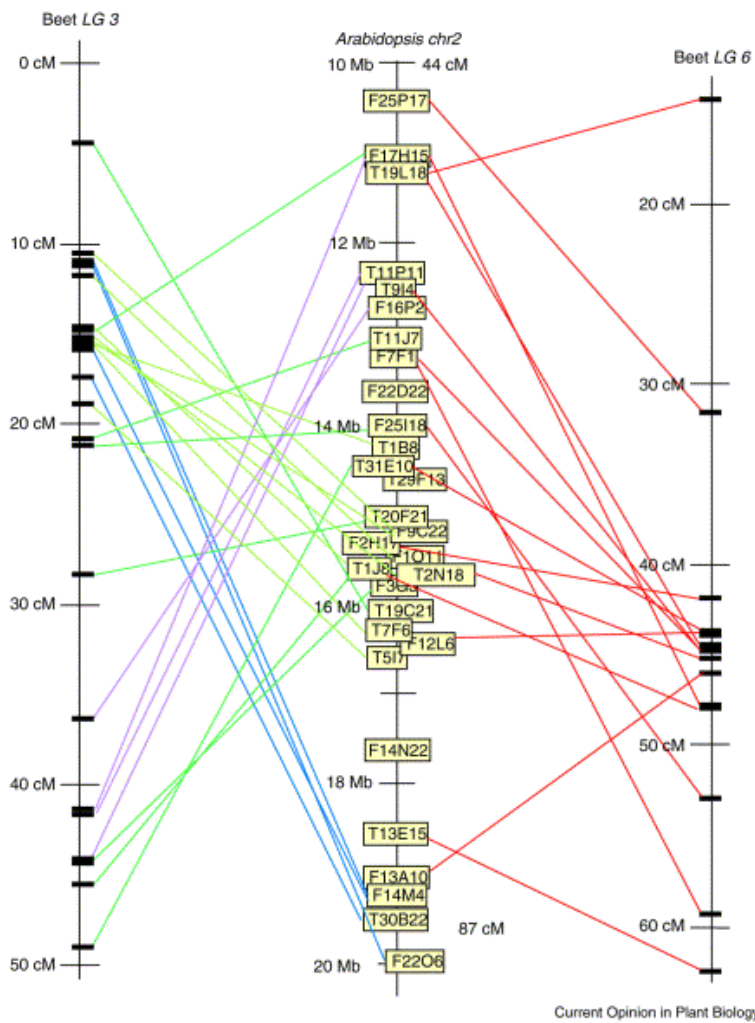
En biologie la comparaison entre organismes a toujours apporté des éléments de réponse à de nombreuses questions fondamentales. Notamment dans le domaine de l'évolution biologique, la comparaison entre espèces a permis la reconstitution de l'histoire des êtres vivants, de leur apparition, de leur adaptation aux environnements les plus divers, du succès évolutif de certains d'entre eux et de la disparition d'autres. Grâce aux progrès de la biologie moléculaire et de la génétique, il est aujourd'hui possible de comparer les génomes entiers de différents organismes, ce qui a mené à l'essor d'un nouveau domaine au sein de la génomique, la science qui étudie la structure, le fonctionnement et l'évolution des génomes. Il s'agit de la génomique comparée, laquelle, grâce à la comparaison des génomes des espèces, vise à découvrir les mécanismes qui ont déterminé la divergence des gènes et des génomes, et qui sont à la base de la diversité de formes, fonctions et adaptation des organismes vivants. Cette discipline s'appuie sur deux conditions :

1. que le contenu et l'ordre des gènes dans le génome des espèces actuelles dérive, à moins de modifications produites au cours de l'évolution, de celui d'un ancêtre commun ; et
2. que nous puissions utiliser les informations sur les espèces actuelles pour rétablir l'ordre et le contenu des gènes du génome ancestral.

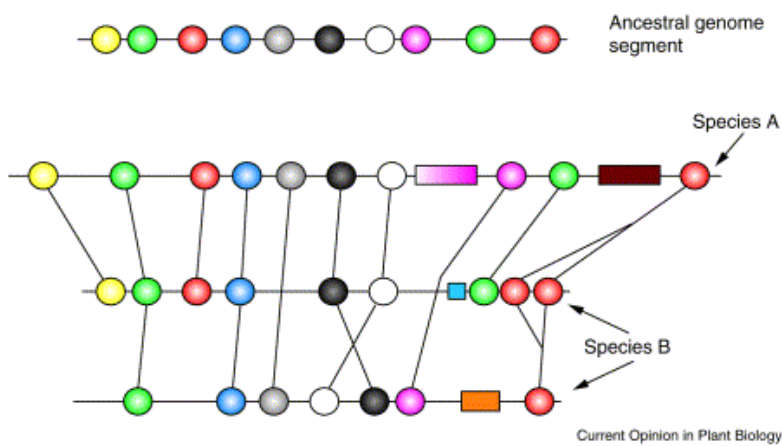
La comparaison de génomes entiers peut être faite à deux échelles, selon que l'on s'intéresse aux cartes génétiques ou physiques. Comme le montre la figure 1, dans le premier cas, les cartes génétiques de deux ou plusieurs espèces phylogénétiquement proches sont alignées à l'aide de marqueurs orthologues et l'unité de distance génétique est le cM (centiMorgan)¹, alors que dans le cas des cartes physiques il s'agit de comparer directement les séquences d'ADN et la distance est mesurée en paires de bases d'ADN (pb). Nous décrirons en détail les principes de la cartographie génétique comparée dans le paragraphe suivant, étant donné que le travail de cette thèse concerne ce niveau de comparaison. Au contraire, nous ne discuterons pas les méthodes de comparaison des séquences, même si nous ferons référence aux résultats obtenus car ils apportent des éléments de discussion importants dans le cadre de l'évolution des génomes.

¹ CentiMorgan (cM) : unité de distance génétique, égale à un pour cent de recombinaison entre deux locus.

Figure 1



Cartographie génétique comparée. Le chromosome 2 d'*Arabidopsis thaliana* est lié aux deux groupes de liaison 3 et 6 de la betterave à sucre. Les marqueurs orthologues (locus qui dérivent d'un même locus ancestral par spéciation) nous permettent de comparer le génome de différentes espèces (Tiré de Barnes et al., 2002).



Cartographie physique comparée. Le même fragment d'ADN a été séquencé chez les deux espèces A et B. Il s'agit d'une région homologe qui dérive du même fragment ancestral. Dans l'espèce B ce fragment a été dupliqué au cours de l'évolution. Par rapport à la région ancestrale des délétions, insertions, inversions sont observés chez les espèces actuelles (Tiré de Barnes et al., 2002).

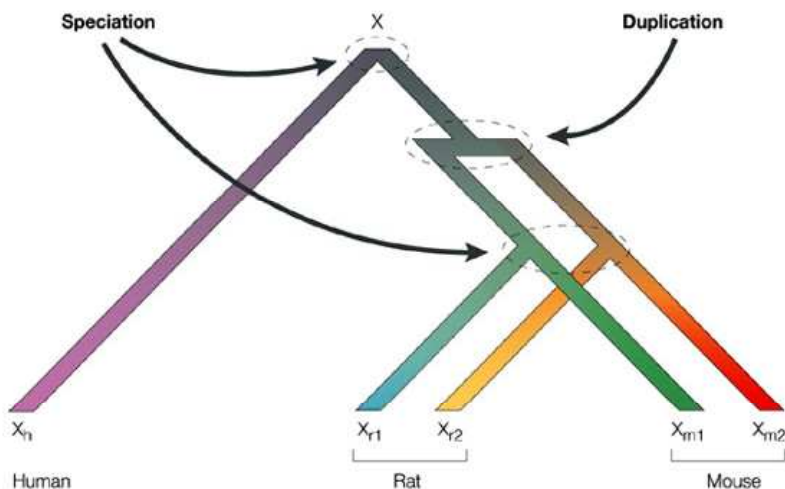
Chez les plantes ces dernières années, des cartes génétiques d'un très grand nombre d'espèces ont été construites, ce qui a permis le développement de la cartographie génétique comparée à partir de la fin des années quatre-vingt (Schmidt, 2000). Quoique seulement le génome de quelques espèces modèles ait été complètement séquencé, le séquençage des EST² a donné un accès assez facile aux séquences de gènes homologues chez plusieurs espèces, permettant des comparaisons au niveau de la séquence nucléotidique. La cartographie génétique comparée et l'alignement des séquences homologues ouvrent la voie à l'étude fascinante de la structure et de la fonction du génome des plantes sous une perspective évolutive. Dans la première partie de la thèse, nous nous focaliserons sur l'état actuel des études de génomique comparée chez différentes familles de plantes, en faisant une synthèse des résultats acquis et des conclusions que nous pouvons tirer concernant la structure, le fonctionnement et l'évolution du génome des plantes. Cela nous permettra de situer le sujet de cette thèse dans le contexte de la génomique comparée des plantes et de formuler les questions de recherche de ce travail à la lumière des résultats obtenus chez d'autres espèces.

A.1.2. Cartographie génétique comparée : définition et intérêt.

Une carte génétique est une représentation du génome d'un organisme faite à l'aide de marqueurs moléculaires. Il s'agit de localiser, tout au long de chacun des chromosomes d'une espèce, des points de repère représentés par ces marqueurs. La distance génétique entre deux marqueurs est calculée grâce à la fréquence de recombinaison entre eux et s'exprime en cM (centiMorgan). Les principes de la cartographie génétique et les différents marqueurs moléculaires utilisés seront discutés en détail dans la partie B. La cartographie génétique comparée se base sur la localisation de marqueurs orthologues chez deux ou plus espèces phylogénétiquement proches. Deux marqueurs sont dits orthologues quand ils dérivent d'un événement de spéciation et non d'un événement de duplication (figure 2a). La duplication et la divergence des séquences est l'un des mécanismes de l'évolution des gènes, qui, avec la spéciation, entraînent la formation de séquences homologues. Par conséquent, nous disons que deux gènes sont homologues quand ils dérivent d'un gène ancestral commun. Si, en particulier, l'homologie est le résultat de la spéciation, nous parlons d'orthologie (homologie par descendance mendélienne) et si, en revanche, l'homologie dérive d'une duplication alors nous parlons de paralogie. Une fois les marqueurs orthologues cartographiés, les deux étapes suivantes d'une étude de cartographie comparée consistent à évaluer la conservation de la

² EST: "expressed sequence tag", fragments de gènes exprimés dans un tissu donné, séquencés à partir de l'ADN complémentaire (ADNc), synthétisé par rétrotranscription de l'ARN.

Figure 2a



Homologue: deux gènes sont homologues quand ils dérivent d'un gène ancestral commun.

Orthologue: quand l'homologie est le résultat de la spéciation (X_h , X_{r1} et X_{m1}).

Paralogue: quand l'homologie est le résultat d'une duplication (par exemple X_{r1} et X_{r2} ou X_h , X_{r2} et X_{m2}).

(http://www.nature.com/nrd/journal/v2/n8/slideshow/nrd1152_bx2.html)

Nature Reviews | Drug Discovery

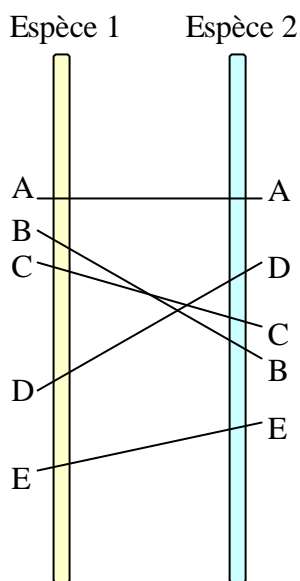


Figure 2b

A gauche cinq gènes orthologues ont été cartographiés chez les deux espèces 1 et 2. Ils sont localisés dans la même région homologue. Il y a donc conservation de la synténie. A droite, les cinq gènes ont aussi conservé le même ordre, donc les deux régions sont colinéaires.

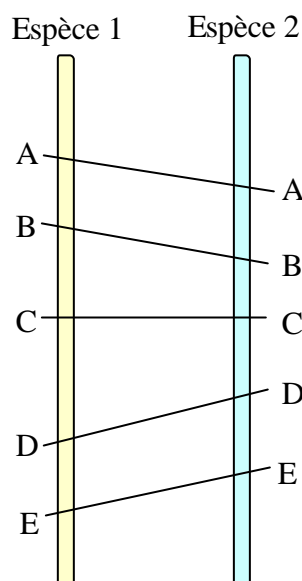
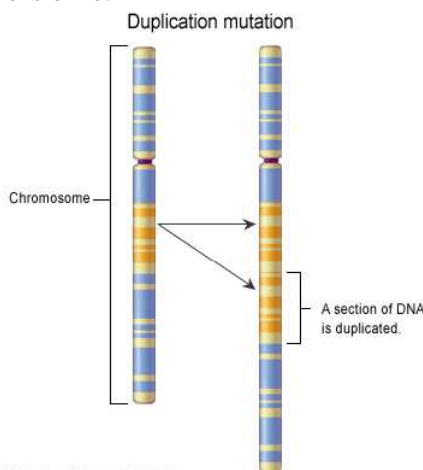
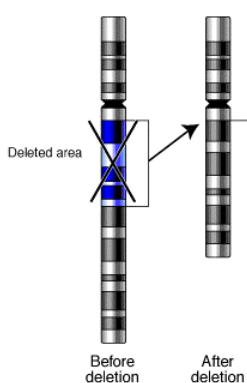
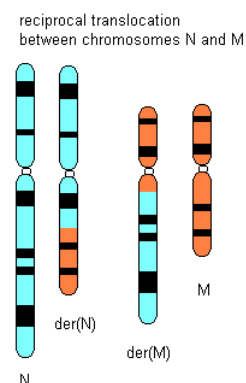


Figure 2c

Trois exemples de réarrangements chromosomiques, intra-chromosome, les deux premiers et inter-chromosomes le troisième.



U.S. National Library of Medicine



liaison génétique entre ces marqueurs et l'ordre de ceux-ci le long d'un chromosome ou d'une région homologue. Nous parlons donc de conservation de la synténie (du grec *syn* = ensemble et *taenia* = ruban) et de la colinéarité, à savoir de la conservation du contenu des gènes au sein d'un chromosome homologue et de leur ordre (Nadeau et Sankoff, 1998 ; figure 2b).

La comparaison des cartes de liaison génétique chez plusieurs espèces animales et végétales a montré que la conservation de la synténie et de la colinéarité diminue avec l'augmentation du temps de divergence entre deux espèces. Au cours de l'évolution, les liaisons génétiques ancestrales sont brisées par des réarrangements intra et inter chromosomiques (délétions, insertions, inversions, translocations ; figure 2c), qui s'accumulent depuis la divergence de deux espèces. La cartographie comparée vise donc à rétablir l'histoire évolutive des génomes en comparant la localisation de gènes homologues dans des espèces différentes. Bien que le principe de la cartographie comparée soit très simple, il reste néanmoins très difficile de disposer de données exhaustives permettant de reconstruire tous les changements chromosomiques qui se sont produits entre deux espèces données. En effet, nous pouvons identifier deux difficultés principales. La première concerne la disponibilité d'un nombre suffisamment élevé de marqueurs orthologues répartis sur tout le génome, et la deuxième la possibilité d'identifier des segments conservés à l'aide de la cartographie génétique. Les familles multigéniques³ et les pseudogènes⁴, par exemple, peuvent parfois empêcher l'identification du bon gène orthologue, surtout si les espèces comparées sont éloignées phylogénétiquement. En effet dans ce cas, il est difficile de distinguer l'érosion de l'identité de séquence due à la spéciation de celle due à la duplication. Concernant la cartographie génétique il ne faut pas oublier, d'une part, que l'estimation de la fréquence de recombinaison et de l'ordre des marqueurs est toujours affectée par une incertitude statistique, et que, d'autre part, l'addition de nouveaux marqueurs à une carte déjà établie entraîne parfois des changements de l'ordre des marqueurs cartographiés auparavant. Il est donc judicieux d'être prudent sur les conclusions qui peuvent être tirées d'une étude de cartographie comparée, en particulier si les marqueurs communs cartographiés entre deux espèces ne sont pas nombreux. Néanmoins, deux applications principales de la cartographie comparée suscitent l'intérêt de la communauté scientifique pour cette thématique : d'une part, la possibilité d'étudier les mécanismes de l'évolution des génomes, et d'autre part la perspective de transférer des informations génétiques et moléculaires d'une espèce à l'autre. Même si la cartographie

³ Famille multigénique: ensemble de gènes ayant de grandes ressemblances fonctionnelles et structurelles. Il s'agit de gènes homologues présents dans le même organisme dérivés par duplication d'un gène ancestral.

⁴ Pseudogène: gène dont la séquence est homologue à celle d'un gène connu, mais qui ne s'exprime pas.

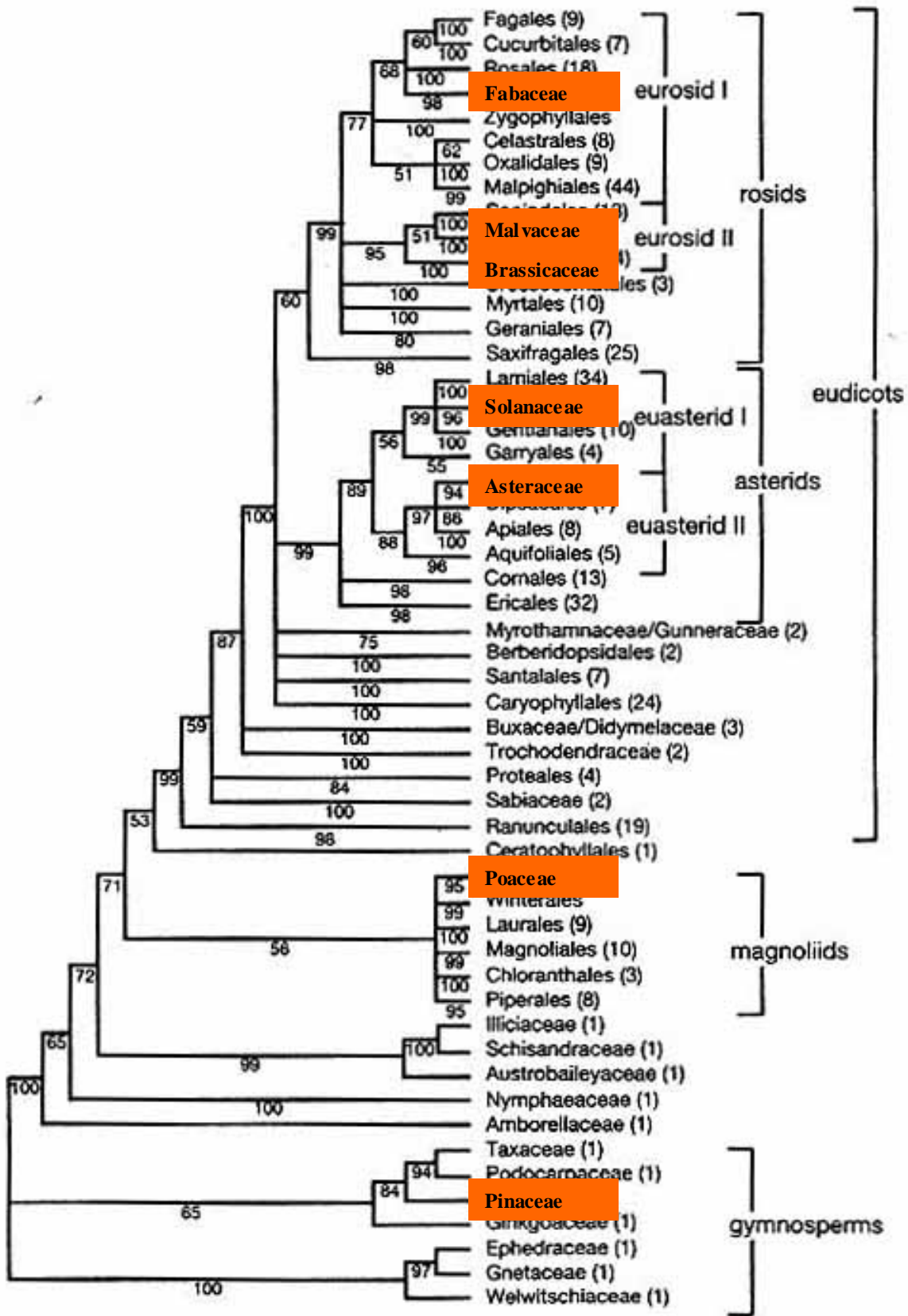
génétiq ue comparée apporte des éclaircissements importants sur la structure, l'organisation et sur le mode et la vitesse d'évolution du génome des plantes (Nadeau et Sankoff, 1998 ; Schmidt, 2000), il devient de plus en plus clair que ce n'est que la comparaison directe des séquences (d'ADN ou protéiques) de génomes entiers qui pourra élucider les processus de l'évolution de ceux-ci et le rôle joué par les réarrangements chromosomiques dans la spéciation (Kondrashov, 1999 ; Ball et Cherry, 2001 ; Mitchell-Olds et Clauss, 2002 ; Sankoff et Nadeau, 2003). En septembre 2001, le Musée d'Histoire Naturelle de New York a annoncé l'ouverture d'une section dédiée à la génomique comparée (Schachter, 2001) afin de reconstruire «l'arbre de la vie» en utilisant les données issues des projets de séquençage. Ainsi, la perspective ambitieuse de construire des arbres phylogénétiques basés sur les données de génomes entiers (Tekaiia et al., 1999) et de comparer des fonctions homologues à l'échelle génomique (Riechmann et al., 2000) devient réalité. Si l'exploitation de la cartographie comparée pour éclaircir les mécanismes de l'évolution des génomes est encore un domaine en développement, en revanche, le transfert des informations génétiques et moléculaires d'une espèce à l'autre compte un très grand nombre d'exemples. Cette approche permet, en effet, de comparer les régions génomiques impliquées dans le contrôle de caractères complexes chez différentes espèces (Paterson et al., 1995) et d'exploiter les connaissances acquises sur une espèce pour, par exemple, identifier un gène orthologue chez une espèce proche (Peng et al., 1999). Comme nous le verrons tout au long de cette thèse, ce transfert de connaissances est d'autant plus intéressant chez les espèces chez lesquelles les ressources moléculaires et génétiques (par exemple la disponibilité de EST) sont encore limitées.

A.1.3. Les études de cartographie génétique et génomique comparées chez les plantes : une perspective phylogénétique.

Les processus de mutation ponctuelle, de délétion, d'insertion et de réarrangement des séquences d'ADN sont la source de l'évolution des chromosomes. L'ensemble de ces changements au cours de l'évolution explique les différences dans l'ordre des gènes, dans le niveau de ploïdie, dans la présence de séquences répétées et dans la taille des génomes des organismes vivants. Il est désormais bien établi que les réarrangements chromosomiques ont joué un rôle dans la spéciation (Rieseberg, 2001 ; Mitchell-Olds et Clauss, 2002) et les études menées chez les mammifères ont permis de démontrer que : 1) l'arbre phylogénétique basé sur la synténie coïncide avec celui issu des études classiques basé sur des données anatomiques et fossiles (Ehrlich et al., 1997) ; 2) la vitesse à laquelle les réarrangements

Figure 3

Phylogénie des gymnospermes et angiospermes (tiré de Soltis et Soltis, 1999). En rouge les familles chez lesquelles des données de génomique comparée sont disponibles et qui seront décrites dans cette thèse.



chromosomiques se produisent varie d'une lignée évolutive à l'autre et d'une région génomique à l'autre (Murphy et al., 2001 ; Eichler et Sankoff, 2003).

Les relations évolutives entre les familles d'angiospermes ont été éclaircies à l'aide de séquences d'ADN chloroplastiques et nucléaires (Soltis et al., 1999). Cela permet aux biologistes d'interpréter les données de génomique comparée chez les plantes dans un contexte évolutif. Par exemple, la tomate (genre *Lycopersicon*) est une espèce qui est très proche phylogénétiquement de la pomme de terre (genre *Solanum*). Il n'est donc pas très étonnant que ces deux genres aient une remarquable conservation de la synténie et de la colinéarité, vu qu'ils partagent un ancêtre commun récent (Soltis et Soltis, 2003). Un autre exemple peut être trouvé chez les céréales ; deux gènes ont été identifiés chez le sorgho dans une région génomique qui a été ensuite comparée avec celle homologue du maïs et du riz. Les deux gènes n'ont été trouvés ni chez le maïs, ni chez le riz. Sachant que le sorgho et le maïs ont divergé du riz il y a 60 millions d'années et que leur divergence remonte à 16,5 millions d'années, les deux gènes sont le résultat d'une insertion chez le sorgho après sa divergence du maïs (Ilic et al., 2003). La phylogénie peut donc aider la génomique comparée de deux manières :

1. elle permet d'avoir à disposition un plan logique de comparaison ; et
2. elle aide à identifier l'apparition de nouvelles fonctions et la conservation d'autres (Laurie et Devos, 2002).

Nous disposons aujourd'hui de données de génomique comparée chez plusieurs familles de gymnospermes et angiospermes (figure 3), ce qui nous offre un échantillonnage assez représentatif permettant de répondre à quelques questions concernant l'évolution du génome des plantes. La conservation de la synténie augmente-t-elle toujours avec la diminution du temps de divergence ? Quels sont les processus principaux responsables de l'évolution du génome des plantes et comment celui-ci est-il structuré ? Nous allons donc considérer les résultats obtenus chez les Pinaceae (gymnospermes), et plusieurs familles d'angiospermes : Poaceae (monocotylédones), Asteraceae, Solanaceae, Brassicaceae, Malvaceae et Fabaceae (eucotylédones), pour donner des éléments de réponse à ces questions.

A1.3.1. Les gymnospermes : l'exemple de la famille des Pinaceae.

Les gymnospermes constituent un groupe de plantes très ancien. Ils seraient apparus à la fin du carbonifère il y a 300 millions d'années, pour ensuite atteindre leur plein développement pendant le jurassique (début du crétacé, entre 200 et 150 millions d'années). Ils comptent parmi les organismes végétaux les plus anciens, les plus grands et les plus hauts de la planète.

Des individus de l'espèce *Pinus aristata* peuvent vivre 5000 années, les séquoia géants (*Sequoiadendron giganteum*) sont les organismes végétaux ayant la plus grande circonférence et des individus appartenant à l'espèce *Sequoia sempervirens* peuvent atteindre plus de 100 m de hauteur. Les gymnospermes possèdent un génome de grande taille (en moyenne 16,99 pg/C⁵), qui peut aller jusqu'à 32,20 pg/C chez le *Pinus nelsonii* (200 fois la taille du génome d'*Arabidopsis thaliana*, 0,15 pg/C) et environ dix fois celui de la plupart des angiospermes. Le nombre de chromosomes ($2n = 2x = 24$) et le karyotype sont conservés chez les conifères, qui ont un génome caractérisé par une faible proportion d'ADN codant (0,1%) et une grande proportion de séquences hautement et moyennement répétées (76%) (Kriebel, 1985). De nombreux rétrotransposons⁶ seraient contenus dans la partie répétée du génome des pins (Elsik et Williams, 2000), tandis que la partie codante de ce génome semble être caractérisée par la présence de familles multigéniques complexes elles-mêmes répétées (Kinlaw et Neale, 1997). La taille du génome des pins a été mise en relation avec des variables environnementales, notamment les précipitations (Wakamiya et al., 1993). Il semblerait que la variation en taille soit corrélée à la taille cellulaire et à la capacité de transport de l'eau le long des trachéides (Wakamiya et al., 1996).

Malgré la grandeur et la complexité du génome des Pinaceae, plusieurs cartes génétiques sont disponibles. Ceci a conduit à la création d'un consortium international (CCGP, « Conifer Comparative Genomics Project », <http://dendrome.ucdavis.edu/Synteny>) afin d'intégrer les résultats obtenus chez différentes espèces, partager les ressources moléculaires (marqueurs moléculaires, pedigree de cartographie, position des QTL⁷) et comparer les cartes génétiques. Les résultats obtenus dans le cadre de ce projet sont discutés en détail par Chagné (2004). Une première comparaison entre les cartes génétiques de *Pinus radiata* et *Pinus taeda* a été faite par Devey et al. (1999) à l'aide de marqueurs RFLP⁸ et SSR⁹. Les 12 groupes de liaison ont pu être associés, ce qui a démontré leur homologie et la conservation de la synténie entre les deux espèces. Cependant, ni les marqueurs RFLP, ni les marqueurs SSR ne se sont révélés être des marqueurs efficaces pour la cartographie comparée chez les conifères. En effet, les marqueurs RFLP peuvent montrer des profils multibandes à cause des familles multigéniques, et les marqueurs SSR sont pour leur part peu transférables entre les espèces de conifères car

⁵ pg/C : pg d'ADN par cellule haploïde. A savoir que 1 pg = 980 Mb, 10⁶ paires de bases.

⁶ rétrotransposon : classe de transposons (fragment d'ADN mobile qui peut se déplacer d'un endroit du génome vers un autre) dont la transposition nécessite la transcription inverse de leur produit de transcription.

⁷ QTL : "Quantitative Trait Locus", locus qui contrôle un caractère complexe.

⁸ RFLP : "Restriction Fragment Length Polymorphism", polymorphisme de la longueur des fragments de restriction.

⁹ SSR : "Simple Sequence Repeat", marqueur microsatellite.

les séquences flanquantes du motif microsatellite ne sont pas suffisamment conservées pour permettre une amplification efficace de locus orthologues dans ce génome énorme. Par conséquent, dans le cadre du projet CCGP il a été décidé d'utiliser des marqueurs de type EST comme point d'ancrage entre les cartes des différentes espèces. Ces marqueurs, comme nous le verrons dans la partie D, sont idéaux pour la cartographie comparée. Dans ce projet, les cartes génétiques de trois espèces de pin ont été alignées avec la carte de l'espèce de référence, celle du *Pinus taeda*. Dix groupes de liaison ont été associés entre le *Pinus taeda* et *Pinus elliottii*, grâce à 60 marqueurs orthologues de type EST, RFLP et isozymes (Brown et al., 2001) ; dix groupes ont été de même associés entre *Pinus taeda* et *Pinus pinaster* (grâce à 32 EST, Chagné et al., 2003) et 12 groupes entre *Pinus taeda* et *Pinus sylvestris* (grâce à 42 EST, Komulainen et al., 2003). Dans tous les cas cités, l'homologie des groupes de liaison, la synténie et la macrocolinéarité ont été démontrés. Aucun réarrangement d'importance majeure n'a été révélé. De plus, Chagné et al. (2003) ont montré la conservation de quelques QTL contrôlant des caractères liés à la qualité du bois entre *P. taeda* et *P. pinaster*. D'après les résultats obtenus jusqu'à présent, depuis sa divergence, il y a 70-100 millions d'années, le génome des Pinaceae est donc resté plutôt stable. Comme cela est envisagé par Chagné (2004), il sera intéressant d'étudier des familles multigéniques entières, pour mieux comprendre la structure de ce génome riche en séquences dupliquées. La cartographie de la famille de l'alcool déhydrogénase (*Adh*) chez *Pinus banksiana* a démontré l'existence de sept gènes différents cartographiés sur deux groupes de liaison (Perry et Furnier, 1996).

A1.3.2. Les monocotylédones : le « système génétique unique » des Poaceae.

La famille des Poaceae comprend environ 10000 espèces, dont des espèces comme l'orge, le maïs, le millet, l'avoine, le riz, le seigle, le sorgho et le blé, qui représentent les céréales de base de la nourriture humaine. Il suffit de réaliser qu'ils constituent 60 pour cent de la production alimentaire mondiale pour comprendre leur importance économique. Il est donc bien compréhensible que de très nombreuses études se soient intéressées à la génétique de ces plantes afin d'améliorer leur productivité, créer des nouvelles variétés résistantes aux maladies, ou plus tolérantes aux stress abiotiques. Le grand intérêt porté pour ces espèces a donc abouti à un énorme développement des connaissances sur leur génome, jusqu'à en faire des espèces modèles pour les études de génomique chez les plantes. Cela a conduit, par exemple, au séquençage complet du génome du riz (Goff et al., 2002 ; Yu et al., 2002). Notamment en cartographie et génomique comparée les études très avancées chez les céréales ont en premier éclairci certaines caractéristiques de la structure et de l'évolution du génome

Figure 4a

Arbre phylogénétique des Poaceae représentant la taille physique des différentes espèces (tiré de Freeling, 2001).

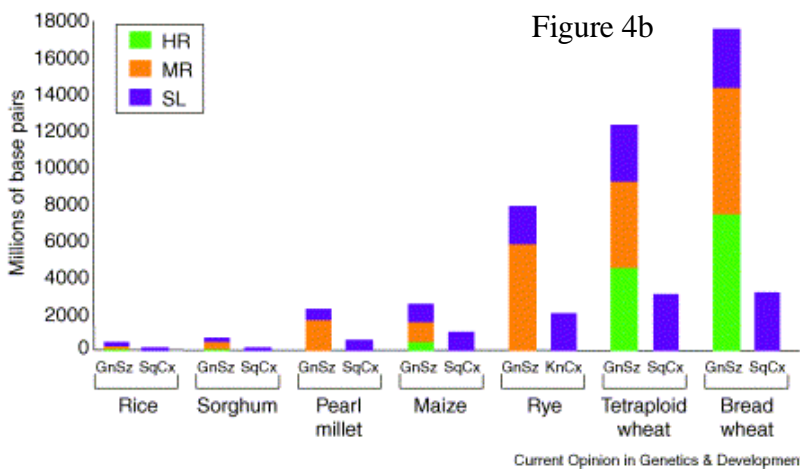
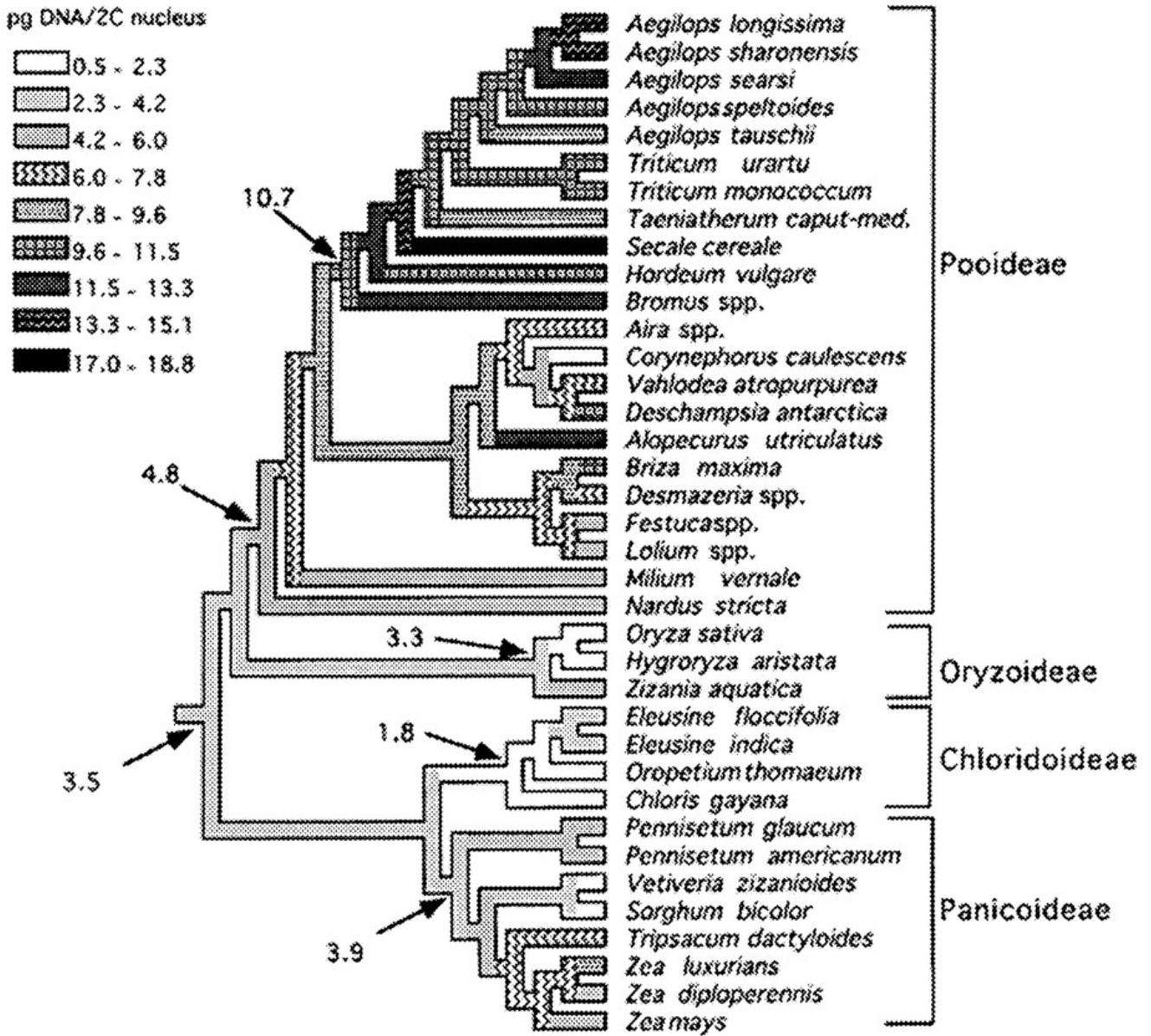


Figure 4b

Structure du génome des Poaceae: en vert les séquences hautement répétées, en orange les séquences moyennement répétées et en bleu les séquences uniques ou faiblement répétées (GnSz : taille du génome ; SqCx : complexité de séquence ; tiré de Paterson et al., 2003).

des plantes (Gaut et al., 2000). De plus, la connaissance des relations phylogénétiques entre les différents taxons de cette famille a sûrement permis une meilleure interprétation des données dérivées de la comparaison des génomes (Laurie et Devos, 2002). A l'intérieur de cette famille il y a quatre tribus principales (figure 4a) : les Pooideae, dont font partie le blé (*Triticum*), le seigle (*Secale*), l'orge (*Hordeum*) et l'avoine (*Avena*) ; les Oryzoideae, qui comprennent le riz (*Oryza*) ; les Chloridoideae et les Panicoideae, dont font partie le millet (*Pennisetum*), le sorgho (*Sorghum*), le maïs (*Zea*) et la canne à sucre (*Saccharum*) (Kellogg, 1998). Les deux premières sont typiques des zones tempérées, alors que les Chloridoideae et les Panicoideae le sont des zones tropicales. La divergence entre les lignées « tempérées » et « tropicales » remonte à il y a 60 millions d'années. A l'exception des Chloridoideae, des programmes de génomique sont en cours chez les autres espèces.

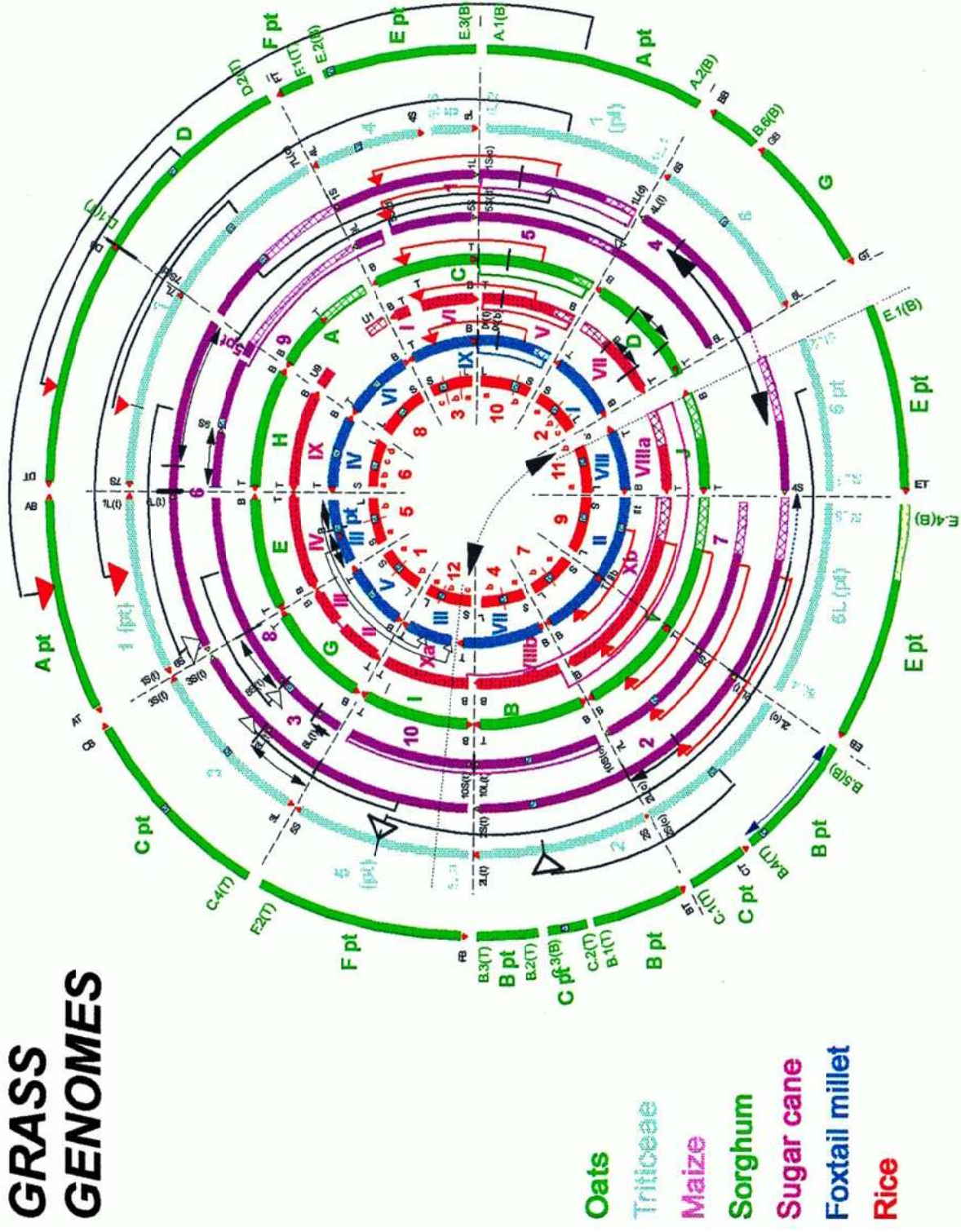
La cartographie génétique comparée. Les caractéristiques les plus étonnantes du génome de ces espèces sont, d'une part, la grande variabilité de sa taille physique (figure 4b, le riz possède un génome haploïde de 430 Mb (10^6 paires de bases) alors que celui du blé hexaploïde fait une taille de 17000 Mb) et, d'autre part, son niveau de ploïdie. Le nombre minimal de chromosomes chez les Pooideae est $x = 7$, mais il y a des espèces diploïdes comme l'orge (*Hordeum vulgare* L., $2n = 14$) et des espèces polyploïdes¹⁰ comme par exemple le blé hexaploïde (*Triticum aestivum* L. $2n = 6x = 42$). Le riz (*Oryza sativa* L.) est diploïde, mais avec un nombre de chromosomes de $n = 12$. Le maïs et le sorgho ont le même nombre de chromosomes ($2n = 20$), en revanche, chez le millet il existe une variabilité entre les espèces (*Pennisetum glaucum* L. : $2n = 2x = 14$, $C = 2,4$ pg ; *Setaria italica* L. $2n = 2x = 18$, $C = 0,54$ pg). Enfin la canne à sucre est certainement le genre avec le génome le plus complexe car nous trouvons une variation entre espèces du nombre de chromosome de 18 à 85.

Malgré l'hétérogénéité du génome de cette famille, les résultats issus de la cartographie comparée ont montré l'incroyable conservation de la synténie et de la colinéarité entre plusieurs espèces, jusqu'à arriver à la description de ce groupe de plantes comme un «système génétique unique» (Bennetzen et Freeling, 1993). La première démonstration de la colinéarité entre les trois génomes A, B et D du blé hexaploïde a été faite par Chao et al., (1989) et cela représente le premier exemple de cartographie comparée chez les Poaceae. Hulbert et al. (1990) ont ensuite démontré que des sondes développées chez le maïs pouvaient s'hybrider

¹⁰ Polyploïdie : processus d'expansion du contenu d'ADN qui détermine la formation d'une cellule ou d'un organisme contenant trois séries ou plus de chromosomes.

Figure 5

Le système génétique unique des Poaceae (tiré de Gale et Devos, 1998).



avec le sorgho, le millet et la canne à sucre. Ils ont ainsi produit le premier alignement des groupes de liaison entre maïs et sorgho, qui montrait la similarité des deux génomes. Ces résultats ont été suivis par de nombreux articles dont nous citerons quelques exemples. Ahn et Tanksley (1993) ont pu aligner les cartes génétiques du riz et du maïs, démontrant que le contenu et l'ordre des marqueurs étaient identiques dans certains chromosomes ou régions et que la plupart des gènes étaient dupliqués chez le maïs, comme cela était attendu du fait des anciennes duplications du génome du maïs. Cette conservation étonnante de l'ordre des gènes dans certaines régions chromosomiques, entre des espèces qui ont plus de 60 millions d'années de divergence, ressort également des comparaisons entre blé et riz (Kurata et al., 1994), blé, orge et seigle (Devos et al., 1993), sorgho et maïs (Pereira et al., 1994), sorgho et canne à sucre (Guimarães et al., 1997 ; Ming et al., 1998), riz et sorgho (Ventelon et al., 2001). L'hypothèse de considérer les Poaceae comme un «système génétique unique», faite par Bennetzen et Freeling en 1993, s'est donc révélée correcte. Gale et Devos (1998) ont pu aligner les génomes du blé, du maïs, du riz, de l'avoine, du sorgho, de la canne à sucre et du millet (figure 5). Trente blocs à peine du génome du riz (Devos et Gale, 2000) sont suffisants pour décrire le génome des autres espèces. Malgré la différence du nombre de chromosomes (de 5 à 12) et du contenu en ADN (de 400 à 6000 Mb) des espèces comparées, peu de réarrangements ont été révélés. Il est à noter que les bornes des trente blocs correspondent à la position des centromères et des télomères, ce qui nous amène à faire l'hypothèse du rôle de ces derniers dans l'évolution des chromosomes (Moore et al., 1997). Il faut, néanmoins, souligner que le nombre de réarrangements entre différents génomes varie beaucoup selon les lignées et qu'il ne serait pas relié à la distance évolutive entre les espèces de manière certaine (Devos et Gale, 2000). Un exemple représentatif est donné par la comparaison du millet avec le riz (Devos et al., 2000). Bien que les deux espèces *Pennisetum glaucum* L. et *Setaria italica* L. (ou bien *S. viridis* L.) soient très proches phylogénétiquement et identiquement éloignées du riz, les deux génomes sont largement réarrangés. De plus, le génome du *P. glaucum* est de même réarrangé par rapport au riz, tandis que le génome de *S. italica* a été facilement intégré dans le modèle du système génétique unique des céréales. Comme nous le verrons aussi dans la famille des Brassicaceae (Lagercrantz, 1998), il existe des lignées qui semblent être caractérisées par une vitesse d'évolution de leurs génomes et de fixation des réarrangements plus élevées.

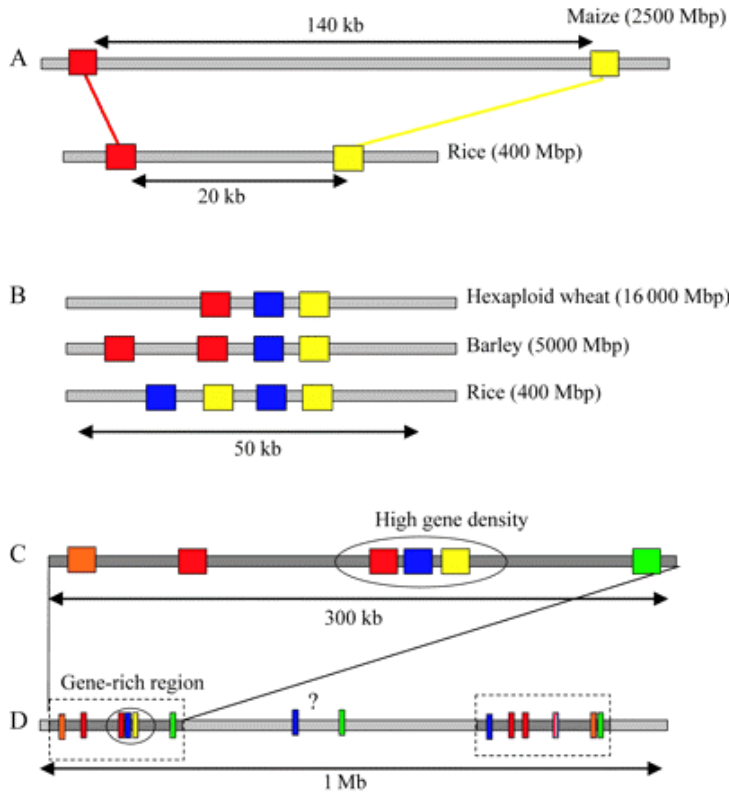
Comparaison des QTL. De la même manière, la comparaison de la position des QTL (Lin et al., 1995 ; Paterson et al., 1995 ; Guimarães et al., 1997 ; Zhang et al., 2001 ; Poncet et al.,

2002 ; Ming et al., 2002 ; Hu et al., 2003), faite à l'aide de la cartographie comparée, n'a fait que renforcer davantage l'idée de la similarité du génome de ces plantes. Nous citerons, comme exemple, le cas de la correspondance des QTL impliqués dans le caractère d'état pérenne entre riz et sorgho (Hu et al., 2003). L'état annuel ou bien pérenne est une stratégie d'adaptation aux changements climatiques saisonniers, qui souvent distingue les espèces de céréales cultivées des espèces sauvages. Des caractères clés pour l'état pérenne ont été localisés sur les cartes du riz et du sorgho. Malgré 60 millions d'années de divergence, dix QTL, sur les 16 identifiés chez le riz, correspondent à neuf QTL du sorgho, sur 11 identifiés dans ce dernier cas. La probabilité que cette correspondance soit due au hasard est de 0,00004, suggérant que le caractère est contrôlé par les mêmes régions chromosomiques chez les deux espèces.

Le riz, espèce modèle pour les céréales. L'extraordinaire synténie, colinéarité et correspondance des QTL à tout de suite conduit à envisager l'utilisation du petit génome du riz (430 Mb), désormais complètement séquencé (Goff et al., 2002 ; Yu et al., 2002), comme génome modèle pour toutes les céréales (Goff, 1999). Une étude a prouvé que le riz n'est pas un ancien polyploïde et seule la duplication d'un chromosome a été identifiée. Cette duplication remonte à 70 millions d'années, avant la divergence des autres céréales, chez lesquelles elle est toujours identifiable (Vandepoele et al., 2003). Le génome du riz semble donc très stable et sa structure serait celle du génome ancestral des céréales (Ilic et al., 2003). Il faut, toutefois, souligner que des études récentes (Paterson et al., 2004 ; Salse et al., 2004) ont démontré la présence de duplications dans le petit génome du riz et des réarrangements entre riz et maïs, qui n'avaient pas été révélées jusqu'à présent.

Une idée séduisante serait de pouvoir utiliser les gènes du riz pour identifier, grâce à la synténie et la colinéarité, les gènes orthologues chez les espèces qui possèdent un génome de grosse taille (comme le blé) et pour lesquelles aucun projet de séquençage n'a été envisagé. Pour que cette possibilité se réalise, il faut que la colinéarité, observée au niveau des marqueurs génétiques utilisés pour la construction de la carte, soit également conservée au niveau des gènes. Dans ce dernier cas nous parlons de microcolinéarité. La comparaison se réalise au niveau de séquence, et il s'agit donc d'une comparaison physique et non génétique. La comparaison de cartes génétiques hautement saturées montre déjà que tout le long du génome il y a non seulement des régions très conservées dans le contenu et l'ordre des gènes, mais aussi des régions dans lesquelles beaucoup de microréarrangements sont identifiables. Ainsi un très grand nombre de microréarrangements a été détecté entre riz et sorgho (Bowers

Figure 6a

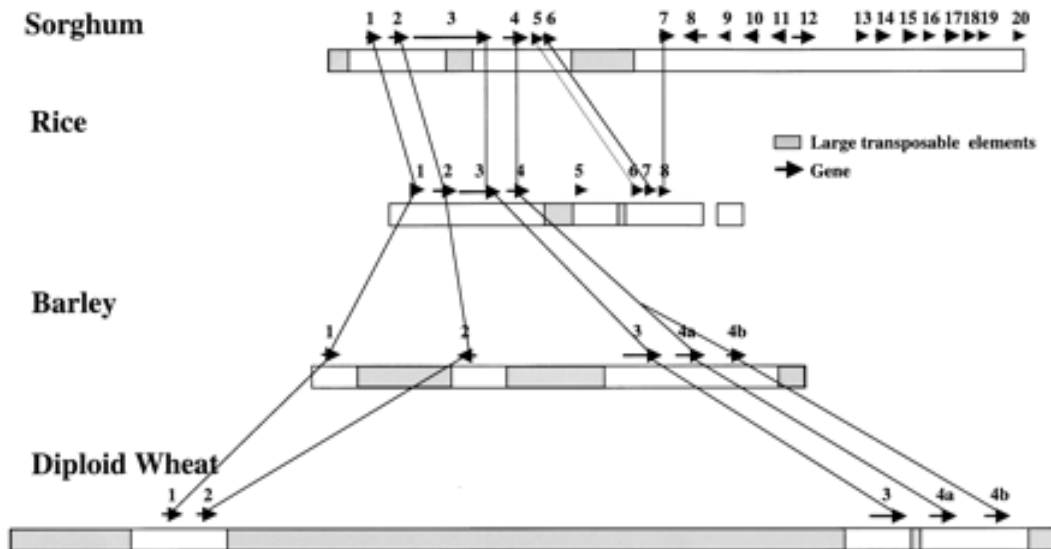


Microcolinéarité (tiré de Feuillet et Keller, 2002):

- A) les régions codantes sont conservées entre maïs et riz, mais la distance entre les gènes est très différente.
- B) Les régions orthologues montrent souvent une densité des gènes similaire, même si la taille des génomes comparés est très différente.
- C et D) Les gènes ne sont pas distribués de façon régulière tout le long du génome. Des régions riches et pauvres en gènes peuvent être identifiés.

Figure 6b

Les régions inter-géniques contiennent, en grande partie, des éléments transposables (tiré de Ramakrishna et al., 2002).



et al., 2003a) et entre riz et blé (La Rota et Sorrells, 2004). Cela pourrait-t-il empêcher d'utiliser le génome du riz comme modèle ?

La comparaison au niveau de séquence. Parmi les premières études de comparaison au niveau de séquence, deux ont montré des résultats opposés. Chen et al. (1997) ont comparé, par cartographie de restriction, la région du génome du maïs comprise entre les locus *sh2/al* et *Adh1* avec la région homologue du sorgho et du riz. Ils ont prouvé que le contenu et l'ordre des gènes étaient conservés chez les trois espèces. En revanche, Tikhonov et al. (1999) ont observé, par séquençage de la région de ce même locus *Adh1*, de nombreuses délétions, insertions et translocations entre maïs et sorgho. En effet, la situation est plutôt complexe. Il est à noter que même quand les mêmes gènes sont présents dans une région homologue, ils peuvent avoir une orientation différente, ou bien être dupliqués dans une espèce et pas dans l'autre. C'est le cas d'une région comparée entre orge et riz (Dubcovsky et al., 2001). De même, la microcolinéarité a été observée pour quatre gènes entre orge, riz, sorgho et blé, mais 48 gènes «snoARN¹¹» sont présents chez le riz et pas chez le sorgho, l'autre espèce à petit génome dans cette comparaison (Ramakrishna et al., 2002). De plus, un des gènes présente une orientation inverse chez l'orge et un autre est dupliqué chez l'orge et le blé. La différence en taille des quatre régions comparées (30 Kb chez le riz, 35 Kb chez le sorgho, 102 Kb, chez l'orge et 215 Kb chez le blé) est due aux éléments transposables. Ceux-ci seraient donc responsables de la différence en taille et en densité de gènes chez les Poaceae (Bennetzen, 2000a ; Bancroft, 2001 ; Feuillet et Keller, 2002). Song et al. (2002) sont arrivés aux mêmes conclusions grâce à la comparaison d'une large région de sorgho (425 Kb) et maïs (435 Kb) avec son homologue chez le riz. Des gènes orthologues conservés sont interposés entre des séquences non conservées contenant un mélange de gènes, éléments transposables et d'autres séquences divergentes (figures 6a et 6b).

Malgré les difficultés évidentes qui peuvent dériver des microréarrangements, l'exploitation de la génomique comparée et de la colinéarité chez ces espèces a permis des applications très intéressantes. Peng et al. (1999) ont, par exemple, montré qu'un gène orthologue chez le blé, le riz et le maïs, contrôlant la hauteur de la plante, est affecté par des mutations dans la région N-terminale de la protéine qui, dans les trois cas, causent des phénotypes équivalents chez les trois espèces et également chez *Arabidopsis*. L'approche de génomique comparée a été utilisée pour saturer une région du blé en utilisant les marqueurs orthologues de riz (Distelfeld

¹¹ snoARN: "small nucleolar RNA", petits ARN nucléolaires qui participent à la modification des ARN ribosomiaux dans le nucléole.

et al., 2004), pour identifier et caractériser une délétion d'un mutant de blé en repérant la région homologue de riz (Sutton et al., 2003) et bien sûr pour le clonage positionnel chez le blé et l'orge (Brueggeman et al., 2002 ; Yan et al., 2003 ; Yan et al., 2004). Par conséquent, une fois la microcolinéarité conservée dans la région d'intérêt, les connaissances moléculaires des espèces à petit génome (riz et sorgho) peuvent être exploitées avec succès chez les espèces à génome moins accessible (blé, maïs, orge, par exemple).

Conclusions. Des conclusions très intéressantes quand à la structure et l'évolution du génome des Poaceae peuvent être tirées des études de génomique comparée menées chez ces espèces :

- la différence en taille et nombre de chromosomes entre les espèces est due à deux processus qui semblent avoir caractérisé l'histoire évolutive de ces génomes : la duplication (de génomes entiers, polyploïdie, ou bien d'une partie du génome) et la mobilité des rétrotransposons (Bevan and Murphy, 1999). Ceux-ci représentent la différence principale entre la taille du génome du maïs et du sorgho. Cinquante pour cent du génome du maïs est représenté par des rétrotransposons et leur nombre aurait augmenté pendant les derniers 5-6 millions d'années après la divergence du sorgho et après l'événement de polyploïdisation qui a caractérisé le génome du maïs il y a 11 millions d'années (Gaut et al., 2000). Soixante-dix pour cent des régions inter-géniques chez les céréales sont représentées par des éléments transposables (Feuillet et Keller, 2002). La polyploïdie est l'un des mécanismes le plus commun de l'évolution des génomes des angiospermes. Masterson (1994) a estimé que 70 pour cent des angiospermes sont polyploïdes. Chez les Poaceae, la cartographie comparée a bien démontré par exemple l'homologie des trois génomes (A, B et D) du blé hexaploïde (Akhunov et al., 2003a) ; l'avoine et la canne à sucre peuvent être polyploïdes ; et des traces d'événements de polyploïdisation sont évidentes chez le maïs (Gaut et al., 2000). A la suite d'un événement de polyploïdisation, le génome montre une tendance rapide à rétablir la diploïdie et probablement la présence de plusieurs copies du même gène favorise l'évolution de certains d'entre eux par divergence. D'après Gaut et al. (2000) même l'activité de rétrotransposition pourrait augmenter dans ces cas, car la présence d'une autre copie du gène permettrait de mieux supporter les mutations dues à la mobilité des rétrotransposons. Si la polyploïdie et les rétrotransposons sont responsables de l'augmentation de la taille du génome chez ces espèces, une étude

récente à permis de prouver le rôle des séquences LTR¹² dans la diminution de la taille génomique (Ma et al., 2004). C'est probablement l'équilibre entre ces processus qui est à la base de l'évolution des ces génomes. En effet, l'augmentation en taille se traduit en l'évolution de nouvelles fonctions par duplication et divergence.

- Malgré l'incroyable conservation de la synténie, de la macrocolinéarité et de la localisation des QTL, de nombreux microréarrangements sont présents, comme nous avons pu le voir. Les inversions, duplications, translocations et délétions sont très fréquentes (Bennetzen et Ramakrishna, 2002 ; Bennetzen et Ma, 2003). Souvent la structure en exons et introns des gènes orthologues est elle même conservée mais les distances inter-géniques ne les sont pas. Une première observation à faire concerne la nature de la partie «non conservée» interposée aux gènes orthologues. Il s'agit évidemment de séquences répétées et de rétrotransposons (Paterson et al., 2003 ; figure 4b). Mais le contenu de ce type de séquences est très différent d'un génome à l'autre et la densité des gènes observée ne semble pas en relation directe avec la densité attendue sur la base de la taille du génome. Il est connu que chez les plantes les génomes sont organisés en régions riches et pauvres en gènes et que les régions pauvres contiennent en grande partie des séquences répétées (Barakat et al., 1998). Chez les Poaceae, tout cela implique que dans les espèces à gros génome les gènes sont plus densément organisés de ce qui serait attendu. Dans les régions riches en gènes on pourrait s'attendre à une moindre différence de la taille des régions inter-géniques entre les espèces à gros et petit génome (Keller et Feuillet, 2000). Tout cela pourrait expliquer la variabilité en microcolinéarité observée dans différentes régions du génome. De plus, récemment, Akhunov et al. (2003b) ont montré chez le blé que la recombinaison a joué un rôle majeur dans l'évolution de la structure de la diversité tout au long du génome. En effet, les régions du centromère à faible recombinaison seraient moins riches en diversité génétique que les régions éloignées du centromère et plus proches du télomère à recombinaison et diversité élevées. Par conséquent, la diversité génétique d'un locus varie en fonction non seulement de son histoire évolutive, mai aussi de sa position le long du chromosome, et les deux choses pourraient être liées (Gaut et al., 2000). Il est à noter que d'après Akhunov et al.

¹² LTR : "Long Terminal Repeat", séquences d'ADN identiques qui se trouvent aux deux extrémités d'un transposon et qui jouent un rôle dans la duplication de ces séquences.

(2003a) même la synténie diminue avec la distance du centromère, et donc elle serait moins conservée dans les régions à fréquence de recombinaison plus élevée.

- La cartographie comparée a aussi permis d'identifier un certain nombre de macroréarrangements à l'intérieur (par exemple des duplications d'une partie d'un chromosome) et entre les génomes (Devos et Gale, 2000). Ces réarrangements ne semblent pas corrélés à la distance phylogénétique entre les espèces et il y a des génomes qui les fixeraient plus rapidement que d'autres. Il est à noter qu'en général les génomes des céréales montrent une vitesse d'évolution plus rapide que les génomes des mammifères et que cela pourrait être lié, d'après certains auteurs, au succès évolutif de cette famille de plantes (Song et al., 2002). Néanmoins il faut souligner que la vitesse d'évolution du génome des Poaceae semble plutôt faible par rapport à celle d'autres familles de plantes (par exemple, les Brassicaceae). Il serait intéressant d'étudier les processus d'adaptation qui ont conduit certaines lignées à avoir une vitesse d'évolution du génome plus importante.
- Enfin, une nouvelle perspective d'étude semble de grand intérêt pour la compréhension de la variabilité de formes et fonctions des céréales et de toutes les plantes. C'est la comparaison des séquences conservées non codantes impliquées dans la régulation de l'expression des gènes. Les premiers résultats chez les céréales semblent indiquer une évolution rapide de ces séquences, ce qui suggère une divergence significative de la régulation de l'expression des gènes orthologues chez les céréales (Kaplinsky et al., 2002 ; Guo et Moose, 2003).

A1.3.3. Les eudicotylédones : Asteraceae, Solanaceae, Brassicaceae, Malvaceae et Fabaceae, cinq exemples pour comprendre la structure du génome des Angiospermes eudicotylédones.

Asteraceae (Compositae). Environ 20000 espèces font partie de cette famille qui regroupe un dixième de toutes les angiospermes connues. Son origine remonte à il y a 30-100 millions d'années en Amérique du Sud, mais ensuite une expansion très rapide a conduit à une distribution cosmopolite (Kesseli et Michelmore, 1996). Malgré l'appartenance à cette famille de plus d'une quarantaine d'espèces d'importance économique, peu d'études ont été jusqu'à présent publiées sur la génomique de ces espèces. Néanmoins, l'énorme diversification de cette famille nous offre un modèle pour des études d'adaptation. Dans ce cadre des résultats

remarquables ont été atteints permettant de décrire quelques mécanismes évolutifs du génome de cette famille.

Le génome des Asteraceae est plutôt grand (de 1,1 pg jusqu'à 10 pg par C) et nous trouvons des espèces diploïdes et polyploïdes (Kesseli et Micheltore, 1996). Le tournesol (*Helianthus annuus*, diploïde, n = 17) et la laitue (*Lactuca sativa*, diploïde, n = 9, 2,3 pg/C) sont les deux espèces les plus étudiées, pour lesquelles des cartes génétiques ont été construites (Landry et al., 1987 ; Kesseli et al., 1994 ; Rieseberg et al., 2003). Dans le cas du tournesol, cette espèce a été prise comme modèle pour étudier le rôle de l'hybridation naturelle dans l'évolution adaptative, en utilisant une approche QTL (Rieseberg et al., 2000 ; Rieseberg et al., 2003). Les résultats ont montré que l'hybridation entre espèces peut jouer un rôle dans l'adaptation à des environnements extrêmes grâce à la combinaison des allèles à effet phénotypique opposé provenant de deux parents.

En revanche, la carte de la laitue a été plutôt utilisée pour la localisation de gènes de résistance aux maladies (gènes de résistance ou gènes R). Les résultats montrent que ces gènes font partie de familles multigéniques, qu'ils sont organisés en clusters et hautement dupliqués. Des éléments transposables ont été identifiés également dans un de ces clusters (Meyers et al., 1998). Des processus de recombinaison inégale¹³ pourraient être impliqués dans la formation de nouveaux variants génétiques de résistance, phénomène déjà observé pour les gènes des immunoglobulines des Vertébrés. La comparaison au niveau de séquence de ces gènes chez le tournesol, la laitue et la chicorée suggère que la duplication des gènes, la modification et la perte des copies représentent les mécanismes principaux de l'évolution de ces séquences (Plocik et al., 2004).

Une étude très intéressante, portant sur le rôle de la duplication du génome dans la diversification écologique, a été conduite par Lawton-Rauh et al. (2003) chez des espèces d'Asteraceae qui font partie de la «Hawaiian silversword alliance». Il s'agit de trente espèces, proches phylogénétiquement, montrant une extraordinaire diversité d'adaptations anatomiques, morphologiques et écologiques, et dérivant du même ancêtre commun. Les espèces sont des allotetraploïdes issues d'un événement d'hybridation interspécifique. L'étude de la diversité nucléotidique de deux gènes impliqués dans la régulation de la diversification florale a pu montrer que différentes forces évolutives peuvent affecter les deux copies d'un

¹³ recombinaison inégal ou crossing over inégal : recombinaison entre deux séquences semblables mais pas homologues et donc mal alignées, conduisant à une duplication ou une délétion. C'est un processus à la base de la formation d'un grand nombre de variantes génétiques.

gène dupliqué. Cela suggère que la polyploïdisation a eu probablement un rôle important dans la radiation adaptative de ces espèces.

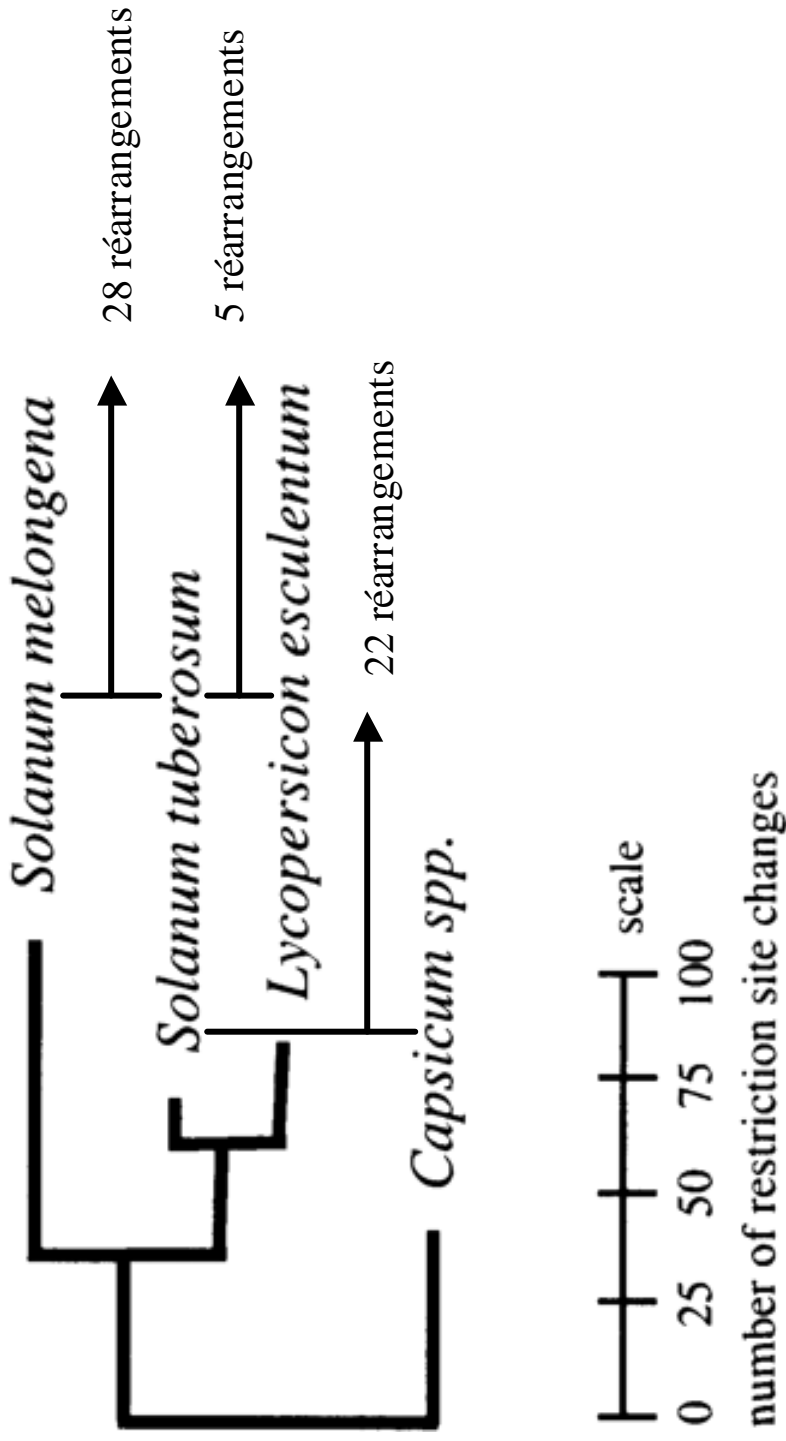
Malgré la disponibilité de cartes génétiques, un seul exemple de cartographie comparée a été publié jusqu'à présent (Burke et al., 2004). Dans ce travail, les microsatellites ont été utilisés pour la comparaison des cartes génétiques de deux espèces de tournesol. Un nombre très élevé de réarrangements chromosomiques a été mis en évidence (5.5-7.3 réarrangements par millions d'années). Cependant, les microsatellites ne constituent pas des bons marqueurs pour la cartographie comparée chez les Asteraceae. Bien que leur génome soit riche en séquences répétées et notamment en SSR (Kesseli et al., 1995), ceux-ci ne sont pas conservés et transférables d'une espèce à l'autre (Whitton et al., 1997). Un projet de génomique comparée, basé sur les EST, est, toutefois, en cours de réalisation chez les Asteraceae (<http://compositdb.ucdavis.edu/>, Michelmores et al., 2001). Ceci devrait permettre de comparer les cartes du tournesol et de la laitue et de conduire des études de microsyténie avec *Arabidopsis*.

Solanaceae. Plus de 90 genres et environ 3000 espèces constituent la famille des Solanaceae. La tomate (*Lycopersicon esculentum*, $2n = 2x = 24$), la pomme de terre (*Solanum tuberosum*, $2n = 4x = 48$), le tabac (*Nicotiana tabacum*, $2n = 4x = 48$), le poivron (*Capsicum annuum*, $2n = 2x = 24$) et l'aubergine (*Solanum melongena*, $2n = 2x = 24$) sont les espèces les plus étudiées de cette famille du point de vue génétique. Le nombre minimal de chromosomes est conservé ($n = 12$), les espèces peuvent être diploïdes, tétraploïdes ou hexaploïdes et la quantité d'ADN du génome haploïde est d'environ 0,7 pg/C pour la tomate, 2,76 pg/C pour le poivron et 1,2 pg/C pour l'aubergine. Le génome de la tomate contient 78% de séquences simple copie. Le reste est constitué par des séquences répétées (ADN ribosomal, TGRI, une séquence de 162 paires de bases répétée en tandem plus de 70000 fois dans le génome et des micro et mini satellites) (Pillen et al., 1996). Les chromosomes du genre *Lycopersicon* sont bien caractérisés au niveau cytologique. Cette connaissance a permis de relier la carte génétique à la carte physique, en montrant que la plupart des 35000 gènes estimés de la tomate seraient contenus dans une partie limitée de chaque chromosome. Les régions du centromère, riches en hétérochromatine, seraient les plus pauvres en gènes (Van der Hoeven et al., 2002).

Le premier travail de cartographie comparée chez les Solanaceae a été publié par Bonierbale et al. (1988) et il s'agit du premier exemple chez les plantes. La comparaison des cartes de la tomate et de la pomme de terre (Bonierbale et al., 1988 ; Tanksley et al., 1992) a mis en

Figure 7

Dendrogramme des Solanaceae basé sur la variation des sites de restriction de l'ADN chloroplastique (tiré de Doganlar et al., 2002a) et nombre de réarrangements détectés grâce à la cartographie comparée.



évidence l'homologie des 12 groupes de liaison, la conservation de l'ordre des marqueurs et une différence remarquable au niveau de la fréquence de recombinaison (1276 cM chez la tomate et 684 cM chez la pomme de terre). Cinq inversions paracentriques¹⁴ qui différencient les deux espèces ont été mises en évidence. Ce type de réarrangement chromosomique semble être en accord avec le karyotype très conservé de la tomate et de la pomme de terre car il ne conduit pas à des échanges inter-chromosomiques. Il est à noter que les points de cassure des cinq inversions sont localisés près du centromère. Chez la tomate, des différences remarquables ont été mises en évidence quand à la fréquence de recombinaison le long du chromosome : il y a forte suppression de la recombinaison au niveau du centromère et dans certains cas également au niveau des télomères. Il est donc probable que les inversions différenciant les deux espèces aient déplacé des gènes de régions à recombinaison élevée vers des régions à recombinaison faible ou vice versa. Cela pourrait avoir des conséquences évolutives importantes pour les gènes impliqués (Tanksley et al., 1992).

La comparaison entre les cartes du poivron, de la tomate, de la pomme de terre et de l'aubergine (Livingstone et al. ; 1999 ; Doganlar et al., 2002a) a permis d'identifier quelques mécanismes d'évolution du génome de cette famille. Une trentaine de cassures, comprenant des translocations, des inversions paracentriques et péricentriques, seraient suffisants pour expliquer les différences entre les quatre génomes. Malgré la taille physique plus élevée chez le poivron et l'aubergine, expliquée par la présence de rétrotransposons chez le poivron, le contenu en gènes est conservé. Les inversions paracentriques semblent avoir eu un rôle primaire dans l'évolution des chromosomes chez les Solanaceae, mais différents types de réarrangements semblent avoir eu lieu à différents moments de l'évolution (Doganlar et al., 2002a). Le nombre de réarrangements semble être en accord avec les distances phylogénétiques entre les espèces (figure 7) et la vitesse d'évolution chromosomique chez les Solanaceae est intermédiaire entre celle plutôt faible des Poaceae et celle élevée des Brassicaceae.

La détection des QTL a fourni au moins trois résultats de grand intérêt : 1) Frary et al. (2000) ont, pour la première fois, caractérisé un QTL au niveau moléculaire chez la tomate. Ils ont démontré que des changements dans la régulation du gène ORFX seraient responsables du QTL *fw2.2*, expliquant 30% de la variabilité en taille du fruit ; 2) la conservation des QTL a été démontrée entre les espèces (Ben Chaim et al., 2001 ; Frary et al., 2003 ; Frary et al., 2004). Dans deux cas (Thorup et al., 2000 ; Doganlar et al., 2002b), les auteurs ont fait

¹⁴ Inversion paracentrique : inversion qui se trouve sur le même bras du chromosome. A l'inverse, une inversion péricentrique implique une cassure de chaque côté du centromère.

l'hypothèse que des variations dans la régulation de gènes orthologues peuvent expliquer la diversité phénotypique chez les différentes espèces de Solanaceae ; 3) enfin, la cartographie comparée des gènes de résistance a mis en évidence la présence de « clusters » de ces gènes maintenus entre les espèces mais avec une spécificité vers des groupes différents de pathogènes (Grube et al., 2000 ; Thabuis et al., 2003).

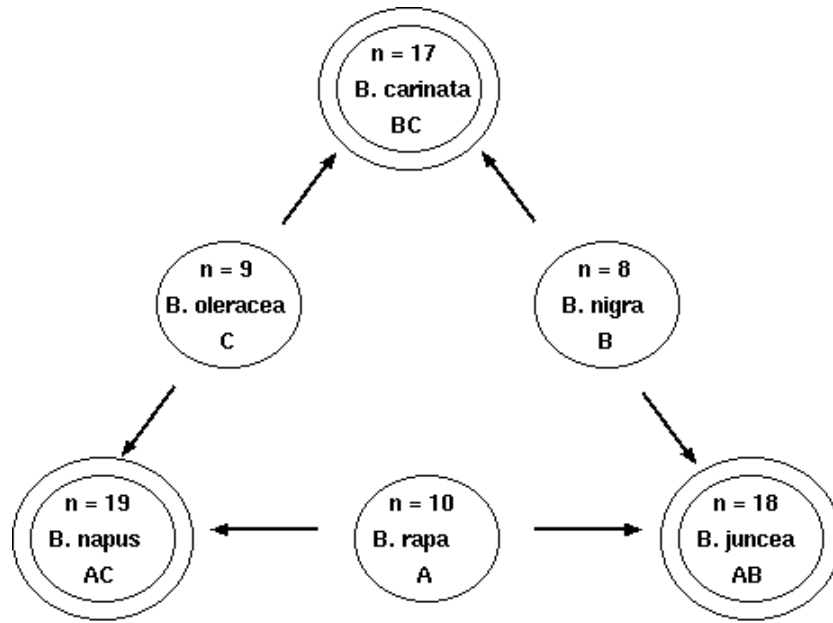
Brassicaceae (Cruciferae). La famille des Cruciferae, comprenant l'espèce modèle *Arabidopsis thaliana* et le genre *Brassica*, a été l'objet de nombreuses études de cartographie comparée, et ce pour deux raisons. D'une part le séquençage du génome de *A. thaliana* (The Arabidopsis Genome Initiative, 2000) a mis à disposition une source importante de séquences transférables à d'autres espèces, notamment les espèces phylogénétiquement proches. Cela a mené à vérifier le niveau de conservation entre les génomes d'*Arabidopsis* et des espèces de *Brassica*. D'autre part, les espèces cultivées du genre *Brassica* possèdent une organisation particulière du génome (figure 8a) : des espèces diploïdes dériveraient d'un ancêtre commun ayant 6 chromosomes (*B. rapa* $2n = 2x = 20$, génome AA ; *B. nigra* $2n = 2x = 16$, génome BB ; *B. oleracea* $2n = 2x = 18$, génome CC ; 0,5-0,7 pg/C). Par hybridation, ils seraient à l'origine des espèces amphidiploïdes¹⁵ (*B. napus* $2n = 4x = 38$, AACCC ; *B. carinata* $2n = 4x = 34$, BBCC ; *B. juncea* $2n = 4x = 36$, AABB ; 1,1-1,3 pg/C). Le genre *Brassica* constitue donc un modèle biologique intéressant pour l'étude de l'évolution du génome par duplication chez les angiospermes (Cheung et Landry, 1996). Malgré la proximité phylogénétique, les espèces de *Brassica* possèdent un génome hautement réarrangé. La comparaison des trois génomes diploïdes (*B. nigra*, *B. oleracea* et *B. rapa*) a permis de mettre en évidence la nature dupliquée de ces génomes (trois copies d'un génome ancestral sont identifiables), et un minimum de 24 réarrangements entre les trois génomes se seraient produits en moins de 10 millions d'années de divergence (Lagercrantz et Lydiate, 1996). Quand une espèce diploïde est comparée à une amphidiploïde (par exemple *B. napus* et *B. oleracea*), l'origine par hybridation de cette dernière est évidente, mais le nombre de réarrangements et leur complexité augmentent (Cheung et al., 1997).

La comparaison entre *A. thaliana* et d'autres espèces de Cruciferae a montré que la macro et microcolinéarité sont très conservées avec *Capsella rubella* (Acarkan et al., 2000), une espèce qui a divergé de *A. thaliana* il y a 6 à 9 millions d'années. Des fragments de plus de 60 cM sont entièrement conservés. En revanche, 26 réarrangements ont été mis en évidence entre *A.*

¹⁵ Amphidiploïde : organisme dérivé par hybridation de deux espèces suivie de la duplication des chromosomes.

Figure 8a

Relations entre les génomes des espèces diploïdes et amphidiploïdes appartenant au genre *Brassica*.

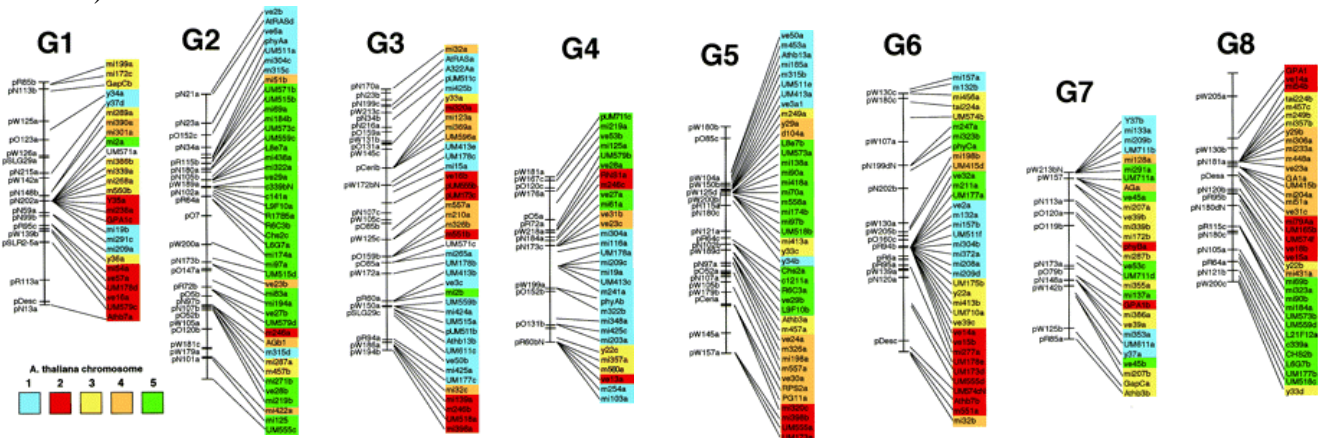


Note: *B. campestris* = *B. rapa*

http://www.umanitoba.ca/faculties/afs/plant_science/courses/39_314/l24/triangleU.gif

Figure 8b

Comparaison entre la carte génétique d'*A. thaliana* et *B. nigra*. En couleur les fragments de 8 chromosomes de *B. nigra* correspondant aux 5 chromosomes d'*A. thaliana* (tiré de Lagercrantz, 1998).



thaliana et *B. oleracea* (Kowalski et al., 1994) et la longueur des fragments colinéaires varie entre 4 et 50 cM. Récemment, des comparaisons faites par le biais des EST n'ont fait que renforcé les résultats obtenues avec les RFLP (Lan et al., 2000 ; Babula et al., 2003 ; Li et al., 2003). Environ 90 réarrangements ont été révélés entre *A. thaliana* et *B. nigra* (Lagercrantz 1998, figure 8b) et les fragments conservés ont une longueur moyenne de 8 cM. Les estimations du temps de divergence entre *Arabidopsis* et *Brassica* varient de 10 à 35 millions d'années. Il est donc évident que la fréquence de réarrangements chromosomiques est très élevée dans ces lignées. Une hypothèse très intéressante a été proposée par Kianian et Quiros (1992) et reprise par Kowalski et al. (1994). D'après ces auteurs, cette famille a évolué au cours de changements climatiques importants survenus il y a 10 millions d'années. Il est possible que l'énorme plasticité de ce génome soit liée à ces conditions environnementales, bien que la fragmentation des populations après la glaciation ait facilité la fixation des réarrangements chromosomiques dans certaines populations isolées.

Le grand nombre de réarrangements et les duplications ont compliqué la recherche et la comparaison des QTL (Osborn et al., 1997 ; Axeisson et al., 2001 ; Lan et Paterson, 2001), du fait que, par exemple, plusieurs QTL peuvent dériver du même gène dupliqué. De la même manière, les études de microsyténie ont mis en évidence que le niveau de conservation du contenu et de l'ordre des gènes varie d'une région à l'autre (O'Neill et Bancroft, 2000 ; Quiros et al., 2001, Ryder et al., 2001). Malgré la présence de zones de microcolinéarité, il reste, néanmoins, difficile de prévoir la position des gènes chez les espèces de *Brassica* à partir d'*Arabidopsis*. Toutefois, il est à noter que les séquences codantes ont une identité nucléotidique très élevée, vu la divergence récente, ce qui permet d'exploiter les séquences d'*Arabidopsis*. En revanche, les gènes de régulation semblent évoluer très rapidement (Lagercrantz et Axelsson, 2000).

L'analyse du génome d'*A. thaliana* et sa comparaison avec le genre *Brassica* ont montré le rôle très important de la duplication des séquences suivie de la perte des copies et de la polyploïdie au cours de l'évolution des génomes des plantes (Blanc et al., 2000 ; Bowers et al., 2003b). Le génome ancestral d'*Arabidopsis* aurait subi quatre événements de duplication il y a 100-200 millions d'années (Vision et al., 2000) et chaque région dupliquée aurait évolué de façon indépendante. Cela explique les microréarrangements observés dans les régions dupliquées (Bancroft, 2001). Le grand nombre de réarrangements entre les génomes d'*Arabidopsis* et *Brassica* serait la conséquence de la polyploïdie, qui, d'après certains auteurs, augmente la vitesse d'évolution des génomes des plantes (Bancroft, 2001). Les changements rapides suite aux événements de polyploïdisation chez les Brassicaceae auraient

provoqué les nombreux translocations, délétions et duplications, qui différencient les deux génomes de *A. thaliana* et *B. oleracea* (Lukens et al., 2003). Contrairement aux Poaceae, le génome d'*Arabidopsis* contient une faible quantité d'éléments transposables et les gènes sont bien repartis dans tout le génome, les régions pauvres en gènes étant concentrées au niveau des centromères et des télomères (Barakat et al., 1998). La différence en taille entre *A. thaliana* et *B. rapa* est principalement due à la duplication plutôt qu'à l'accumulation de séquences répétées (Jackson et al., 2000). Dans les deux familles des Poaceae et des Cruciferae, des mécanismes différents semblent donc avoir caractérisé l'histoire évolutive de leurs génomes.

Malvaceae. Le genre *Gossypium*, auquel appartient le coton, est certainement le plus étudié du point de vue génétique au sein de cette famille. Des espèces diploïdes ($n = 13$) et polyploïdes (par exemple *Gossypium hirsutum*, $2n = 4x = 52$) font partie de ce genre. L'apparition du coton tetraploïde, qui contient deux génomes dits A et D, remonte à 1 à 2 millions d'années par un événement de polyploïdisation suite à la migration sur le continent américain de l'ancêtre diploïde, portant le génome A. Les deux génomes A ($C = 1,9$ pg) et D ($C = 1,0$ pg) ont donc évolué sur deux continents différents, et leur divergence est estimée de 6 à 11 millions d'années (Wendel et Cronn, 2003). Du fait de sa polyploïdisation récente, ce genre constitue un modèle d'étude de ce processus important dans l'évolution du génome des plantes. La construction de cartes génétiques et la comparaison des génomes A et D (Reinisch et al., 1994 ; Brubaker et al., 1999 ; Rong et al., 2004) ont montré que : 1) les deux génomes A et D sont globalement conservés ; 2) malgré la différence de taille physique, les distances génétiques sont comparables entre les génomes A et D, ce qui veut dire que la différence en séquences répétées n'a aucune influence sur la fréquence de recombinaison. En revanche, des différences de taux de recombinaison sont évidentes entre les génomes diploïde et allopolyploïde ; 3) deux translocations ont été mises en évidence entre les génomes A tetraploïde et A diploïde. Une vingtaine d'inversions dans l'ordre des marqueurs différencient les deux génomes diploïde et tetraploïde ; 4) plusieurs duplications suggèrent un événement de polyploïdisation plus ancien (il y a 25 millions d'années). Peu de réarrangements se seraient donc produits après l'allopolyploïdisation chez le coton contrairement à ce qui s'est passé chez les Brassicaceae (Liu et al., 2001). La polyploïdisation aurait, en tous cas, offert la possibilité d'un potentiel d'adaptation plus élevé avec la fusion des caractéristiques de deux génomes (Jiang et al., 1998) et grâce à la présence du même gène en plusieurs copies, qui

peuvent évoluer de façon indépendante et être régulés différemment (Cronn et al., 1999 ; Adams et al., 2003).

Fabaceae. La famille des Fabaceae, ou légumineuses, est caractérisée par une grande variété d'espèces (environ 17000) et plusieurs espèces d'importance économique font partie de ce groupe : le soja (*Glycine max*, $2n = 4x = 40$), le haricot (*Phaseolus vulgaris*, $2n = 2x = 22$), le pois (*Pisum sativum*, $2n = 2x = 14$), la luzerne (*Medicago sativa*, $2n = 4x = 32$), l'arachide (*Arachis hypogaea*, $2n = 4x = 40$), le pois chiche (*Cicer arietinum*, $2n = 2x = 16$), la fève (*Vicia faba*, $2n = 2x = 16$). La fixation de l'azote au niveau des nodosités, induites par le rhizobium, est caractéristique de cette famille et de nombreuses études visent à comprendre les aspects moléculaires et le contrôle génétique de ce processus (Kathryn et al., 2003).

Le nombre de chromosomes, le niveau de polyploïdie et la taille physique du génome (d'environ 0,5 pg/C chez le haricot et jusqu'à 12,5 pg/C chez la fève) varient entre les espèces (Young et al., 1996). Les premières études de cartographie ont mis en évidence la nature dupliquée du génome du soja et la conservation de la fonction des gènes dupliqués (Shoemaker et al., 1996). Malgré la duplication du génome du soja, des segments conservés ont été identifiés entre ce génome et celui de *Phaseolus vulgaris* et *Vigna radiata*, deux génomes diploïdes phylogénétiquement proches (Boutin et al., 1995 ; Lee et al., 2001). Les segments conservés sont présents deux fois chez le soja. De plus, la comparaison avec *Arabidopsis* montre que les mêmes fragments sont en partie conservés chez cette espèce et ils sont restés conservés depuis la divergence d'*Arabidopsis* des légumineuses, il y a 90 millions d'années. Récemment, l'espèce diploïde *Medicago truncatula* a été choisie comme modèle chez les légumineuses pour son génome relativement petit (environ 0,5 pg/C), le status de plante annuelle, l'auto-fertilité, la variabilité naturelle élevée et la proximité phylogénétique avec les autres espèces de légumineuses (Kathryn et al., 2003 ; Choi et al., 2004). Des études de microsyténie entre *M. truncatula*, la soja et le pois ont montré la possibilité de transfert d'information moléculaire entre les trois espèces (Gualtieri et al., 2002 ; Yan et al., 2003 ; Yan et al., 2004).

D'autres résultats de cartographie génétique comparée chez les Rosaceae (Dirlewanger et al., 2004), l'*Eucalyptus* (Marques et al., 2002), le peuplier (Cervera et al., 2001) ont montré de même la conservation de la synténie et des QTL chez des espèces proches phylogénétiquement.

A1.3.4. Comparaison entre taxa éloignés phylogénétiquement.

Paterson et al. (1996) ont estimé que de 43 à 58% de fragments chromosomiques d'une longueur de 3 cM ou moins, devrait rester colinéaire entre monocotylédones et eucotylédones. Cette étude, basée sur les premières données partielles de cartographie comparée, constitue la première tentative de comparaison entre taxa éloignés phylogénétiquement. Selon cette étude, une fois ces régions colinéaires identifiées, les séquences connues d'*Arabidopsis* pourraient être exploitées pour repérer les séquences orthologues dans l'espèce d'intérêt. Cette perspective a mené au développement de toute une série d'études visant à démontrer l'existence de régions colinéaires entre monocotylédones et eucotylédones. La nature dupliquée du génome d'*Arabidopsis*, la faible identité de séquence avec le riz et le maïs et les fréquents microréarrangements entre *Arabidopsis* et les deux monocotylédones ont tout de suite découragé les chercheurs (Devos et al., 1999 ; Liu et al., 2001 ; Brendel et al., 2002 ; Salse et al., 2002 ; van Buuren et al., 2002). Malgré l'identification de certaines régions microcolinéaires entre le riz et *Arabidopsis* (Mayer et al., 2001, Salse et al., 2002), ce dernier ne constitue pas un bon génome modèle pour le monocotylédones. En revanche, la comparaison de la structure en exons et introns de gènes homologues montre la colinéarité des exons, ce qui justifie l'utilisation du génome d'*Arabidopsis* pour l'annotation de celui du riz (Liu et al., 2001).

Au contraire, la comparaison d'*Arabidopsis* avec d'autres espèces eucotylédones a permis d'identifier quelques régions colinéaires et d'utiliser ce génome pour la cartographie fine chez la tomate (Oh et al., 2002) et le clonage positionnel chez les légumineuses (Stracke et al., 2004). Néanmoins, les résultats suggèrent que cette colinéarité reste limitée. Par exemple, la comparaison entre *Arabidopsis* et les légumineuses (Fabaceae) a d'une part mis en évidence l'identification d'une région de colinéarité avec le soja (Grant et al., 2000). Cette région de colinéarité n'a pas été confirmée dans une comparaison avec *Medicago truncatula* (Zhu et al., 2003). Cela suggère une variabilité du niveau de la conservation de la colinéarité d'une région à l'autre du génome et d'une espèce à l'autre. D'autre part, les traces de duplications segmentales fréquentes, suivies de réarrangements chromosomiques et d'une évidente divergence des régions dupliquées, ont mis en cause l'utilisation d'*Arabidopsis* comme modèle pour toutes les légumineuses. De plus les gènes impliqués dans le processus de fixation de l'azote, caractéristique des Fabaceae, semblent n'avoir pas d'homologues chez *Arabidopsis* (Zhu et al., 2003).

De la même manière, la comparaison entre *Arabidopsis* et la tomate a mise en évidence que :

- 1) la colinéarité entre espèces éloignées phylogénétiquement est conservée à courte distance,

mais 2) certains gènes peuvent être absents ou dupliqués. Deux conséquences très intéressantes peuvent être tirées de ces études : 1) la duplication de vastes portions de génome suivie de la perte sélective de gènes et la divergence des séquences semble avoir joué un rôle de grande importance au cours de l'évolution des génomes des eucotylédones ; 2) le génome d'*Arabidopsis*, considéré petit et simple, s'est révélé très complexe à cause de plusieurs événements de duplication et son exploitation comme génome modèle est loin d'être facile. La différence en taille des séquences intergéniques entre *Arabidopsis* et la tomate n'est pas due à la présence de rétrotransposons. Il est possible que cette différence dérive surtout d'une différence de vitesse d'accumulation ou délétion de séquences non codantes entre les deux espèces. Il est à noter que cette différence pourrait être liée à une pression sélective vers une différente taille optimale des cellules dans une espèce à état pérenne (comme la tomate dans son habitat naturel) et une annuelle (*Arabidopsis*) (Ku et al., 2000 ; Rossberg et al., 2001).

Il est évident que les données encore incomplètes (Delseny, 2004) ne permettent pas de vérifier de manière exhaustive l'hypothèse de Paterson et al. (1996) selon laquelle il existerait des fragments de génome conservés (SCEUS¹⁶) dans toutes les angiospermes. Néanmoins, Fulton et al. (2002) ont identifié un ensemble de gènes, dits COS¹⁷, qui sont restés très conservés au cours de l'évolution car ils codent pour des fonctions impliquées dans le métabolisme et les processus de base du fonctionnement d'une cellule. Ces gènes ont une vitesse d'évolution moins rapide que les gènes de régulation comme certains facteurs de transcription (Van der Hoeven et al., 2002) et donc la cartographie des orthologues chez plusieurs espèces, même éloignées phylogénétiquement, devrait être plus simple grâce à la conservation de séquence plus élevée. La cartographie de gènes très conservés chez la pomme de terre (Gebhardt et al., 2003) et plusieurs espèces d'eucotylédones (*Beta vulgaris*, *Helianthus annuus*, *Prunus* spp., *Arabidopsis thaliana*, Dominguez et al., 2003) a mis en évidence la présence de blocs de gènes conservés dans toutes les espèces. L'hypothèse d'un avantage sélectif offert par ces arrangements de gènes, qui devraient représenter des « opérons » chromosomiques, a été proposée (Lee et al., 2001 ; Gebhardt et al., 2003). Cependant, il n'est pas encore possible de démontrer que cette hypothèse soit plus probable que la simple conservation de blocs de gènes due au hasard. Des données plus complètes, (vastes régions génomiques analysées, gènes de nature différente, plus d'espèces) et donc plus significatives du point de vue statistique, pourront éclaircir ce point.

¹⁶ SCEUS : "Shortest Conserved Evolutionary Unit Sequences".

¹⁷ COS : "Conserved Ortholog Set markers".

Tableau 1 Fréquence de réarrangements chromosomiques pour différentes taxa (tiré de Paterson et al., 1996).

Taxon	Génes (nombre de chromosomes)	Réarrangements chromosomiques	Temps de divergence estimé (MA ¹)	Réarrangements par chromosome par MA ¹
Dicotylédones				
Solanaceae	<i>Lycopersicon/Solanum</i> (12)	5	10	0.042
	<i>Lycopersicon/Capsicum</i> (12)	33	40 (±10)	0.069
Brassicaceae	<i>Arabidopsis/Brassica</i> (5/9)	26	<10	0.52
Malvaceae	<i>Gossypium</i> spp. (13)	9	4-11	0.11
Monocotylédones				
Poaceae	<i>Zea/Sorghum</i> (10)	≥14	24	0.058
	<i>Zea/Oryza</i> (10/12)	23	66	0.035
	<i>Oryza/Triticum</i> (12/7)	≥20	66	0.043
	<i>Triticum/Secale</i> (7/7)	13	6	0.31
Mammalia				
Primate/Rodentia	<i>Homo/Mus</i> (24/20)	138	100	0.069
Moyenne				0.14 (±0.06)

1/MA : millions d'années.

A.1.4. Les principaux résultats et les questions issues des études de comparaison des génomes chez les plantes.

Malgré l'hétérogénéité des données (grande variabilité du nombre de marqueurs orthologues utilisés dans chaque comparaison) et l'absence de données exhaustives (tous les réarrangements ne sont pas détectés, la microcolinéarité est étudiée dans des portions très limitées du génome), les résultats des études de génomique comparée nous apportent des éléments très importants pour la compréhension de la structure et de l'évolution du génome des plantes. Nous essaierons de répondre donc à quelques questions en nous basant sur les résultats précédemment décrits.

- **Le nombre de réarrangements augmente-t-il avec le temps de divergence entre les espèces comparées ?** En général, plus les espèces sont éloignées plus le nombre de réarrangements augmente, mais il n'y a pas une forte corrélation entre le temps de divergence et le nombre de réarrangements. Ceci peut-être dû aux données incomplètes de cartographie comparée et des erreurs sur l'estimation du temps de divergence. Cependant, certaines lignées montrent beaucoup plus de réarrangements, en dépit du court temps de divergence. C'est, notamment, le cas des Brassicaceae (Lagercrantz et Lydiate, 1996) et du millet (Devos et al., 2000). Le tableau 1, tiré de Paterson et al. (1996), montre le nombre de réarrangements par chromosome par millions d'années chez plusieurs espèces.
- **Quels types de réarrangements observe-t-on?** Les réarrangements chromosomiques majeurs semblent moins fréquents que les microréarrangements (Bennetzen, 2000b), cela explique la conservation de la synténie et de la colinéarité observée souvent lors de la comparaison des cartes génétiques chez les espèces proches phylogénétiquement. Parmi les réarrangements majeurs observés, les inversions paracentriques sont les plus fréquentes. Celles-ci n'impliquent pas le centromère et elles permettent, donc, une conservation du karyotype entre les espèces.
- **Quelles régions du génome sont impliquées dans ces réarrangements?** Les points de cassure des inversions observées sont souvent localisés près du centromère. Des séquences répétées associées au télomère ont été trouvées près du centromère et vice versa (Bennetzen, 1998). Les locus près du centromère ont tendance à avoir une

fréquence de recombinaison et un niveau de diversité génétique plus faible que les locus près du télomère (Gaut et al., 2000). D'après ces auteurs, dans les régions à faible recombinaison, le déséquilibre de liaison¹⁸ provoqué, par exemple, par l'effet de la sélection sur un locus devrait concerner une région plus grande que celle affectée au niveau du télomère. Tout cela implique que le niveau de diversité génétique varie en fonction de la position du gène sur le chromosome et qu'il aurait des régions plus pauvres en diversité. Il est donc évident que le déplacement des locus d'une région à faible recombinaison vers une région à recombinaison élevée entraîne des conséquences évolutives importantes.

- **Le gènes sont-ils distribués de façon régulière dans le génome et leur position le long d'un chromosome est-elle importante ?** Comme nous avons pu le voir précédemment, l'identification de blocs de gènes conservés depuis longtemps (Dominguez et al., 2003) chez plusieurs espèces d'angiospermes suggère la conservation de groupes particuliers de gènes. Deux questions se posent alors: 1) y aurait-il un effet de position ? 2) est-il imaginable que des blocs de gènes puissent constituer des unités d'évolution du génome ? Les gènes ne sont pas distribués régulièrement sur le génome. Ils sont absents ou rares dans les régions riches en hétérochromatine (télomères et centromères, Bennetzen, 1998). Si nous considérons, de plus, la variation en fréquence de recombinaison et diversité génétique observée tout le long du chromosome, nous pouvons faire l'hypothèse que la position a un effet important sur la vitesse d'évolution des gènes. En revanche, même si des blocs conservés depuis longtemps ont été identifiés, leur raison d'être n'est pas connue. Il n'y a pas des données qui puissent confirmer l'hypothèse de blocs fonctionnels mélangés au cours de l'évolution (Paterson et al., 2000).
- **Comment pouvons-nous expliquer l'énorme variation en taille du génome des plantes ?** La taille physique du génome des plantes varie de 38 Mb (*Cardamine amara*, Cruciferae) jusqu'à 87000 Mb (*Fritillaria assyriaca*, Liliaceae). Cette énorme variation est expliquée par le niveau de ploïdie et le contenu en séquences répétées, particulièrement en rétrotransposons. La duplication (du génome entier ou de gènes individuels) et l'expansion des rétrotransposons expliquent l'augmentation en taille du

¹⁸ Déséquilibre de liaison : situation dans laquelle deux allèles de deux locus distincts sont plus fréquemment associés dans une même population que ne le voudrait le hasard.

génomique (Bennetzen, 1998), alors que des processus comme la recombinaison illégitime¹⁹ (Devos et al., 2002) sont responsables de sa réduction.

- **Pourquoi la polyploïdie et les rétrotransposons ont-ils joué un rôle tellement important au cours de l'évolution des génomes des plantes ?** D'une part, parce que la duplication des gènes permet la création de « matériel moléculaire » pour l'évolution de nouvelles fonctions. Même si cette dernière possibilité ne semble pas fréquente (Lynch et Conery, 2000), les génomes sont sujets à des réarrangements importants après des événements de polyploïdie. Ces réarrangements vont souvent rétablir un état diploïde et ils détermineraient l'ensemble des microréarrangements que nous pouvons observer actuellement. La polyploïdie est donc considérée comme un phénomène qui augmente la vitesse d'évolution du génome (Bancroft, 2001 ; Barnes, 2002). D'autre part, l'action mutagène des transposons est bien connue. Leurs mouvements peuvent générer toute sorte de mutation, dans des séquences codantes ou régulatrices. Ils représentent aussi des séquences homologues dispersées sur le génome, cela peut provoquer des réarrangements chromosomiques par recombinaison illégitime (Heslop-Harrison, 2000 ; Bennetzen, 2000a).
- **Ya-t-il des différences entre les familles ?** L'hétérogénéité des données et le niveau très différent des analyses ne permettent pas de faire une comparaison fiable. Par exemple, les gymnospermes ont un génome apparemment stable, mais le faible nombre de marqueurs utilisé n'est pas suffisant pour mettre en évidence des réarrangements. Néanmoins, les Poaceae, les Solanaceae et les Brassicaceae, trois familles suffisamment étudiées, montrent des différences importantes. Une taille physique très variable, associée à la présence d'éléments transposables, est observée chez les Poaceae. Peu de réarrangements caractérisent cette famille. Le même niveau de séquences répétées est identifiable chez les Solanaceae et les Brassicaceae (20% à peu près), les éléments transposables sont absents ou peu nombreux. Toutefois, le génome des Solanaceae est peu réarrangé, comme les Poaceae, alors que celui des Brassicaceae est caractérisé par un très grand nombre de réarrangements.

¹⁹ Recombinaison illégitime : recombinaison entre séquences sans homologie majeure.

- **Perspectives d'étude encore ouvertes.** **1)** Les plantes sont adaptées à des environnements très contrastés, certaines ont subi des processus importants de domestication, l'hybridation entre les espèces est fréquente et le temps d'une génération à l'autre ainsi que leur aire de répartition varient énormément. Certaines variations des caractéristiques des génomes des plantes pourraient être directement liées à l'adaptation. Barbara McClintock (1984) a formulé une hypothèse selon laquelle des stress environnementaux pourraient être à la base de changements génomiques, mais peu d'études dans ce sens ont été menées (Wakamiya et al., 1993). **2)** Une attention particulière dans la comparaison des génomes devrait être donnée aux séquences de régulation de l'expression des gènes et aux processus épigénétiques (Mitchell-Olds et Clauss, 2002). **3)** L'évolution du génome se réalise au niveau de la séquence d'ADN, des gènes, des fragments de chromosomes et des chromosomes. C'est pourquoi pour comprendre les changements évolutifs d'une espèce à l'autre, lors d'une comparaison des génomes, tous les niveaux d'évolution devraient être pris en compte (Heslop-Harrison, 2000).

Figure 9a

Classification des Fagacées proposée par Nixon (1989) basée sur la morphologie. Entre parenthèses le nombre d'espèces.

Fagaceae

Fagoideae :

- Fagus* (12)
- Quercus* (450-600)
- Trigonobalanus* (1)

Castaneoideae :

- Chrysolepsis* (2)
- Castanea* (7)
- Castanopsis* (120)
- Lithocarpus* (300)

Figure 9b

Classification du genre *Quercus* proposée par Camus (1936-1954).

Genre : *Quercus*

Sous-genre :

- Cyclobalanopsis*
- Euquercus*

Section :

- Cerris*
- Mesobalanus*
- Lepidobalanus*
- Macrobalanus*
- Protobalanus*
- Erythrobalanus*

Figure 9c

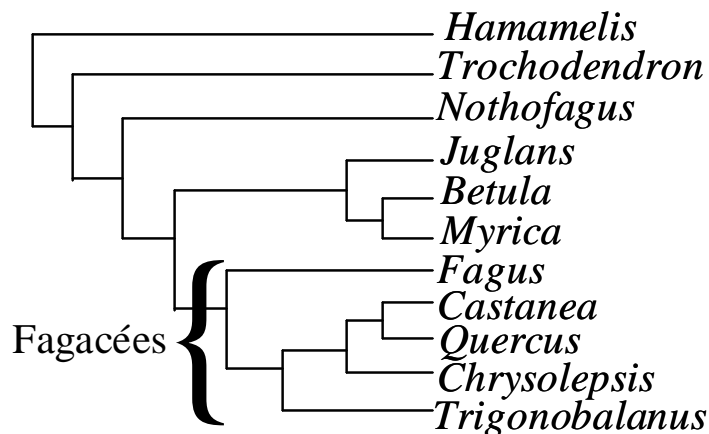
Les sept espèces de châtaignier.

Genre : *Castanea*

- C. sativa*
- C. dentata*
- C. mollissima*
- C. crenata*
- C. seguinii*
- C. henryi*
- C. pumila*

Figure 9d

Hypothèse de phylogénie pour les Fagacées (Manos et Steele 1997, Manos *et al.* 2001).



A.2. LES FAGACÉES.

Les deux espèces *Quercus robur* L. et *Castanea sativa* Mill. feront l'objet de cette étude. Le chêne pédonculé et le châtaignier européen sont deux espèces forestières majeures en Europe et dans tout l'hémisphère Nord, qui appartiennent à la famille des Fagacées. Une présentation de cette famille, avec une attention particulière aux deux espèces qui nous intéressent, sera faite dans ce paragraphe. Nous ferons référence dans cette partie aux résultats et à la synthèse bibliographique faite par Xu (2004).

A.2.1. Taxonomie.

La famille des Fagacées, appartenant à l'ordre des Fagales, contient sept genres : *Castanea*, *Castanopsis*, *Chrysolepsis*, *Fagus*, *Lithocarpus*, *Quercus* et *Trigonobalanus* (Figure 9a). Environ 1000 espèces font partie de cette famille. Cette classification est principalement basée sur des caractères morphologiques de la fleur et du fruit. En effet, les genres *Fagus*, *Quercus* et *Trigonobalanus*, regroupés dans la sous-famille des Fagoideae, possèdent des fleurs mâles pendantes et la pollinisation est anémophile. En revanche, les genres *Castanea*, *Castanopsis*, *Chrysolepsis*, *Lithocarpus*, à pollinisation entomophile, possèdent les fleurs mâles érectiles et sont regroupés dans la sous-famille des Castaneoideae (Manos et al., 2001).

Le chêne pédonculé (*Q. robur* L.) appartient donc au genre *Quercus*, qui est un des genres dominant de cette famille, avec 400-600 espèces selon les auteurs. La classification la plus exhaustive est celle proposée par Camus (1936-1954), qui distingue les sous-genres *Euquercus* et *Cyclobalanopsis*. Dans le sous-genre *Euquercus*, nous trouvons les deux sections de *Lepidobalanus* et *Erythrobalanus*, dont font partie les chênes blancs et les chênes rouges (Figure 9b).

Le châtaignier européen (*C. sativa* Mill.) appartient à l'autre sous-famille, celle des Castanoideae. Le genre *Castanea* contient sept espèces, dont quatre, le châtaignier européen (*C. sativa*), le châtaignier américain (*C. dentata*), le châtaignier chinois (*C. mollissima*) et le châtaignier japonais (*C. crenata*) sont les plus connues car il s'agit d'espèces cultivées, économiquement importantes (Figure 9c). Les autres espèces, *C. seguinii*, *C. henryi* et *C. pumila*, ne sont pas exploitées par l'homme (Rutter et al., 1990).

A.2.2. Origine et phylogénie.

Les nombreux restes fossiles ont permis de dater l'apparition des Fagaceae à la transition Secondaire-Tertiaire, il y a environ 80-60 millions d'années. Le genre *Fagus* se serait séparé

Figure 10a

Aire de répartition des Fagacées (tiré de Manos et Steele, 1997).



Figure 10b

Aire de répartition du chêne pédonculé en Europe.

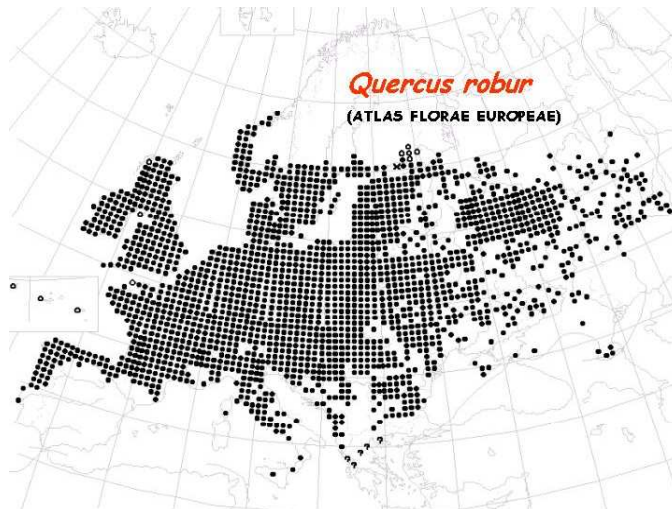
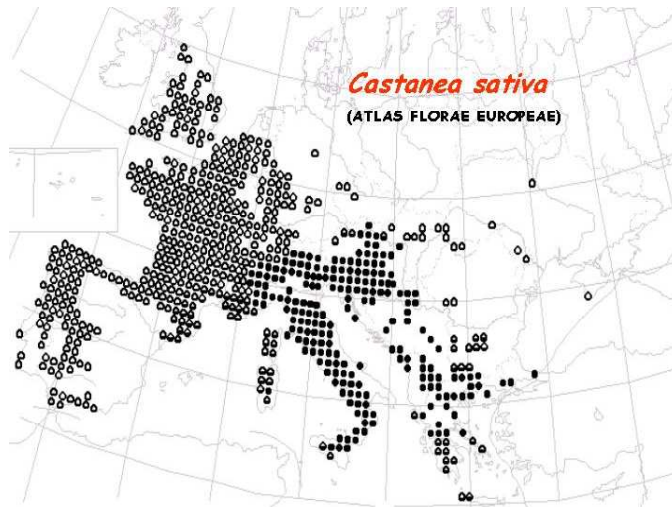


Figure 10c

Aire de répartition du châtaignier européen.



des autres genres de Fagaceae précocement, en revanche une polytomie caractériserait l'origine des genres *Castanea*, *Castanopsis*, *Chrysolepsis*, *Lithocarpus*, et *Quercus*. Ces derniers se seraient diversifiés de façon rapide et simultanée (Manos et Stanford, 2001 ; Xu, 2004). Le temps de divergence minimum entre les deux genres *Quercus* et *Castanea* a été estimé à 60 millions d'années, les deux genres sont phylogénétiquement très proches (Figure 9d). Les premiers restes fossiles de Fagaceae ont été trouvés en Europe et en Amérique du Nord. Pour le genre *Quercus* des restes fossiles ont été trouvés en Amérique du Nord, en Europe et en Asie. Cela rend difficile établir avec certitude l'origine géographique et temporelle de ce genre. D'après les résultats de Xu (2004), l'origine asiatique serait la plus probable, suivie d'une colonisation successive en Europe et Amérique du Nord des deux sections *Cerris* et *Lepidobalanus*. Le genre *Castanea*, apparu il y a environ 60 millions d'années, semble avoir colonisé l'hémisphère boréal avant la séparation définitive de l'Europe de l'Amérique du Nord (Krebs et al., 2004).

A.2.3. Distribution géographique.

Comme le montre la figure 10a, les espèces de Fagacées dominent l'hémisphère Nord de la planète (Manos et Steele, 1997). Elles se trouvent notamment dans la région tempérée de cet hémisphère, dans toute la région méditerranéenne et les régions montagneuses de l'Asie du Sud, de l'Amérique centrale et du Sud. Les genres *Quercus*, *Castanea* et *Fagus* sont distribués dans les trois continents, au contraire les autres genres (*Lithocarpus*, *Chrysolepsis*, *Trigonobalanus* et *Castanopsis*) ne sont pas présents en Europe. Le genre *Quercus*, notamment le sous-genre *Euquercus*, occupe vastes zones de toute l'aire de répartition des Fagacées. Au contraire le sous-genre *Cyclobalanopsis* est situé en Asie. Deux espèces dominent en Europe, le chêne pédonculé (*Q. robur* L., figure 10b) et le chêne sessile (*Q. petraea* Matt.).

Quatre espèces de châtaignier sont natives d'Asie, *C. mollissima*, *C. seguinii*, *C. henryi* et *C. crenata*, deux d'Amérique du Nord, *C. dentata* et *C. pumila*, et, enfin, le châtaignier européen (*C. sativa*) couvre une partie de l'Europe et notamment le bassin méditerranéen (Figure 10c). Son origine est probablement turque et caucasienne (Zohary et Hopf, 1988). Cependant, plusieurs refuges, pendant la dernière glaciation, ont été identifiés pour cette espèce dans la région turque et caucasienne ainsi que dans les péninsules italienne et ibérique (Krebs et al., 2004).

A.2.4. Intérêt écologique et économique.

Le genre *Quercus* joue en rôle majeur du point de vu écologique simplement pour l'extension de sa répartition dans l'hémisphère Nord et la variété des environnements occupés. Le chêne entretient toute une série d'associations mutualistes avec des champignons ectomycorhiziens, des insectes qui forment des galles et des vertébrés qui mangent les graines (Manos et Stanford, 2001 et les références citées). Du point de vu économique, le chêne est une essence feuillue commercialisée pour son bois.

De son côté le châtaignier a été longuement exploité pour le bois et le fruit. L'influence de l'homme a donc joué un rôle très important dans l'histoire de cette espèce (Conedera et al., 2004). Notamment, dans le bassin méditerranéen cette espèce a été introduite et cultivée par les Romains, surtout pour le bois. Ensuite, le châtaignier, nommé « l'arbre à pain », a connu une importante diffusion due à la valeur nutritionnelle de son fruit. Même si des forêts « naturelles » existent dans le bassin méditerranéen et en Turquie, il s'agit d'une espèce profondément liée à l'activité de l'homme.

A.2.5. Connaissances sur le génome des Fagacées et contexte de la thèse.

Les études de cartographie génétique et génomique chez les Fagacées étaient encore peu développées en 1998, quand le travail de cartographie génétique sur châtaignier, qui sera décrit dans la partie B de cette thèse, a été commencé. Deux cartes génétiques ont été publiées à la fin des années quatre-vingt dix. Kubisiak et al. (1997) ont utilisé un croisement interspécifique entre *C. dentata* et *C. mollissima* pour construire la première carte génétique d'une espèce de Fagacées. La faible saturation de cette carte n'a pas permis d'obtenir une bonne estimation de la taille génétique du génome du châtaignier. Une carte du chêne pédonculé a été publiée par Barreneche et al. (1998). Les données de liaison génétique ont permis d'estimer la taille génétique de cette espèce à 1200 cM (unités Kosambi²⁰). Plus récemment une carte pour l'espèce *Fagus sylvatica* L. a été publiée par Scalfi et al. (2004). Les auteurs ont estimé la taille génétique de cette espèce à environ 1100 cM (unités Kosambi).

La taille physique du génome de *Q. robur* et *C. sativa* et le nombre de chromosomes sont conservés. Les deux espèces sont diploïdes ($2n = 2x = 24$) et le contenu d'ADN par cellule haploïde est de 0,94 pg pour *Q. robur* (Zoldos et al., 1998) et 0,81 pg pour *C. sativa* (Brown et Siljak-Yakovlev, communication personnelle). Zoldos et al. (1998) ont montré que la

²⁰ Unité Kosambi : unité de cartographie génétique calculée selon la formule de Kosambi, qui sera décrite dans la partie B.

variation de la taille physique du génome chez plusieurs espèces de chêne est très faible (de 0,94 à 1 pg/C). Tout cela suggère un génome plutôt stable chez les Fagacées. La taille limitée de ce génome (six fois celui d'*Arabidopsis*, 0,15 pg/C), représente sans doute un avantage pour des études de génomique.

Le travail de cartographie comparée chez les Fagacées qui fait l'objet de cette thèse a son origine dans l'exigence de partager les connaissances moléculaires assez limitées chez cette famille. Plusieurs facteurs ont contribué au développement de ce projet :

- disponibilité de marqueurs microsatellites développés chez plusieurs espèces de chêne et chez le châtaignier européen (cf. paragraphe D.2.1).
- Disponibilité d'une banque d'EST développée chez le chêne sessile (*Q. petraea*) (Jérémy Derory, thèse en cours : Génomique et diversité du débourrement chez les chênes, Université de Bordeaux I).
- Disponibilité des résultats d'une recherche QTL contrôlant des caractères d'adaptation chez le chêne pédonculé (*Q. robur*) (Saintagne, 2003).

A.3. OBJECTIFS DE LA THÈSE.

Ce travail de thèse représente une étude pionnière de cartographie génétique comparée chez les Fagacées. Deux intérêts majeurs sont à l'origine de cette étude :

- les deux espèces étudiées sont proches phylogénétiquement, le karyotype est conservé, la taille physique des deux génomes est similaire. Tout cela pris en compte et à la lumière des résultats obtenus chez les autres espèces de plantes, nous devrions nous attendre à une conservation de la synténie entre ces deux espèces. Si la conservation de la synténie est vérifiée, les connaissances moléculaires et génétiques acquises sur une des deux espèces pourront être transférées à l'autre. Cet objectif est très important vu les ressources moléculaires encore limitées chez les deux espèces et de manière plus générale chez les Fagacées.
- Comme nous le verrons dans la partie C, l'étude des caractères d'adaptation chez les arbres forestiers est un domaine de grand intérêt. Les prévisions d'un changement global du climat ont, en effet, conduit les chercheurs à s'intéresser aux processus génétiques et moléculaires qui interviennent dans l'adaptation des plantes à leur environnement. Trois caractères adaptatifs différents seront considérés dans ce travail. L'objectif est de comparer le contrôle génétique de ces caractères et vérifier si les mêmes régions génomiques sont impliquées chez les deux espèces étudiées. Tout cela nous permettra de fournir des éléments de réflexion sur l'évolution de ces caractères complexes.

Pour atteindre les deux objectifs décrits, nous avons choisi la stratégie suivante :

- construction d' une carte génétique du châtaignier.
- Recherche de QTL contrôlant les caractères d'adaptation chez le châtaignier.
- Cartographie de marqueurs moléculaires orthologues chez le chêne et le châtaignier et comparaison des deux cartes génétiques.
- Construction de deux cartes consensus une pour le chêne et une pour le châtaignier, contenant les marqueurs utilisés pour la détection des QTLs et les marqueurs orthologues.

- Projection des QTL identifiées chez les deux espèces sur la carte consensus.
- Comparaison de la position des QTL par rapport aux marqueurs communs aux deux espèces.

Les questions auxquelles cette thèse tentera de répondre sont donc les suivantes :

- La taille génétique du génome du châtaignier est-elle comparable à celle du chêne ?
- Quelles sont les caractéristiques du contrôle génétique des caractères adaptatifs chez le châtaignier (nombre de locus impliqués, variance phénotypique expliquée, distribution dans le génome) ?
- Quels types de marqueurs moléculaires sont transférables entre le chêne et le châtaignier ?
- Les groupes de liaison sont-ils homologues ?
- La synténie et la colinéarité sont-elles conservées ?
- Le modèle du système génétique unique des Poaceae est-il applicable aux Fagacées ?
- Le contrôle génétique des caractères d'adaptation est-il similaire chez les deux espèces ?
- Les régions génomiques contrôlant ces caractères sont-elles conservées ?
- Quelles conclusions peuvent être tirées concernant l'évolution des caractères d'adaptation et comment appliquer les résultats obtenus à l'étude de la diversité adaptative chez ces espèces ?

B. Cartographie génétique du châtaignier

Still, human geneticists will need patience, persistence, luck, and a willingness to settle for a less than complete genetic map. Filling the last few gaps of the map will be a slow, frustrating enterprise.

(Lange K. et Boehnke M. (1982) How many polymorphic genes will it take to span the human genome? Am. J. Hum. Genet. 34: 842-845)

B.1. LA CARTOGRAPHIE GÉNÉTIQUE CHEZ LES ARBRES FORESTIERS.

Le croisement contrôlé et la construction de la carte de liaison génétique du châtaignier européen ont été réalisés à l'Istituto di Biologia Agroambientale e Forestale (IBAF) du CNR (Porano, Italie), sous la direction de Fiorella Villani. Ce travail, que nous avons commencé en 1998, a été intégré dans le projet CASCADE (« Securing gene conservation, adaptive and breeding potential of a model multipurpose tree species (*Casatanea sativa*) in a changing environment », <http://soi.cnr.it/~chestnut>), coordonné par Fiorella Villani. Dans le cadre de ce projet, une collaboration avec l'Equipe de Génétique de l'UMR BIOGECO de l'INRA de Pierroton a été établie. Cette collaboration a mené au développement du projet de cartographie comparée entre chêne et châtaignier sous la direction d'Antoine Kremer. L'analyse des données pour construire la carte génétique du châtaignier a été réalisée avec l'aide de Christophe Plomion. Ce travail a fait l'objet d'une publication présentée en annexe I. Dans ce paragraphe, après avoir décrit les principes de construction d'une carte génétique, nous ferons une synthèse des principaux résultats et nous décrirons l'évolution de la carte génétique de *C. sativa*. Deux nouvelles cartes seront présentées par rapport à celle publiée en 2001: la carte utilisée pour la détection des QTL et la carte, contenant les marqueurs orthologues, utilisée pour la comparaison avec la carte du chêne.

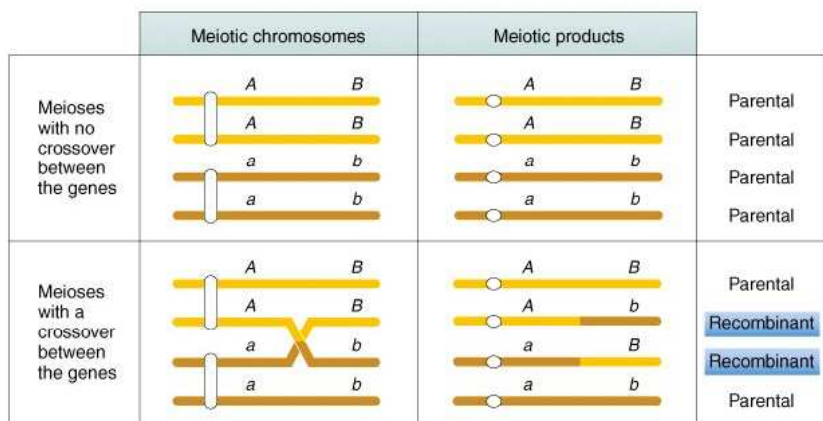
B.1.1. Principes de la cartographie génétique.

Une carte génétique est une représentation des chromosomes d'une espèce, faite à l'aide de marqueurs génétiques. Un marqueur génétique est une « caractéristique » de l'organisme (morphologique, biochimique, moléculaire), qui montre une ségrégation mendélienne, polymorphe et facile à suivre dans chaque individu, par exemple, d'une descendance issue d'un croisement contrôlé. Les marqueurs utilisés aujourd'hui sont essentiellement des marqueurs moléculaires. De nombreux types de marqueurs moléculaires sont employés en cartographie : les marqueurs dominants, qui identifient plusieurs locus à la fois (RAPD²¹, AFLP²¹, ISSR²¹) et les marqueurs codominants et locus-spécifiques (RFLP, SSR). Comme nous le verrons plus loin, le type de marqueurs utilisés est très important en cartographie.

La première carte génétique date du début du siècle passé et a été construite au sein du laboratoire du grand généticien Morgan. Ce travail a été fait sur la drosophile par un thésard

²¹ RAPD : "Random Amplified Polymorphic DNA" ; AFLP : "Amplified Fragment Length Polymorphisms" ; ISSR : "Inter Simple Sequence Repeats", ces marqueurs moléculaires seront décrits dans la partie B3.

Encadré 1 : "Crossing-over" et taux de recombinaison.



http://fig.cox.miami.edu/Faculty/Dana/F05_08.JPG

Lors de la méiose l'échange entre chromosomes homologues (dit « crossing-over ») donne lieu à la formation de gamètes parentaux et recombinés. Le taux de recombinaison entre deux locus, r , est une mesure indirecte de la distance entre eux.

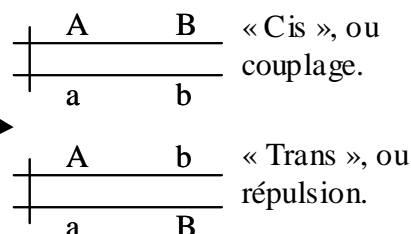
Considérons un croisement « backcross » :

parents AAbb x aabb

F₁ AaBb, deux configurations ou phases sont possibles →

« Backcross » AaBb x aabb, qui donne lieu à n individus.

Dans le cas « cis », les génotypes attendus et observés sont les suivants :



Génotypes	AaBb	Aabb	aaBb	aabb
Effectifs théoriques	$n(1-r)/2$	$nr/2$	$nr/2$	$n(1-r)/2$
Effectifs observés	a	b	c	d

Si les deux locus A/a et B/b ségrègent de façon indépendante, alors $r = 0,5$ et les quatre génotypes auront la même fréquence : $n/4$, n étant le nombre total d'individus. En revanche, si les deux locus sont liés, les effectifs attendus dépendent du taux de recombinaison qui est estimé de la façon suivante : génotypes recombinés/ n , c'est à dire $r = (b+c)/n$, dans le cas de couplage décrit au-dessus.

Tableau 2 : Taille physique et génétique de 15 espèces de plantes (tiré de Chagné et al., 2002).

Espèce	Taille physique du génome haploïde (Mb)	Taille génétique (cM) (MapMaker)	Nombre de chromosomes (n)	Taille moyenne des chromosomes (cM)	Rapport taille physique/génétique (Mb/cM)
<i>Arabidopsis thaliana</i>	150	675	5	135	0,22
<i>Prunus persica</i>	300	712	8	90	0,42
<i>Oryza sativa</i>	150	1490	12	125	0,3
<i>Populus deltoides</i>	550	2300	19	121	0,23
<i>Eucalyptus grandis</i>	600	1370	11	125	0,43
<i>Brassica rapa</i>	650	1850	10	185	0,35
<i>Quercus robur</i>	900	1200	12	100	0,75
<i>Lycopersicon esculentum</i>	980	1280	12	107	0,76
<i>Solanum tuberosum</i>	1540	1120	12	93	1,37
<i>Zea mays</i>	2500	1860	10	186	1,34
<i>Lactuca sativa</i>	2730	1950	9	217	1,4
<i>Triticum tauschii</i>	4200	1330	7	190	3,15
<i>Hordeum vulgare</i>	5500	1250	7	178	4,4
<i>Pinus taeda</i>	21000	1700	12	141	12,35
<i>Pinus pinaster</i>	25500	1850	12	154	13,78

de Morgan, Sturtevant (1913), et les marqueurs génétiques étaient représentés par des marqueurs morphologiques. Sturtevant eut la brillante idée de lier le concept de recombinaison génétique avec la distance physique qui sépare deux marqueurs (deux locus) sur un même chromosome. Au cours de la méiose, qui donne lieu aux gamètes, des échanges entre les chromosomes homologues se produisent : il s'agit du « crossing-over », représenté schématiquement entre deux locus A et B dans l'encadré 1. Le phénomène est appelé recombinaison génétique et le calcul de la fréquence de cette recombinaison (r) permet d'estimer la distance entre locus. En effet, cette fréquence sera d'autant plus élevée que les deux locus seront éloignés l'un de l'autre sur le chromosome. Comme le montre l'encadré 1, r est estimé comme le rapport entre les gamètes de type recombiné et ceux de type parental. Si $r = 0$, les locus sont totalement liés et aucun gamète de type recombiné se produit lors de la méiose. Si $0 < r < 0,5$, les deux locus sont liés. La valeur de r est une mesure de cette liaison et est estimé par la proportion de gamètes recombinés. Enfin, si $r = 0,5$, les deux locus ségrègent de façon indépendante, soit ils se trouvent sur deux chromosomes différents, soit sur le même chromosome mais à une distance suffisante pour qu'il y ait autant de gamètes recombinés que de gamètes parentaux. La relation entre la distance génétique (calculée à partir du taux de recombinaison) et la distance physique (calculée en paires de bases) varie d'une espèce à l'autre et d'une région à l'autre du génome (tableau 2).

La construction d'une carte génétique demande, donc, trois éléments indispensables :

1. une descendance dans laquelle nous pouvons suivre la ségrégation de marqueurs génétiques ;
2. des marqueurs génétiques pour caractériser les individus de la descendance ;
3. des logiciels d'analyse statistique des données de ségrégation.

Les applications de la cartographie génétique sont nombreuses. Comme nous avons pu le voir dans l'introduction, la disponibilité d'une carte génétique permet d'identifier les régions génomiques contrôlant les caractères complexes, de cartographier des gènes d'intérêt, de comparer les génomes d'espèces proches. Une carte génétique représente la base pour étudier la structure, le fonctionnement et l'évolution du génome (Liu, 1998).

B.1.2. Ségrégation mendélienne, estimation de la liaison génétique et de la distance entre marqueurs.

L'établissement de la carte génétique consiste à : 1) vérifier la ségrégation mendélienne des marqueurs utilisés ; 2) détecter la liaison génétique entre marqueurs et réunir donc les marqueurs en groupes de liaison, chaque groupe représente un chromosome ; 3) ordonner les

Test du χ^2 pour vérifier la ségrégation mendélienne.

Dans le cas d'un « backcross » et utilisant des marqueurs dominants (par exemple RAPD) la ségrégation mendélienne est vérifiée de la façon suivante :

soit O_1 le nombre observé d'individus qui présentent la bande amplifiée et O_0 le nombre d'individus sans la bande ;

le nombre attendu dans chaque classe (E_1 et E_0) est $n/2$, étant n le nombre total d'individus ;

alors, $\chi^2 = (O_1 - E_1)^2 / E_1 + (O_0 - E_0)^2 / E_0 = (O_1 - O_0)^2 / n$. Pour $\alpha = 0,05$, $\chi^2_{(1 \text{ ddl})} = 3,84$.

Dans les cas d'un pseudo testcross (cf. B.1.7.) en configuration intercross (cf. tableau 3):

$E_1 = 3/4 N$, $E_0 = 1/4 N$ et $\chi^2 = (O_1 - E_1)^2 / E_1 + (O_0 - E_0)^2 / E_0 = (O_1 - 3O_0)^2 / 3n$.

Estimation de la liaison, r , par la méthode du maximum de vraisemblance.

Pour estimer r la méthode du maximum de vraisemblance est souvent utilisée (Allard, 1956). Soit r le taux de recombinaison entre deux locus donnés, m_1, m_2, \dots, m_t les effectifs théoriques des individus dans les classes de ségrégation $1, \dots, t$ et a_1, \dots, a_t le nombre d'individus observés dans ces classes. L'équation de vraisemblance s'écrit de la façon suivante :

$$L = a_1 \log(m_1) + a_2 \log(m_2) + \dots + a_t \log(m_t).$$

Pour obtenir l'équation de vraisemblance proprement dite il faut maximiser cette expression par rapport à r , c'est à dire il faut annuler sa dérivée par rapport à r :

$$dL/dr = a_1 [d\log(m_1)/dr] + a_2 [d\log(m_2)/dr] + \dots + a_t [d\log(m_t)/dr] = 0.$$

Une des solutions de cette équation est la valeur de r que nous voulons estimer. La solution de cette équation est souvent obtenue par des méthodes itératives, par exemple l'algorithme EM (« Expectation-Maximization », Dempster et al., 1977). Dans le cas d'un « backcross », la solution est, toutefois, très simple.

Considérons les effectifs théoriques et observés de l'encadré 1. L'équation à résoudre sera :

$$dL/dr = a [d\log[(n(1-r)/2)]/dr] + b [d\log(nr/2)/dr] + c [d\log(nr/2)/dr] + d [d\log[(n(1-r)/2)]/dr] = 0.$$

Sachant que : $d\log[(n(1-r)/2)]/dr = -1/(1-r)$ et $d\log(nr/2)/dr = 1/r$, alors

$$-a/(1-r) + b/r + c/r - d/(1-r) = 0.$$

Enfin, sachant que $a + b + c + d = n$,

$r = (b + c)/n$, qui représente une estimation du taux de recombinaison.

marqueurs au sein de chaque groupe de liaison ; 4) estimer la distance en cM entre les marqueurs.

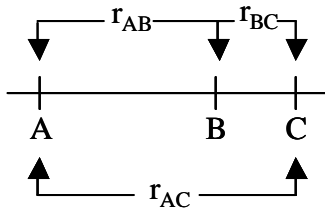
Un test χ^2 est utilisé afin de vérifier la ségrégation mendélienne (encadré 2). Les marqueurs pour lesquels le test est significatif sont dits « distordus ». La distorsion de ségrégation peut avoir soit une cause biologique (par exemple, liaison avec des locus létaux) ou une cause statistique (faible nombre d'individus utilisé pour le calcul). Les marqueurs « distordus » peuvent être intégrés dans une carte génétique, si leur positionnement n'influence pas la fiabilité statistique de l'ordre des autres marqueurs.

La deuxième étape consiste à tester la liaison entre deux locus, pour ensuite réunir les marqueurs en groupes de liaison. Le test le plus utilisé est celui du « LOD score » (Morton, 1955), qui est égal au logarithme décimal du rapport des vraisemblances des deux hypothèses : celle de liaison et celle d'indépendance entre les deux locus considérés ($r = 0,5$). Ce test nous permet de calculer combien de fois l'hypothèse de liaison est plus « probable » que celle de non liaison. La méthode du maximum de vraisemblance pour estimer la liaison r dans un cas de « backcross » est présentée dans l'encadré 2. Une fois le taux de recombinaison, r , estimé, celui-ci est utilisé pour calculer la distance génétique en centiMorgan entre deux locus.

B.1.3. Fonctions de cartographie.

La fréquence de recombinaison est une mesure indirecte des phénomènes de « crossing-over » entre deux locus et ne tient pas compte que des recombinaisons impaires, tandis que les événements pairs ne sont pas considérés. A courte distance, r est un bon estimateur de la distance, mais à cause des échanges multiples, la relation entre le taux de recombinaison et la distance n'est pas linéaire. Il faut, donc, une fonction de cartographie qui puisse tenir compte du nombre moyen de points d'échange entre deux locus. C'est pourquoi la distance génétique est mesurée en cM (centiMorgan, en l'honneur du généticien Morgan) et pas directement en r . La fonction d'Haldane (1919) et celle de Kosambi (1944) sont les deux fonctions de cartographie les plus utilisées. L'encadré 3 montre les caractéristiques de ces deux fonctions. Comme le montre l'encadré 3, la fonction de Kosambi, à la différence de celle d'Haldane, tient compte de l'interférence, c'est à dire la probabilité qu'un « crossing-over » dans une position quelconque soit dépendante d'un échange à une position différente. La fonction de Kosambi s'approche mieux de la réalité biologique que celle de Haldane. Toutefois, sachant que l'interférence n'est pas constante le long du chromosome, aucune fonction de cartographie ne donne une estimation précise de la distance génétique (Crow, 1990).

Encadré 3 : Fonctions de cartographie.



Considérons trois locus A, B et C liés sur le même fragment d'ADN comme le montre le schéma à côté. Soit r le taux de recombinaison et c le coefficient de coïncidence, c'est à dire le rapport entre le nombre réel de « crossing-over » sur le nombre théorique de « crossing-over » calculé sous l'hypothèse d'absence d'interférence.

$$r_{AC} = r_{AB} (1 - r_{BC}) + r_{BC} (1 - r_{AB}) = r_{AB} - r_{AB} r_{BC} + r_{BC} - r_{AB} r_{BC} = r_{AB} + r_{BC} - 2 r_{AB} r_{BC}$$

Ils est donc évident que r n'est pas additif et qu'il est important de tenir compte des double recombinants. De plus, il faudrait considérer l'influence d'un « crossing-over » sur la probabilité d'en avoir un autre à courte distance, c'est à dire le phénomène d'interférence. Pour cela nous pouvons introduire c , le coefficient de coïncidence et récrire l'expression de la façon suivante :

$$r_{AC} = r_{AB} + r_{BC} - 2cr_{AB}r_{BC}$$

$$r_{AC} - r_{AB} = r_{BC}(1 - 2cr_{AB})$$

considérant $r_{AC} - r_{AB} = \Delta r$ et r_{BC} faible, donc égal à la distance réelle ($r_{BC} = \Delta m$), nous pouvons calculer cette distance m en fonction de r :

$$\Delta r = \Delta m (1 - 2cr_{AB})$$

$$dm/dr = 1/(1 - 2cr)$$

Si r est faible et en absence d'interférence ($c = 1$), alors $dm/dr = 1$ et donc r coïncide avec la distance m .

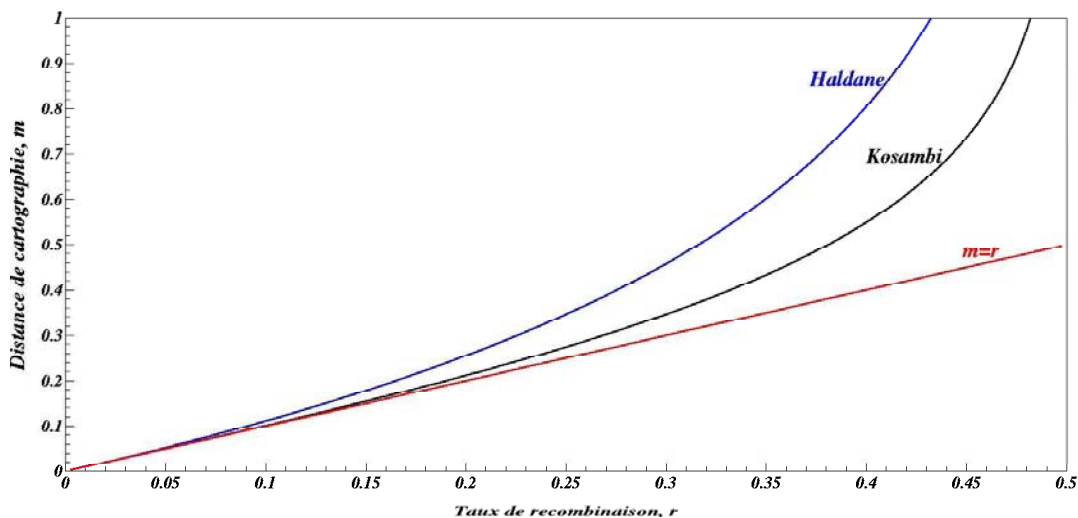
Pour r quelconque et en absence d'interférence ($c = 1$), $dm/dr = 1/(1 - 2r)$. L'intégration de cette expression donne la fonction de cartographie d'Haldane (1919) :

$$m = -\frac{1}{2} \ln (1 - 2r).$$

Pour r quelconque et en présence d'interférence, nous pouvons obtenir la fonction de cartographie de Kosambi (1944). Dans la fonction de Kosambi $c = 2r$. Cela implique qu'en absence d'interférence ($c = 1$), $r = 0,5$ et que quand l'interférence est maximale ($c = 0$), $r = 0$. En intégrant l'expression $dm/dr = 1/(1 - 4r^2)$ nous avons la fonction de Kosambi :

$$m = \frac{1}{4} \ln [(1+2r)/(1 - 2r)].$$

La distance génétique est représentée par m et s'exprime en centiMorgan (cM), les deux formules donnent la valeur en Morgan. Comme le montre le graphique au-dessous, à courte distance les deux fonctions donnent des valeurs similaires de distance génétique. La relation linéaire entre r et m est conservée pour r petit.



B.1.4. Détermination de l'ordre des marqueurs.

Une fois les groupes de liaison établis, l'étape suivante consiste à ordonner les marqueurs au sein de chaque groupe. Pour m marqueurs nous avons $m!/2$ ordres possibles. Pour obtenir l'ordre le plus probable, la méthode est toujours celle du maximum de vraisemblance. En appliquant cette méthode, pour chaque carte possible la probabilité que les données conduisent à cette carte est calculée. Cette probabilité représente la vraisemblance de la carte. La carte retenue sera celle qui montre la vraisemblance la plus forte. Les algorithmes de calcul de vraisemblance d'une carte peuvent varier d'un logiciel à l'autre. En général, l'analyse est toujours une analyse en trois points et plus (« multipoints »), qui prend en compte les doubles recombinaisons. Ensuite, la distance en cM est estimée grâce aux fréquences de recombinaison obtenues de la carte la plus probable. Il est à noter que : 1) plus il y aura des marqueurs entre deux locus, plus le taux de recombinaison observé augmentera et la distance se rapprochera de la distance réelle ; 2) à courte distance (par exemple, $r < 0,10$), l'imprécision sur l'estimation du vrai ordre augmente et des effectifs importants (500-1000 individus), typiques de la cartographie génétique fine, sont nécessaires pour que la probabilité de ne pas se tromper soit élevée (Liu, 1998).

B.1.5. Influence du type de marqueurs et de la taille de la population.

En général, le type de descendance (« backcross », F₂, etc.), le type de marqueurs (dominant ou codominant), le type de ségrégation (1:1, 1:2:1, 3:1, 1:1:1:1), la phase des marqueurs (cis ou trans, cf. encadré 1) et le nombre d'individus utilisé pour l'établissement de la carte génétique influencent la précision de l'estimation de la fréquence de recombinaison et la probabilité de choisir le vrai ordre des marqueurs. Chacune de ces influences est détaillée par Liu (1998). Nous rappellerons ici qu'il est possible de calculer l'information du taux de recombinaison en fonction du type de ségrégation pour différents couples de marqueurs. Comme le montre la figure 11, les couples de marqueurs qui ségrègent 1:2:1/1:2:1 ou 1:1/1:1 sont les plus informatifs.

Quel est le nombre minimal d'individus à utiliser pour obtenir une carte génétique fiable ? La réponse à cette question nous permet de trouver un compromis entre une détection efficace de la liaison génétique entre marqueurs et leur ordonnancement et le minimum de réactions de génotypage à faire. Ce nombre dépend du type de descendance et du type de marqueurs. Pour un « backcross », 100 individus sont largement suffisants pour détecter la liaison (Liu, 1998). Ce nombre permet d'avoir un ordonnancement de marqueurs assez fiable pour des intervalles

d'environ 5-10 cM et il n'est pas trop élevé du point de vue pratique (génotypage des marqueurs sur la descendance).

B.1.6. Saturation de la carte et taille du génome.

Une carte génétique est saturée quand : 1) le nombre de groupes de liaison est égal au nombre haploïde de chromosomes de l'espèce ; 2) tout autre marqueur ajouté à la carte est lié génétiquement à l'un des groupes de liaison. Beckmann et Soller (1983) ont proposé la formule suivante pour calculer le nombre de marqueurs nécessaires à la saturation d'une carte génétique :

$$n = \ln(1-p) / \ln(1-2m/L)$$

où n est égal au nombre de marqueurs nécessaires afin qu'un pourcentage p donné de marqueurs ne soit éloigné de plus de m cM d'un autre marqueur sur un génome de taille L. Lange et Boehnke (1982) ont suggéré de multiplier cette valeur par 1,2-1,3 car la formule de Beckmann et Soller ne prend pas en compte la distribution souvent hétérogène des marqueurs sur les génomes. Pour le châtaignier, si nous considérons p = 0,95 (saturation de 95%), m = 10 cM et L = 1200 cM (100 cM par chromosome), alors n varie entre 178 et 223 (la dernière valeur étant obtenue en multipliant 178 par 1,25).

Pour une carte non saturée, la taille génétique du génome est estimée à partir des données de ségrégation. Chakravarti et al. (1991) ont proposé une formule pour cette estimation qui a été souvent utilisée, notamment pour l'estimation de la taille génétique du génome de *Q. robur* (Barreneche et al., 1998) et que nous avons aussi utilisé pour le châtaignier. Il faut souligner que cette estimation est basée sur l'hypothèse d'une distribution aléatoire des marqueurs sur le génome et tous les chromosomes sont supposés de taille égale. La formule de Chakravarti est la suivante :

$$G_{(z)} = 2MX_{(z)} / K_{(z)}$$

où $G_{(z)}$, est la taille du génome en cM estimée à un $LOD \geq z$, M le nombre de méioses informatives, $M = n(n-1) / 2$ où n = nombre de marqueurs, $X_{(z)}$ la distance maximale observée entre deux locus liés et $K_{(z)}$ le nombre de couples de marqueurs liés à un $LOD \geq z$.

B.1.7. La stratégie de double pseudo test-cross.

Les arbres forestiers sont des espèces sauvages, allogames, qui ont des temps de générations très longs, des populations nombreuses et qui généralement sont caractérisés par une dépression de consanguinité très forte due au fardeau génétique. Cela implique que les croisements et les stratégies classiquement utilisés en cartographie génétique ne s'adaptent

Figure 11

Information du taux de recombinaison, r , (définie comme l'inverse de la variance de r) en fonction du type de ségrégation pour différents couples de marqueurs.

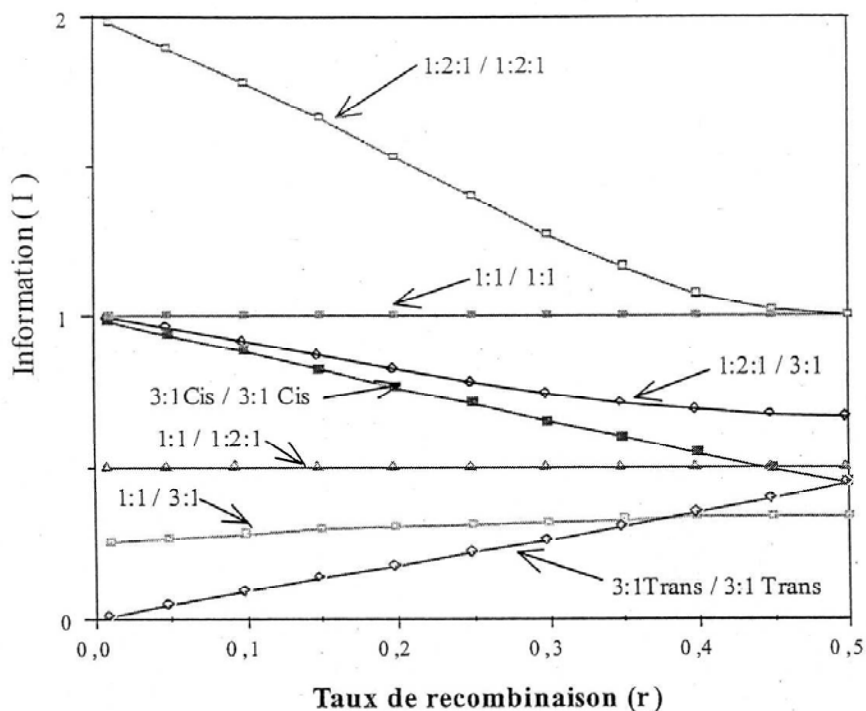


Tableau 3 : Configurations possibles de marqueurs co-dominants et dominants dans une descendance de plein-frères ou demi-frères chez les espèces forestières allogames.

Génotypes parentaux (f x m)	Marqueurs co-dominants	Marqueurs dominants	Type de ségrégation	Possibilité de cartographie
aa x aa	aa	Aa	pas de ségrégation	aucune
ab x aa	ab:aa=1:1	Aa:aa=1:1	"testcross/backcross"	f
aa x ab	aa:ab=1:1	aa:Aa=1:1	"testcross/backcross"	m
ab x ab	aa:ab:bb=1:2:1	A_:aa=3:1	"intercross"	f et m
ab x bc	ab:ac:bb:bc=1:1:1:1	-	"outcross"	complètement informatif
ab x cd	ac:ad:bc:bd=1:1:1:1	-	"outcross"	complètement informatif

f, femelle
m, mâle

pas à ces organismes (Ahuja, 2001). Néanmoins, l'hétérozygotie élevée de ces populations et la possibilité d'obtenir des descendance nombreuses de plein-frères ont mené au développement d'une stratégie, dite de double pseudo test-cross, très utilisée chez les arbres forestiers. Cette stratégie a été appliquée pour la première fois chez l'*Eucalyptus* par Grattapaglia et Sederoff (1994). Elle est basée sur l'utilisation de marqueurs dominants et de croisements F_1 . Les fragments amplifiés présents chez un des deux parents et absents chez l'autre, qui ségrègent 1:1 dans la descendance, sont dit marqueurs en configuration « testcross » (tableau 3) et ils permettent d'établir deux cartes distinctes : une carte pour le parent femelle et une carte pour le parent mâle. Les deux cartes peuvent être reliés grâce aux marqueurs (fragments amplifiés) présents chez les deux parents et qui ségrègent 3:1 dans la descendance, dits en configuration « intercross ». Quand des marqueurs codominants sont utilisés, d'autres types de ségrégation sont possibles (tableau 3), 1:1:1:1 et 1:2:1. Ces derniers aussi permettent de relier les deux cartes parentales. Les marqueurs qui ségrègent dans la descendance sont dits informatifs. Il faut souligner que les différents types de ségrégation et la phase (cf. encadré 1) souvent inconnue donnent lieu à un mélange des marqueurs très hétérogène en ce qui concerne l'information (figure 11). Cette stratégie a été utilisée pour construire plusieurs cartes chez les arbres forestiers. Une liste des cartes publiées chez ces espèces est disponible sur le site : <http://dendrome.ucdavis.edu>.

B.1.8. Les logiciels « MapMaker » et « JoinMap ».

Plusieurs logiciels ont été développés afin de construire des cartes génétiques. Une liste est disponibles sur le site <http://dendrome.ucdavis.edu>. Nous avons utilisé deux logiciels différents : « MapMaker » et « JoinMap ». Le premier est disponible sur le site <ftp://ftp-genome.wi.mit.edu/distribution/software/mapmaker3/> (Lander et al., 1997). Il s'agit d'un logiciel très utilisé qui permet de faire l'analyse de liaison et d'ordonnement des marqueurs avec la possibilité de vérifier l'ordre des marqueurs utilisant l'algorithme EM (« Expectation-Maximization », Dempster et al., 1977). La limite de ce logiciel est que seuls les marqueurs en configuration « testcross » peuvent être considérés. Au contraire, le logiciel « JoinMap » est payant (<http://www.joinmap.nl>, Stam, 1993 ; Stam et Van Ooijen, 1995), mais il permet d'intégrer l'analyse des marqueurs 3:1 et 1:2:1 et ainsi de construire des cartes dites « consensus » qui sont le résultat d'une analyse prenant en compte les données de ségrégation de deux parents. Généralement, les distances en cM obtenues utilisant « MapMaker » sont plus larges que celles calculées par « JoinMap ». En effet, l'algorithme utilisé pour l'estimation des distances est différent entre les deux logiciels. « MapMaker »

estime la distance en utilisant le critère de maximum de vraisemblance, alors que « JoinMap » utilise un critère de minimisation de la distance totale d'un groupe de liaison (Brigitte Mangin, communication personnelle).

Le logiciel « MapChart » a été utilisé pour la création des figures relatives à ces cartes génétiques et aux QTL (Voorrips, 2002).

B.2. LES POPULATIONS BURSA ET HOPA ET LE CROISEMENT CONTROLÉ.

B.2.1. Les études de diversité génétique chez le châtaignier européen.

Castanea sativa est la seule espèce native de châtaignier en Europe. Son centre d'origine a été localisé dans la région caucasienne et dans l'est de la Turquie (Zohary et Hopf, 1988). Cette espèce, présente en Europe au cours du tertiaire, a disparu de toute l'Europe du Nord-Ouest durant les périodes de glaciations du pléistocène. Elle n'aurait survécu que dans des refuges limités situés en Turquie et dans la région caucasienne (Huntley et Birks, 1983), et dans les péninsules italienne et ibérique (Krebs et al., 2004). Les premiers restes fossiles, qui démontrent la distribution de cette espèce liée à l'activité de l'homme, ont été identifiés en Turquie, en Grèce, en Bulgarie et en Italie et ils remontent à 3700-3600 ans (Conedera et al., 2004). D'après Huntley et Birks (1983), il y a eu deux périodes d'expansion du châtaignier vers l'Europe : 1) la première, datée de 5000 ans, est due à l'activité des populations agricoles du néolithique ; 2) la deuxième expansion plus importante vers le bassin méditerranéen aurait eu lieu grâce aux Romains, qui exploitaient déjà cette espèce pour le bois et son fruit. Néanmoins, les Romains n'auraient pas cultivé cette espèce de façon intensive. C'est durant le moyen-âge que la culture du châtaignier pour son fruit est devenue plus importante (Conedera et al., 2004).

L'histoire du châtaignier européen aide à comprendre les résultats des études de diversité génétique basées sur les isozymes (Villani et al., 1991 ; Villani et al., 1994 ; Machon et al., 1996). Ces études ont mis en évidence une différenciation génétique très élevée entre les populations de la Turquie orientale et les populations européennes. Selon ces études, les populations provenant de la Turquie occidentale sont plus proches des populations italiennes que de celles originaires de la Turquie de l'est (Villani et al., 1991 ; Villani et al., 1994). Ces résultats sont en accord avec l'hypothèse selon laquelle la Turquie orientale serait le centre d'origine de cette espèce, qui aurait été cultivée d'abord en Turquie occidentale pour ensuite être introduite dans tout le bassin méditerranéen. Par le biais de marqueurs chloroplastiques, Fineschi et al. (2000) ont proposé la péninsule ibérique, en plus de la Turquie, comme une deuxième région refuge pendant la dernière glaciation. Récemment les résultats du projet CASCADE (<http://soi.cnr.it/~chestnut>) ont mis en évidence trois « pools » génétiques différents : 1) un groupe génétique en Turquie orientale ; 2) un deuxième groupe grec ; 3) et un troisième groupe comprenant le reste de l'Europe avec de faibles variations d'est en ouest.

La culture de cette espèce dans tout le bassin méditerranéen a sûrement eu de profondes conséquences sur sa diversité génétique.

B.2.2. Identification de deux populations de châtaignier adaptées à deux environnements contrastés.

La Turquie n'est pas seulement le centre d'origine de *C. sativa*, mais aussi une région de grand intérêt du point de vue climatique pour cette espèce. En effet, d'est en ouest, le climat passe d'une région humide et fraîche, la région de la mer Noire, à une région sèche et plus chaude, la région méditerranéenne. Les populations de châtaignier de l'est et de l'ouest de la Turquie sont donc soumises à des conditions climatiques opposées. Pigliucci et al., (1990) ont montré que la différenciation génétique entre les populations de l'est et de l'ouest est en partie expliquée par des variables climatiques, notamment les précipitations. Il est à noter que la quantité de pluie est un des facteurs qui limitent l'expansion du châtaignier. Cette particularité des populations turques a conduit à toute une série d'études qui ont mené à l'identification des deux écotypes²² adaptés à des conditions environnementales opposées. La différenciation entre les populations orientales et occidentales a également été mise en évidence au niveau morphologique et physiologique par Villani et al. (1992). Lauteri et al. (1997) ont ensuite comparé les populations orientales et occidentales dans le même environnement au sein de tests de provenance. Plusieurs caractéristiques physiologiques liées à l'efficacité d'utilisation de l'eau de la plante ont montré une forte variabilité. Deux populations, dites Bursa et Hopa du nom de leur lieu d'origine (figure12), ont montré un comportement contrasté. L'adaptation de ces deux populations à l'environnement méditerranéen (Bursa) ou humide (Hopa) a été l'objet de la thèse de Maria Cristina Monteverdi (2004), dont nous rappellerons ici les résultats principaux :

- l'écotype Hopa est adapté à un climat humide. En conditions de sécheresse, pour éviter la perte d'eau, il ferme les stomates. Le contrôle stomatique est donc prépondérant et il peut tolérer de courtes périodes de sécheresse. Les individus appartenant à cette population montrent, en revanche, des caractéristiques typiques d'une adaptation à des conditions de lumière limitée : par exemple, une croissance juvénile en hauteur élevée, des feuilles larges et limitées en épaisseur.
- l'écotype Bursa est adapté à un climat sec. Il montre des adaptations structurelles à cet environnement qui lui permettent de tolérer des conditions de sécheresse à long terme. Les individus appartenant à cette population ont une sensibilité stomatique faible

²² Écotype : unité biologique différenciée par adaptation à un type de milieu.

Figure 12

Le croisement contrôlé



Deux individus appartenant aux deux populations Bursa et Hopa ont été choisis comme parent femelle et mâle du croisement. Les paramètres climatiques (pluie et température) des deux régions sont reportés à gauche.



Fleur femelle (à gauche) et mâle (à droite) de *C. sativa*.



Le croisement, fait en 1998, a donné à l'automne 224 châtaignes. A gauche le parent femelle en juillet 1998.

Les 186 individus obtenus ont poussé dans des pots et ils sont restés en pépinière de 1999 à 2001.



En 2002, les plantes ont été transférées sur le terrain. A droite un de 152 individus qui ont survécu.



associée à une faible efficacité d'utilisation de l'eau, en comparaison à des individus Hopa dans le même environnement sec. Cette caractéristique dérive de leur adaptation structurelle à la disponibilité en eau limitée. En effet, ils montrent une efficacité hydraulique remarquable, des racines profondes, des feuilles épaisses et étroites.

B.2.3. Le croisement contrôlé.

Le croisement contrôlé utilisé dans cette thèse a été réalisé entre deux individus de 9 ans appartenant aux populations Bursa et Hopa (figure 12). Les deux individus faisaient partie de la collection de l'institut IBAF de Porano en Italie. Le choix des parents du croisement a été fait sur les bases suivantes:

- la stratégie de double pseudo test-cross est efficace si les deux parents ont une hétérozygotie élevée. Le profil isoenzymatique des individus a été analysé pour sélectionner les plus hétérozygotes. Il est à noter que les deux populations présentent une diversité génétique intra population élevée (Villani et al., 1991).
- Les individus avec le maximum de fleurs femelles ont été considérés pour le croisement afin d'obtenir une descendance nombreuse.
- Les paramètres physiologiques des individus (notamment, la valeur de la discrimination isotopique du carbone treize, qui sera décrit dans la partie C) ont été considérés afin de choisir deux individus extrêmes. Cela, tout en considérant l'hétérozygotie élevée des individus, devrait nous assurer une bonne efficacité de détection des QTL.
- L'individu Bursa a été choisi comme parent femelle et l'individu Hopa comme parent mâle, sachant que des croisements faits auparavant avaient indiqué que cette configuration était la plus efficace (Villani et Cherubini, communication personnelle).

Le croisement a été fait au printemps 1998. Les branches de l'individu Bursa portant des fleurs femelles ont été privées des chatons et enfermées dans des poches pour éviter le contact avec le pollen de l'air. Les chatons de l'individu Hopa ont été récoltés et le pollen distribué sur les fleurs femelles trois fois pendant dix jours. A l'automne, 224 châtaignes ont été récoltées et plantées en pot à la pépinière de l'Istituto di Biologia Agroambientale e Forestale. Un petit morceau de l'endosperme de chaque individu de la descendance a été utilisé pour contrôler, à l'aide des isozymes, le génotype et vérifier la bonne réussite du croisement. Le même tissu a été utilisé pour extraire l'ADN selon la méthode de Doyle et Doyle (1987). En 1999, 2000 et 2001 les plantes ont été laissées dans des pots (figure 12) et dans les conditions environnementales naturelles. Cependant, elles ont été arrosées régulièrement pendant l'été.

En 2002, les individus de la descendance ont été transférés sur le terrain du même institut. Parmi les 186 individus qui avaient germé, 152 ont survécu et ils constituent la famille de plein-frères issue de ce croisement.

L'expérience de Marcello Cherubini a été d'une importance fondamentale pour la réalisation du croisement. Il a aussi contrôlé la bonne réussite du croisement grâce aux isozymes. Claudia Mattioni a participé à la réalisation du croisement, la récolte et l'extraction de l'ADN. Marcello Cherubini, Giovanni de Simoni et Maria Cristina Monteverdi m'ont aidé dans ce travail aussi bien à la pépinière que sur le terrain.

B.3. LES MARQUEURS MOLÉCULAIRES UTILISÉS.

B.3.1. Le choix des marqueurs.

Trois critères ont guidé le choix des marqueurs, étant donné l'équipement limité du laboratoire de biologie moléculaire de l'institut IBAF de Porano :

- aucune connaissance préalable sur les séquences de l'espèce ;
- la simplicité du protocole ;
- le coût faible.

Les marqueurs RAPD (Williams et al., 1990) et les ISSR (Zietkiewicz et al., 1994) ont été choisis, alors que les AFLP (Vos et al., 1995) n'ont pas été considérés en raison de la lourdeur de la technologie requise et de leur coût élevé. Les RAPD et les ISSR sont des marqueurs dominants. Toutefois, la stratégie double pseudo test-cross est efficace avec ce type de marqueurs. Dans les deux cas, il s'agit d'une amplification de plusieurs locus à la fois, dont la séquence n'est pas connue (marqueurs anonymes). Chaque marqueur est caractérisé par la taille du fragment amplifié et le polymorphisme est de type présence/absence. Néanmoins, le principe des deux techniques est différent. Une seule amorce courte, souvent 10 pb, de séquence arbitraire est utilisée dans l'amplification des marqueurs RAPD. L'amorce s'hybride à chaque fois qu'il y aura dans l'ADN une séquence complémentaire (ou comportant un nombre limité de mésappariements). Si deux sites d'hybridation sont proches l'un de l'autre (par exemple, à moins de 3000-2500 pb) en direction opposée, l'amplification aura lieu. Si un de deux sites est absent dans un individu, il n'y aura pas d'amplification et le polymorphisme présence/absence sera observé (figure 13a). Le principe des ISSR s'inspire de celui des RAPD mais l'amorce est de nature différente. Il s'agit d'une séquence répétée (tableau 5) qui cible donc des microsatellites. La PCR amplifie les séquences flanquantes du microsatellite. L'amorce peut aussi avoir deux ou trois bases arbitraires à son extrémité 3', ce qui garantit une hybridation moins aspécifique. Cette technique permet aussi de développer des marqueurs SSR à partir des bandes amplifiées (Lian et al., 2001). Un profil ISSR obtenu pour la descendance est montré dans la figure 13a. Une séparation sur gel d'agarose est suffisante pour révéler le polymorphisme.

Les marqueurs SSR et les STS²³ dérivées des EST, qui ont été introduits par la suite, seront décrits dans la partie D.

²³ STS : "Site Tagged Sequences", fragments d'ADN caractérisés par séquence.

B.3.2. L'optimisation du protocole.

Les marqueurs RAPD ont parfois présenté des problèmes de reproductibilité (Skov, 1998), liés à la multiplicité des sites d'amplification potentielles dans le génome. En effet, pendant l'amplification, les différents sites d'hybridation de l'amorce sont en compétition entre eux, la moindre variation des conditions de réaction peut avoir des conséquences sur cette compétition et aboutir à des profils légèrement différents. Pour cette raison, il est très important d'optimiser le protocole d'amplification. Après avoir essayé différentes concentrations de MgCl₂, dNTP, amorce, ADN et enzyme (Taq polymérase) nous avons choisi le protocole reporté dans la figure 13a, car il permettait d'avoir le même profil d'amplification en utilisant différentes extractions d'ADN et des protocoles d'amplification légèrement différents. La concentration en MgCl₂, ADN et enzyme sont les facteurs qui ont le plus d'influence sur le profil d'amplification. Le cycle de PCR choisi correspond au cycle classique de PCR (Sambrook et al., 1989), à l'exception de la température d'hybridation qui est de 36 °C, plutôt faible à cause de l'amorce très courte. Il est à noter que les 45 cycles et l'étape d'élongation de 2 minutes permettent d'amplifier efficacement les produits les plus longs. Le même protocole a été utilisé au départ pour l'amplification des ISSR. Cependant, la diminution de la quantité d'ADN et l'addition au mélange de réaction de 2,5% de formamide ont été nécessaires pour diminuer le bruit de fond de la réaction du au nombre élevé des réactions aspécifiques (« smearing ») et obtenir des profils nets. Le cycle de PCR utilisé est celui de Tsumura et al. (1996). La séparation des bandes amplifiées a été faite sur gel d'agarose, 1,4% pour les RAPD et 2% pour les ISSR. Le coût moyen d'une réaction de génotypage a été de 0,25 euro et le temps nécessaire à la construction de la carte génétique a été de huit mois.

B.3.3. La fiabilité des marqueurs.

Bien que les ISSR soient généralement plus reproductibles que les RAPD (Tsumura et al., 1996 ; Fang et al., 1997 ; Nagaoka et Ogihara, 1997), une fois le protocole optimisé, les deux types de marqueurs se sont révélés également fiables. Le même protocole RAPD a été utilisé dans deux laboratoires différents et aucune variation importante du profil n'a été observée. Les ISSR ont été utilisés dans le cadre du projet CASCADE pour des études de diversité génétique chez le châtaignier européen par Claudia Mattioni. Une bande amplifiée chez un individu appartenant aux populations européennes a été purifiée, clonée et séquencée. Sa séquence coïncide parfaitement avec la séquence du fragment correspondant isolé dans les individus de la descendance (figure 13b).

B.4. LA CARTE GÉNÉTIQUE DU CHÂTAIGNIER.

B.4.1. Le criblage des marqueurs informatifs et le génotypage.

Le criblage des marqueurs et le génotypage des individus de la descendance ont été faits en trois étapes successives :

- criblage des marqueurs informatifs sur les deux parents et six individus F_1 . L'objectif de cette étape est de sélectionner, en faisant le minimum de réactions de génotypage, les amorces RAPD et ISSR qui sont informatives dans le croisement. Cela veut dire qu'il faut retenir les amorces qui amplifient des marqueurs ségrégeant dans des proportions 1:1 ou 3:1. La probabilité que sur n individus il en y ait n_1 avec la bande amplifiée et n_0 sans la bande suit une distribution binomiale. Par conséquent, dans le cas d'un marqueur en configuration « testcross » et utilisant 6 individus F_1 , le risque de se tromper, c'est à dire de considérer un marqueur monomorphe alors qu'il est polymorphe, est égale à $\alpha = (1/2)^6 = 0,016$. Dans le cas d'un marqueur en configuration « intercross », ce risque devient plus important : $\alpha = (3/4)^6 = 0,178$. Nous avons choisi de faire le criblage sur 6 individus car le nombre de réactions de génotypage n'est pas trop élevé et, en même temps, la probabilité de retenir les marqueurs « testcross » est satisfaisante. En effet, ces derniers sont les marqueurs utilisés pour la construction des deux cartes parentales et donc ceux qui contribuent à la saturation du génome. En revanche, les marqueurs « intercross » sont utilisés pour relier les deux cartes. Dans cette première phase, nous avons donc décidé de courir le risque de perdre quelques-uns de ces marqueurs, en faisant le criblage sur 6 individus.
- Les amorces qui ont donné des marqueurs informatifs ont été choisies afin de génotyper la descendance. Le génotypage a été fait sur 90 individus. Considérant les 6 individus de l'étape de criblage, au total 96 individus ont été utilisés pour la construction de la carte. Comme nous l'avons rappelé dans la section B.1.5, une population de 100 individus est suffisante pour la construction d'une carte fiable du point de vue statistique. Le génotypage a été fait d'abord sur 45 individus.
- L'extraction de l'ADN et le génotypage des 45 individus restants ont été faits dans un second temps. Nous avons choisi cette stratégie pour deux raisons : 1) l'amplification des marqueurs RAPD et ISSR faite sur trois échantillons indépendants (dont nous avons fait l'extraction de l'ADN à trois moments différents) permet avec grande fiabilité de ne garder que les marqueurs reproductibles ; 2) une version préliminaire de la carte génétique établie avec 51 individus nous a permis d'exclure de la dernière

Tableau 4. Amorces ISSR sélectionnées pour le génotypage.

Nombre de l'amorce ^a	Séquence (5'→3') ^b	Température d'hybridation (°C)	Marqueurs informatifs par amorce
810	GAGAGAGAGAGAGAT	50	5
814	CTCTCTCTCTCTA	50	4
815	CTCTCTCTCTCTG	52	3
818	CACACACACACACAG	52	4
820	GTGTGTGTGTGTGTC	52	2
823	TCTCTCTCTCTCCTC	52	2
824	TCTCTCTCTCTCTCG	52	2
825	ACACACACACACACT	50	2
828	TGTGTGTGTGTGTGA	50	3
834	AGAGAGAGAGAGAGYT	52	3
836	AGAGAGAGAGAGAGYA	52	4
841	GAGAGAGAGAGAGAYC	52	3
843	CTCTCTCTCTCTCTRA	52	5
844	CTCTCTCTCTCTCTRC	52	5
845	CTCTCTCTCTCTCTRG	52	2
848	CACACACACACACARG	52	3
852	TCTCTCTCTCTCTCRA	52	2
858	TGTGTGTGTGTGTGTRT	52	4
873	GACAGACAGACAGACA	48	3
879	CTTCACTTCACTTCA	38	1
880	GGAGAGGAGAGGAGA	48	1
881	GGTGGGGTGGGGTG	54	2

a: Ce nombre correspond à celui donné aux amorces dans la série #9 obtenue de UBCBL (University of British Columbia Biotechnology Laboratory).

b: Y = C, T; R = A, G.

Tableau 5. Nombre et type de ségrégation des marqueurs utilisés pour la construction de la carte.

	Testcross femelle (ségrégation 1:1)	Testcross mâle (ségrégation 1:1)	Intercross (ségrégation 3:1 or 1:2:1*)	Total
RAPD	143	128	40	311
ISSR	41	19	5	65
Isozymes	3	1	1*	5
Total	187	148	46	381
« Distordus » ($\alpha=0.05$)	18 (9.6%)	15 (10.1%)	5 (10.9%)	38 (10.0%)

Tableau 6. Comparaison entre les marqueurs RAPD et ISSR.

	Nombre total d'amorces utilisées	Marqueurs cartographiés	Marqueurs informatifs/amorce	Marqueurs testcross	Marqueurs intercross
RAPD	120	311	2.6	271 (87%)	40 (13%)
ISSR	22	65	3.0	60 (92%)	5 (8%)

Tableau 7. Comparaison entre la carte femelle et mâle.

	Femelle (Bursa)	Mâle (Hopa)		Femelle (Bursa)	Mâle (Hopa)
Nombre de marqueurs	233 (187 testcross)	194 (148 testcross)	Longueur estimée du génome (Chakravarti)	947	1054
Groupes de liaison (n=12)	12	12	Saturation	76%	68%
Nombre de marqueurs jalon	92	95	Distance moyenne (cM) entre deux marqueurs jalon and intervalle de variation	9.0 (1.0-34.2)	8.7 (1.0-35.0)
Longueur totale de la carte (Kosambi)	720	721	Longueur minimale et maximale des groupes et moyenne (\pm écart type).	10-108 (60 \pm 28)	15-95 (60 \pm 22)

étape du génotypage une quinzaine d'amorces redondantes, qui donnaient des marqueurs fortement liés à d'autres obtenus avec des amorces différentes.

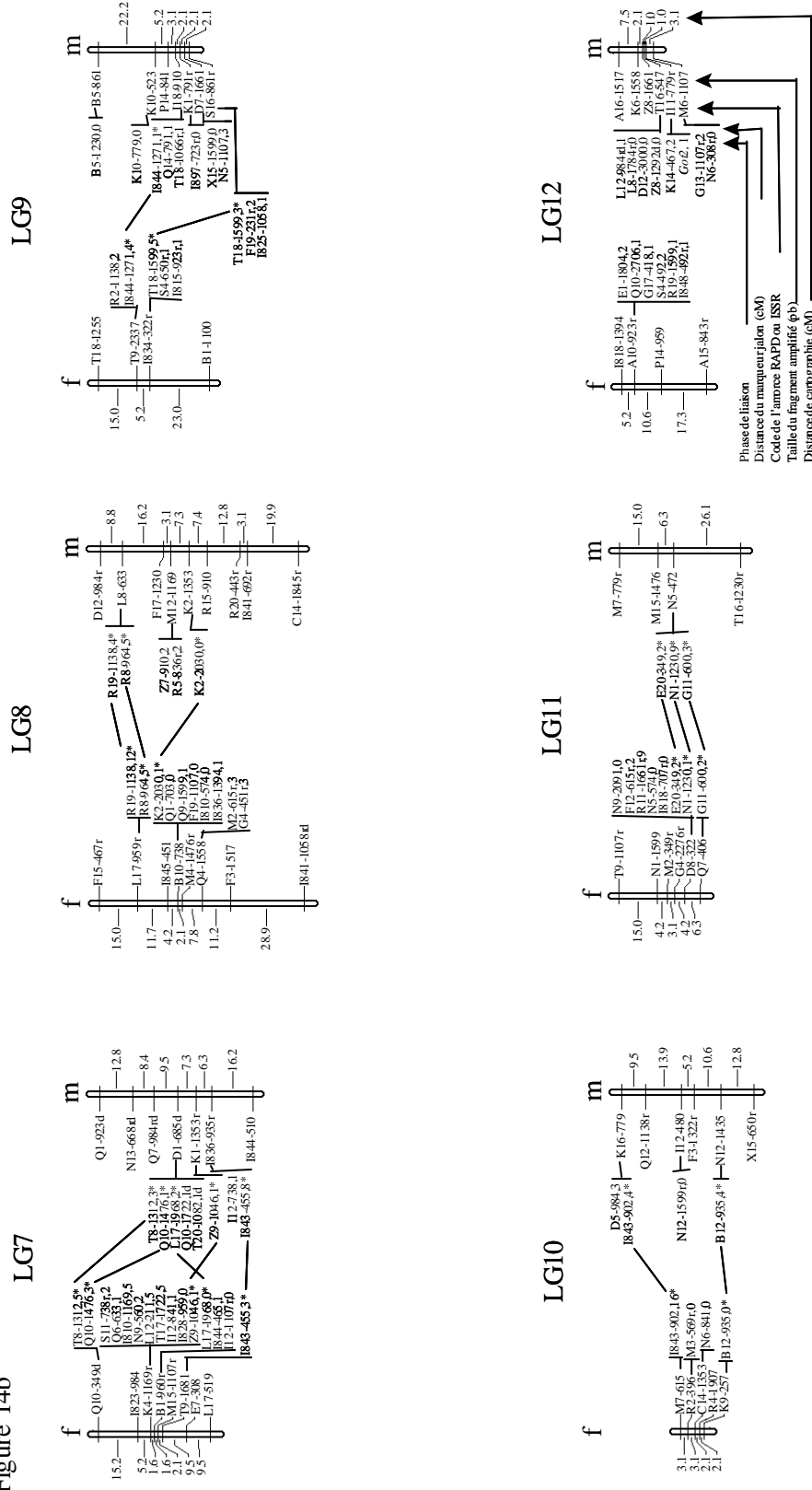
Au total 340 amorces RAPD et 100 ISSR ont été testées. Parmi les 340 amorces RAPD, 198 (58%) ont permis de révéler au moins un marqueur informatif. Au total 450 marqueurs informatifs ont ainsi été obtenus. Cent vingt amorces ont été sélectionnées pour le génotypage de la descendance sur la base du rapport du nombre des marqueurs informatifs par amorce, de la reproductibilité et de la redondance. Cinquante pour cent des amorces ISSR a donné un profil clair, 22 ont été choisis pour le génotypage (tableau 4) pour les mêmes raisons décrites ci-dessus. Les amorces contenant le motif AT n'ont jamais donné un profil clair. Il est probable que ces amorces auto-complémentaires rendent l'amplification moins efficace. Les marqueurs obtenus et le type de ségrégation sont reportés dans le tableau 5. Dix pour cent des marqueurs ont montré une distorsion significative ($\alpha = 0,05$). Dans un premier temps, ces marqueurs ont été exclus de l'analyse. Le tableau 6 montre une comparaison entre les marqueurs RAPD et ISSR. Malgré le taux de marqueurs informatifs par amorce plus élevé dans le cas des ISSR, cette différence n'est pas significative. Les deux types de marqueurs semblent équivalents, toutefois il faut noter que les marqueurs ISSR ont souvent montré un polymorphisme plus élevé que les RAPD (Tsumura et al., 1996 ; Nagaoka et Ogihara, 1997). La carte a finalement été construite en utilisant 381 marqueurs.

B.4.2. La construction de la carte.

Les deux cartes femelle (187 marqueurs « testcross ») et mâle (148 marqueurs « testcross ») ont été établies à l'aide des logiciels « MapMaker » et « JoinMap ». La stratégie de cartographie a consisté en quatre étapes :

- les marqueurs ont été regroupés en groupes de liaison. Pour cela, l'option « group » du logiciel « MapMaker » a été utilisée. Une valeur de LOD de 3 à 6 et une fréquence de recombinaison maximale de 0,40 ont été testées. Douze groupes de liaison ont ainsi été obtenus pour les deux cartes parentales à toutes les valeurs de LOD. Un LOD = 4 a donc été choisi afin d'avoir le maximum de marqueurs liés et en même temps un risque plus faible d'obtention de fausses liaisons. Peu de marqueurs n'ont pas été liés suggérant une bonne saturation (0 chez la femelle et 10 chez le mâle) et l'addition des marqueurs « distordus » n'a pas affecté le résultat de cette analyse.
- Les options « compare » et « ripple » ont été utilisées pour déterminer l'ordre des marqueurs le plus probable. L'option « ripple », grâce à des permutations de l'ordre des marqueurs, permet de calculer combien de fois un ordre donné est plus

Figure 14b



« probable » qu'un autre. Les marqueurs qui n'ont pas pu être positionnés avec fiabilité ($LOD > 2$, ordre choisi au moins 100 fois plus probable que celui permuté), ont été enlevés dans un premier temps. Les marqueurs ordonnés de cette manière constituent la carte dite « jalon ». Tous les autres marqueurs ont été positionnés en utilisant l'option « near » près des marqueurs jalons (figure 14).

- La troisième étape a consisté à comparer l'ordre obtenu avec celui donné par « JoinMap ». A chaque fois qu'une incohérence a été relevée, l'option « error detection » du logiciel « MapMaker » a été utilisée pour révéler des erreurs de génotypage (fréquence trop élevée de doubles recombinaisons). Les gels d'agarose ont été contrôlés et l'analyse d'ordonnancement répétée. Lincoln et Lander (1992) ont montré que les erreurs de génotypage peuvent avoir une influence importante sur l'ordre des marqueurs, mais aussi sur la longueur totale de la carte. De plus, un ordre fiable est une caractéristique très importante pour la détection des QTL.
- Enfin, la dernière étape est faite à l'aide de « JoinMap » qui permet, grâce à la cartographie des marqueurs 3:1, de relier les deux cartes parentales. Comme le montre la figure 11, l'information des marqueurs 1:1/3:1 est très faible. Pour cette raison, suivant la suggestion de Debener et Mattiesch (1999), seuls les marqueurs « intercross » qui montraient au moins trois liaisons ont été intégrés dans la carte. Les 37 marqueurs « intercross » ont permis de relier onze groupes sur douze.

Le tableau 7 et la figure 14 montrent cette première version de la carte génétique de *C. sativa*. La saturation, calculée comme le rapport entre la distance totale observée et la longueur du génome estimée comme décrit auparavant (cf. B.1.6), est de 76% pour la carte femelle et 68% pour la carte mâle. Ce niveau de saturation est très proche de celui obtenu chez *Q. robur* (75%) par Barreneche et al. (1998). Utilisant 187 marqueurs « testcross » chez la femelle et 148 chez le mâle, la saturation obtenue est plus faible que celle prévue par Beckmann et Soller (1983). Cependant, comme cela apparaît évident sur la figure 14, la plupart des groupes de liaison montrent une région de concentration des marqueurs. La distribution non homogène des marqueurs tout au long du groupe de liaison a souvent été observée, par exemple chez les Solanaceae (Tanksley et al., 1992). Ce phénomène est la conséquence directe du fait que la fréquence de recombinaison n'est pas égale dans les différentes régions du chromosome. Comme nous avons pu le voir dans l'introduction, la recombinaison est élevée dans les régions riches en gènes et faible dans les régions riches en hétérochromatine, par exemple les centromères (Thuriaux, 1977). Aucune information n'est disponible concernant la position des centromères chez le châtaignier, par conséquent il n'est pas possible de distinguer si cette

distribution non homogène est due à la proximité du centromère ou plutôt à une tendance des marqueurs à cibler des régions particulières du génome.

La taille génétique estimée du génome du châtaignier est de 947 cM pour la femelle et 1054 cM pour le mâle. Malgré les valeurs plus faibles que ceux reportés pour le chêne par Barreneche et al. en 1998 (1192 cM pour la femelle et 1235 cM pour le mâle), la taille génétique des deux espèces est similaire. Ce résultat est très important pour la comparaison des deux génomes car, sachant que les deux espèces ont aussi une taille physique proche (0,94 pg/C chez le chêne et 0,81 pg/C chez le châtaignier), le rapport entre taille physique et génétique est de même similaire. En moyenne, ce rapport est de 0,77 Mb/cM chez le chêne et 0,81 Mb/cM chez le châtaignier. Cette valeur similaire, proche de celle de la tomate (cf. tableau 2), amène à deux réflexions intéressantes :

- le rapport n'est pas trop élevé ; théoriquement des approches de cartographie fine et clonage positionnel sont envisageables chez les deux génomes ; chez les arbres forestiers, le problème majeur, pour réaliser la cartographie fine, reste la nécessité d'avoir des effectifs importants.
- la similarité de la taille physique du génome entre les deux espèces a une conséquence importante : si la synténie, la colinéarité et le contenu en gènes sont conservés entre chêne et châtaignier, alors les deux génomes devraient aussi avoir une proportion de séquences répétées similaire. Cela impliquerait que la différenciation entre les deux génomes n'a pas eu lieu par expansion de ce type de séquences.

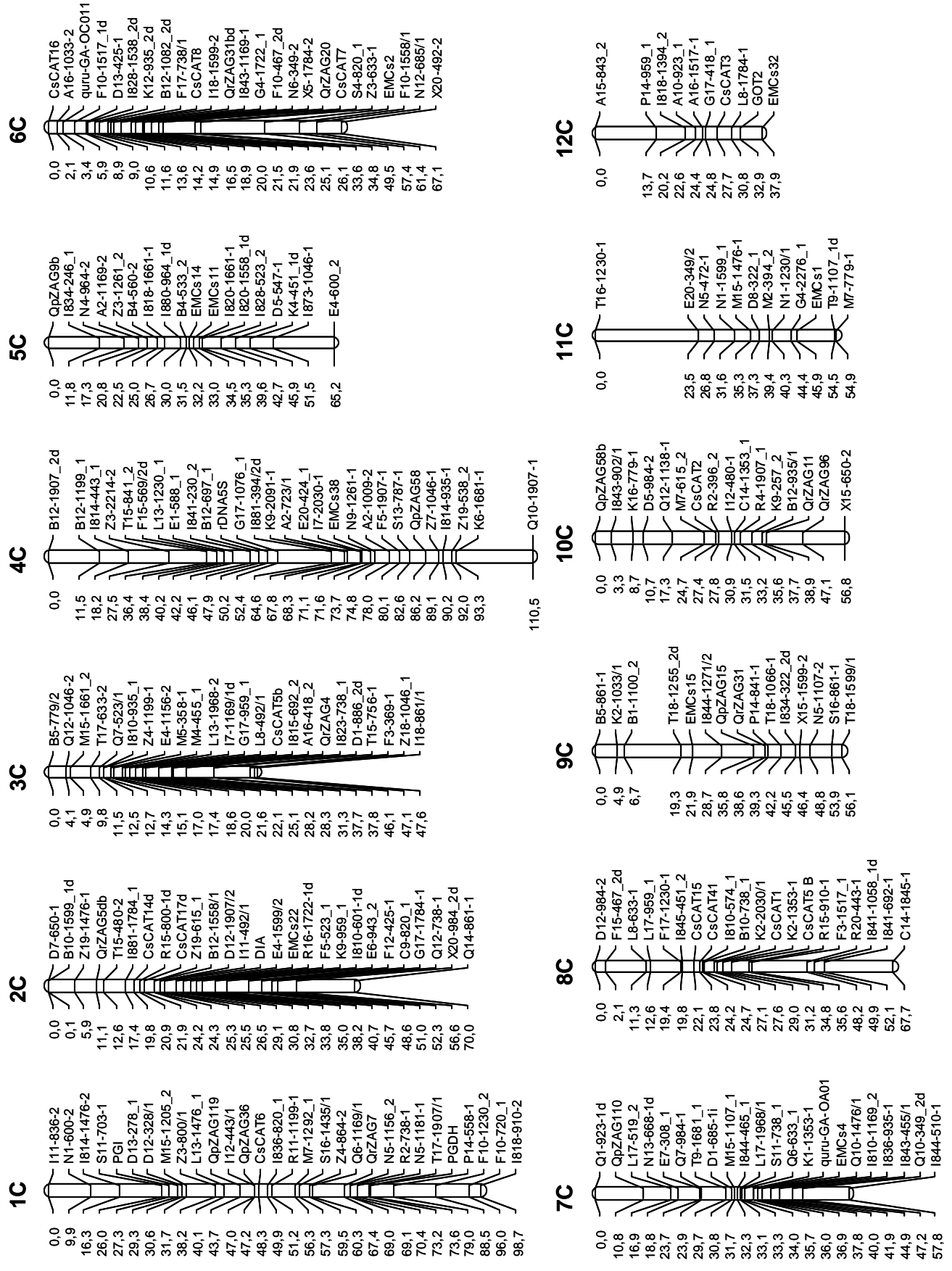
B.4.3. L'évolution de la carte du châtaignier.

Dans la partie D, nous discuterons l'utilisation de marqueurs SSR et STS (dérivés des EST) pour la comparaison de la carte génétique du châtaignier avec celle du chêne. Le résultat de l'intégration des SSR dans la carte du châtaignier sera, toutefois, présentée ici, car cela nous permet de commenter l'évolution de la carte par rapport à la première version décrite ci-dessus et d'introduire la carte utilisée pour la détection des QTL, qui sera présentée dans la partie C.

La carte génétique du châtaignier contient à présent 517 marqueurs : 311 RAPD, 65 ISSR, 51 AFLP, 5 isozymes, 39 SSR et 46 STS, dont 42 dérivés des EST. Au cours de cette thèse la carte a été enrichie en marqueurs et un nombre d'individus plus important a été génotypé avec un sous échantillonnage de marqueurs. La carte a subi trois changements importants:

- 51 marqueurs AFLP (Vos et al., 1995), génotypés sur 90 individus, ont été intégrés, en utilisant le protocole mis au point sur chêne par Stéphanie Mariette et Catherine

Figure 17. Carte consensus contenant tous les marqueurs utilisés dans la recherche des QTL.



Bodénès et décrit dans Gerber et al. (2000). Les 51 marqueurs AFLP ont été obtenus grâce à 4 combinaisons amorce/enzyme (ou PEC, « primer/enzyme combinations »). Le test a été fait pour vérifier si les AFLP pouvaient cibler des régions différentes du génome du châtaignier (donc augmenter la saturation), et surtout pour comparer l'efficacité par rapport à celle des RAPD et ISSR. La figure 15 (les AFLP en violet) montre la carte consensus femelle/mâle obtenue à l'aide du logiciel « JoinMap », sans optimisation de l'ordre des marqueurs, utilisant un seuil de LOD = 6 et un taux de recombinaison maximale de $r = 0,40$. Les 51 marqueurs sont distribués sur tout le génome et n'ont pas contribué de façon importante à l'augmentation de la saturation. En revanche, l'efficacité plus élevée de cette technique, en comparaison à celle des RAPD et des ISSR, est très évidente : 12,75 bandes informatives par gel contre 2,6 pour les RAPD et 3,0 pour les ISSR.

- Afin de détecter les QTL, un sous échantillon de marqueurs bien répartis sur tout le génome a été sélectionné à partir de la première version de la carte. Ces marqueurs ont été génotypés sur 186 individus. La carte de détection des QTLs contient aussi les 39 marqueurs SSR qui ont permis de relier le douzième groupe de liaison entre le parent femelle et mâle. La figure 16 montre les deux cartes jalons femelle et mâle et la figure 17 la carte consensus obtenue grâce aux marqueurs « intercross ». Les seuils de LOD = 8,0, $r = 0,28$ et $\text{LOD} > 3,0$ pour l'option « ripple », rendent cette carte plus robuste du point de vue statistique. Cela dérive de l'effectif plus important. La carte femelle contient 109 marqueurs (66 RAPD, 19 ISSR, 2 isozymes et 22 SSR) et a une longueur de 848,6 cM ; la carte mâle contient 108 marqueurs (76 RAPD, 11 ISSR, 1 isozyme et 20 SSR) et a une longueur de 832,9 cM. La saturation est augmentée : 84% pour la femelle et 75% pour le mâle. Les estimations de la taille génétique sont de 1009 cM pour la femelle et 1108 cM pour le mâle. Ces estimations coïncident pour trois valeurs de LOD (3, 4 et 5) et ils ne s'éloignent pas trop de l'estimation faite avec la première version de la carte. Il faut aussi noter que l'ordre des marqueurs n'a pas changé (à l'exception de quelques marqueurs proches l'un de l'autre dans les groupes 2, 11 et 12), ce qui valide la fiabilité de l'ordre obtenu lors de la première version de la carte.
- Enfin, des marqueurs STS ont été intégrés. Cela a permis d'introduire des gènes, dont la fonction est connue, qui peuvent représenter des gènes candidats. La nouvelle carte consensus contenant ces derniers marqueurs sera présentée dans la partie D et sera comparée à celle du chêne.

C. Détection de QTL impliqués dans les caractères d'adaptation chez le châtaignier

It is almost impossible with any brevity to exemplify the notion of adaptation. Just because adaptation consists, even in the simplest cases, in a multiplicity of correspondences between one sufficiently complicated system, the organism itself, and another equally complicated, the environment in which it finds itself. It is, indeed, just this multiplicity that makes the thing recognizably adaptive.

(Fisher R.A. (1934) Adaptation and mutations. School Science Review 15: 294-301)

C.1. LES CARACTÈRES ADAPTATIFS.

C.1.1. L'adaptation et les caractères adaptatifs.

L'adaptation est le changement structurel et/ou fonctionnel qui rend un organisme ou une population mieux capable de survivre dans un environnement (Koski et al., 1997). Il y a deux types d'adaptation. D'une part l'adaptation physiologique qui, traditionnellement, est le changement phénotypique d'un individu, donc d'un seul génotype, suite au changement des conditions environnementales. Aujourd'hui le terme de plasticité phénotypique est souvent préféré pour indiquer ce type d'adaptation. D'autre part, l'adaptation génétique est le changement évolutif de la structure génétique d'une population en réponse à la sélection naturelle. Au cours de cette thèse nous utiliserons le mot adaptation pour adaptation génétique et plasticité phénotypique pour la réponse d'un génotype au changement des conditions environnementales qui l'entourent. L'adaptation et la spéciation sont deux processus clé dans la théorie de l'évolution biologique de Darwin (1859).

En 1859 Darwin élabore la théorie de l'évolution des espèces fondée sur la sélection naturelle, c'est à dire sur la capacité différentielle des individus de se reproduire et donc de transmettre leurs gènes à la génération suivante. Cette aptitude reproductive d'un génotype donné est appelée fitness, ou valeur adaptative, et elle dépend des qualités intrinsèques du génotype dans un environnement donné. Tout caractère qui augmente la fitness d'un individu est un caractère adaptatif. Les caractères adaptatifs, grâce à l'action de la sélection naturelle, déterminent l'adaptation des populations à l'environnement. Un caractère, qui est adaptatif dans un environnement donné, peut, bien sûr, avoir un effet défavorable sur la fitness dans un environnement différent et donc perdre dans cette dernière condition l'attribut d'adaptatif.

L'évolution des caractères adaptatifs et les bases génétiques de l'adaptation sont loin d'être éclaircies. Deux théories ont été proposées : le modèle infinitésimal de Fisher (1930) et celui oligogénique de Orr (1998). D'après le modèle infinitésimal, un caractère adaptatif résulte des effets alléliques à un grand nombre de loci, chaque locus contribuant pour une faible part au caractère. Le modèle de Fisher constitue la synthèse néo-darwiniste du débat sur le rôle des micro et macromutations dans l'évolution. Bien que le rôle évolutif des mutations à effet fort ait été reconnu par certains (Bateson, 1913 ; Morgan 1932), la vision néo-darwiniste selon laquelle beaucoup de mutations à effet faible auraient joué un rôle majeur dans l'adaptation, résumée par l'expression « *natura non facit saltum* », s'est largement imposée jusqu'à présent.

Au contraire, selon le modèle oligogénique de Orr, des locus à effet fort joueraient un rôle très important au cours de l'évolution qui amène une population vers l'optimum phénotypique dans un environnement donné. Suivant ce dernier modèle la distribution des effets des allèles fixés au cours de l'adaptation est exponentielle (peu de locus à effet fort et beaucoup plus de locus à effet faible et/ou modéré seraient impliqués dans l'adaptation). Orr et Coyne (1992), ont mis en évidence plusieurs points faibles dans la théorie de Fisher. Du point de vue théorique il y a au moins trois points à considérer : 1) comme le note Kimura (1983), Fisher montre que les mutations à effet faible ont une probabilité plus élevée d'être favorables que les mutations à effet fort. Cependant, il ne dit pas si les micro mutations ont aussi une probabilité plus élevée de se fixer ; 2) Fisher parle de changements lents et progressifs vers l'optimum d'adaptation, mais ne considère pas la possibilité de brusques changements environnementaux, auxquels, souvent, les organismes doivent faire face ; 3) enfin, il considère qu'une mutation quelconque affectant n'importe quel caractère a le même effet sur la fitness.

Kimura (1983) a ensuite montré que même si les micro mutations sont plus favorables, il est moins probable qu'elles puissent se fixer. En tenant compte des résultats de Kimura, Orr (2000) a, enfin, proposé son modèle selon lequel, les macromutations peuvent parfois jouer un rôle dans l'adaptation grâce notamment à une probabilité plus élevée de se fixer. Dans des conditions environnementales particulières, ce type de mutations permettrait plus rapidement de s'approcher de l'optimum phénotypique. Toutefois, puisque les macromutations ont une probabilité faible d'être favorables, l'adaptation aurait lieu surtout grâce aux mutations à effet intermédiaire. De plus, l'adaptation est caractérisée par beaucoup de caractères qui interagissent l'un avec l'autre afin de donner le phénotype optimal dans un environnement donné et à cause de l'effet pleiotrope²⁴, une mutation peut être favorable pour un caractère mais défavorable pour un autre. Par conséquent, les organismes les plus complexes sont aussi les plus lents à s'approcher de l'optimum phénotypique dans un environnement donné et donc à s'adapter.

Même si certains résultats obtenus lors d'études de détection des QTL semblent en accord avec la théorie de Orr (par exemple la distribution des effets des QTL, comme nous pourrions le voir par la suite), et des mutations à effet fort ont été décrites comme responsables de la variation quantitative, la compréhension des bases génétiques de l'adaptation des populations à leur milieu reste encore très fragmentaire (Barton et Keightley, 2002). Orr et Coyne (1992)

²⁴ Effet pleiotrope : effet d'un gène sur plusieurs caractères différents.

ont suggéré une stratégie possible pour identifier le nombre et la contribution des gènes impliqués dans l'adaptation. Il faudrait tout d'abord disposer d'une analyse génétique qui puisse révéler les mutations (ou mieux, des allèles) à effet faible et à effet fort ; et il faudrait étudier des caractères qui soient clairement liés à l'adaptation. Ce type d'étude devrait être mené dans des populations naturelles. L'étude de recherche de QTL qui a été conduite lors de cette thèse chez le châtaignier suit ces suggestions.

Trois caractères adaptatifs seront considérés dans le cadre de cette thèse : la phénologie des bourgeons, l'efficacité d'utilisation de l'eau et la croissance, dont les caractéristiques essentielles seront décrites dans les paragraphes suivants.

C.1.2. La phénologie du bourgeon (débourrement et arrêt de croissance).

Dans les zones tempérées, caractérisées par une alternance des saisons, l'interruption de la croissance à l'approche de l'hiver est une stratégie d'adaptation au froid des arbres. Cette interruption de l'activité des méristèmes des bourgeons et l'arrêt de la croissance végétative qui en suit est appelé dormance. Deux périodes de la phénologie des bourgeons de l'arbre marquent la fin et le début de cette période de dormance. A l'automne, l'inhibition de l'activité de division cellulaire des bourgeons, due à la réduction de la longueur du jour, permet à la plante de survivre face aux basses températures de l'hiver. Ce processus détermine la formation du bourgeon d'hiver et il est appelé arrêt de croissance. En revanche, la fin de la période de dormance et la reprise de l'activité végétative au printemps sont caractérisées par le débourrement. La température est le facteur environnemental principal, qui contrôle le processus de débourrement. En réalité, de nombreux facteurs internes et environnementaux interviennent dans ces deux processus (Rohde et al., 2000). Lang (1987) a distingué trois types différents de dormance : l'écodormance, l'interruption de la croissance causée par les conditions environnementales défavorables ; la paradormance, l'inhibition de l'activité de division cellulaire au niveau des bourgeons latéraux, provoquée par le bourgeon apical ; et, enfin, l'endodormance, c'est à dire la dormance contrôlée par des facteurs internes au bourgeon dormant, qui à ce stade ne reprend pas sa croissance même en conditions favorables.

La régulation du débourrement et de l'arrêt de croissance est un compromis entre la survie et la productivité saisonnière d'un arbre. Un arrêt de croissance précoce est une garantie d'une bonne acclimatation au froid de l'hiver, en dépit d'une saison de croissance plus courte. Par ailleurs, un débourrement précoce permet une saison de croissance plus longue, mais une

gelée tardive peut provoquer des dégâts importants qui pourraient compromettre la saison de croissance. En général les arbres forestiers manifestent une variabilité clinale de la date de débourrement, comme l'ont montré de nombreux tests de provenances. Les populations provenant de latitudes et altitudes plus élevées, caractérisées par des saisons de croissance plus courtes, montrent un arrêt de croissance plus précoce que les populations provenant de latitudes et altitudes plus basses (Howe et al., 2000 et les références citées). La relation est plus complexe pour le débourrement, mais une différence significative est toujours évidente entre populations originaires d'environnements contrastés du point de vue géographique et climatique (Jermstad et al., 2001a ; Howe et al., 2003). Chez le chêne, les populations provenant de latitudes plus septentrionales sont toujours plus tardives que celles provenant de latitudes plus méridionales (Ducousso et al., 1996). Chez les espèces de chêne, le débourrement est également sujet à une pression de sélection biotique à cause de certaines espèces d'insectes défoliateurs (Tikkanen et Tiitto, 2003). Ces résultats démontrent qu'il s'agit, donc, de deux caractères adaptatifs très importants qui au cours de l'évolution des peuplements forestiers ont été soumis à l'action de la sélection naturelle.

C.1.3. La discrimination isotopique du carbone 13.

La discrimination isotopique du carbone 13 est une mesure indirecte de l'efficacité d'utilisation de l'eau (WUE, de l'anglais « water use efficiency ») de la plante (Farquhar et al., 1989). A l'échelle foliaire, l'efficacité d'utilisation de l'eau instantanée est le rapport entre la biomasse produite ou assimilation (A) et l'eau utilisée (E) par la plante ou transpiration. Ce rapport, A/E, est un indicateur important de l'adaptation des plantes à la sécheresse. Comment est-il lié à la discrimination isotopique du carbone 13 ? Farquhar et al. (1989) ont modélisé cette relation en montrant que pour les plantes en C₃, les processus foliaires qui déterminent l'efficacité d'utilisation de l'eau correspondent à ceux qui sont impliqués dans la discrimination entre les deux isotopes, ¹²C et ¹³C, accomplie par la plante. En effet, les plantes en C₃ discriminent l'isotope lourd du carbone, ¹³C, lors de l'assimilation photosynthétique du CO₂. Cette discrimination se réalise principalement au niveau de deux étapes : 1) la diffusion du CO₂ de l'air ambiant vers l'intérieur de la feuille à travers les stomates (discrimination de 4,4 ‰) ; et 2) la carboxylation par l'enzyme RuBisCO (discrimination de 30 ‰).

Il faut savoir que :

- 1) la composition isotopique relative en ¹³C et ¹²C d'un échantillon est définie par :

$$\delta^{13}\text{C} = (R_e/R_b) - 1 \quad (1)$$

où Re et Rb sont les rapports molaires isotopiques, $R = {}^{13}\text{C}/{}^{12}\text{C}$, dans l'échantillon et un standard conventionnel (Belemnite Pee Dee) ;

2) l'effet isotopique est défini par le rapport

$$\alpha = R_{\text{source(ou substrat)}}/R_{\text{produit}} \quad (2)$$

et la discrimination isotopique (Δ) par l'équation

$$\Delta = \alpha - 1 \quad (3)$$

et donc, en combinant les équations 1, 2 et 3, la discrimination est égale à :

$$\Delta = (\delta^{13}\text{C}_{\text{source}} - \delta^{13}\text{C}_{\text{produit}})/(1 + \delta^{13}\text{C}_{\text{produit}}).$$

En sachant que $\delta^{13}\text{C}_{\text{produit}}$ est négligeable par rapport à l'unité, nous pouvons écrire l'approximation suivante:

$$\Delta \cong \delta^{13}\text{C}_{\text{source}} - \delta^{13}\text{C}_{\text{produit}}.$$

Dans le cas de la discrimination des plantes en C_3 , $\Delta \cong \delta^{13}\text{C}_{\text{air}} - \delta^{13}\text{C}_{\text{matière organique des plantes}}$.

Par exemple $\Delta \cong (-8\text{‰}) - (-27\text{‰}) \cong 19\text{‰}$.

Suivant le modèle de Farquhar et al., (1989) pour les plantes en C_3 , la discrimination Δ a été mise en relation avec la diffusion du CO_2 à travers les stomates et la carboxylation dans les chloroplastes selon l'équation suivante :

$$\Delta = a + (b - a) C_i/C_a \quad (4)$$

où a est la discrimination contre le ${}^{13}\text{C}$ durant la diffusion du CO_2 à travers les stomates (4,4‰), b la discrimination durant la carboxylation (28‰), C_i et C_a les concentrations en CO_2 dans les espaces intercellulaires et l'air.

D'ailleurs, $\text{WUE} = A/E$ est le rapport entre le flux de CO_2 (taux d'assimilation de CO_2 , A) et de H_2O (taux de transpiration, E) à travers les stomates. Ces flux sont décrits par :

$$A = g(C_a - C_i)/1,6 \quad (5)$$

$$E = g(W_i - W_a) \quad (6)$$

avec $C_a - C_i$ et $W_i - W_a$ les différences de concentrations en CO_2 et H_2O entre l'air et les espaces intercellulaires et g la conductance stomatique à la diffusion de vapeur d'eau. En combinant les équations 4, 5, et 6 nous avons, enfin, la relation entre WUE et Δ :

$$\text{WUE} = A/E = [C_a (b - \Delta)]/[1,6(W_i - W_a)(b - a)].$$

Entre les deux paramètres il y a donc une relation directe qui permet de déterminer les plantes plus ou moins efficaces dans l'utilisation de l'eau par une simple mesure de la signature isotopique du carbone de la matière organique végétale. Il a été, en effet, montré, pour plusieurs espèces herbacées cultivées ainsi que pour les arbres forestiers, que WUE et Δ sont

des caractères adaptatifs qui présentent des différences entre populations ou écotypes adaptés à des conditions contrastées de disponibilité en eau (voir Brugnoli et Farquhar, 2000 pour une revue). Δ est donc un caractère, dont l'héritabilité a été montrée, qui nous permet de rendre compte des différences d'adaptation à la sécheresse entre géotypes, populations et espèces. Ce caractère est très complexe et de nombreux déterminants environnementaux et génétiques expliquent sa variation phénotypique (Brugnoli et Farquhar, 2000). Certains d'entre eux seront considérés plus loin afin de comprendre les résultats de la détection des QTL que nous présentons dans cette thèse.

C.1.4. La croissance.

La croissance saisonnière d'un arbre peut être évaluée en mesurant la hauteur et le diamètre au début et à la fin de la saison végétative. La croissance en hauteur et en diamètre est une composante essentielle de la productivité en biomasse d'un arbre. En effet, il s'agit du résultat de l'assimilation en carbone et de sa répartition dans les différentes parties de la plante. Il est à noter que l'assimilation et la répartition du carbone sont contrôlées et affectées par plusieurs voies biochimiques, processus physiologiques et caractéristiques anatomiques de la plante, ainsi que par différentes variables environnementales comme la température, la concentration de CO₂ dans l'air, la disponibilité d'eau et des nutriments (Kirschbaum, 2000). Il est donc évident qu'il s'agit de caractères composés et complexes, pour lesquels il est souvent difficile d'identifier les déterminants génétiques et environnementaux qui les régissent (Bradshaw et Stettler, 1995). Toutefois, une proportion de la variation phénotypique, très variable selon les espèces et les différentes estimations, est sous contrôle génétique et la nature adaptative de ces caractères est évidente. Il suffit de penser que la croissance en hauteur dans la forêt naturelle est l'un des caractères les plus importants dans la compétition pour la lumière. Comme pour le Δ , nous ne considérerons que certains aspects de la croissance en hauteur et en diamètre des arbres pour essayer d'éclaircir les résultats obtenus dans la détection des QTL.

C.1.5. L'étude des caractères adaptatifs chez les arbres forestiers dans le cadre des changements climatiques globaux.

Au cours des 100 dernières années la concentration de carbone de l'air est passée de 280 à 370 parties par million et la température a augmenté d'environ 0,5°C (IPCC, 2001). Ce réchauffement du climat de la planète n'est pas sans précédents, mais la vitesse, à laquelle il se produit, est sans doute exceptionnelle. C'est pourquoi depuis une vingtaine d'années les chercheurs s'intéressent de près aux effets écologiques du changement climatique (Walther et

al., 2002). Ils ont montré que les plantes, les animaux et les écosystèmes sont déjà en train de répondre à ces changements (Barber et al., 2000 ; Parmesan et Yohe, 2003 ; Root et al., 2003). Les changements en cours et prévus qui concernent les écosystèmes forestiers vont de l'augmentation du CO₂ et de la température, bien sûr, à l'augmentation de la longueur de la saison de croissance, à des conditions hydriques plus sèches, à des changements concernant la disponibilité de nutriments. De plus, la fréquence d'événements climatiques extrêmes (tempêtes et tornades, par exemple), des incendies et des épidémies d'insectes et de champignons semble augmenter en fréquence par rapport au passé. Le débourrement, la croissance et l'efficacité d'utilisation de l'eau sont donc trois caractères qui seront affectés par les pressions de sélection dérivant des changements climatiques. En général les environnements présentant les conditions climatiques optimales de croissance pour une essence donnée se déplaceront vers le nord et dans des zones plus élevées. Bien que la réponse des arbres à ces changements ne soit pas complètement prévisible (Saxe et al., 2001), il semble que les arbres ne seront pas affectés de manière catastrophique, même si une perte en biodiversité semble probable. Étant donné le temps de génération long des espèces forestières et la rapidité des changements climatiques en cours, la plasticité phénotypique et la possibilité de migration vers des environnements plus favorables semblent deux des stratégies adaptatives les plus efficaces pour ces espèces.

Dans le cadre des changements climatiques globaux, l'étude des caractères adaptatifs chez les espèces forestières est extrêmement précieuse. Elle permet d'évaluer le potentiel adaptatif d'une espèce donnée et, en même temps, de fournir des stratégies possibles de conservation génétique. Afin de réaliser ces objectifs il est fondamental de comprendre comment un caractère adaptatif évolue et quels sont les gènes responsables de la variabilité adaptative d'une espèce.

C.2. LA DÉTECTION DES QTL.

C.2.1. La détection des QTL : l'étude du contrôle génétique des caractères complexes. Comment peut-elle aider à la compréhension des bases génétiques de l'adaptation ?

Les caractères, contrôlés par plusieurs locus, qui manifestent une variation phénotypique continue, montrant souvent une distribution normale, sont appelés caractères quantitatifs ou complexes et leur hérédité est définie comme polygénique. Le phénotype de ces caractères est déterminé (en proportions variables selon le type de caractère) par le génotype et l'environnement. Grâce aux approches statistiques de la génétique quantitative, il est possible d'estimer la proportion de la variabilité phénotypique expliquée par les effets des gènes, ce qui correspond à l'héritabilité d'un caractère quantitatif. Chacun des locus qui contribuent à la variabilité génétique de ces caractères est appelé QTL, en anglais « Quantitative Trait Locus ».

La génétique quantitative nous offre une vision générale des propriétés génétiques de ces caractères (nombre approximatif, proportion de la variance génétique expliquée par les effets additifs, de dominance et d'épistasie), mais c'est grâce à la détection des QTL que nous pouvons déterminer avec plus de précision le nombre de locus en cause, leur position dans le génome et leurs effets (Tanksley, 1993 ; de Vienne et Causse, 1997). Ce type d'analyse consiste à décomposer les caractères quantitatifs en leurs composantes mendéliennes, les QTL, et elle est basée sur la recherche systématique de déséquilibre de liaison (ou association statistique) entre les allèles aux marqueurs cartographiés et les locus contrôlant les caractères. Bien que la première association entre un marqueur morphologique et un locus contrôlant un caractère quantitatif remonte au début du siècle passé (Sax, 1923) et que le principe de la détection des QTL ait été formulé en 1961 par Thoday, ce n'est qu'à partir de la fin des années 80 que la disponibilité d'un grand nombre de marqueurs moléculaires qui ont permis la construction de cartes génétiques et le développement des méthodes statistiques d'analyse des données ont favorisé l'essor des études de recherche de QTL (Tanksley, 1993).

Les éléments nécessaires pour pouvoir détecter des QTL sont :

- une descendance en ségrégation comprenant plusieurs centaines de plein-frères ;
- une carte génétique construite à partir de cette descendance ; la construction de la carte nécessite le « génotypage » de chaque individu de la descendance pour un ensemble de marqueurs bien répartis sur le génome. La carte génétique doit être de bonne

qualité : elle doit être saturée et comprendre des marqueurs régulièrement distribués sur les groupes de liaison (tous les 10-20 cM) ;

- la valeur du caractère quantitatif mesurée pour chacun des individus de la descendance ; la distribution des valeurs phénotypiques doit être normale, car les méthodes d'analyse statistique utilisées se basent sur l'hypothèse de normalité ;
- un système de traitement des données qui puisse mettre en évidence les marqueurs liés statistiquement au caractère et estimer les paramètres génétiques des QTL trouvés (nombre, position, effets, interactions).

La détection des QTL a abouti dans plusieurs cas au clonage du gène impliqué. Une vingtaine de gènes chez les animaux (Glazier et al., 2002) et une dizaine chez les plantes (Paran et Zamir, 2003) ont été ainsi clonés à partir de leur localisation sur le génome préalablement identifiée par la recherche de QTL. Chez les plantes modèles, notamment chez *Arabidopsis*, l'essor des nouvelles technologies de génotypage à haut débit, la disponibilité d'une vaste collection de SNP²⁵, de mutants et de lignées récombinantes (RIL, « Recombinant Inbred Lines », en anglais), conduiront sans doute dans les prochaines années à la caractérisation moléculaire de plusieurs QTL (Borevitz et Chory, 2004). Si la détection des QTL est réalisable dans la plupart des espèces végétales, le clonage du gène responsable du QTL est beaucoup moins accessible. Par exemple, chez les arbres forestiers les effectifs limités, souvent autour de 150-200 individus, ne permettent pas, comme nous le verrons dans le paragraphe C.2.3, ni d'estimer de façon précise la position des QTL ni d'avoir accès à la cartographie fine. De plus, la disponibilité de séquences de type EST reste limitée et ces séquences ne constituent qu'une couverture très partielle du génome. Nous ne disposons que très rarement de cartes physiques du génome associées aux cartes génétiques. Enfin, chez les conifères la taille énorme du génome empêche toute stratégie de clonage positionnel. Néanmoins, même si l'identification du gène reste encore difficile et peu précise, il y a au moins deux raisons pour lesquelles la détection des QTL chez les arbres forestiers peut conduire à des résultats intéressants :

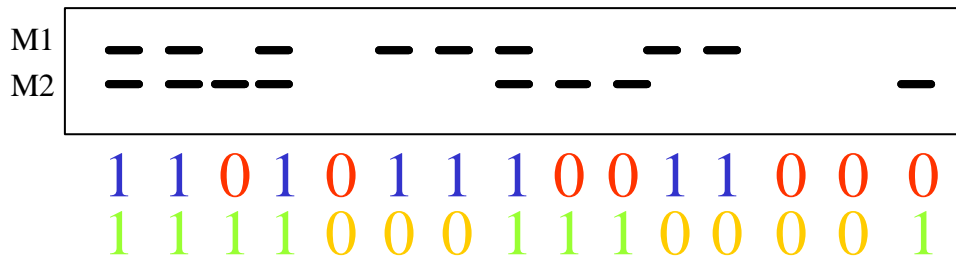
- la détection des QTL contrôlant les caractères adaptatifs peut aider à éclaircir une des questions concernant les bases génétiques de l'adaptation : combien y-a-t-il de locus impliqués et s'agit-il de locus à effet fort ou bien faible ? Les locus impliqués dans le processus d'adaptation sont-ils conservés entre les populations et les espèces ? L'approche QTL peut de la même manière être utilisée chez les arbres forestiers pour

²⁵ SNP: "Single Nucleotide Polymorphism", polymorphisme d'un seul nucléotide.

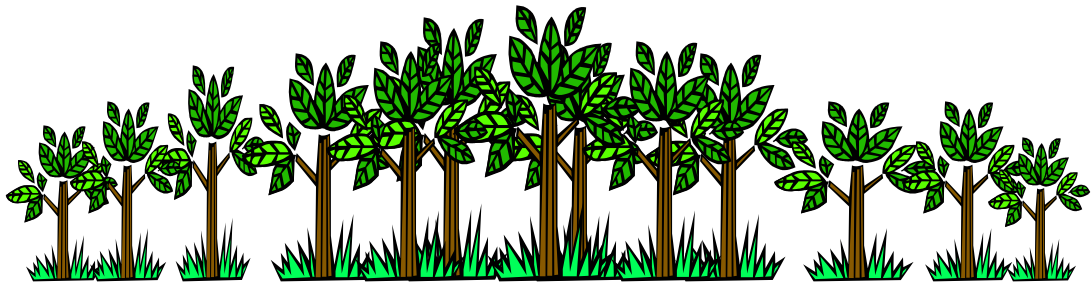
Figure 18 (modifiée de Costa, 1999)

Analyse QTL : recherche d'une liaison statistique entre un marqueur polymorphe et la variabilité d'un caractère quantitatif.

Génotypes à deux marqueurs, M₁ et M₂, des individus d'une descendance issue d'un croisement contrôlé.

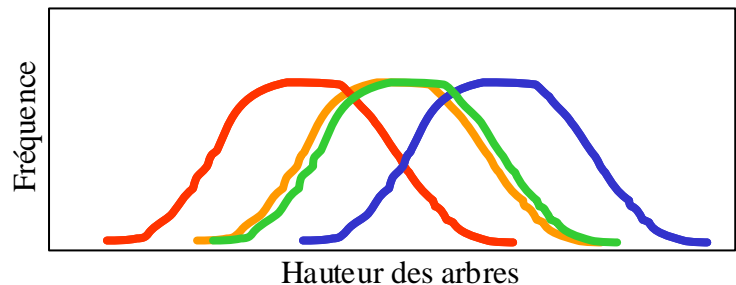


Le caractère quantitatif est mesuré dans tous les individus de la descendance



S'il y a des différences significatives entre la moyenne phénotypique des individus avec la bande et celle des individus sans la bande alors le marqueur est associé à un QTL.

$$\begin{aligned} \overline{X}_0 &\neq \overline{X}_1 \text{ QTL...oui} \\ \overline{X}_0 &= \overline{X}_1 \text{ QTL...non} \end{aligned}$$



Ségrégation des allèles à un QTL dans une descendance F₁ sous le modèle de double pseudo test-cross :

$$\text{♀} = Q_1Q_2 \quad \text{♂} = Q_3Q_4$$

La ségrégation des allèles de chacun des deux parents se réalise sous l'effet de bruit de fond des allèles de l'autre :

$$\text{ségrégation maternelle } (Q_1Q_3 + Q_1Q_4) - (Q_2Q_3 + Q_2Q_4)$$

$$\text{ségrégation paternelle } (Q_3Q_1 + Q_3Q_2) - (Q_4Q_1 + Q_4Q_2).$$

identifier les régions génomiques responsables de la spéciation (Saintagne et al., 2004). Cette application de la détection des QTL au domaine de la génétique évolutive suscite un grand intérêt, chez les plantes en général (Mauricio 2001) et notamment chez les espèces forestières.

- Le génome du peuplier sera bientôt complètement séquencé constituant ainsi une espèce « modèle » en génomique des arbres forestiers (Brunner et al., 2004). Ceci est particulièrement intéressant dans le cas des Fagacées. En effet, le peuplier appartient à l'ordre des Malpighiales qui fait partie de la même section (Eurosid I) des Fagacées. Chez le peuplier plusieurs études de détection des QTL contrôlant les caractères d'adaptation, notamment la phénologie des bourgeons, ont été menées et plusieurs gènes impliqués sont déjà connus (Frewen et al., 2000 ; Rohde et al., 2002).

C.2.2. Les approches statistiques pour la détection des QTL et le choix des niveaux de signification des QTL.

Le principe de détection des QTL est simple : il s'agit de rechercher les marqueurs dont la ségrégation des allèles dans la descendance montre une liaison statistique significative avec la valeur phénotypique d'un caractère donné chez les individus. Comme le montre la figure 18, dans le cas d'un marqueur dominant ségrégeant 1:1 dans une famille F_1 il s'agit de tester (par exemple par ANOVA) si la différence entre les moyennes phénotypiques des individus présentant la bande amplifiée et de ceux ne présentant pas la bande est significative. Ce type d'analyse est appelé analyse simple marqueur car les marqueurs sont pris en compte individuellement. La figure 18 montre aussi la ségrégation des allèles à un QTL dans le même type de descendance et les classes de génotypes sous le modèle de double pseudo test-cross. Bien que le principe soit très simple, les méthodes statistiques de détection des QTL sont plutôt complexes. Nous ne considérerons pas les détails statistiques des différentes approches, mais nous nous focaliserons sur la description et l'évolution des méthodes d'analyse en considérant les avantages et les défauts de chacune. Récemment, une synthèse a été faite par Doerge (2002) et nous ferons référence à cet article dans la partie qui suit.

Les différentes méthodes proposées pour la détection des QTL sont : l'analyse simple marqueur (SMA, « Single-Marker Analysis », Edwards et al., 1987), l'analyse par intervalle (IM, « Interval Mapping », Lander et Botstein, 1989), l'analyse par intervalle composite (CIM, « Composite Interval Mapping », Zeng, 1994 ; MQM « Multiple QTL Mapping »,

Jansen, 1993) et l'analyse par intervalle multiple (MIM, « Multiple Interval Mapping », Korol et al., 1998a ; Kao et al., 1999).

L'analyse par simple marqueur, que nous avons présenté auparavant, considère les marqueurs individuellement et le test est fait par ANOVA, t-test ou régression. L'hypothèse nulle est l'indépendance des moyennes phénotypiques des individus de la classe génotypique. Cette hypothèse est rejetée quand la différence entre les moyennes est significative. Dans ce dernier cas le marqueur est lié à un QTL contrôlant le caractère étudié. L'avantage de ce type d'analyse est sa simplicité. En effet, n'importe quel logiciel de statistique de base peut être utilisé, il s'agit d'une analyse qui n'est pas affectée de façon importante par l'écart à la normalité des données phénotypiques. De plus, si des effectifs importants et la correction de Bonferroni du niveau de signification pour les tests multiples sont utilisés, les résultats de ce type d'analyse sont presque toujours validés par d'autres méthodes. Cependant, cette analyse ne permet pas de localiser la position des QTL sur le génome, les estimations des effets ne sont pas précises et il est pratiquement impossible d'estimer le nombre de QTL car nous ne connaissons pas la portion de génome couverte par les marqueurs testés.

L'analyse par intervalle, faite souvent par maximum de vraisemblance (Lander et Botstein, 1989), permet de tester l'hypothèse de présence d'un QTL non seulement au niveau de marqueurs individuels mais aussi au niveau de l'intervalle entre eux. La détection des QTL est donc faite à l'aide d'une carte génétique. A chaque point régulièrement espacé entre deux marqueurs, la vraisemblance de l'hypothèse de présence d'un QTL est écrite et comparée à celle d'absence d'un QTL. La valeur de LOD, le logarithme décimal du rapport des deux maximum de vraisemblance, est la statistique utilisée pour établir la signification du QTL. Afin d'avoir une bonne efficacité de détection des QTL la carte doit être saturée, contenir peu de données manquantes, des intervalles de 10 à 20 cM et être construite avec un effectif important (Darvasi et al., 1993 ; Darvasi et Soller, 1994). Rebai et al. (1995) ont montré que l'analyse par intervalle est légèrement plus puissante que l'analyse simple marqueur à courtes distances (moins de 20 cM), mais la différence d'efficacité entre les deux analyses augmente avec la distance entre marqueurs. L'analyse par intervalle ne teste qu'un QTL à la fois et souvent la présence de deux QTL sur le même chromosome montrant le même effet peut aboutir à la détection d'un faux QTL au milieu des deux. Si, en revanche, les deux QTL proches ont un effet opposé, le risque est de n'en voir aucun.

L'analyse par intervalle composite (Zeng, 1994 ; Jansen, 1993) permet de s'affranchir des effets des autres QTL détectés dans le génome quand le test de présence d'un QTL est fait à une position donnée. Au niveau pratique cela veut dire qu'une première analyse

d'identification des QTL permet de choisir certains QTL comme covariables à utiliser lors d'une deuxième analyse plus détaillée qui vise à la localisation précise des QTL. Ce type de stratégie permet de se libérer de l'effet des QTL à effet majeur et d'améliorer la précision de la localisation des QTL dans le cas de présence de deux QTL sur le même chromosome.

L'analyse par intervalle composite ne considère pas les interactions entre les QTL, qui sont en revanche, considérées dans celle par intervalles multiples (Korol et al., 1998a ; Kao et al., 1999). Toutefois, l'introduction des interactions entre les QTL complique énormément l'analyse à cause du nombre élevé de QTL potentiels et de leurs interactions. La solution heuristique qui a été proposée consiste dans une première étape à détecter les QTL par intervalle composite, ensuite construire un modèle avec les QTL détectés et leurs interactions et enfin rechercher les interactions significatives. Toutefois, les QTL ayant des effets d'épistasie mais qui ne sont pas détectés comme QTL individuels dans la première étape, ne seront pas identifiés par cette analyse.

Il est intéressant de noter que les méthodes statistiques de détection des QTL, qui continuent à s'améliorer afin d'étudier les interactions entre les différents QTL détectés, peuvent être appliquées également à l'analyse des données d'expression des gènes. Cette application est de grand intérêt car elle permet en association avec la technologie des filtres d'ADN complémentaire à haute densité (« microarrays », en anglais) d'identifier des polymorphismes de niveau d'expression (EPL, « expression-level polymorphism ») et des régions du génome contenant des groupes de gènes qui peuvent interagir entre eux (Doerge, 2002).

Une fois le choix de la méthode d'analyse fait, une question importante dans la détection des QTL est de définir le seuil statistique pour le risque de première espèce. Les deux conclusions fausses auxquelles nous pouvons arriver sont : 1) d'une part détecter un QTL quand il n'est pas présent (faux positif, erreur de première espèce) ; 2) d'autre part, ne pas détecter un QTL alors qu'il est présent (faux négatif, erreur de deuxième espèce). Dans le cas de la détection des QTL le risque de première espèce est déterminé par le choix du niveau de signification (α). Au contraire le risque de deuxième espèce dépend des conditions expérimentales (par exemple, la taille de l'effectif et l'effet plus ou moins grand du QTL) et il constitue la puissance de l'expérience.

Quel risque de première espèce choisir ? En général, nous voulons un risque de 5% au niveau de l'expérience, dans ce cas particulier donc au niveau du génome. En effet le problème est plutôt complexe car si la valeur de LOD d'un QTL pris individuellement est proportionnelle à une distribution χ^2 , cela n'est pas vrai pour un ensemble de distribution de LOD. En raison

Encadré 4

Calcul des niveaux de signification des QTL au niveau des chromosomes.

Pour chaque carte l'erreur de première espèce au niveau de marqueur a été calculée selon la formule :

$$\alpha_{\text{marq}} = 1 - (1 - \alpha_{\text{g}})^{(1/M)}$$

où α_{marq} = erreur au niveau du marqueur,
 α_{g} = erreur au niveau du génome et M =
 nombre des marqueurs sur le génome.

Parent	α_{g}	α_{chr}	Nombre marqueurs	α_{marq}
femelle	0,05	0,0043	109	0,00047
mâle	0,05	0,0043	108	0,00047
femelle	0,10	0,0087	109	0,00097
mâle	0,10	0,0087	108	0,00098
femelle	0,15	0,0135	109	0,00149
mâle	0,15	0,0135	108	0,00150
femelle	0,20	0,0184	109	0,00205
mâle	0,20	0,0184	108	0,00206
femelle	0,40	0,0417	109	0,00468
mâle	0,40	0,0417	108	0,00472

des tests multiples effectués, de la liaison génétique, à cause de laquelle les tests ne sont pas indépendants, de l'écart de la normalité des données phénotypiques, de la proportion de données manquantes, de la distorsion de ségrégation et de la distribution non homogène des marqueurs sur le génome, la distribution théorique de la statistique utilisée n'est pas connue. Il est cependant très important de pouvoir associer à la valeur de LOD obtenue un niveau de signification pour pouvoir comparer ensuite les résultats entre expériences différentes (Lander et Kruglyak, 1995 ; van Ooijen , 1999). Churchill et Doerge (1994) ont proposé une méthode par ré-échantillonnage souvent utilisée afin de calculer si la valeur de LOD n'a pas été obtenue par chance. Suivant cette méthode, les couples des génotypes et des valeurs phénotypiques de chaque individu sont brisés et le jeu de données est ré-échantillonné : les valeurs phénotypiques sont redistribuées aux hasard sur tous les individus dont les génotypes à tous les marqueurs sont maintenus. Les données permutées sont analysées avec la même stratégie utilisée pour l'analyse des données originales et les valeurs de LOD obtenues sont comparées à la valeur de LOD du jeu de données expérimentales. Cela nous donne l'erreur de première espèce associée au QTL. L'encadré 4 montre l'erreur de première espèce calculé dans le cas du châtaignier.

Pour les QTL significatifs les paramètres génétiques à estimer sont : 1) la position du QTL sur le chromosome et l'intervalle de confiance de cette position, 2) l'effet de substitution allélique du QTL et 3) la proportion de variance phénotypique expliquée par le QTL. La position la plus probable du QTL coïncide avec la position sur le chromosome qui correspond à la valeur de LOD plus élevée. L'intervalle de confiance de cette position peut être calculé par « bootstrap » (Visscher et al., 1996). Cette méthode est souvent utilisée quand l'effectif est inférieur à 500 individus, la distribution des caractères peut s'écarter de la normalité et les marqueurs ne sont pas distribués de façon homogène sur le génome. En ce qui concerne l'effet de substitution il s'agit d'estimer l'effet sur la valeur phénotypique de la substitution d'un de deux allèles au QTL avec l'autre. Dans le cas d'un backcross et en utilisant des marqueurs dominants cet effet peut être calculé comme : $d = X_{Qq} - X_{qq}$, étant X_{Qq} et X_{qq} , les valeurs moyennes du caractère dans les deux classes de génotypes possibles. Dans la même situation la proportion de variance phénotypique expliquée par le QTL est égale à PEV (« Phenotypic Explained Variance ») = $\frac{1}{4}(d^2/\sigma_{ph}^2)$, où d est l'effet de substitution allélique et σ_{ph}^2 la variance phénotypique du caractère dans la descendance.

C.2.3. Limites de la détection des QTL.

La détection des QTL chez les animaux et chez les plantes a permis pour la première fois d'avoir accès aux « facteurs mendéliens » qui contrôlent les caractères complexes. Comme cela a été rappelé précédemment, dans certains organismes modèles comme *Arabidopsis* (Borevitz et Chory, 2004) et la souris (The Complex Trait Consortium, 2003) il est désormais possible de cloner les gènes responsables de certains QTL contrôlant les caractères complexes. Il est, néanmoins, très important de noter que de nombreux paramètres influencent les résultats de la détection de QTL et que dans la plupart des espèces nous sommes loin de la situation idéale des organismes modèles. Il est essentiel de prendre en compte les facteurs influençant la détection des QTL afin d'optimiser les conditions expérimentales et de bien interpréter les résultats de l'analyse.

La réussite et la puissance d'une expérience de détection de QTL dépendent de plusieurs facteurs : 1) la présence d'allèles ségrégeant au QTL dans la descendance utilisée ; 2) le type de descendance ; 3) l'héritabilité du caractère ; 4) le nombre des locus qui contrôlent le caractère ; 5) la valeur des effets de QTL individuels ; 6) la taille de la population en ségrégation. Nous commenterons certains de ces limites qui, comme nous le verrons plus loin, influencent les résultats de la détection des QTL que nous décrirons dans la partie C.3.

L'effectif de la descendance et la valeur des effets des QTL sont deux facteurs qui influencent de façon importante la puissance de détection des QTL. Van Ooijen (1992) a démontré que dans le cas d'une descendance « backcross », la probabilité de détecter un QTL qui explique 5% de la variance phénotypique varie de 0,11 à 0,84 en passant d'un effectif de 100 à 400 individus. La même probabilité varie de 0,41 à 1,00 pour un QTL expliquant 10% de la variance phénotypique. Il est évident qu'avec les effectifs généralement utilisés (100-200 individus) les QTL à effet très faible ne seront pas détectés.

L'effectif a aussi des conséquences très importantes sur la précision de la localisation des QTL et l'estimation de l'effet (Beavis, 1995). En général, les intervalles de confiance sur la position du QTL sont larges, de l'ordre de 10-20 cM et l'imprécision augmente avec la diminution des effectifs. Cela vient du fait que les événements rares de recombinaison, nécessaires pour bien distinguer les effets du QTL, ne sont détectés qu'en utilisant beaucoup d'individus. De plus, l'effectif faible provoque systématiquement une surestimation des effets des QTL. Par conséquent, les valeurs des effets ne peuvent être comparées entre expériences que si les effectifs utilisés sont du même ordre de grandeur.

L'héritabilité du caractère est un autre facteur qui joue aussi sur la puissance de détection des QTL. Pour les caractères à faible héritabilité si l'effectif est autour de 100-200 individus, il est

Tableau 8. Quelques études de recherche de QTL chez les arbres forestiers.

Espèce	Caractère et héritabilité	Effectif et type de descendance	Méthode statistique	Seuil de signification	Nombre de QTL	Variance phénotypique expliquée	Stabilité des QTL	Référence
<i>Pinus taeda</i> L.	Densité du bois	177 (F ₂)*	ANOVA	0,05	5	23% (5 QTL ensemble)	-	Groover et al. 1994
<i>Pinus pinaster</i> Ait.	Hauteur	126 (F ₂)	ANOVA	0,001-0,05	6 (3 âges)	6-20%	Aucun QTL stable	Plomion et al. 1996
<i>Pinus radiata</i>	Volume et diamètre	93 (F ₁)	CIM	0,05 (α _q)	12 volume et 21 diamètre (4 âges)	6,8-30%	45% des QTL sont stables	Enebiri et al. 1998
<i>Pinus taeda</i> L.	Hauteur et diamètre	84-171 (F ₂)*	MIM**	0,005-0,0001	13 hauteur (6 âges et 2 sites), 8 diamètre (2 âges)	7,3-30%	Peu de QTL stables	Kaya et al. 1999
<i>Pinus sylvestris</i>	Arrêt de croissance et résistance au froid	300-400 (F ₁)	Approche Bayésien	-	4 arrêt de croissance, 7 résistance au froid	4,8-13% (arrêt de croissance) 0,01-11,7% (résistance au froid)	-	Hurme et al. 2000
<i>Pinus sylvestris</i>	Croissance et résistance au froid	94 (F ₁)	IM	-	12 croissance (5 caractères) 4 résistance au froid (2 caractères)	9,3-15,9% (croissance) 11,3-22,0% (résistance au froid)	-	Lerceteau et al. 2000
<i>Pinus taeda</i> L.	Qualité du bois (0,20-1,00)	172 (F ₂)*	MIM**	0,005-0,01	5,8 caractère en moyenne	5,4-15,7%	Certains stables, d'autres no (spécifiques d'une âge)	Sewell et al. 2000
<i>Pinus pinaster</i> Ait.	δ ¹³ C (0,17) et diamètre (0,19)	202 (F ₂)	CIM	0,05 (α _q) et 0,05 (α _c)	8 (δ ¹³ C) et 6 (diamètre)	4,7-12,4% (δ ¹³ C) et 5,9-18,1% (diamètre)	-	Brendel et al. 2002
<i>Pinus palustris</i> Mill. x <i>P. elliotii</i> Engl.	Hauteur (0,47-0,68) et diamètre	258 (BC ₁ , trois sites)	ANOVA/IM	0,005 (ANOVA)	1,05 caractère en moyenne	3,6-11,0%	Peu stables d'un site à l'autre	Weng et al. 2002
<i>Pinus taeda</i> L.	Qualité du bois	172-457 (3 pedigrees F ₂ * et différentes âges)	MIM**	0,01-0,05 (α _c)	Pedigree 172 (Sewell et al. 2000) ; 6,3 caractère (pedigree 457) ; 4,0 caractère (pedigree 445)	1,7-5,7% (pedigree 457) et 1,8-4,4% (pedigree 445)	62% stables au cours des âges différentes, 30% entre les pedigrees	Brown et al. 2003
<i>Pseudotsuga menziesii</i>	Résistance au froid (0,17-0,31)	224 (F ₂)	MIM**	0,005-0,01	4,3 caractère	2,0-9,8%	-	Jermstad et al. 2001a
<i>Pseudotsuga menziesii</i>	Debournement	78-190 (F ₂ , deux sites)	MIM**	0,005-0,01	4,1 caractère/année (4 années différentes)	1,2-11,6%	Stables entre les années, no entre les deux sites	Jermstad et al. 2001b
<i>Pseudotsuga menziesii</i>	Debournement et arrêt de croissance	environ 400 (F ₂)	MIM**	0,005-0,01	4-11 caractère	0,7-9,5%	Stables entre les deux sites	Jermstad et al. 2003
<i>Eucalyptus grandis</i> x <i>E. urophylla</i>	Propagation végétative	122 (F ₁)	IM	0,01	6,7 caractère	7,3-21,0%	-	Grattapaglia et al. 1995
<i>Eucalyptus grandis</i> x <i>E. urophylla</i>	Croissance (0,1-0,3), densité du bois.	201 (F ₁)	IM	0,1 (α _q)	10 densité du bois, 3 croissance (trois âges)	5,1-14%	68% stables entre deux âges	Verhaegen et al. 1997
<i>Eucalyptus tereticornis</i> x <i>E. globulus</i>	Propagation végétative (0,4)	895 (221 génotypes extrêmes F ₁)	ANOVA, IM, CIM	0,005 (ANOVA) 0,1-0,01 (α _c)	3,3 caractère/deux années	2,6-17,0%	Quelques QTL stables entre les deux années	Marques et al. 1999
<i>Eucalyptus nitens</i>	Hauteur (0,19) et surface de la feuille (0,36)	465 (deux pedigrees F ₂ *)	ANOVA	0,05	3 (hauteur) et 2 (surface de la feuille)	10,3-14,7% (hauteur) et 9,8-11,6% (surface de la feuille)	-	Byrne et al. 1997
<i>Salix viminalis</i> x <i>S. schweinfurthii</i>	Hauteur (0,54-0,78) et diamètre (0,81)	94 (F ₁)	IM/CIM	0,05	1,2 caractère/trois âges	14-22%	Peu des QTL stables entre les âges	Tsarouhas et al. 2002
<i>Salix viminalis</i> x <i>S. schweinfurthii</i>	Debournement (0,43-0,72)	92 (F ₁)	IM/CIM	0,05	1,5 caractère/deux âges	12-24%	-	Tsarouhas et al. 2003
<i>Populus trichocarpa</i> x <i>P. deltoides</i>	Croissance (0,5-0,8) et debournement (0,98)	55 (F ₁) (répétitions)	IM	0,05	1-2 (croissance) et 5 debournement	24,4-33,4% (croissance) et 28,7-51,5% (debournement)	-	Bradshaw et Stettler, 1995
<i>Populus trichocarpa</i> x <i>P. deltoides</i>	Croissance	55 (F ₂) (répétitions)	Régression linéaire	0,001	0-5 caractère	10-48,6%	-	Wu 1998a
<i>Populus trichocarpa</i> x <i>P. deltoides</i>	Croissance	55 (F ₂) (répétitions)	IM	0,05 (α _q)	4 (trois âges)	20-70%	-	Wu et al. 1998
<i>Populus trichocarpa</i> x <i>P. deltoides</i>	Debournement (0,94) et arrêt de croissance (0,91)	346 (F ₂)	IM/CIM	0,05 (α _q)	9 (debourement) et 4 (arrêt de croissance)	5,9-16,8% (debourement) et 6,0-12,2% (arrêt de croissance)	-	Frewen et al. 2000
<i>Quercus robur</i> L.	Morphologie foliaire	89-184 (F ₁) (répétitions)	CIM	0,05 (α _c)	1-3 caractère	6,0-22,9%	-	Saintagne et al. 2004
<i>Quercus robur</i> L.	Debournement (0,15-0,51), hauteur (0,04-0,23)	89-220 (F ₁) (répétitions)	CIM	0,05 (α _{c+q})	12 debourement (3 sites), 2-4 hauteur	3,1-10,7% (debourement), 4,1-18,7% (hauteur)	Debournement : 6 stables entre deux sites et 1 dans les trois sites	Saintagne et al., 2003

*pedigree de trois générations.

**MIM selon Knott et al., 1997.

indispensable de mener la détection des QTL dans des environnements différents et avoir plusieurs répétitions afin d'obtenir une bonne estimation des effets et du nombre minimal de QTL impliqués. En effet, même avec un effectif de 1000 individus si les QTL sont nombreux et l'héritabilité autour de 0,30 la puissance de détection reste faible, 50% des QTL sera détecté dans une expérience individuelle.

Enfin, le nombre de QTL contrôlant le caractère influence aussi la puissance de détection en fonction de la taille de la population et l'héritabilité du caractère (Beavis, 1995). Pour donner un exemple, en utilisant un effectif de 100 individus, même avec un caractère très héritable (0,90) contrôlé par 10 QTL, seulement 40% des QTL seront détectés. Si le caractère est contrôlé par 40 QTL, moins de 10% des QTL seront détectés. Si à tout cela nous ajoutons la variation des QTL contrôlant le caractère en fonction des environnements différents et à différents âges, il est évident qu'une expérience individuelle de recherche de QTL conduite sur 100-200 individus ne donne pas une vision exhaustive du contrôle génétique d'un caractère complexe.

C.2.4. Les études de détection des QTL chez les arbres forestiers.

Les résultats principaux de la détection des QTL chez les arbres forestiers ont été synthétisés par Sewell et Neale (2000) et van Buijtenen (2001). Le tableau 8 montre une liste d'études de détection des QTL menées chez des espèces forestières différentes, pour plusieurs caractères, en utilisant différents types de descendance et des effectifs très variables. En général, le temps de génération long des arbres forestiers empêche d'avoir des pedigrees optimaux pour la détection des QTL. Toutefois, le niveau d'hétérozygotie élevé de ces espèces et des stratégies comme le double pseudo test-cross permettent d'obtenir facilement des pedigrees en ségrégation pour les caractères quantitatifs. Les descendance de type F_1 et F_2 sont les plus utilisées. La croissance, la phénologie des bourgeons, les caractères liés à la qualité du bois, la propagation végétative et l'efficacité d'utilisation de l'eau ont été étudiés chez les espèces forestières. Il s'agit de caractères à héritabilité très variable. En général, la phénologie des bourgeons montre l'héritabilité la plus élevée (jusqu'à 0,98), mais les estimations varient beaucoup. Par exemple, les caractères de croissance en hauteur et diamètre peuvent avoir des héritabilités de 0,1 à 0,8. La valeur de l'héritabilité dépend de la méthode utilisée pour son estimation et de l'environnement dans lequel elle est estimée, il est donc très difficile de pouvoir généraliser par caractère. Malgré l'hétérogénéité des effectifs et des méthodes d'analyse, il y a au moins trois observations à faire sur les données synthétisées dans le tableau 8 :

- 1) le nombre de QTL détectés est faible. Même dans le cas du débourrement qui souvent est très héritable, un maximum de 11 QTL ont été détectés en utilisant un effectif de 400 individus (Jermstad et al., 2003) et deux sites de détection. Pour les autres caractères le nombre de QTL est rarement au-dessus de 5. A l'exception de deux cas, les expériences avec 5 QTL ou plus ont été menées sur un effectif supérieur à 150 individus. Les deux exceptions concernent deux caractères à héritabilité moyenne et élevée et dans un cas la détection des QTL a été faite sur plusieurs clones de la descendance (Grattapaglia et al., 1995 ; Bradshaw et Stettler, 1995). En revanche, Byrne et al. (1997) ont trouvé peu de QTL par caractère bien que l'effectif soit de 465 individus divisé en deux familles. L'analyse par ANOVA et la faible héritabilité, du moins en ce qui concerne la croissance en hauteur, peuvent expliquer ce résultat.
- 2) Les contributions des QTL à la variance phénotypique demeurent faibles. Nous avons déjà évoqué que cette contribution était sensible à l'effectif de la descendance. Le faible effectif provoque systématiquement une surestimation de ces effets. Tous les cas présentant au moins une estimation au-dessus de 25%, pour un QTL individuel, concernent des effectifs inférieurs à 100 individus. Par ailleurs, la diminution des valeurs des effets des QTL avec un effectif important est d'autant plus évidente dans le papier de Brown et al., (2003). En utilisant 457 et 445 individus les effets individuels des QTL ne se situent jamais au-dessus de 5,7%.
- 3) La stabilité des QTL entre âges et environnements différents semble liée au type de caractère. En général, les QTL impliqués dans le débourrement sont plus stables que les QTL impliqués dans la croissance. Il faut aussi noter que, notamment pour la croissance, le nombre de QTL détectés augmente avec les sites et les âges testés. Cela suggère que le contrôle génétique de la croissance est plus dépendant des conditions expérimentales.

Il est évident que la comparaison entre études différentes est très difficile car les estimations des paramètres génétiques dépendent de façon importante des conditions expérimentales et de l'échantillonnage. De plus, l'effectif optimal pour avoir une détection efficace des QTL se situe autour de 1000 individus, loin des effectifs souvent disponibles chez les arbres forestiers.

Etant donné ces limites, une stratégie possible chez les arbres forestiers afin de pouvoir obtenir des résultats plus complets est la suivante :

- effectuer la détection des QTL sur plusieurs années et dans plusieurs environnements ;

- utiliser la multiplication végétative des descendants pour améliorer la précision des mesures phénotypiques ;
- décomposer les caractères complexes en composantes à héritabilité plus élevée et étudier un caractère avec plusieurs mesures différentes ;
- comparer les résultats obtenus pour le même caractère entre espèces différentes.

Cette stratégie devrait permettre, par exemple, de comparer les estimations des effets du même QTL détecté plusieurs fois, de distinguer entre QTL stables et spécifiques d'une condition expérimentale donnée et d'avoir une estimation plus précise du nombre des QTL.

C.3. DÉTECTION DES QTL IMPLIQUÉS DANS LA PHÉNOLOGIE DES BOURGEONS, LA DISCRIMINATION ISOTOPIQUE DU CARBONE 13 ET LA CROISSANCE CHEZ LE CHÂTAIGNIER.

Le travail de recherche de QTL chez le châtaignier s'est déroulé à l'Istituto di Biologia Forestale e Agroambientale du CNR de Porano, où la famille F₁ de châtaignier est installée et au Laboratoire de Génétique de l'UMR BIOGECO de l'INRA de Pierroton, où les analyses des données ont été réalisées avec l'aide de David Pot et Christophe Plomion. Maria Cristina Monteverdi a réalisé les observations phénologiques en 2000 et 2001 et elle m'a toujours aidée et conseillée lors des mesures phénotypiques. Marco Lauteri a dirigé les mesures de discrimination isotopique et il m'a assistée dans l'interprétation physiologique des résultats. Ce travail a fait l'objet d'une publication présentée en annexe II.

C.3.1. Les mesures phénotypiques : 2000, 2001 et 2002.

Les caractères adaptatifs étudiés sont : 1) la phénologie des bourgeons (nous avons mesuré le débourrement et l'arrêt de croissance) ; 2) la discrimination isotopique du carbone 13 ; 3) la croissance (nous avons mesuré le diamètre et la hauteur au début et à la fin de la saison de croissance végétative). Les mesures ont été effectuées au cours de trois années : 2000, 2001 et 2002 qui correspondent à la deuxième, troisième et quatrième saison végétative des plantes, à l'exception de l'arrêt de croissance qui a été mesuré une seule année, en 2002. Le mot mesure sera souvent utilisé pour indiquer un caractère mesuré dans une année donnée. Les raisons du choix de ces trois caractères et des trois années de mesures sont les suivantes :

- comme le montrent les figures 19a et 19b, Lauteri et al. (1999) ont étudié les deux populations Bursa et Hopa dans un environnement commun et ils ont montré que les paramètres physiologiques liés aux caractères étudiés sont différents entre les deux écotypes. Il s'agit de trois caractères adaptatifs qui ont probablement subi chez ces deux populations des pressions sélectives vers deux « optima » phénotypiques différents. De plus, ces trois caractères sont souvent étudiés chez les arbres forestiers. Par conséquent, l'aspect clairement adaptatif de ces caractères et la possibilité de faire ensuite des comparaisons avec d'autres espèces a guidé le choix.
- Vu la puissance limitée d'une expérience individuelle de recherche de QTL, surtout avec un effectif faible, les mesures ont été faites durant trois années. Cette stratégie devrait permettre de vérifier la stabilité des QTL d'une année sur l'autre. En même

Figure 19

a) Comparaison de trois caractères adaptatifs étudiés chez les deux populations Bursa et Hopa dans un environnement commun.

b) Caractéristiques physiologiques opposées des deux populations Bursa et Hopa résumées schématiquement.

Lauteri *et al.* 1999. Moyenne \pm SE.

	BURSA	HOPA
Débourrement (distance julienné)	97.3 \pm 2.2	105.6 \pm 2.5
Delta (feuilles ‰)	17.4 \pm 0.5	16.8 \pm 0.2
Diamètre (mm)	7.0 \pm 0.7	6.6 \pm 0.5
Hauteur (cm)	13.4 \pm 1.6	16.3 \pm 1.4

	Bursa	Hopa
Delta (1/WUE)	↑	↓
Photosynthèse	↑	↓
Croissance juvénile	↓	↑
Croissance adulte	↑	↓
Débourrement	précoce	tardif

c)



Stades de développement du bourgeon de châtaignier durant le débourrement (photos aimablement fournies par Josefa Fernández López). Le stade indiqué par la flèche est celui noté dans les deux mesures "bud" et "bud70" décrites dans le texte.

d)

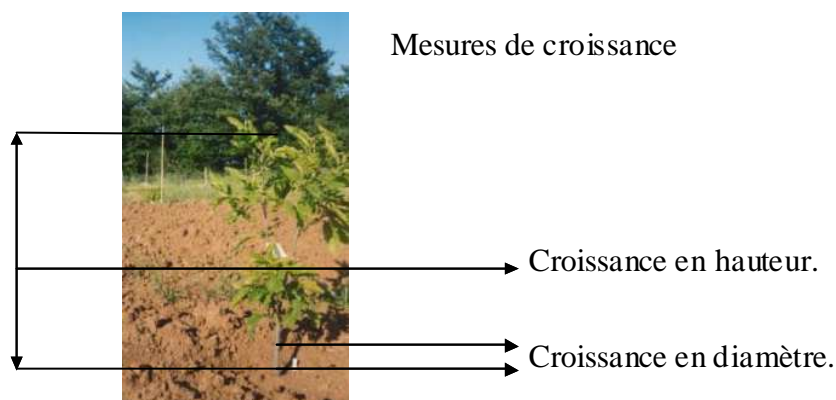


Stades de formation du bourgeon d'hiver (photos aimablement fournies par Josefa Fernández López). Le stade 4 correspond au bourgeon en dormance.

Légende 1-11 (voir texte) : 1 = plante en pleine activité végétative ; 2 = 0-10 % des bourgeons en phase 2-3 ; 3 = 10-20 % des bourgeons en phase 2-3...11 = 90-100 % des bourgeons en phase 2-3.

Légende 1-5 (voir texte) : 1 = plante en pleine activité végétative ; 2 = bourgeon terminal est en phase 3, mais la plupart des bourgeons sont en phase 1 ; 3 = bourgeon terminal en phase 3 et les bourgeons restant en phase 2 et 3 ; 4 = bourgeon terminal en phase 3-4 ; 5 = bourgeon terminal en phase 4.

e)



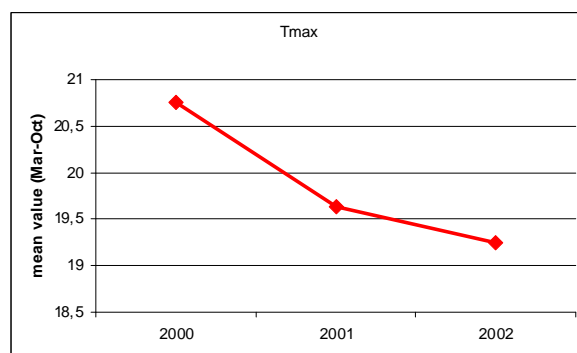
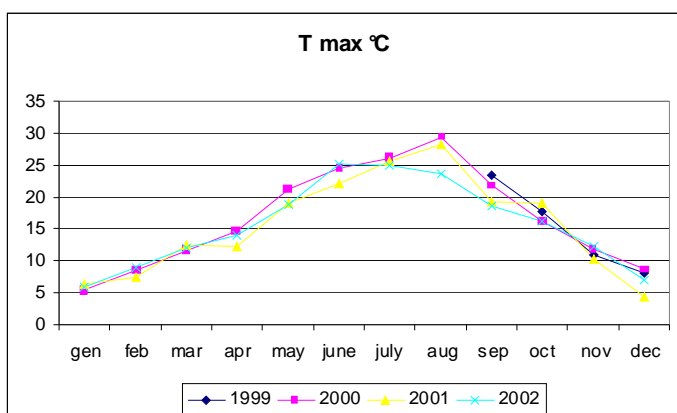
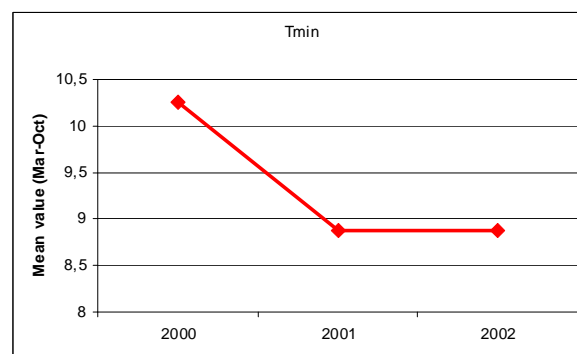
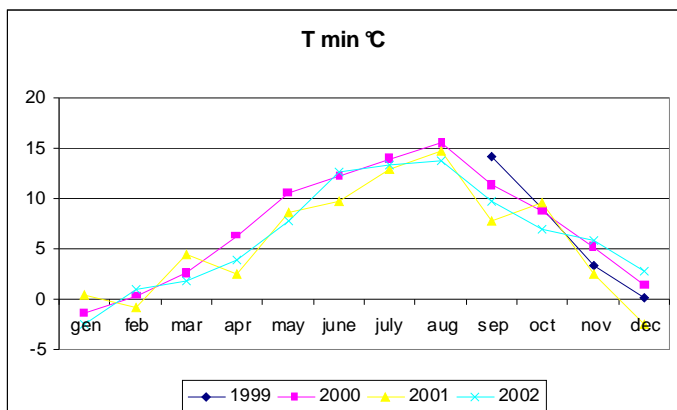
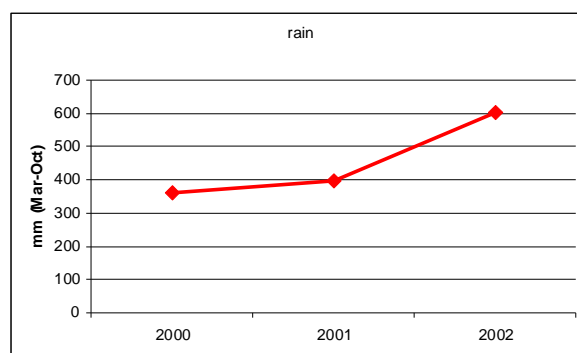
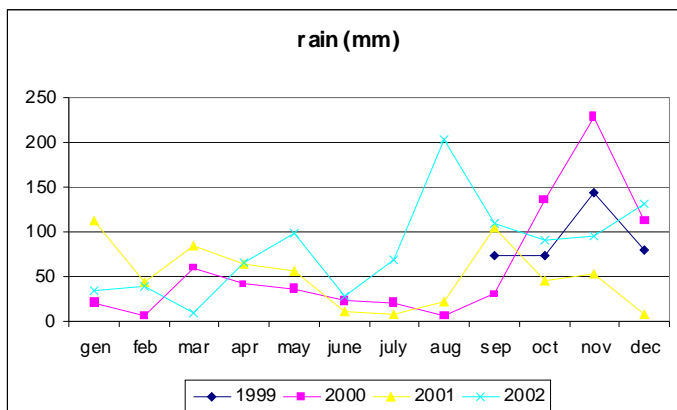
temps, certains locus pourraient être spécifiques d'une année (donc d'un stade de développement, d'un environnement particulier).

Deux types de mesures ont été faits pour le débourrement, une pour l'arrêt de croissance, une pour le carbone 13 et plusieurs pour la croissance :

- **Phénologie.** La date de débourrement a été notée à deux étapes du processus : 1) l'ouverture du premier bourgeon de la plante quelle que soit la position du bourgeon (ce caractère est appelé « bud », pour tous les caractères le nom est suivi des chiffres 2000, 2001 et 2002 correspondant aux trois années de mesures) et 2) l'ouverture de 70% des bourgeons de la plante (ce caractère est appelé « bud70 »). Dans les deux cas l'étape de bourgeon ouvert est celle indiquée par la flèche dans la figure 19c et la valeur utilisée dans l'analyse est le nombre de jours passés entre le premier janvier et la date d'ouverture. Ces deux mesures ont été faites d'une part parce qu'au stade juvénile les châtaigniers ont souvent plusieurs rejetons et, d'autre part, parce que la durée séparant l'ouverture du premier bourgeon de celles des autres bourgeons peut varier considérablement. L'arrêt de croissance (appelé « budset ») n'a été mesuré qu'en 2002. D'août à octobre l'état des bourgeons a été régulièrement observé et les mesures ont été faites à quatre dates fixées, une fois que 30% des plantes ont montré les signes de l'arrêt de croissance sur le bourgeon terminal. La figure 19d montre quatre stades de l'arrêt de croissance chez le châtaignier. Pendant trois des quatre observations un score de 1 à 5 a été donné à chaque individu, correspondant à des stades différents de développement des arbres. Les cinq stades sont décrits dans la figure 19d. Au contraire, lors de la première mesure, à cause de la variabilité phénotypique élevée pour ce caractère, les 5 stades n'étaient pas suffisants pour bien représenter la variabilité entre individus. Pour cette raison une échelle de 1 à 11 a été utilisée (figure 19d). La somme des points donnés durant les quatre journées a été utilisée pour l'analyse.
- **Discrimination isotopique du carbone 13.** Cette mesure a été faite à la fin de la saison végétative (en octobre) sur les feuilles. Un échantillon représentatif des feuilles de chaque individu a été récolté, séché et réduit en poudre. Trois sous-échantillons de chaque individu ont été analysés indépendamment et la valeur moyenne a été utilisée pour l'analyse. La composition isotopique a été mesurée à l'aide d'un spectromètre de masse (CF-IRMS, ISOCHROM II, VG-ISOTECH, Middlewich, Cheshire, UK), en utilisant le standard conventionnel PBD (Belemnite Pee Dee). La discrimination a été

Figure 20

Données pluviométriques et de température (maximum et minimum) au cours des trois années de mesure. A gauche moyennes mensuelles, à droite moyennes de mars à octobre.



calculée comme décrit dans le paragraphe C.1.3 et ce caractère, mesuré durant les trois années, est appelé Δ ou « delta ».

- **Croissance.** La croissance en hauteur et en diamètre a été mesurée en 2000, 2001 et 2002. Comme le montre la figure 19e, la hauteur a été mesurée sur la tige principale de chaque individu (cette mesure est appelée « height »). Le diamètre a été mesuré à la base de la tige principale et en 2001 et 2002 également à 10 cm du sol. Les deux mesures sont appelés « diameter » et « diamb ». La croissance (différence entre la mesure d'octobre et celle d'avril) et la valeur totale de la hauteur (« heighttot ») et du diamètre (« diametertot », « diambtot ») ont été utilisées pour les analyses. Le diamètre a été mesuré à deux niveaux parce que les rejets au pied des plantes empêchent une mesure précise du diamètre de base pour certains individus. Les analyses ont été faites sur les données de croissance et les valeurs totales afin de vérifier si les mêmes QTL sont détectés. En effet, Conner et al. (1998) ont montré que les observations faites sur plusieurs saisons de végétation peuvent détecter les QTL spécifiques d'un stade particulier de développement, alors que l'analyse faite avec les valeurs totales de croissance peut détecter les QTL contrôlant la vigueur de la plante.

L'effectif utilisé est de 150 individus, à l'exception de 6 cas particuliers (tableau 9). En effet, le débourrement a été mesuré sur 170 individus en 2000 (il faut rappeler qu'au départ il y avait 186 individus et qu'une trentaine des plantes sont mortes durant les trois années). Au contraire, quatre mesures de croissance ont été faites sur environ 130 individus car certains individus ont été exclus de l'étude si à l'automne la tige mesurée au printemps était endommagée ou si des rejets à la base de la tige principale empêchaient la mesure du diamètre. Il est important de noter que les plantes ont du faire face à des conditions environnementales très différentes au cours des trois années d'observation. Tout d'abord, de 1999 à 2001 les plantes étaient en pot à la pépinière et au cours de l'hiver 2001 elles ont été transférées sur le terrain. La figure 20 montre trois paramètres climatiques (la pluviométrie, la température maximale et minimale) relevés au cours des trois années et les moyennes de mars à octobre (les mois de la saison végétative) au cours de chaque année. Il est évident qu'au cours des trois années les plantes ont du faire face à des conditions climatiques très différentes. Comme nous le verrons par la suite, cela peut expliquer certains des résultats obtenus.

La figure 21 montre la distribution des valeurs phénotypiques de six mesures. Même si dans 7 cas (tableau 9) la distribution s'écartait de façon significative de la distribution normale, nous

Tableau 9

Description des caractères mesurés au cours des trois années et valeurs des statistiques de base.

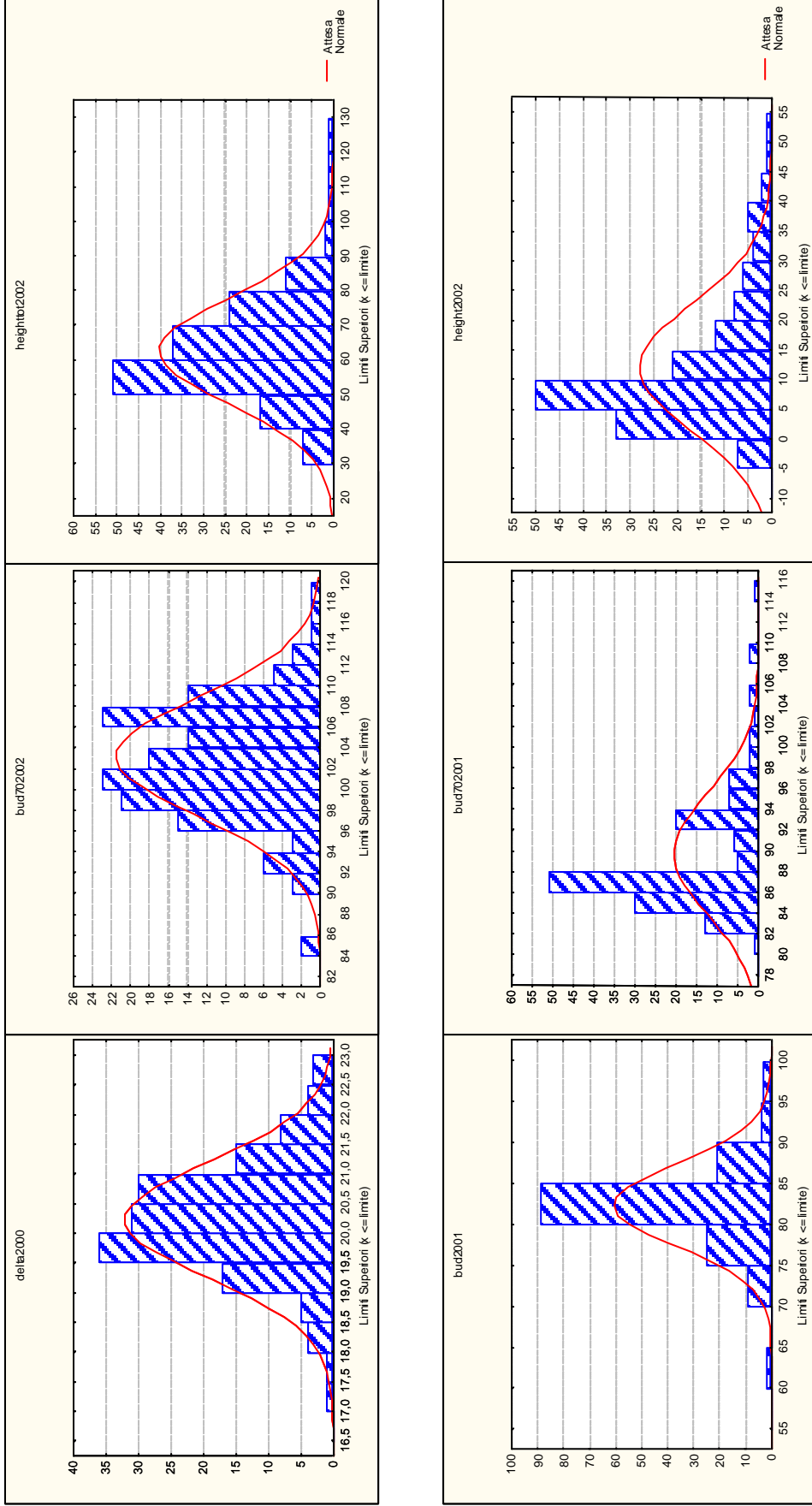
Nom du caractère (unités)	Définition du caractère	N	Moyenne (ET)	Min	Max	CV (%)	Normalité ⁽¹⁾	Effet bloc ⁽¹⁾	Effet année ⁽¹⁾
Δ2000 (%)	discrimination isotopique du carbone 13 2000	155	20.2 (0.96)	17.1	22.8	4.75	ns	ns	***(Δ)
Δ2001	discrimination isotopique du carbone 13 2001	152	20.1 (1.23)	15.8	22.4	6.12	ns	ns	
Δ2002	discrimination isotopique du carbone 13 2002	152	18.8 (0.94)	15.3	20.8	5.00	ns	ns	
bud2000 (Julian days)	débourement, premier bourgeon 2000	174	91 (4.94)	66	104	5.43	*	ns	*** (bud)
bud2001	débourement, premier bourgeon 2001	153	82 (5.02)	62	100	6.12	*	ns	
bud2002	débourement, premier bourgeon 2002	153	91 (5.74)	78	105	6.31	ns	ns	
bud702000	débourement, 70% des bourgeons 2000	172	102 (4.52)	89	113	4.43	*	ns	*** (bud70)
bud702001	débourement, 70% des bourgeons 2001	150	90 (5.78)	82	115	6.42	*	ns	
bud702002	débourement, 70% des bourgeons 2002	153	103 (5.66)	86	119	5.50	ns	ns	
budset2002 (score)	arrêt de croissance 2002	151	19 (4.99)	6	26	26.26	*	**	
diameter2000 (mm)	accroissement en diamètre au niveau du terrain 2000	137	4.2 (2.00)	0.6	12.7	47.62	ns	ns	*** (diameter)
diameter2001	accroissement en diamètre au niveau du terrain 2001	153	2.9 (1.47)	0.0	7.7	50.69	ns	ns	
diameter2002	accroissement en diamètre au niveau du terrain 2002	150	6.4 (3.35)	0.9	17.4	52.34	ns	***	
diameterb2001	accroissement en diamètre à 10 cm de hauteur 2001	153	2.4 (1.02)	0.3	5.5	42.50	ns	ns	
diameterb2002	accroissement en diamètre à 10 cm de hauteur 2002	151	4.4 (2.26)	0.8	13.4	51.36	ns	***	
height2000 (cm)	accroissement en hauteur 2000	135	16.6 (7.80)	3.8	53.1	46.99	ns	ns	*** (height)
height2001	accroissement en hauteur 2001	152	11.4 (7.04)	0.8	44.2	61.75	*	ns	
height2002	accroissement en hauteur 2002	150	12.0 (10.67)	0.0	51.8	88.92	*	*	
diametero2000 ⁽²⁾ (mm)	valeur du diamètre au niveau du terrain mesuré en octobre 2000	136	10.7 (2.12)	6.2	21.3	19.81	ns	ns	
diametero2001	valeur du diamètre au niveau du terrain mesuré en octobre 2001	153	13.4 (2.24)	7.4	21.3	16.72	ns	ns	
diametero2002	valeur du diamètre au niveau du terrain mesuré en octobre 2002	151	19.7 (4.21)	10.3	31.6	21.37	ns	*	
heighto2000 ⁽²⁾ (cm)	valeur de la hauteur mesurée en octobre 2000	133	36.2 (8.36)	17.9	72.1	23.09	ns	ns	
heighto2001	valeur de la hauteur mesurée en octobre 2001	153	49.7 (11.50)	26.2	103.7	23.14	ns	ns	
heighto2002	valeur de la hauteur mesurée en octobre 2002	152	62.6 (15.09)	34.0	125.8	24.11	ns	ns	
diametero2001 ⁽²⁾ (mm)	valeur du diamètre à 10 cm mesuré en octobre 2001	153	9.5 (1.81)	5.1	16.9	19.05	ns	ns	
diametero2002	valeur du diamètre à 10 cm mesuré en octobre 2002	152	13.9 (3.24)	6.5	24.5	23.31	ns	*	

N: nombre d'individus; ET: écart type; CV: coefficient de variation.

(1) Test Kolmogorov-Smirnov pour l'écart de la normalité. ns: non significatif, * significatif à 5%, ** à 1%, *** à 0,1%. L'effet année a été testé pour les caractères mesurés au cours des trois années, mais (2) il n'a pas été testé pour les valeurs totales de croissance à cause de leur autocorrélation.

Figure 21

Distribution des valeurs phénotypiques de six caractères mesurés au cours des trois années. Bud2001, bud702001 et height2002 s'écartent significativement d'une distribution normale.



avons toujours utilisé les données brutes pour les analyses. Même si théoriquement les méthodes d'analyse statistique sont basées sur l'hypothèse de distribution normale dans les classes génotypiques des QTL, il a été, en effet, montré que les résultats de la détection des QTL ne changent pas de façon significative après la transformation des données. Le tableau 9 montre les mesures effectuées et les valeurs des statistiques de base pour chaque mesure. Les caractères montrent une variabilité élevée dans la descendance et tous les caractères mesurés durant les trois années ont des moyennes significativement différentes d'une année à l'autre (testé par ANOVA). Etant donné les conditions variables des trois années de mesures et la variation ontogénique, il ne faut pas oublier qu'il s'agit d'arbres en phase juvénile, ce résultat est tout à fait compréhensible. Nous décrivons dans le prochain paragraphe la stratégie suivie pour tenir compte de cet effet. Six caractères mesurés sur le terrain en 2002 ont montré un effet bloc significatif (testé par ANOVA), par conséquent les données ont été ajustées à cet effet avant de faire la détection des QTL.

La figure 22 montre les corrélations phénotypiques entre les caractères. Une analyse de classification hiérarchique a été faite entre les valeurs phénotypiques de tous les caractères : la distance euclidienne a été calculée entre les caractères et le résultat est montré sous forme de dendrogramme dans la figure 22a. Ce type d'analyse nous permet de voir rapidement les caractères qui sont les plus corrélés entre eux. Plusieurs aspects peuvent être soulignés (figure 22b) :

- en général, les données de débourrement ne sont pas fortement corrélées avec les données de Δ et croissance. En revanche, une corrélation significative et négative est souvent observée entre Δ et les mesures de croissance. Cela suggère que les individus avec une efficacité d'utilisation faible (donc Δ plus élevé) montrent aussi une croissance plus faible et vice versa;
- les deux mesures de débourrement (« bud » et « bud70 ») sont toujours corrélées significativement durant la même année. Cela veut dire que globalement les plantes les plus précoces pour l'ouverture du premier bourgeon, sont aussi celles qui ouvrent plus précocement le reste des bourgeons. Les données des années 2000 et 2001 ne sont pas corrélées, alors que la corrélation est significative entre 2001 et 2002. La température significativement plus élevée en 2000, par rapport aux deux années suivantes, peut expliquer en partie ce résultat ;

- l'arrêt de croissance est faiblement corrélé avec le débourrement en 2002, en revanche, il est corrélé de façon significative avec les données de croissance de la même année, en accord avec la tendance selon laquelle un individu qui forme précocement le bourgeon d'hiver (il aura un score plus élevé, dans le système de mesure utilisé), montrera aussi une croissance inférieure ;
- les données de Δ sont corrélées entre les trois années et, comme nous l'avons déjà rappelé, elles sont négativement corrélées avec les données de croissance ;
- enfin, les données de croissance ne sont pratiquement pas corrélées entre les trois années, mais au cours de la même année la croissance en hauteur est toujours corrélée à la croissance en diamètre.

Nous discuterons plus en détail quelques-unes de ces corrélations à la lumière des résultats de la détection des QTL.

C.3.2. Le logiciel MultiQTL et la stratégie d'analyse des données.

Le logiciel MultiQTL (Britvin et al. 2001, <http://esti.haifa.ac.il/~poptheor>) a été utilisé pour la détection des QTL. Ce logiciel possède plusieurs avantages et au moment de l'analyse il était disponible gratuitement. Ce logiciel permet de tester plusieurs modèles statistiques de détection des QTL, de l'analyse par marqueur individuel, au CIM et dans sa version plus récente le MIM. Le logiciel possède trois modèles d'analyse intéressants : l'hypothèse de présence de deux QTL sur le même groupe de liaison peut être testée (Ronin et al., 1999) ; les données dérivant de différents environnements peuvent être analysées à la fois (Korol et al., 1998b) ; et enfin, le logiciel peut analyser les caractères corrélés selon une option « MultiTraits » qui devrait théoriquement permettre d'identifier les QTL montrant un effet pleiotrope (Korol et al., 2001). MultiQTL permet aussi de calculer le niveau de signification par test de permutation selon Churchill et Doerge (1994) et l'intervalle de confiance par « bootstrap » (Visscher et al. 1996). Le logiciel est très convivial et rapide pour les modèles à un seul QTL et un seul caractère. Dans ce dernier cas l'analyse CIM, le test de permutation et le « bootstrap » pour un caractère peuvent demander moins d'une heure.

Vu les conditions très variables des trois années de mesures et la variabilité attendue dérivant du développement ontogénique des plantes, l'analyse QTL a été réalisée de la façon suivante:

- pour les caractères montrant un effet année significatif (tableau 9) un modèle ANOVA a été testé afin d'intégrer cet effet dans l'analyse :

$$Y_{ijk} = \mu + \text{Année}_i + M_j + \text{Année}_i * M_j + \epsilon_{ijk},$$

Tableau 10

Phénologie : résultats de la détection des QTL.

Carte	Caractère	LG ⁽¹⁾	Position (cM) ⁽²⁾	LOD	Valeur P (signification %) ⁽³⁾	d ⁽⁴⁾	PEV (%) ⁽⁵⁾
femelle	bud2000	3	3.3	2.82	0.0009 (5%)	-2.837	8.1
		12	38.4	3.05	0.0000 (5%)	2.785	7.9
	bud2001	1	28.3	3.66	0.0000 (5%)	2.844	8.5
		7	43.7	1.93	0.0196 (20%)	-2.074	4.5
	bud2001	9	23.9	2.37	0.0047 (10%)	2.286	5.6
	bud2002	1	22.7	4.61	0.0000 (5%)	4.095	12.2
		9	23.9	3.05	0.0009 (5%)	2.995	6.8
		12	19.9	3.00	0.0009 (5%)	2.903	6.3
	bud702001	4	22.1	3.48	0.0009 (5%)	-3.727	10.3
		6	12.8	2.92	0.0000 (5%)	-3.435	8.8
	bud702002	1	27.1	3.48	0.0000 (5%)	2.998	6.7
		2	89.3	3.62	0.0000 (5%)	3.436	8.5
		3	48.2	3.91	0.0000 (5%)	3.039	7.1
		5	25.5	3.41	0.0000 (5%)	-2.988	6.7
		9	23.9	6.32	0.0000 (5%)	3.889	11.6
		12	19.9	1.99	0.0121 (20%)	2.046	3.2
	budset2002	2	64.7	3.12	0.0019 (5%)	2.822	8.9
		8	33.2	2.18	0.0131 (15%)	2.335	6.1
	bud	1	20.6	5.42	0.0000 (5%)	2.699	11.8
		4	33.5	2.50	0.0084 (10%)	-1.573	4.2
		9	23.9	4.50	0.0000 (5%)	2.184	8.2
		12	40.5	5.30	0.0000 (5%)	2.510	10.7
	bud70	4	19.2	3.31	0.0009 (5%)	-2.002	7.1
9		23.9	2.35	0.0084 (20%)	1.736	5.3	
12		24.1	3.61	0.0000 (5%)	2.119	8.0	
mâle	bud2002	6	45.5	2.36	0.0084 (10%)	2.849	6.3
		11	9.6	2.62	0.0056 (15%)	3.214	7.8
	bud702001	6	21.6	2.23	0.0121 (15%)	3.057	6.9
		9	28.6	2.49	0.0028 (5%)	-3.228	8.6
	bud702002	11	8.0	2.41	0.0047 (10%)	3.235	8.0
	budset2002	6	50.8	3.44	0.0009 (5%)	2.916	9.6
		9	21.7	6.64	0.0000 (5%)	-3.870	17.1
	bud	3	42.1	2.18	0.0140 (20%)	1.848	5.8
		6	41.0	3.39	0.0000 (5%)	2.165	8.1
	bud70	2	45.7	2.08	0.0178 (15%)	1.896	6.3

(1) LG : groupe de liaison.

(2) Position : distance en cM (Kosambi) de l'extrémité du groupe de liaison.

(3) Valeur P au niveau du groupe de liaison (1000 permutations) et erreur de première espèce au niveau de génome (%).

(4) d: effet de substitution allélique.

(5) PEV: proportion de la variance phénotypique expliquée par le QTL.

Tableau 11
Croissance : résultats de la détection des QTL.

Carte	Caractère	LG ⁽¹⁾	Position (cM) ⁽²⁾	LOD	Valeur P (signification %) ⁽³⁾	d ⁽⁴⁾	PEV (%) ⁽⁵⁾
femelle	height2000	2	93.1	2.87	0.0037 (5%)	4.867	9.6
		7	0.0	3.82	0.0000 (5%)	-5.294	11.8
		10	10.7	2.72	0.0000 (5%)	-4.419	8.0
		11	15.4	2.20	0.0065 (15%)	3.919	6.2
	height2001	1	0.0	2.74	0.0056 (5%)	-3.729	7.2
	height2002	8	78.4	3.33	0.0019 (5%)	-8.356	17.0
	heighttot2000	10	17.5	1.94	0.0093 (20%)	-4.344	7.0
	heighttot2001	1	3.4	2.83	0.0037 (5%)	-6.667	8.3
	heighttot2002	1	6.1	4.93	0.0000 (5%)	-10.290	11.5
		10	23.4	3.18	0.0000 (5%)	-7.947	7.1
	diameter2001	11	6.5	2.01	0.0056 (15%)	-0.784	7.1
	diamb2001	11	4.5	2.21	0.0056 (15%)	-0.540	7.2
	diameter2002	7	48.9	2.48	0.0047 (5%)	1.684	7.7
	diamb2002	5	0.0	2.75	0.0047 (10%)	1.171	7.8
	diametertot2001	10	23.4	2.35	0.0028 (5%)	-1.090	5.9
	diambtot2001	10	23.4	2.05	0.0093 (20%)	-0.881	5.9
	diambtot2002	6	23.5	2.2	0.0121 (15%)	-1.589	6.4
		10	23.4	1.95	0.0110 (20%)	-1.458	5.5
	height	1	33.8	3.09	0.0009 (5%)	-2.876	7.3
		8	74.3	2.56	0.0065 (10%)	-3.490	10.6
10		23.4	2.49	0.0009 (5%)	-2.506	5.5	
mâle	height2002	11	17.8	3.46	0.0000 (5%)	-7.020	11.5
	diameter2001	7	40.2	2.78	0.0019 (5%)	-0.829	8.1
	diamb2001	7	39.3	2.33	0.0028 (5%)	-0.503	6.3
	diameter2002	6	15.6	2.08	0.0121 (15%)	-1.879	9.8
		9	32.6	2.14	0.0140 (20%)	1.468	5.9
	diametertot2001	3	33.7	2.43	0.0047 (10%)	-1.201	7.1
	diambtot2001	3	33.7	3.57	0.0000 (5%)	-1.166	10.3

Tableau 12
Discrimination isotopique du carbone 13 : résultats de la détection des QTL.

Carte	Caractère	LG ⁽¹⁾	Position (cM) ⁽²⁾	LOD	Valeur P (signification %) ⁽³⁾	d ⁽⁴⁾	PEV (%) ⁽⁵⁾
femelle	Δ 2000	3	18.6	3.30	0.0019 (5%)	-0.531	7.3
		4	0.0	4.62	0.0000 (5%)	-0.615	10.1
		7	27.6	2.16	0.0093 (10%)	0.415	4.6
		12	16.5	2.10	0.0131 (20%)	0.425	5.8
	Δ 2001	3	53.6	5.40	0.0000 (5%)	-0.894	13.2
		6	23.5	3.37	0.0009 (5%)	0.667	7.3
		7	12.7	2.41	0.0075 (10%)	0.560	5.1
	Δ 2002	8	54.7	3.39	0.0019 (5%)	0.739	8.8
		1	5.1	4.51	0.0000 (5%)	0.661	12.2
	Δ	1	23.5	3.25	0.0009 (5%)	0.497	7.9
		3	57.7	3.80	0.0000 (5%)	-0.498	8.1
		4	0.0	2.08	0.0140 (15%)	-0.359	4.1
		6	55.4	2.92	0.0037 (5%)	0.430	6.0
		7	16.3	2.18	0.0140 (15%)	0.366	4.5
		mâle	Δ 2000	3	39.7	2.40	0.0019 (5%)
4	25.3			2.11	0.0093 (10%)	-0.432	4.9
10	11.6			4.19	0.0000 (5%)	0.631	10.4

où Y_{ijk} est la valeur phénotypique de l'individu k , portant l'allèle j , mesuré l'année i , μ est la moyenne globale, $Année_i$ l'effet de l'année i , M_j l'effet du marqueur j , $Année_i * M_j$ l'effet d'interaction année marqueur et ϵ_{ijk} la variabilité résiduelle. Les données ont été ajustées à cet effet.

- La détection des QTL a été réalisée en utilisant les données brutes année par année et les données ajustées à l'effet année (les QTL détectés dans ce dernier cas sont appelés « bud », « bud70 », « Δ », « height » « diameter »). Cette stratégie a été choisie afin de pouvoir détecter et les QTL spécifiques d'une année et les QTL stables. En effet, les données ajustées devraient nous permettre de nous affranchir des conditions environnementales variables.

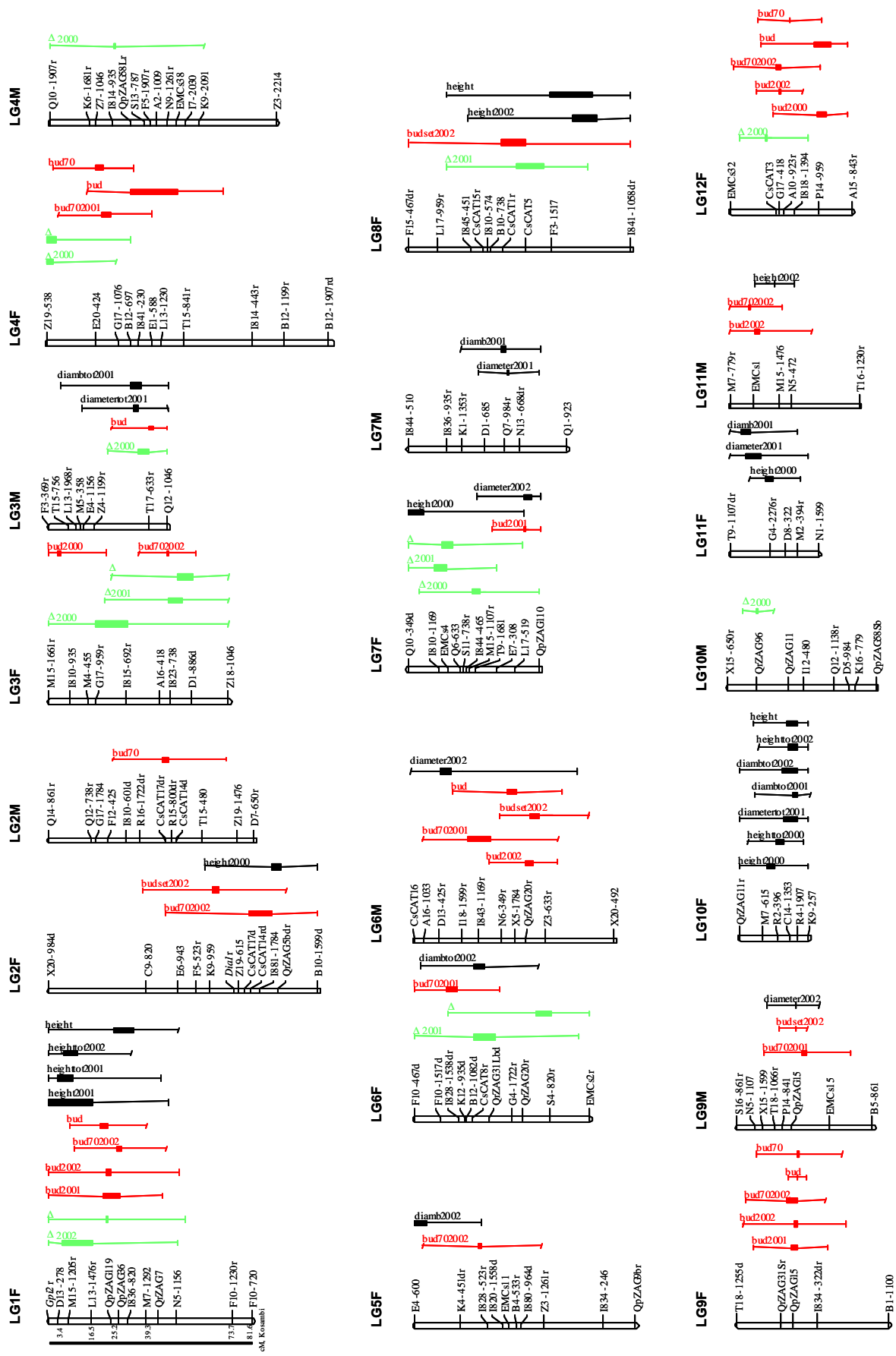
Pour chaque caractère la stratégie d'analyse choisie est la suivante :

- le CIM a été utilisé pour la détection des QTL car il permet une détection plus efficace par rapport aux autres modèles statistiques ;
- deux types de QTL ont été retenus, les QTL « significatifs », pour lesquels le seuil de détection est fixé à 5% au niveau de génome et les QTL « putatifs », pour lesquels le même seuil est fixé à 20%. Cette stratégie a été souvent suggérée car elle permet de retenir aussi les QTL, pour lesquels le risque de se tromper est plus élevé, mais qui pourraient, par exemple, se révéler plus significatifs dans une autre expérience (Lander et Kruglyak, 1995).
- l'intervalle de confiance (à 95%), l'effet de substitution allélique et la variance phénotypique expliquée par le QTL ont été calculés comme décrit dans le paragraphe C.2.2. Il est très important de noter que, sous la stratégie de double pseudo test-cross, la direction (« + » ou « - ») de l'effet de substitution allélique peut être comparée au sein du même groupe de liaison mais il n'est pas possible de comparer la direction des effets entre QTL qui se trouvent sur des groupes de liaison différents car la procédure de construction de la carte génétique ne permet pas de savoir lequel des deux chromosomes homologues présents dans un génome diploïde a été considéré à chaque fois.

C.3.3. Résultats de la détection des QTL.

En utilisant la stratégie décrite ci-dessus, 80 QTL, 35 impliqués dans la phénologie, 28 dans la croissance et 17 dans la discrimination isotopique du carbone 13, ont été détectés (tableau 10, 11 et 12, et figure 23). Soixante-deux QTL proviennent des données étudiées année par année et 18 des données ajustées à l'effet année. Cinquante QTL sont significatifs ($\alpha_g = 5\%$) et 30

Figure 23 QTL détectés chez le châtaignier. Les QTL contrôlant la discrimination isotopique du carbone 13 sont représentés en vert, en rouge les QTL impliqués dans la phénologie et en noir les QTL impliqués dans la croissance. Pour chaque QTL nous avons représenté l'intervalle de confiance et la position la plus probable du QTL (rectangle en gras, position du pic de LOD donné par l'analyse CIM et après 1000 ré-échantillonnages réalisés par bootstrap).



putatifs ($\alpha_g = 20\%$). Cependant parmi les 30 QTL putatifs, 25 sont « colocalisés » avec d'autres QTL significatifs pour le même caractère, suggérant qu'il ne s'agit pas de faux positifs. Comme le montre la figure 23, plusieurs QTL, détectés dans des années différentes ou bien en utilisant les données ajustées à l'effet année, « colocalisent ». Il s'agit probablement du même QTL détecté plusieurs fois. Considérant que les QTL, qui se trouvent dans la même région du groupe de liaison et qui montrent le même signe de l'effet de substitution allélique, peuvent correspondre au même QTL, alors nous avons détecté 14 QTL impliqués dans le débourrement, 12 dans la croissance et 8 dans la discrimination isotopique du carbone 13. Toutefois, les intervalles de confiance très larges empêchent de définir avec certitude le concept de « QTL unique » quand plusieurs QTL se positionnent dans la même région. Il est aussi probable que deux QTL impliqués dans le même caractère soient présents sur le même groupe de liaison. Cette possibilité a été testée en utilisant l'option disponible dans le logiciel MultiQTL mais aucun test significatif n'a été mis en évidence. Sur les douze groupes de liaison, l'hypothèse de 2 QTL n'a été jamais significativement plus probable que celle d'un seul QTL. Ce résultat ne veut pas dire qu'il n'y a pas 2 QTL contrôlant le même caractère sur le même groupe de liaison mais que, probablement, cette expérience n'est pas suffisamment puissante (effectif faible) pour tester cette hypothèse.

C.3.3.1. Phénologie des bourgeons.

Les 35 QTL détectés pour la phénologie (tableau 10), 31 pour le débourrement et 4 pour l'arrêt de croissance, sont distribués sur 11 groupes de liaison et expliquent entre 4,2 et 17,1% de la variance phénotypique. Les QTL impliqués dans « bud » et « bud70 » « colocalisent » sur les groupes 1F, 4F, 6M, 9F, 11M et 12M et la direction de l'effet de substitution allélique est la même au sein de chaque « colocalisation ». Vu la corrélation phénotypique significative entre les deux mesures de débourrement, il est probable qu'il s'agisse des mêmes QTL. Cependant, même si les QTL impliqués dans « bud » et « bud70 » « colocalisent » dans la plupart des cas, il est possible de noter quelques différences. La plus intéressante concerne le groupe de liaison 3 de la carte femelle. Sur ce groupe deux QTL impliqués dans le débourrement ont été détectés en 2000 (« bud2000 ») et 2002 (« bud702002 »). Leur position est bien distincte et les effets de substitution allélique sont de signe opposé. Il est donc probable qu'il s'agisse de deux QTL différents qui contrôlent deux stades du processus de débourrement. D'autres cas concernent les groupes 2F et 5F (« bud70 ») et le groupe 7F (« bud »). Cette hypothèse de QTL agissant sur des stades différents du débourrement ne peut être vérifiée pour l'instant. En général, la détection des QTL faite en utilisant plusieurs méthodes ou paramètres différents

pour mesurer un caractère constitue une bonne stratégie de validation de la localisation des QTL, comme le montrent les résultats de la détection des QTL impliqués dans le débourrement. Cela sera encore plus évident pour les QTL de croissance.

Parmi les 4 QTL détectés pour l'arrêt de croissance, 3 « colocalisent » avec des QTL impliqués dans le débourrement. Comme nous le verrons plus loin ces « colocalisations » laissent penser que les deux processus partagent les mêmes voies biochimiques.

C.3.3.2. Croissance.

En considérant tous les mesures effectuées, 28 QTL ont été détectés pour la croissance en diamètre et en hauteur (tableau 11). Etant donné que 16 paramètres sur plusieurs années ont été utilisés pour la détection des QTL, le nombre des QTL par caractère mesuré est très faible, mais en accord avec ce qui a été trouvé dans d'autres exemples de recherche de QTL impliqués dans la croissance (tableau 8). Les QTL sont distribués sur 10 groupes de liaison et ils expliquent entre 5,5 et 17,0% de la variance phénotypique. Un cluster de QTL contrôlant la croissance en hauteur est évident sur le groupe 1 femelle. Le groupe 10, toujours de la carte femelle, contient plusieurs QTL de croissance en hauteur et diamètre détectés sur les trois années. Les QTL de croissance et ceux correspondant à la hauteur totale et au diamètre à la fin de la saison végétative ne sont « colocalisés » que sur ces deux groupes de liaison. Peu de « colocalisations » entre les QTL contrôlant la croissance en diamètre et en hauteur ont été détectées.

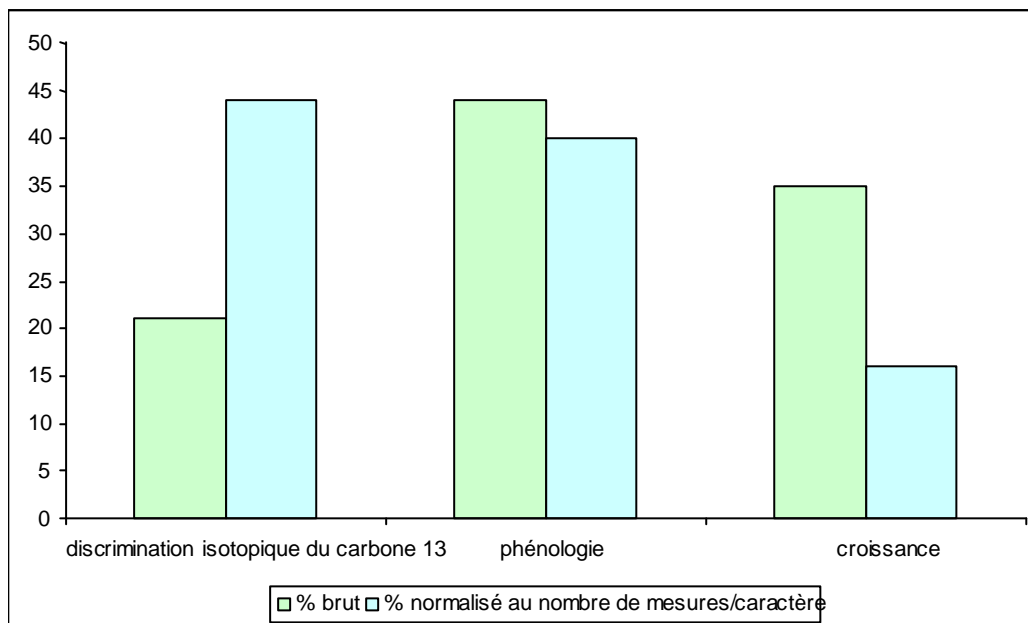
C.3.3.3. Discrimination isotopique du carbone 13.

Au cours des trois années, 17 QTL ont été détectés pour la discrimination isotopique du carbone 13 (tableau 12) sur 8 groupes de liaison différents. Ils expliquent entre 4,1 et 13,2% de la variance phénotypique. La plupart des QTL ont été détectés en 2000 et 2001, un seul QTL a été détecté en 2002.

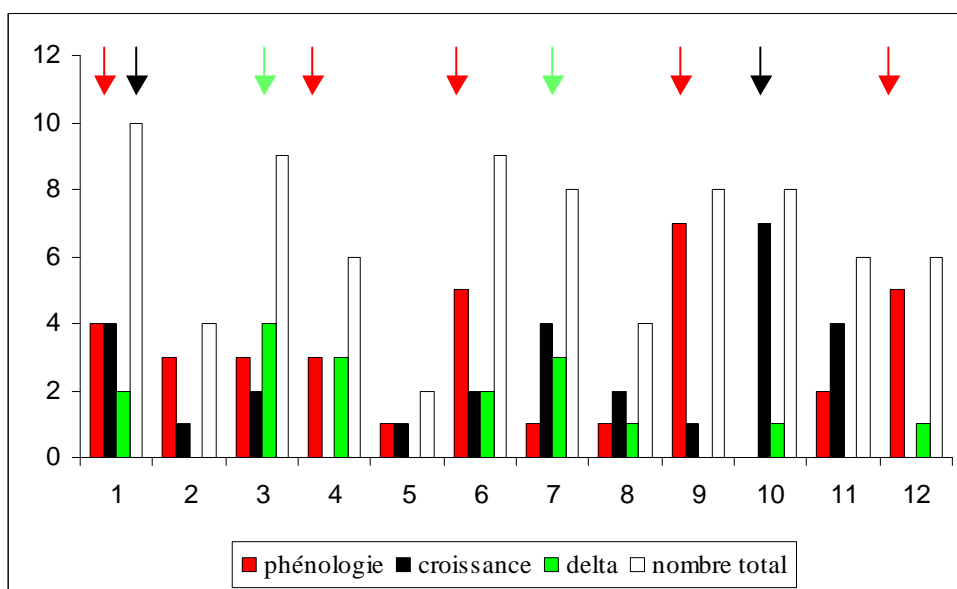
Comme le montre la figure 24a, le nombre de QTL détectés par type de caractère (phénologie, discrimination isotopique du carbone 13 et croissance) est très faible pour la croissance. En effet, sur le total des QTL identifiés, 21% sont des QTL impliqués dans la discrimination isotopique, 44% dans la phénologie et 35% dans la croissance. Néanmoins, les 17 QTL impliqués dans la discrimination isotopique ont été détectés grâce aux mesures sur les trois années et l'analyse des données ajustées à l'effet année. Par conséquent, nous avons une moyenne de 4,25 QTL (17/4) par mesure. Le même raisonnement pour la phénologie et la

Figure 24

a) Proportion des QTL détectés pour chaque caractère.



b) Distribution des QTL sur les douze groupes de liaison.



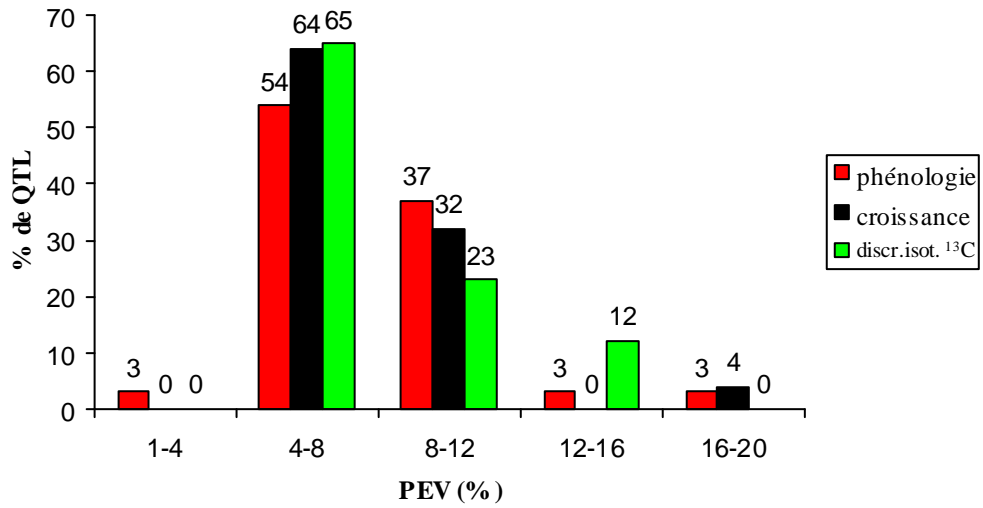
↓ QTL stables pour le débourrement.

↓ QTL stables pour la croissance.

↓ QTL stables pour la discrimination isotopique du carbone 13.

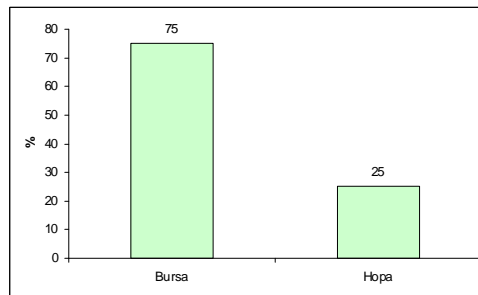
Figure 24

c) Variance phénotypique expliquée par les QTL.

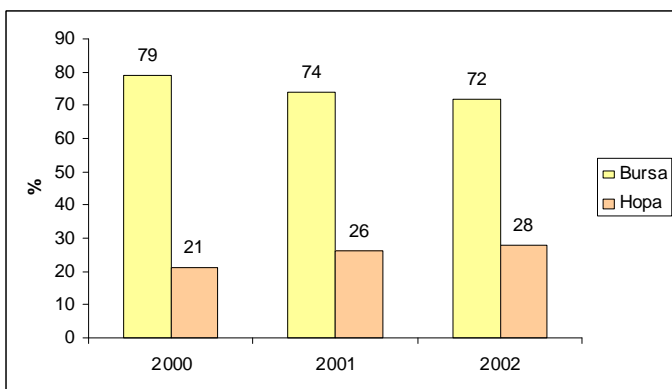


d) Proportion de QTL détectés sur les deux cartes parentales.

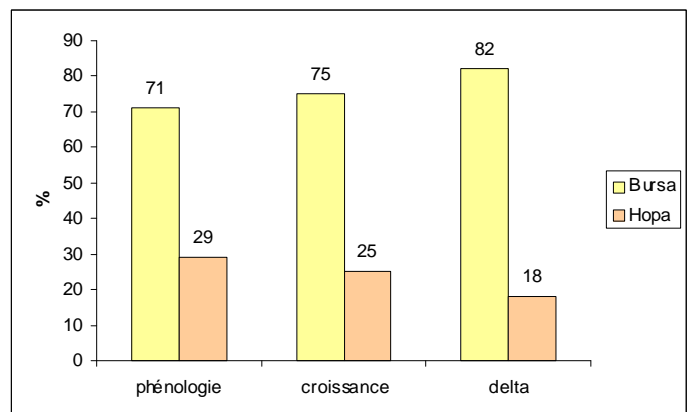
Nombre total de QTL



Au cours des trois années.



Pour les trois caractères.



croissance met en évidence que cette dernière semble un caractère plutôt difficile à décomposer en facteurs mendéliens. Même si les estimations des héritabilités sont très variables pour les caractères de croissance, il s'agit souvent de valeurs très faibles. De plus, les facteurs biologiques qui interviennent dans la croissance en hauteur ou en diamètre sont nombreux et variables selon les environnements. Tout cela peut expliquer le faible nombre de QTL détectés par caractère dans le cas de la croissance (figure 24a).

L'analyse des données ajustées à l'effet année permet d'identifier les mêmes QTL détectés grâce à l'analyse année par année, à l'exception de deux QTL : « bud70 » et « bud » sur les groupes 2 et 3 de la carte mâle. Il est intéressant de noter que dans les deux cas un QTL contrôlant le même caractère est présent sur l'homologue femelle. Il est probable que la ségrégation des allèles du parent male n'ait pas donné lieu à un effet détectable car le bruit de fond de la ségrégation des allèles du parent femelle était trop important. Cette hypothèse semble être confirmée par le fait qu'en effet deux QTL significatifs à 40% au niveau du génome sont identifiables dans les deux cas (résultats non présentés).

C.3.3.4. Distribution sur le génome et stabilité des QTL.

La figure 24b montre la distribution des QTL sur les 12 groupes de liaison du châtaignier. Tous les groupes de liaison contiennent au moins un QTL. Ce résultat est plutôt attendu vu la complexité des caractères. Les groupes de liaison contenant le plus grand nombre de QTL contrôlant le même caractère sont ceux qui pourraient contenir les QTL stables. Nous considérerons un QTL stable quand il a été détecté au moins pendant deux années différentes et/ou confirmé par l'analyse sur les données ajustées. Donc dans les figures 23 et 24b, des QTL stables peuvent être identifiés pour les trois caractères étudiés. Le groupes 1F, 4F, 6M, 9F et 12F possèdent des QTL stables pour le débourrement, les groupes 1F et 10F pour la croissance, et les groupes 3F et 7F pour la discrimination isotopique du carbone 13. L'identification de QTL stables permet d'identifier les régions génomiques qui contribuent, dans plusieurs conditions environnementales, à la variance phénotypique du caractère. Le groupe 9F contient, par exemple, 5 QTL impliqués dans le débourrement détectés dans deux années différentes et l'analyse des données ajustées à l'effet année permet une localisation plus précise de ce QTL. Le groupe homologue de la carte mâle contient aussi deux QTL un pour le débourrement et un autre pour l'arrêt de croissance. Il est donc probable que dans cette région il y ait des gènes impliqués dans la régulation de base de l'activité des bourgeons.

La stratégie de détection des QTL au cours de plusieurs années ou dans des environnements différents permet d'une part de vérifier les QTL détectés (Sewell et Neale, 2000), d'autre part

de distinguer deux types de QTL : ceux qui sont toujours exprimés et ceux qui semblent spécifiques d'un environnement ou d'un stade de développement particulier. Dans le cas du châtaignier, il semble que le débourrement soit un caractère contrôlé par plus de QTL stables que la croissance et la discrimination isotopique du carbone 13. Jermstad et al. (2001a ; 2003) ont démontré chez *Pseudotsuga menziesii* que les QTL contrôlant le débourrement sont stables au cours des années et entre deux sites différents. Le résultat obtenu sur châtaignier semble être en accord avec ces conclusions. Il ne faut pas oublier que le débourrement est un caractère à héritabilité modérée ou élevée et comme nous avons pu le voir la puissance de détection des QTL augmente avec l'héritabilité. Plus un caractère est héritable, plus il y aura des chances de détecter des QTL et des QTL stables car les effets de l'environnement sont moins importants. Au contraire, la plupart des QTL impliqués dans la croissance et la discrimination isotopique du carbone 13 semblent spécifiques d'une année. En ce qui concerne la croissance, il semble que le contrôle génétique de la croissance juvénile des arbres se modifie au cours des années (Kremer 1992), d'ailleurs, plusieurs auteurs ont montré que les QTL impliqués dans la croissance varient d'un stade de développement à l'autre (Plomion et al., 1996 ; Emebiri et al., 1998 ; Tsarouhas et al., 2002). Ces QTL, qui interviennent dans un stade de développement particulier et non dans un autre, sont appelés épigénétiques (Wu et al., 2002). Il faut aussi ajouter que la croissance et la discrimination isotopique du carbone 13 sont deux caractères très complexes, souvent à héritabilité faible pour lesquels les arbres montrent une plasticité phénotypique élevée, ce qui pourrait expliquer l'implication de QTL différents selon les conditions environnementales.

C.3.3.5. Variance phénotypique expliquée par les QTL.

Une première question à laquelle nous essaierons de répondre est : quelle est la contribution des QTL détectés dans cette étude à la variation du caractère phénotypique ? Pour les trois caractères étudiés des estimations très variables des héritabilités ont été calculées chez les arbres forestiers : de 0,5 à 0,9 pour le débourrement (Bradshaw et Stettler, 1995 ; Jermstad et al., 2001a), de 0,20 à 0,80 pour les paramètres de croissance (Bradshaw et Stettler, 1995 ; Byrne et al., 1997) et de 0,20 à 0,50 pour la discrimination isotopique (Brendel et al., 2002). Si nous considérons les mesures qui ont permis de détecter le plus grand nombre de QTL expliquant la proportion la plus élevée de variance phénotypique (bud702002 femelle, PEV=43,8%, height2000 femelle PEV=35,6% et delta2001 femelle PEV=34,4%), alors les QTL détectés expliquent de 49 à 88 % de la variance génétique du débourrement, de 45 à 100% de celle de la croissance et de 69 à 100% de celle du delta. Chez le chêne, les valeurs

d'héritabilité varient de 0,15 à 0,51 pour le débourrement, de 0,32 à 0,80 pour la discrimination isotopique du carbone 13 et de 0,04 à 0,23 pour la croissance en hauteur (Saintagne, 2003). Les estimations d'héritabilité disponibles chez le châtaignier sont de 0,48 à 0,89 pour la croissance en hauteur (Pliura et Eriksson, 2002) et 0,31 pour la discrimination isotopique (Lauteri et al., 2004). La valeur pour le delta est proche des estimations obtenues pour d'autres espèces forestières, en revanche les estimations pour la croissance en hauteur semblent plus élevées que celles observées dans d'autres exemples. Ces résultats semblent indiquer une efficacité élevée de détection des QTL. Néanmoins, le calcul contient des biais importants : tout d'abord pour les trois expériences considérées la PEV totale est simplement la somme des PEV de chaque QTL, donc elle risque d'être surestimée. De plus, les valeurs de variance phénotypique expliquée par les QTL sont surestimées à cause de l'effectif plutôt faible de cette expérience. Zeng et al., (1999) ont montré que cette surestimation est d'autant plus importante en cas d'héritabilité faible. Il est donc probable que le calcul soit vraisemblable pour le débourrement, mais fortement surestimé pour la discrimination isotopique et probablement pour la croissance.

Le biais dans l'estimation de l'effet des QTL est encore plus évident si nous considérons le même QTL détecté plusieurs fois. Les 5 QTL impliqués dans le débourrement détectés à la même position sur le groupe 9 femelle ont une valeur de PEV qui varie de 5,3 à 11,6%. Cette variation va de 7,2 à 11,5% pour les 4 QTL de la hauteur sur le groupe 1 femelle. Il est clair que seul un effectif important permettra une estimation plus précise de ces valeurs et donc de la proportion de QTL, contrôlant le caractère, qui ont été détectés. Il ne faut pas oublier que l'effet d'un QTL peut également changer d'une condition expérimentale à l'autre car un même gène peut être plus ou moins impliqué dans la détermination du phénotype d'un caractère complexe dans des environnements différents. La comparaison des valeurs des effets des QTL entre le chêne et le châtaignier nous permettra de vérifier si les QTL conservés ont le même effet relatif.

La deuxième question à laquelle nous essaierons de répondre est la suivante : y a-t-il plus de QTL à effet fort ou à effet faible ? Malgré le biais sur les estimations des effets, en considérant les valeurs pour tous les caractères et tous les QTL nous pouvons avoir une idée de la distribution relative des effets des QTL comme le montre la figure 24c. Cette distribution a une forme de L, comme cela a déjà été montré par plusieurs études de recherche de QTL (Tanksley, 1993 ; Kearsey et Farquhar, 1998). La plupart des QTL détectés montrent un effet faible ou modéré, et peu de QTL à effet fort sont identifiés. La figure 24c indique

aussi que la détection des QTL faite avec un effectif de 150-200 individus ne permet pas de détecter les QTL à effet très faible. Cette distribution des valeurs des effets semble en accord avec la théorie de Orr (1998). Les caractères adaptatifs seraient contrôlés par plusieurs facteurs à effet faible ou modéré, mais des gènes à effet fort semblent toujours présents. Malgré les limites liées à l'analyse QTL, particulièrement l'impossibilité de détecter tous les QTL contrôlant un caractère avec une seule expérience et la surestimation des effets, cette distribution en forme de L a été observée dans la majorité des cas et il est donc probable qu'il y ait une raison biologique.

C.3.3.6. La plupart des QTL ont été détectés chez le parent femelle : adaptation de cet écotype à la sécheresse ?

La figure 24d montre la proportion de QTL détectés sur la carte femelle et mâle. Il est évident qu'il existe une différence significative : 75% des QTL ont été détectés chez le parent femelle et 25 % chez le parent mâle. Cette différence est toujours significative pour les trois caractères et pendant les trois années (figure 24d). Bien qu'il existe une différence en saturation entre la carte femelle (84%) et la carte mâle (75%), cela n'est pas suffisant pour expliquer la différence de QTL détectés. Comme nous avons pu le voir dans la partie B.2., les deux parents de la famille utilisée pour la détection des QTL sont originaires de deux régions turques contrastées du point de vue climatique. En particulier, le parent femelle, Bursa, vient de la région méditerranéenne et appartient à une population adaptée à des conditions de sécheresse. La détection des QTL au cours des trois années a été faite dans un environnement typiquement méditerranéen. La pluviométrie annuelle du site Bursa (700 mm par année, Lauteri et al., 1997) est très proche de celle relevée en Italie en 2000, 2001 et 2002 (figure 20). Il est donc probable que les allèles du parent femelle aient contribué majoritairement à la variance phénotypique de la famille de plein-frères dans un environnement similaire à celui dont il est originaire. Il s'agit d'une hypothèse qu'il faudrait vérifier, par exemple, en plaçant une copie de la même famille dans un environnement humide proche de l'environnement d'origine du parent mâle, Hopa. Dans ce dernier cas et en supposant que le parent mâle soit également hétérozygote aux QTL contrôlant les caractères impliqués, nous devrions observer la plupart des QTL sur le génome paternel. Récemment, Verhoeven et al. (2004) ont utilisé une stratégie de recherche de QTL pour aborder le problème des bases génétiques de la différenciation adaptative entre populations. Dans leur étude, deux écotypes d'orge sauvage adaptés à deux environnements différents ont été utilisés pour établir un pedigree de cartographie. Des lignées dérivées de ce croisement ont été utilisées afin de détecter les QTL

impliqués dans des caractères liés à la fitness. La détection des QTL a été faite dans les deux environnements d'origine des deux écotypes. Les résultats ont montré que : 1) les lignées dérivant du croisement sont moins performantes que les écotypes d'origine dans les deux environnements ; 2) les QTL impliqués dans les caractères liés à la fitness, concordant dans les deux essais, montrent une valeur différente de leur effet dans les deux environnements, mais ils n'ont pas un effet opposé entre les deux sites ; 3) les allèles à effet favorable sur la fitness montrent un effet majeur dans le site d'origine de l'écotype duquel ils dérivent ; 4) quelques QTL ont montré des effets dans un seul de deux environnements. Selon les auteurs, il semble que l'action combinée de plusieurs allèles à différents locus soit le mécanisme responsable de la différenciation adaptative. Même si l'étude réalisée sur le châtaignier ne peut pas être complètement comparée au travail de Verhoeven et al. (2004), leurs conclusions sont très intéressantes car elles démontrent l'effet favorable des allèles du génotype adapté à un environnement donné dans le site d'origine. Vu les limites de la détection des QTL dans une famille F_1 (figure18), nous pouvons supposer que dans le cas du châtaignier les allèles favorables du parent femelle dans un environnement méditerranéen auraient masqué les allèles du parent mâle moins favorables dans cet environnement. Une observation intéressante à ce propos est la suivante : 35% (31/89) des QTL détectés par CIM chez le parent femelle n'ont pas été retenus car ils n'étaient pas significatifs à 20% au niveau de génome ; cette proportion est plus élevée chez le parent mâle (49%, 17/35), ce qui pourrait indiquer chez ce parent une proportion plus élevée d'allèles aux QTL qui n'arrivent pas à avoir un effet significatif dans la descendance.

C.3.3.7. « Colocalisations » des QTL : pour le même caractère et entre caractères.

Plusieurs groupes de liaison contiennent des QTL impliqués dans les trois caractères qui semblent localisés dans la même région (Figure23, groupe 1 femelle). Il est clair que les intervalles de confiance très larges et la présence de QTL sur tous les groupes de liaison donnent lieu à des « colocalisations » simplement par chance. Néanmoins, la relation biologique entre certains caractères (par exemple, le débourrement et l'arrêt de croissance) et les corrélations phénotypiques observées entre certains caractères peuvent mener à l'identification de « colocalisations » intéressantes du point de vue biologique, dues probablement à la pleiotropie des gènes impliqués dans ces caractères corrélés.

Un premier type de « colocalisations » intéressantes sont celles relatives aux QTL détectés en utilisant soit des mesures différentes du même caractère (par exemple les deux mesures de débourrement) soit des mesures de caractères liés du point de vue biologique (par exemple,

débourrement et arrêt de croissance). Nous avons déjà évoqué la coïncidence entre les QTL détectés avec la mesure « bud » et « bud70 » du débourrement. La « colocalisation » entre les QTL impliqués dans l'arrêt de croissance et dans le débourrement est aussi très intéressante. Comme nous l'avons déjà souligné, trois des 4 QTL détectés pour l'arrêt de croissance « colocalisent » avec des QTL impliqués dans le débourrement : 2F, 6M et 9M (figure 23). Nous savons que le débourrement est principalement contrôlé par la température ; en revanche, la lumière (et donc la durée du jour) est le facteur essentiel intervenant dans l'arrêt de croissance (Howe et al., 2000). Malgré la diversité des signaux environnementaux auxquels les deux processus répondent, il paraît évident que certaines voies biochimiques contrôlant l'activité des bourgeons doivent être communes dans les deux processus. Frewen et al. (2000) chez le peuplier ont en effet trouvé trois QTL impliqués dans l'arrêt de croissance qui sont proches de QTL contrôlant le débourrement. Bien que la corrélation phénotypique entre le débourrement et l'arrêt de croissance soit faible en 2002, toutefois dans les trois « colocalisations » trouvées les QTL impliqués dans les deux caractères ont le même signe de l'effet de substitution allélique concordant avec la corrélation phénotypique positive. Cela veut dire que si les trois QTL ont un effet pleiotrope sur les deux caractères, l'effet des allèles à ces QTL va dans le sens d'un débourrement tardif (score élevé) et un arrêt de croissance précoce (score élevé) ou bien vice versa. Il est alors probable que des gènes impliqués, par exemple, dans la régulation (activation et/ou inhibition) de l'activité mitotique du bourgeon soient responsables de ces QTL.

Concernant la croissance, plusieurs mesures ont été faites et comme nous avons pu le voir peu de QTL par mesure ont été détectés et la plupart des QTL sont spécifiques d'une année. Il est, néanmoins, possible d'identifier des QTL que nous pouvons appeler de vigueur, comme les QTL du groupe 1F et 10F. Surtout dans le cas du groupe 10, pendant les trois années, plusieurs QTL impliqués dans la croissance en hauteur et en diamètre ont été détectés dans cette région ; les signes des effets concordent, suggérant un QTL unique ou bien plusieurs QTL liés agissant dans le même sens et contrôlant des processus généraux impliqués dans la croissance des arbres.

Au contraire, dans la plupart des cas, l'absence de correspondance entre les QTL impliqués dans la croissance en hauteur et celle en diamètre, en dépit de leur corrélation phénotypique significative, suggère que le contrôle génétique de ces deux processus est différent. Ce résultat peut être expliqué en considérant que les deux méristèmes impliqués dans les deux types de croissance sont différents (méristème primaire dans le cas de la croissance en hauteur

et méristème secondaire pour le diamètre) et donc il est probable que plusieurs voies biochimiques spécifiques soient activées dans les deux situations.

Nous discuterons dans les paragraphes suivants de la « colocalisation » entre QTL contrôlant des caractères différents qui ont montré une corrélation phénotypique significative. En effet, dans ce cas les « colocalisations » entre QTL peuvent être attendues car une corrélation génétique peut exister entre les caractères quantitatifs. Dans ce dernier cas les QTL « colocalisés », qui probablement auront un effet pleiotrope, doivent montrer les signes des effets de substitution en accord avec celui de la corrélation.

Delta et croissance. Une corrélation phénotypique négative a été observée entre les données de croissance et la discrimination isotopique du carbone 13 (delta). Lauteri et al. (2004) ont montré qu'au stade juvénile il existe une corrélation génétique forte et significative entre delta et croissance chez des populations de châtaignier provenant d'environnements contrastés. Les résultats de la détection des QTL montrent que 5 QTL impliqués dans l'efficacité d'utilisation de l'eau sont sur un groupe qui contient aussi des QTL impliqués dans la croissance. Dans 4 cas, 1F, 6F, 7F (delta/hauteur) et 8F le signe entre les effets des QTL est opposé en accord avec la corrélation négative observée entre les deux caractères. Il faut noter que la dernière « colocalisation » concerne le parent mâle, précisément le groupe 3M, mais la direction de l'effet de substitution coïncide entre les QTL. Même si la relation entre croissance et efficacité d'utilisation de l'eau est plutôt complexe (Brugnoli et Farquhar, 2000), la corrélation négative révélée dans cette descendance semble avoir des bases génétiques. L'adaptation au climat sec du parent femelle pourrait expliquer cette corrélation. En effet, les individus de la population Bursa montrent une capacité photosynthétique, une conductance stomatique et du mésophylle plus élevées que les individus de la population Hopa dans le même environnement méditerranéen (Lauteri et al., 1997). En revanche, ils ont une croissance plus faible de la partie aérienne dans la phase juvénile (Lauteri et al., 1997, Monteverdi, 2004). La relation négative entre croissance et discrimination isotopique du carbone 13 est en accord avec un modèle d'adaptation structurelle de la plante à des conditions de sécheresse. Selon ce modèle l'écotype Bursa durant sa phase juvénile favorise la répartition du carbone au niveau racinaire. La capacité photosynthétique élevée serait donc liée à cette répartition racinaire en dépit d'une croissance aérienne plus faible. Le développement des racines pour un approvisionnement d'eau plus efficace constitue une adaptation à long terme à la sécheresse. La faible efficacité d'utilisation de l'eau (delta élevé) et la faible croissance aérienne correspondraient donc à ce type d'adaptation. Les 4

« colocalisations » observées chez le parent femelle entre l'efficacité de l'utilisation de l'eau et la croissance seraient les combinaisons alléliques de l'adaptation génétique de cet écotype à la sécheresse.

Croissance et arrêt de croissance. En 2002 une corrélation phénotypique négative entre la croissance et l'arrêt de croissance au niveau des bourgeons a été mise en évidence. La formation précoce des bourgeons d'hiver entraîne une croissance mineure car la saison végétative est plus courte. Dans les trois « colocalisations » observées en 2002 entre les QTL contrôlant la croissance et l'arrêt de croissance au niveau des bourgeons (6M, 8F et 9M), les effets de substitution ont un effet opposé en accord donc avec la corrélation phénotypique et suggérant une corrélation génétique entre les deux caractères. L'interruption de la croissance en automne est une stratégie adaptative très importante pour la résistance au froid de l'hiver, car les individus qui ont un arrêt précoce de la croissance seront aussi mieux adaptés au froid, en dépit d'une croissance plus faible. Au contraire, les génotypes qui continuent à maintenir les bourgeons actifs plus longtemps montreront une croissance plus importante mais seront sujets aux dégâts du froid précoce (Howe et al., 2000). Les plantes ont besoin de trouver un compromis entre les deux tendances qui leur garantira de survivre pendant l'hiver. La « colocalisation » des QTL impliqués dans les deux caractères met en évidence des bases génétiques communes pour les deux processus. Ces régions génomiques sont celles sur lesquelles agit la sélection afin d'obtenir les combinaisons alléliques les plus adaptées à un environnement donné. Il est intéressant de noter que deux de ces « colocalisations » ont été détectées sur la carte mâle et la « colocalisation » est très précise dans le cas du groupe 9M. Comme il est évident dans la figure 20, en 2002 la pluviométrie pendant la saison végétative a été significativement plus élevée qu'en 2001 et 2000. Il est probable que dans ces conditions environnementales la plasticité phénotypique des individus de la descendance ait été déterminée principalement par les allèles dérivant de l'individu Hopa adapté génétiquement à ce type de conditions climatiques. Malheureusement, les observations d'arrêt de croissance n'ont été faites qu'en 2002 et la relation avec les conditions climatiques est une simple observation. Toutefois, les résultats obtenus permettent de faire des hypothèses qui pourront être vérifiées par la suite.

C.3.3.8. Conclusions : les résultats semblent indiquer une adaptation locale des deux parents à leur environnement d'origine.

Cette détection des QTL chez le châtaignier est limitée essentiellement par deux facteurs :

- 1) l'effectif faible qui conduit principalement à des estimations biaisées des effets des QTL d'une part et d'autre part ne permet pas de détecter les QTL à effet très faible ;
- 2) la complexité des caractères adaptatifs étudiés. Cette complexité ne permet pas en mesurant un paramètre physiologique unique de bien rendre compte de la variabilité phénotypique du caractère dans la descendance et en particulier, pour la discrimination isotopique du carbone 13 et la croissance, il est pratiquement impossible d'arriver à décomposer tous les facteurs mendéliens contrôlant le caractère qui ségrégent dans cette famille de plein-frères. L'imprécision des mesures, par exemple, a des effets sur la précision de localisation des QTL et donc sur les intervalles de confiance. De plus, il ne faut pas oublier qu'il s'agit de plantes jeunes et que, par exemple, la croissance juvénile est soumise à un contrôle génétique différent de celle d'une plante adulte.

Toutefois, certaines observations d'intérêt peuvent être faites :

- les trois caractères montrent beaucoup plus des QTL à effet faible ou modéré que fort, mais des QTL à effet fort sont toujours détectés. Cela est en accord avec la théorie de Orr (1998), selon laquelle les allèles à effet fort auraient joué un rôle important durant l'adaptation à cause de leur probabilité plus élevée de se fixer et d'approcher l'optimum phénotypique.
- Même si le débourrement est le caractère qui possède le nombre le plus élevé de QTL stables, tous les caractères montrent des QTL stables au cours des trois années. Ce résultat met en évidence que la recherche de QTL permet assez facilement d'identifier des régions constamment impliquées dans le contrôle génétique des caractères étudiés. Toutefois, plusieurs QTL exprimés dans des conditions particulières pourraient constituer la base de la plasticité phénotypique des individus face à des conditions environnementales particulières.
- Il semble évident, comme suggéré par Orr et Coyne (1992), que les populations naturelles adaptées à des conditions environnementales contrastées constituent un bon matériel pour mener des études portant sur les bases génétiques de l'adaptation. Les résultats semblent confirmer l'adaptation génétique des deux populations Hopa et Bursa à deux environnements contrastés. Comme souligné par Verhoeven et al. (2004), l'adaptation locale des populations naturelles semble se fonder sur de nombreux caractères. Les combinaisons alléliques aux locus contrôlant ces caractères détermineraient la plasticité phénotypique des individus. Les allèles aux différents locus pourraient contribuer plus ou moins au phénotype selon l'environnement dans lequel se trouve l'individu. La corrélation génétique entre la discrimination isotopique

du carbone 13 et la croissance met également en évidence que l'adaptation se réalise grâce à l'interactions de plusieurs locus agissant sur différents caractères (Orr, 1998).

Pour pouvoir vérifier les hypothèses faites trois directions de recherche semblent envisageables :

- 1) il serait intéressant de tester la même famille de châtaignier dans des conditions environnementales différentes, en particulier dans un environnement humide similaire à celui d'origine du parent Hopa. Pour cela il faudrait avoir des clones des individus de la famille...
- 2) Les caractères de croissance et de discrimination isotopique du carbone 13 devraient être décomposés dans des caractères plus « simples » et héréditaires, afin de pouvoir identifier quel processus physiologique particulier est responsables de chaque QTL. Par exemple, dans le cas de la discrimination isotopique du carbone 13, il serait intéressant de pouvoir effectuer des mesures de conductance stomatique, d'assimilation, de photosynthèse afin d'identifier l'effet sous-jacent à chaque QTL.
- 3) La cartographie de gènes candidats fonctionnels et expressionnels pourrait aider à éclaircir aussi les voies biochimiques et les différents processus physiologiques qui se cachent derrière les QTL détectés.

D. Cartographie génétique comparée chez les Fagacées

« ...*Mission impossible*... »

D.1. LA CARTOGRAPHIE GÉNÉTIQUE COMPARÉE CHEZ LES FAGACÉES : OBJECTIFS ET ACQUIS.

Le projet de cartographie génétique comparée entre le chêne pédonculé et le châtaignier européen a démarré en 2000 dans le cadre du projet CASCADE (<http://soi.cnr.it/~chestnut>). Teresa Barreneche a optimisé le transfert des marqueurs SSR entre les deux espèces et ce travail ainsi que la cartographie génétique des SSR a fait l'objet d'une publication présentée en annexe III. Ensuite, les séquences EST développés par Jérémy Derory (thèse en cours : Génomique et diversité du débourrement chez les chênes, Université de Bordeaux I) ont été exploitées afin de les cartographier chez les deux espèces. Une dizaine de EST cartographiés sur chêne par Ilga Porth (Porth et al., soumis, a,b) ont été également cartographiés chez le châtaignier. Caroline Morera-Dutrey a participé à la cartographie des EST dans le cadre de son stage de licence en Biologie et Biochimie Appliquée (Université de Bordeaux II). La publication concernant ce travail, qui est en cours de préparation, est présentée en annexe IV.

Comme nous avons pu le voir en détail dans l'introduction de cette thèse, la cartographie comparée donne la possibilité d'aborder l'étude des changements évolutifs des génomes et de mieux comprendre leur structure et leur fonctionnement. En particulier, les objectifs ponctuels à atteindre dans le cadre de ce travail sont :

- évaluer deux types de marqueurs moléculaires (SSR et STS dérivés des EST) pour la cartographie comparée chez les Fagacées ;
- cartographier les marqueurs orthologues chez les deux espèces afin de comparer les cartes génétiques du chêne et du châtaignier ;
- évaluer le niveau de conservation de la synténie et de la colinéarité afin de comprendre si le transfert de connaissances moléculaires et génétiques est possible entre ces deux espèces ;
- comparer la position des QTL impliqués dans les caractères adaptatifs afin de les valider et d'étudier le contrôle génétique de ces caractères chez deux espèces qui ont divergé il y a 60 millions d'années.

D.1.1. La carte génétique du chêne pédonculé.

Comme nous l'avons indiqué dans le paragraphe A.2.5., la carte génétique du chêne pédonculé a été publiée par Barreneche et al. en 1998. La descendance F₁ utilisée pour la construction de la carte dérive d'un croisement entre deux individus provenant l'un (le parent

femelle, 3P) de la station de recherche de Pierroton et l'autre (le parent mâle, A4) d'Arcachon (Barreneche et al., 1998). La famille est constituée de 278 individus, dont 94 ont été utilisés pour la construction de la carte génétique. Un programme de bouturage a permis d'avoir plusieurs copies clonales des descendants de cette famille, qui ont été utilisées pour la détection des QTL. Une caractéristique très intéressante de la carte génétique du chêne est la présence de marqueurs comme les SSR et le locus 5S de l'ADN ribosomal, potentiellement transférables chez le châtaignier. Dix-huit microsatellites ont été cartographiés sur les douze groupes de liaison du chêne dans cette première version de la carte. Certains de ces SSR avaient été déjà testés chez plusieurs espèces appartenant à la famille des Fagacées et 47% des microsatellites testés avaient amplifié chez le châtaignier (Steinkellner et al., 1997a). De plus, la saturation de cette carte est satisfaisante (75%) et comparable à celle du châtaignier, elle possède douze groupes de liaison et une nouvelle version est disponible dans laquelle des marqueurs AFLP ont été intégrés permettant d'atteindre une saturation proche de 80% (Barreneche et al., en préparation). La nouvelle version de la carte a été également utilisée pour la détection des QTL contrôlant les caractères adaptatifs dans le cadre de la thèse de Caroline Saintagne (2003). Cette nouvelle version de la carte sera donc utilisée pour la comparaison avec la carte du châtaignier et elle est présentée dans la partie E.

D.1.2. Les marqueurs orthologues.

Afin de pouvoir comparer les deux cartes génétiques, des marqueurs orthologues doivent être cartographiés chez les deux espèces. Deux types de marqueurs ont été testés : les SSR et les STS qui dérivent d'une banque EST.

Les microsatellites sont composés de courtes séquences (de 1 à 6 nucléotides) répétées en tandem et ils sont distribués sur tout le génome (Powell et al., 1996). Le motif répété peut être parfait, par exemple $(GA)_n$, imparfait, par exemple $(GA)_nTC(GA)_n$, ou bien composé, par exemple $(GA)_n(CA)_n$. Le génome des plantes est riche en séquences répétées qui peuvent être utilisées comme marqueurs moléculaires (Morgante et Olivieri, 1993). En effet, les séquences flanquant le motif répété sont utilisées pour mettre au point des amorces qui permettent l'amplification par PCR de ces locus. Les microsatellites sont codominants, très polymorphes, distribués de façon homogène sur le génome et ils possèdent un grand nombre d'allèles à chaque locus. Morgante et al. (2002) ont montré que la fréquence des microsatellites est inversement proportionnelle à la taille du génome et cette fréquence est constante dans les régions transcrites du génome, suggérant une association préférentielle des séquences répétées

avec les régions à faible copie du génome. Il est désormais clair que ces séquences répétées jouent un rôle fonctionnel important dans l'organisation de la chromatine, la régulation du métabolisme de l'ADN (réplication, recombinaison, réparation²⁶) et la régulation de l'expression des gènes (Li et al., 2002). Cela pourrait expliquer leur association avec la fraction fonctionnelle du génome.

Les SSR ont été testés dans le cadre de ce travail pour deux raisons :

- 1) d'une part, 18 SSR étaient déjà présents dans la carte génétique du chêne et l'article de Steinkellner et al. (1997a) démontrait qu'il était possible de les amplifier chez le châtaignier. Les SSR avaient été déjà utilisés pour la cartographie comparée chez les arbres forestiers (Marques et al., 2002) et de nombreux auteurs avaient abordé le sujet de la transférabilité des SSR chez plusieurs groupes de plantes, par exemple chez les Citrineae (Kijas et al., 1995), les Brassicaceae (Kresovich et al., 1995 ; van Treuren et al., 1997 ; Plieske et Struss, 2001 ; Saal et al., 2001), l'*Eucalyptus* (Byrne et al., 1996), les Gramineae (Brown et al., 1996 ; Brondani et al., 2001 ; Guyomarc'h et al., 2002 ; Matsuoka et al., 2002), les Asteraceae (Whitton et al., 1997), les Legumineae (Peakall et al., 1998), les Pinaceae (Fisher et al., 1998 ; Echt et al., 1999 ; Karhu et al., 2000 ; Hodgetts et al., 2001 ; Kutil et Williams, 2001 ; Rajora et al., 2001 ; Sheperd et al., 2002), les Cucurbitaceae (Danin-Poleg et al., 2000), les Vitaceae (Di Gaspero et al., 2000 ; Rossetto et al., 2002) et les Rosaceae (Yamamoto et al., 2001 ; Dirlwanger et al., 2002). Le développement de marqueurs SSR à partir de banques d'ADN génomique et complémentaire est souvent un travail long et lourd. C'est pourquoi la possibilité de transférer des SSR entre espèces phylogénétiquement proches est une stratégie intéressante. Néanmoins, les articles cités au-dessus, ont clairement démontré que l'efficacité de transfert des SSR dépend de la distance évolutive entre les espèces, de la complexité du génome et du type de microsatellites. Plus les espèces sont éloignées phylogénétiquement, moins efficace sera le transfert de locus SSR. Ce transfert est d'autant plus difficile dans le cas des espèces à génome grand et répété, comme les conifères. Enfin, les microsatellites contenant des motifs tri-nucléotide semblent plus conservés entre les espèces et donc plus faciles à transférer. Il est à noter que la seule amplification d'un locus n'implique pas que le motif répété soit conservé. Le transfert de microsatellites doit être démontré par séquençage (Chen et al., 2002).

²⁶ MMR : "Mismatch Repair System". Système de réparation des mutations.

- 2) D'autre part, il est évident que, une fois certains locus SSR transférés entre *Quercus* et *Castanea*, ils pourraient être utilisés comme outils moléculaires communs chez les deux genres et dans la famille des Fagacées. Ces marqueurs pourraient, donc, constituer non seulement des points d'ancrage pour des projets de cartographie comparée au sein de cette famille, mais aussi des marqueurs très polymorphes et codominants à exploiter dans les études de diversité génétique, flux des gènes et phylogénie des différentes espèces.

Les EST sont des séquences de fragments de gènes exprimés, qui, comme nous l'avons décrit en détail dans l'introduction, ont été souvent utilisés dans les projets de cartographie comparée chez plusieurs espèces de plantes. Chez les conifères Brown et al. (2001) ont montré que les EST sont beaucoup plus efficaces que les SSR pour la cartographie génétique comparée. Trois caractéristiques essentielles les rendent très précieux :

- 1) il s'agit de séquences codantes et donc beaucoup plus conservées entre les espèces que les séquences flanquantes des microsatellites. Pour cette raison, il est plus facile de les transférer entre espèces proches phylogénétiquement.
- 2) Il s'agit de gènes, souvent à fonction connue, qui peuvent constituer des gènes candidats. Leur cartographie permet d'établir des cartes génétiques fonctionnelles utiles dans l'identification de régions génomiques contrôlant des caractères d'intérêt.
- 3) Les séquences EST peuvent être très nombreuses car l'essor des projets de transcriptomique a mis à disposition un nombre important de ces séquences pour plusieurs espèces.

Les séquences EST développées chez le chêne ont été donc exploités pour la cartographie génétique comparée entre le chêne et le châtaignier.

Tableau 13. Liste de locus SSR. Résultats de l'amplification et de la cartographie.

Locus ^a	Motif répété	Amplification <i>Q. robur</i> ^b	Amplification <i>C. sativa</i> ^b	Type de ségrégation et groupe de liaison, <i>Q. robur</i> ^c	Type de ségrégation et groupe de liaison, <i>C. sativa</i> ^c
QpZAG1/5	(GT) ₁ (GA) ₂	2	0	1:1:1:1-LG7	
QpZAG3/62	(AG) ₁	2	0	1:1:1:1-LG10	
QpZAG3/64	(AG) ₂	2	2	1:2:1-LG2	non informatif (M)
QpZAG9*	(AG) ₂	2	2	1:1:1:1-LG7	1:1:1:1-LG5
QpZAG15	(AG) ₂	2	2	1:2:1-LG9	1:1:1:1-LG9
QpZAG16	(AG) ₁	2	2	1:1:1:1-LG6	non informatif (M)
QpZAG36	(AG) ₁	2	2	1:1:1:1-LG2	1:1:1:1-LG1
QpZAG46	(AG) ₃	2	0	1:1:1:1-LG2	
QpZAG58**	(GA) ₃	2	2	1:1:1:1-LG5	LG4(l)-LG10(sb)
QpZAG102	(AC) ₂ AA(AG) ₃	2	0	1:1:1:1-LG9	
QpZAG104	(AG) ₁ AT(GA) ₃	2	1	1:1-LG2	
QpZAG108	(AG) ₃	2	1	1:1:1:1-LG2	
QpZAG110	(AG) ₃	2	2	1:1:1:1-LG8	1:1-LG7
QpZAG119	(GA) ₂	2	2	1:1:1:1-LG2	1:1:1:1-LG1
QrZAG2**	(TC) ₂₂	2	0	1:1:1:1-LG10	
QrZAG4	(GA) ₃ GT(GA) ₃	2	2	1:1-LG11	1:2:1-LG3
QrZAG5**	(TC) ₂₃	2	2	1:1:1:1-LG8(S)-LG6(L)	1:1-LG2
QrZAG7	(TC) ₁₇	2	2	1:1:1:1-LG2	1:1:1:1-LG1
QrZAG8	(TC) ₂₁	2	1	1:1-LG12	
QrZAG11	(TC) ₂₂	2	2	1:1:1:1-LG10	1:1:1:1-LG10
QrZAG15	(GA) ₁	2	0	1:1:1:1-LG9	
QrZAG18**	(GA) ₁	2	0	1:1-LG11	
QrZAG19	(GA) ₇	2	0	1:1-LG9	
QrZAG20	(TC) ₁₈	2	2	1:1-LG1	1:1:1:1-LG6
QrZAG25	(GA) ₂	2	1	1:1:1:1-LG10	
QrZAG30	(GA) ₂	2	0	1:1-LG12	
QrZAG31**	(GA) ₁	2	2	1:1:1:1-LG9	1:1-LG9(s) 1:1-LG6(Lb)
QrZAG39	(TC) ₃₄	2	0	1:1:1:1-LG5	
QrZAG49	(GA) ₂	2	0	1:1:1:1-LG7	
QrZAG58	(GA) ₂	2	0	1:1:1:1-LG2	
QrZAG65	(TC) ₂ (TA) ₁₀	2	1	1:1:1:1-LG10	
QrZAG73	(TC) ₄₃	2	1	1:1-LG5	
QrZAG74	(GA) ₂	2	1	1:1:1:1-LG1	
QrZAG87	(TC) ₂₀	2	1	1:1:1:1-LG2	
QrZAG90	(GA) ₃	2	1	1:1-LG9	
QrZAG96	(TC) ₂₀	2	2	1:1-LG10	1:1-LG10
QrZAG101	(TC) ₂ (AC) ₁₅	2	2	1:1-LG1	non informatif (M)
QrZAG102	(GA) ₂	2	0	1:1-LG3	
QrZAG103	(TC) ₁₅	2	1	1:1-LG10	
QrZAG108	(GA) ₁ (GGGA) ₃	2	0	1:1:1:1-LG7	
QrZAG111**	(TC) ₁₉	2	0	1:1:1:1-LG11	
QrZAG112	(GA) ₂	2	2	1:1-LG12	non informatif (P)
QrZAG121**	(GA) ₂	2	2	1:1:1:1-LG9	non informatif (P)
MSQ4	(GA) ₁	2	2	1:1:1:1-LG4	non informatif (M)
MSQ13	(GA) ₁	2	2	1:1:1:1-LG6	non informatif (M)
MSQ16	-	2	1	1:1:1:1-LG9	
MicJ-AG22	-	2	2	1:1:1:1-LG9	non informatif (M)
quru-GA-0A01	(GA) ₁	2	2	1:1-LG8	1:1:1:1-LG7
quru-GA-0C11	(GA) ₁	2	2	1:1:1:1-LG1	1:1:1:1-LG6
quru-GA-0C19	(GA) ₁	2	1	1:1-LG8	
quru-GA-0M05	(GA) ₂	2	0	1:1:1:1-LG10	
quru-GA-0M07	(GA) ₁	2	2	1:1:1:1-LG9	non cartographié
quru-GA-1F02	(GA) ₁	2	2	1:1:1:1-LG6	non informatif (M)
EMCs1**	(GGC) ₂ GGTTGC(GGC) ₃	2	2	1:1-LG7	1:1-LG11
EMCs2**	(GGC) ₇	2	2	non informatif (M)	1:1-LG6
EMCs4**	(GCG) ₇	2	2	non informatif (M)	1:1-LG7
EMCs10	(CA) ₈	2	2	non informatif (M)	non informatif (P)
EMCs11	(GGC) ₆	1	2		1:1:1:1-LG5
EMCs12	(CCA) ₅	2	1	non informatif (M)	
EMCs13	(CAG) ₁₀	2	2	1:1-LG1	non informatif (M)
EMCs14	(GAG) ₁ (GGC) ₆	2	2	non informatif (M)	1:2:1-LG5
EMCs15	(CCA) ₉	2	2	1:1-LG9	1:1-LG9
EMCs17	(CCA) ₅	0	2	non informatif (P)	non informatif (P)
EMCs22	(GA) ₁	1	2	1:1-LG2	
EMCs32	(GA) ₁	1	2	1:1:1:1-LG12	
EMCs38	(GA) ₁	1	2	1:1-LG4	
CsCAT 1**	(TG) ₂ TA(TG) ₂₄	1	2	1:1-LG8	
CsCAT 2	(AG) ₁	0	2	1:1:1:1-LG10	
CsCAT 3	(AG) ₂₀	0	2	1:1:1:1-LG12	
CsCAT 5**	(GA) ₂₀	2	2	non informatif (P)	1:1:1:1-LG8(S)-1:2:1-LG3(L)
CsCAT 6	(AC) ₂ AT(AC) ₄	2	2	non informatif (P)	1:1:1:1-LG1
CsCAT 7**	(TG) ₈ CG(TG) ₁	2	2	non informatif (P)	1:1-LG6
CsCAT 8	(GT) ₁ (GA) ₂₀	0	2		1:1-LG6
CsCAT 13	(CT) ₁₄	0	2		non informatif (P)
CsCAT 14	(CA) ₂	2	2	1:1-LG4	1:1:1:1-LG2
CsCAT 15	(TC) ₁₂	2	2	non informatif (P)	1:1:1:1-LG8
CsCAT 16	(TC) ₂₀	2	2	non informatif (M)	1:1-LG6
CsCAT 17*	(CA) ₁ 9A(CA) ₂ AA(CA) ₃	2	2	1:1-LG10	1:1:1:1-LG2
CsCAT 20	(AC) ₄ (AT) ₂ A(AT) ₂	2	2	non informatif (M)	non informatif (M)
CsCAT 27	(AC) ₁	2	2	non informatif (M)	non informatif (M)
CsCAT 32	(AC) ₁ TC(AC) ₃	2	2	non informatif (P)	non informatif (P)
CsCAT 39	(AG) ₂	2	2	non informatif (P)	non informatif (P)
CsCAT 41	(AG) ₁	1	2		1:1:1:1-LG8

^aSSR chêne : locus Qp (*Q. petraea*) et Qr (*Q. robur*) (Steinkellner et al. 1997b, Kampfer et al. 1998, Lexer 1999); locus MSQ (Dow et al. 1995, Dow et Ashley 1996); locus MicJ-AG22 (Kawahara, communication personnelle); locus quru (Aldrich et al. 2003). SSR châtaignier : locus EMCs (Buck et al., 2003); locus CsCAT (Marinoni et al., 2003).

^b2: amplification claire et intense, 1: amplification légère ou donnant un profil complexe (multi-band), 0: pas d'amplification.

*la taille des bandes amplifiées ne correspond pas entre les deux espèces.

**Les locus QrZAG2, QrZAG5, QrZAG18, QrZAG111 et QrZAG121 chez le chêne et EMCs1, EMCs2, EMCs4, CsCAT1, CsCAT5, CsCAT7, QpZAG58 et QrZAG31 chez le châtaignier, montrent deux ou trois niveaux d'amplification.

^cnon informatif (P) : le locus montre deux allèles différents chez les deux parents mais la configuration n'est pas informative. Non informatif (M) : les deux parents sont monomorphes pour le même allèle.

D.2. LE TRANSFERT DE MICROSATELLITES ENTRE CHÊNE ET CHÂTAIGNIER.

D.2.1. Le microsatellites disponibles.

Nous avons testé deux séries de microsatellites : une série développée chez différentes espèces de *Quercus* (Dow et al., 1995 ; Dow et Ashley, 1996 ; Steinkellner et al., 1997b ; Kampfer et al., 1998 ; Lexer, 1999 ; Aldrich et al., 2002) et une autre série développée chez *Castanea sativa* Mill. (Buck et al., 2003 ; Marinoni et al., 2003). Le tableau 13 montre la liste des microsatellites testés. Cinquante-trois SSR locus dérivent du chêne et 30 du châtaignier. Tous les microsatellites ont été testés chez les deux espèces en utilisant les deux parents des croisements et six individus de la descendance afin d'identifier les locus informatifs au sein des deux pedigrees de cartographie. Le travail de transfert des SSR d'un genre à l'autre a consisté en trois étapes : dans un premier temps les microsatellites qui montraient un profil clair (une bande amplifiée nette) et qui étaient polymorphes dans la descendance ont été sélectionnés ; 96 individus des deux descendances ont été génotypés avec ces locus et les SSR ont été ainsi intégrés dans les deux cartes génétiques ; enfin, un allèle de chaque locus a été cloné et séquencé. L'identité de séquence élevée entre les deux espèces et la localisation sur les groupes homologues constituent une vérification de l'orthologie de ces marqueurs.

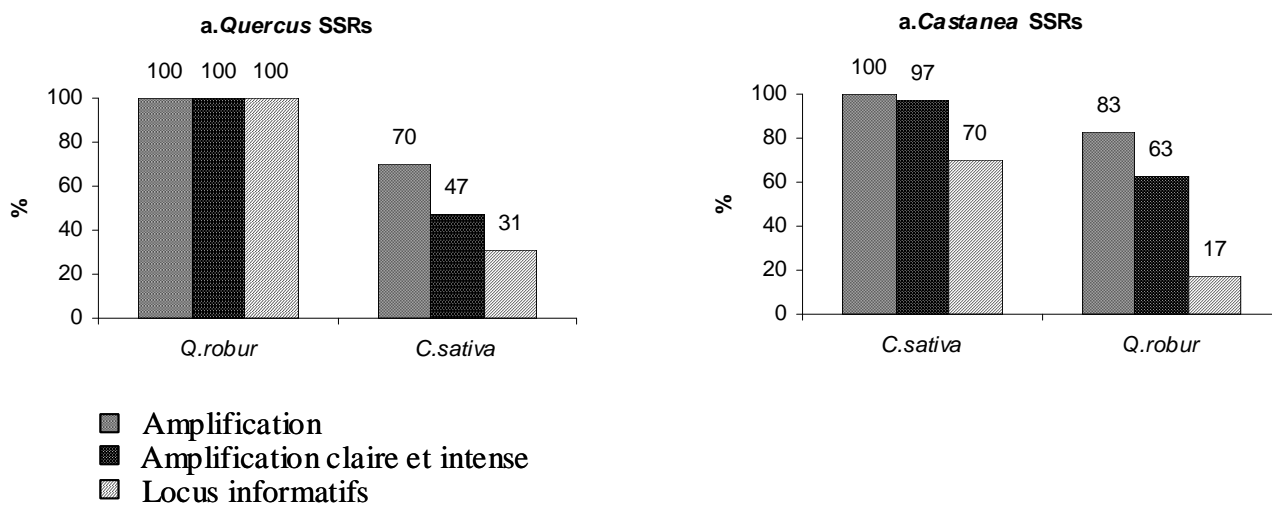
D.2.2. Le transfert entre les deux genres.

Les locus SSR ont été amplifiés grâce au protocole suivant :

- la réaction de PCR a été faite dans un volume total 20 µl, contenant 2 µl de tampon 10X (Invitrogen, Carlsbad, Calif., USA), 2 mM de MgCl₂, 200 µM de chaque dNTP, 0,2 µM de chaque amorce, 5% de DMSO (diméthylsulfoxyde) 5-10 ng d'ADN génomique et 0,5 unités de *Taq* Polymerase (Invitrogen, Carlsbad, Calif., USA).
- Le cycle d'amplification a consisté en 10 minutes à 95°C, 26 cycles de 50 secondes à 94°C, 45 secondes à la température d'hybridation, 1 minute et 30 secondes à 72°C et enfin un cycle de 8 minutes à 72 °C (Botta et al., 1999).

Chaque couple d'amorces a été testé chez les deux espèces à la température d'hybridation recommandée pour l'espèce dans laquelle le locus avait été développé (que nous appellerons espèce source). Si l'amplification dans l'espèce non-source (espèce dans laquelle le locus est transféré) donnait lieu à un produit faible mais de la même taille que dans l'espèce source, nous avons optimisé la réaction de PCR en agissant sur deux paramètres : la température

Figure 25. Efficacité de transfert des SSR entre les deux genres *Quercus* et *Castanea*.



d'hybridation, qui a été baissée de 3-5°C, et la quantité d'ADN, qui a été augmentée jusqu'à dix fois car nous avons observé une amélioration du rendement de la réaction d'amplification dans l'espèce non-source avec une quantité d'ADN cible majeure. La diminution de la température d'hybridation permet de diminuer la spécificité de cette étape de la réaction et d'avoir une hybridation efficace même en présence de misappariements au niveau des séquences des amorces. En général, une fois la température d'hybridation optimisée, la quantité d'ADN de 5-10 ng était toujours optimale pour l'amplification du locus dans l'espèce non-source. Les SSR codés 2 dans le tableau 13 ont donné un produit clair et intense et ils ont été utilisés par la suite. Aucun effort d'optimisation n'a été fait pour les SSR qui donnaient un profil multi-bande dans l'espèce non-source, c'est à dire la plupart des SSR codés 1 dans le même tableau. Il faut noter que dans trois cas (QpZAG9, QrZAG5 et CsCAT17) la taille de la bande amplifiée varie entre les deux espèces alors même que le produit est intense et net dans les deux cas. De plus, 5 locus chez le chêne et 8 chez le châtaignier (tableau 13) montrent deux ou trois niveaux d'amplification. Etant donné qu'il s'agit de bandes bien différenciées en taille, ces locus ont été considérés pour la cartographie. Les produits de PCR ont été séparés sur un gel d'acrylamide 6% dénaturant et visualisés soit grâce au séquenceur Li-cor (modèle 4000 et 4000L IRD40, Li-cor Inc, Lincoln, Neb., USA) soit par coloration au nitrate d'argent (Streiff et al., 1998).

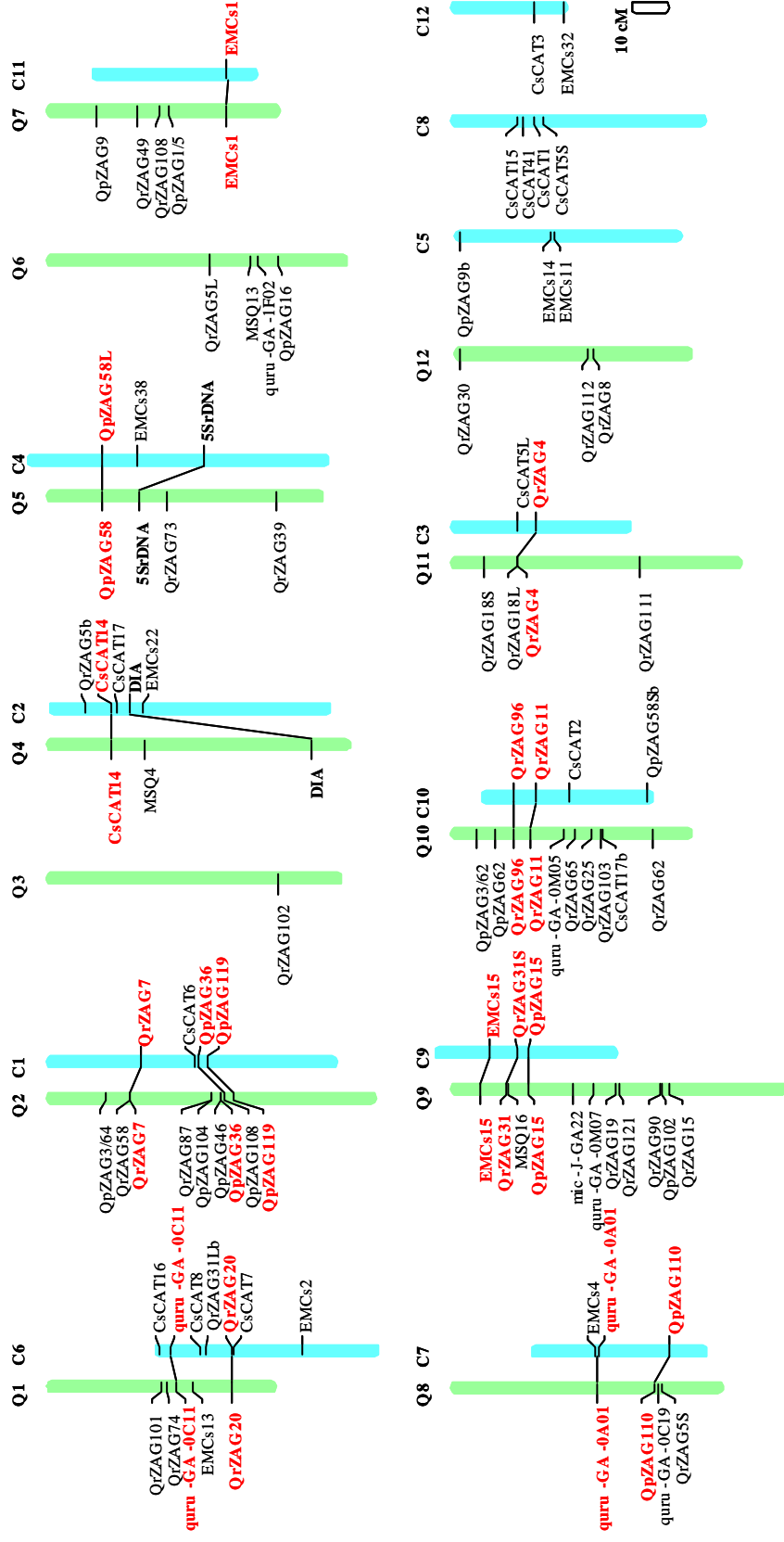
La figure 25 montre l'efficacité d'amplification des microsatellites et la proportion des locus informatifs dans les deux pedigrees. Il faut noter que les SSR chêne sont tous informatifs dans la descendance de cette espèce car ils ont été sélectionnés au préalable sur la base de cette caractéristique. Il est évident que dans les deux cas il est possible d'amplifier des locus dans l'espèce non-source (47% des locus du chêne ont donné une bonne amplification chez le châtaignier et 63% des locus châtaignier ont bien amplifié chez le chêne). Néanmoins, la proportion de locus amplifiés chez l'espèce non-source qui sont informatifs dans la descendance est plutôt faible : 31% ont pu être cartographiés chez le châtaignier et seul 17% chez le chêne. Il y a, au moins, deux observations d'intérêt à faire en ce qui concerne ces résultats. La première regarde évidemment l'efficacité de ces marqueurs pour la cartographie génétique comparée. Comme déjà observé (Brown et al., 2001), les microsatellites sont faiblement conservés chez les espèces phylogénétiquement éloignées. Ils sont, donc, très efficaces pour aligner des cartes génétiques différentes au sein de la même espèce mais moins intéressants entre genres. Toutefois, un certain nombre de ces locus peut être transféré facilement, ce qui reste très intéressant car le développement de marqueurs SSR est souvent lourd et laborieux. La deuxième observation concerne la différence significative de locus

informatifs chez l'espèce non-source entre les deux séries de SSR. En effet, sur 19 (63%) microsatellites du châtaignier qui ont bien amplifié chez le chêne, 5 seulement (17%) étaient informatifs dans la descendance du chêne. En revanche, 31% des locus du chêne ont pu être cartographiés chez le châtaignier. Une explication à cette différence peut être trouvée dans la nature du motif répété des microsatellites. Nous pouvons noter dans le tableau 13 que la plupart des microsatellites du chêne contiennent des motifs di-nucléotide, alors que sur les 19 locus de châtaignier qui ont bien amplifié chez le chêne, 12 contiennent dans l'espèce source soit un motif tri-nucléotide soit un motif imparfait ou bien composé. La plupart d'entre eux sont monomorphes ou non informatifs chez le chêne. D'ailleurs, 7 locus du chêne testés sur châtaignier contenaient des motifs imparfaits ; 5 n'ont pas pu être amplifiés et sur les 2 restants, un seulement a pu être cartographié chez le châtaignier.

Des résultats similaires ont été obtenus chez d'autres espèces.

- Les études sur plusieurs espèces d'eucaryotes ont montré que les microsatellites contenant des motifs tri-nucléotide sont plus faciles à transférer entre espèces (Wang et al., 1994 ; Young et al., 2000 ; Morgante et al., 2002 ; Scotti et al., 2002 ; Chagné et al., sous presse) que les microsatellites contenant des motifs di-nucléotide, imparfaits ou bien composés (Kutil et Williams, 2001). Les motifs tri-nucléotides sont associés plus souvent que les autres types de motifs aux régions fonctionnelles du génome, soit dans la partie codante des gènes, soit dans des régions de régulation (Tóth et al., 2000 ; Young et al., 2000). C'est pourquoi ils ont une séquence plus conservée entre les espèces et donc il est plus facile de les transférer. Il est, en effet, probable que les séquences SSR dans les régions codantes et de régulation des gènes soient soumises à des pressions selectives plus importantes que les microsatellites dans d'autres régions (Li et al., 2004). Pour la même raison, ils seraient moins polymorphes que les microsatellites contenant des motifs di-nucléotide (Kutil et Williams, 2001 ; Rajora et al., 2001 ; Sheperd et al., 2002). Cela pourrait expliquer notre résultat : les SSR tri-nucléotide du châtaignier seraient donc plus facilement transférables chez le chêne mais probablement moins polymorphes.
- Concernant les microsatellites avec un motif imparfait ou composé, nous avons également observé une diminution de l'efficacité de transférabilité. Une explication possible est fournie par Taylor et al. (1999). Selon ces auteurs, lorsqu'une interruption se produit dans un motif microsatellite, le phénomène de « slipped-strand mispairing », qui est la cause essentielle des mutations de taille au niveau du microsatellite, serait moins probable. Cela conduirait d'une part à une diminution du niveau de

Figure 26a. Locus SSR cartographiés chez le chêne (vert) et chez le châtaignier (bleu). En rouge les locus SSR communs entre les deux espèces. Les groupes de liaison sont ordonnés en prenant la carte génétique du chêne pédonculé comme référence.



polymorphisme du locus et d'autre part à une réduction du nombre de répétitions. Les deux processus devraient aboutir à la « mort » du motif microsatellite. Les mutations (insertions et délétions), qui se produiraient à la suite de la première interruption du motif dues probablement au « crossing-over » inégal, seraient responsables de la conservation plus faible de ce type de microsatellites entre les espèces (Kruglyak et al., 1998 ; Kutil et Williams, 2001).

D.2.3. La cartographie génétique des microsatellites et la première comparaison des cartes du chêne et du châtaignier.

Les microsatellites qui étaient informatifs dans les deux pedigrees ont été cartographiés en utilisant 96 individus de chaque descendance. Le logiciel « MapMaker » a été utilisé pour cartographier les marqueurs ségrégeant 1:1 et 1:1:1:1 et le logiciel « JoinMap » pour les marqueurs ségrégeant 1:2:1 (tableau 13). Un seuil de LOD = 6.0 a été choisi dans les deux cas. La figure 26a montre les locus SSR cartographiés chez le chêne (Barreneche et al., 1998 ; Barreneche et al., en préparation) et le châtaignier. Soixante SSR, dont 5 transférés du châtaignier, sont distribués sur les 12 groupes de liaison du chêne. Parmi les 39 locus cartographiés chez le châtaignier, 17 proviennent du chêne. Dans le cas aussi du châtaignier, chaque groupe de liaison contient au moins un SSR. La distribution homogène des SSR sur le génome permet d'obtenir facilement des points d'ancrage sur chaque groupe de liaison. Les 39 microsatellites cartographiés chez le châtaignier européen ont été suffisants pour relier les 12 groupes de liaison de cette espèce avec ceux obtenus dans la carte génétique construite en utilisant un croisement interspécifique du châtaignier américain, *C. dentata*, et chinois, *C. mollissima* (Paul Sisco, « The American Chestnut Foundation », communication personnelle). Cela démontre clairement l'utilité des marqueurs microsatellites pour la cartographie comparée chez des espèces très proches phylogénétiquement. Les microsatellites du châtaignier ont été utilisés dans le cadre du projet CASCADE pour des études de flux de gènes et d'analyse globale de la diversité génétique de populations naturelles (<http://soi.cnr.it/~chestnut>). La position sur la carte génétique a guidé le choix des locus à utiliser dans ce but car les locus liés génétiquement peuvent donner des résultats biaisés.

Dix-neuf SSR ont pu être cartographiés chez les deux espèces. Seize de ces SSR, le locus *Dia* (l'isozyme diaphorase, Barreneche et al., 1998 ; Annexe I) et le locus 5SrDNA (Annexe III) ont permis de relier 9 des 12 groupes de liaison des deux espèces. Parmi les 9 groupes associés, 7 contiennent au moins deux marqueurs et l'ordre est conservé entre les trois marqueurs présents sur les groupes Q2-C1 et Q9-C9 (figure 26a). Il s'agit très probablement

de groupes homologues entre les deux espèces. Ces locus constituent les premiers points d'ancrage pour la cartographie génétique comparée chez les Fagacées.

Comme le montre la figure 26b il y a trois cas qui méritent une attention particulière :

- 1) trois microsatellites (QpZAG9, QrZAG5 et CsCAT17) amplifient une bande de taille différente chez les deux espèces et ils se trouvent sur des groupes qui ne semblent pas homologues. Ces marqueurs sont appelés b dans l'espèce non source (par exemple, le locus QrZAG5 sera appelé QrZAG5b chez le châtaignier). Ces locus sont-ils des paralogues ?
- 2) Certains locus, comme QrZAG2, QrZAG5, QrZAG18, QrZAG111 et QrZAG121 chez le chêne et EMCs1, EMCs2, EMCs4, CsCAT5, CsCAT1 et CsCAT7 chez le châtaignier montrent deux ou trois niveaux d'amplification au sein de la même espèce. Aucun des deux fragments amplifiés par le locus QrZAG5 chez le chêne ne correspond au niveau amplifié chez le châtaignier. A l'exception de QrZAG5 et CsCAT5, les différents niveaux d'amplification montrent le même profil de ségrégation, suggérant qu'il s'agit des mêmes locus dupliqués en tandem. En revanche, les deux niveaux d'amplification de QrZAG5 et CsCAT5 identifient deux locus indépendants sur les groupes Q6 et Q8 dans le cas de QrZAG5 et sur les groupes C3 et C5 dans le cas de CsCAT5 (figure 26a et 26b). Les locus sont distingués par la lettre L ou S afin d'indiquer la taille du fragment amplifié (L « large » et S « small »). S'agit-il de locus dupliqués au sein du même génome ?
- 3) Enfin, deux locus QpZAG58 et QrZAG31, qui amplifient une seule bande chez le chêne, montrent une bande additionnelle chez le châtaignier. Dans ce dernier cas, la ségrégation est indépendante entre les deux niveaux. Deux locus, dont un orthologue à celui du chêne, pour chacun de ces SSR ont été identifiés chez le châtaignier (QpZAG58L sur le groupe C4, QpZAG58S sur le groupe10, QrZAG31L sur le groupe C6 et QrZAG31S sur le groupe C9). S'agit-il de locus dupliqués chez le châtaignier après sa divergence du chêne ?

Il est évident, d'après ces résultats de cartographie de locus SSR, qu'une même paire d'amorces peut amplifier deux locus différents. Les microsatellites comme les gènes peuvent donc être organisés en familles « multi-géniques ». Plusieurs auteurs ont déjà mis en évidence des locus SSR multi-copie (Fisher et al., 1998 ; Peakall et al., 1998 ; Karhu et al., 2000) et Röder et al. (1998) ont démontré que les mêmes amorces SSR peuvent amplifier des locus non homologues dans deux génomes différents. D'ailleurs, chez les conifères Elsiek et Williams (2001) ont montré que les motifs (ATG)_n sont souvent associés à des motifs tri-

nucléotide riches en CG et ils semblent organisés en familles sur la base des séquences des régions flanquantes.

Lors du transfert de microsatellites d'une espèce à l'autre, il est recommandé de cartographier le locus afin de démontrer qu'il s'agit bien du même locus. Pour démontrer l'orthologie des marqueurs cartographiés chez les deux espèces et éclaircir les relations entre les locus dupliqués, nous avons séquencé un allèle de chaque locus et calculé l'identité de séquence entre chêne et châtaignier.

D.2.4. La comparaison des locus transférés au niveau de séquence.

Un allèle de chaque locus a été cloné (TOPO TA-cloning kit, Invitrogen, Carlsbad, California, USA) et séquencé (Dye Primer Cycle Sequencing Core kit, PE Biosystems, Warrington, UK). Les séquences des deux espèces ont été alignées grâce au logiciel ALIGN (http://www.infobiogen.fr/services/analyseq/cgi-bin/alignn_in.pl). Le tableau 14 montre les résultats des alignements de séquence. L'orthologie semble évidente pour 14 des ces locus. En effet, l'identité de séquence varie de 86,3 à 97,4% (tableau 14 et figure 27a), le motif répété est toujours présent et les résultats de cartographie ont montré qu'ils se trouvent sur des groupes apparemment homologues. Au contraire, les trois locus (QpZAG9, QrZAG5 et CsCAT17), qui ont amplifié des fragments de taille différente chez les deux espèces, ont une identité de séquence plus faible, de 45,2 à 77,4%, suggérant qu'il s'agit de locus paralogues. La figure 27a et b montre l'alignement de séquence de deux locus QrZAG20 et QrZAG5, le premier conservé et le deuxième probablement paralogue entre chêne et châtaignier.

La comparaison des séquences a pu mettre en évidence des points très intéressants :

- tout d'abord, même si les séquences flanquantes sont bien conservées, le motif répété dans l'espèce non-source est souvent modifié (imparfait ou composé) ou plus court que dans l'espèce source (tableau 14). Cette tendance a été déjà observée chez plusieurs espèces (van Treunen et al., 1997 ; Peakall et al., 1998 ; Di Gaspero et al., 2000 ; Kutil et Williams, 2001) et il semble être lié au phénomène de sélection des SSR les plus polymorphes chez l'espèce dans laquelle ils ont été développés. En effet, lors du développement de marqueurs SSR, les locus les plus polymorphes sont souvent choisis. Ces locus sont aussi, en général, ceux qui contiennent les motifs les plus longs et parfaits. Mais les locus les plus polymorphes chez une espèce ne sont pas forcément les plus polymorphes chez une autre. Donc, lors du transfert des microsatellites d'une espèce à l'autre, la probabilité d'avoir des motifs plus courts,

imparfaits ou composés, et donc moins polymorphes, est plus élevée que dans l'espèce source.

- Dans quatre cas (QrZAG4, QrZAG7, QrZAG31 et CsCAT14) des délétions, de 4 à 13 paires de bases, ont été observées au niveau des régions flanquantes dans l'espèce non-source. Ces délétions expliquent l'identité de séquence plus faible entre les deux espèces par rapport aux autres locus orthologues. Il est important de noter que le polymorphisme du locus SSR transféré peut venir soit du nombre des répétitions du motif répété, soit de mutations dans les régions flanquantes. Il est donc évident que des phénomènes d'homoplasie²⁷ peuvent se produire. Il faut savoir que les mécanismes de l'évolution des microsatellites sont très variables selon le type de motif répété (Orti et al., 1997) et que la vitesse de mutation varie entre locus contenant motifs parfaits ou bien imparfaits et/ou composés et d'une espèce à l'autre (Kruglyak et al., 1998). Si un locus est transféré d'une autre espèce et s'il est utilisé pour estimer des paramètres, par exemple, de différenciation génétique, il serait nécessaire de vérifier que la variation du nombre de répétitions du motif microsatellite est à la base du polymorphisme du locus pour ne pas avoir des estimations biaisées à cause de l'homoplasie (van Treunen et al., 1997 ; Karhu et al., 2000). Le risque d'homoplasie, d'autant plus évident lors du transfert d'un locus SSR d'une espèce à l'autre, a été clairement démontré par Chen et al. (2002).
- La comparaison de séquence entre les différents niveaux d'amplification d'un même locus a montré qu'il s'agit de locus paralogues et donc probablement dupliqués du même locus ancestral. La figure 28a montre les deux niveaux d'amplification du locus CsCAT5 chez le châtaignier qui présentent une identité de 67%. Les locus QpZAG58 et QrZAG31 sont également deux cas intéressants. Les deux locus amplifient deux bandes chez le châtaignier, mais une seule bande chez le chêne. Un des deux niveaux d'amplification du châtaignier correspond à celui du chêne et les deux locus QpZAG58L (C4) et QpZAG31S (C9) semblent orthologues chez les deux espèces. En effet, les allèles de ces deux locus (tableau 14) montrent une identité de séquence plutôt élevée. En revanche, les locus QpZAG58LSb (C10) et QrZAG31L (C6) montrent une identité de séquence plus faible (de 54% à 78%) quand les séquences flanquantes sont comparées à celles du chêne et du deuxième locus révélé chez le châtaignier (figure 28b). Dans les deux cas il est intéressant de considérer l'évolution

²⁷ Homoplasie : processus pour lequel deux allèles identiques se produisent par une "histoire mutationnelle" différente.

du motif répété. En ce qui concerne le locus QpZAG58, le motif parfait du chêne n'est pas retrouvé chez le locus orthologue du châtaignier qui contient, au contraire, un motif imparfait avec une longue interruption (tableau14). Le motif est encore plus court et transformé dans le deuxième locus du châtaignier (QpZAG58Sb), dont les deux allèles, ségrégeant dans la descendance, diffèrent par deux bases à un motif mono-nucléotide, (A)₉ et (A)₁₁ (figure 28c). Une situation similaire a été observée pour le locus QrZAG31. Le motif est beaucoup plus court chez le châtaignier dans le locus orthologue et il disparaît pratiquement dans le locus dupliqué. Même s'il s'agit de simples observations, cette évolution du motif répété de ces locus SSR semble en accord avec les hypothèses faites par Taylor et al. (1999) décrites dans le paragraphe D.2.2. Comme le montre la figure 28c, les résultats des comparaisons de séquence permettent de distinguer des événements de duplication avant et après la divergence des deux genres.

Bien que *Q. robur* et *C. sativa* appartiennent à deux genres différents, le transfert de microsatellites a été relativement efficace par rapport à d'autres espèces (cf. paragraphe D.1.2.). Les locus transférés constituent les premiers points d'ancrage pour la cartographie génétique comparée chez les Fagacées. L'utilité de ces locus a été démontrée par l'association de la carte génétique du châtaignier européen avec celle issue du croisement interspécifique entre le châtaignier américain et chinois (Paul Sisco, communication personnelle). Néanmoins, nous avons pu démontrer que même dans le cas de SSR orthologues, le motif répété dans l'espèce non-source peut être différent de celui de l'espèce source (QpZAG119 et CsCAT14, tableau 14). L'utilisation donc de ces locus pour des études, par exemple, de phylogénie ou diversité génétique, au sein de la famille des Fagacées, nécessite de disposer de la séquence des allèles afin de démontrer qu'il n'y a pas des phénomènes d'homoplasie. De plus, vu le rôle fonctionnel de certains SSR Li et al. (2002) ont montré que la neutralité n'est pas une caractéristique générale de ces marqueurs. L'évolution complexe du motif répété et le rôle fonctionnel des SSR rendent l'utilisation des ces marqueurs beaucoup plus compliquée que nous le pensions.

D.3. LA CARTOGRAPHIE DES EST CHEZ LE CHÊNE ET LE CHÂTAIGNIER.

D.3.1. Les ressources en EST.

Deux séries d'EST ont été utilisées dans le cadre de cette thèse :

- 1) Une étude de transcriptomique a été développée par Ilga Porth chez le chêne afin d'identifier les gènes différentiellement exprimés en réponse au stress osmotique (Porth et al., soumis, a). Pour 10 de ces EST, une recherche de polymorphisme SNP a été conduite chez les parents ayant donné lieu aux familles utilisées pour la cartographie du chêne et du châtaignier. Parmi ces 10 EST, 8 ont été cartographiées chez le chêne (Porth et al., soumis, b). Les dix EST ont donc été cartographiées chez le châtaignier au cours de cette thèse et elles ont permis d'identifier 7 locus orthologues supplémentaires entre les deux espèces. Les dix gènes provenant de cette étude ont été utilisés pour tester différentes techniques de cartographie des EST : deux techniques de génotypage SNP basées sur les systèmes Victor (PerkinElmer) et MegaBACE (Amersham Biosciences) et deux techniques basées sur une séparation électrophorétique sur gel de polyacrylamide (SSCP²⁸ et DGGE²⁹). Ces techniques seront décrites dans le paragraphe suivant.
- 2) Dans le cadre d'une thèse réalisée au laboratoire par Jérémy Derory (Génomique et diversité du débourrement chez les chênes, Université de Bordeaux I) une étude a été menée sur les gènes différentiellement exprimés au cours du débourrement. Environ 800 séquences EST sont disponibles et consultables sur Internet à l'adresse suivante <http://cbi.labri.fr/outils/SAM/COMPLETE/index.php>, sous le nom de projet « *Quercus petraea* bud ESTs » (Derory et al., soumis). Soixante-treize EST ont été utilisées afin de développer des marqueurs STS pour la cartographie comparée chez le chêne et le châtaignier. De plus, 13 gènes candidats identifiés dans le cadre de la thèse de Jérémy Derory ont aussi été intégrés dans le projet de cartographie. Au total, 86 EST ont donc été testées. Les EST sont décrites dans le tableau 15. Dans ce tableau elles sont organisées en trois « projets » : les projets « Chênedeb » et « SSHChêne2 » correspondent à deux ensembles indépendants de séquences sur lesquelles un traitement bioinformatique a été appliqué ; le troisième projet correspond aux 13 gènes

²⁸ SSCP : "Single Strand Conformation Polymorphism".

²⁹ DGGE : "Denaturing Gradient Gel Electrophoresis".

(8 EST et 4 gènes candidats fonctionnels) étudiés par Jérémy Derory. Il s'agit d'EST très intéressantes car elles correspondent à des gènes candidats à la fois fonctionnels et expressionnels. Comme nous le verrons dans la partie E de cette thèse, la « colocalisation » des QTL impliqués dans le débourrement avec certaines EST nous permettra d'identifier des gènes candidats positionnels pour ce caractère.

L'objectif premier de la cartographie des EST est d'augmenter le nombre de marqueurs orthologues afin de vérifier si les douze groupes de liaison entre chêne et châtaignier sont homologues. En effet, les microsatellites nous ont permis de relier 7 groupes de liaison et deux autres ont été liés uniquement par un seul marqueur SSR. Avec les marqueurs EST, nous essaierons de répondre à deux questions supplémentaires : existe-il des macroréarrangements entre le génome du chêne et du châtaignier et quel est le niveau de conservation de la synténie et de la colinéarité?

D.3.2. La cartographie des EST : le génotypage SNP et les techniques SSCP et DGGE.

Afin de cartographier les EST, il est tout d'abord nécessaire de choisir la méthode pour révéler les mutations éventuellement présentes dans le fragment de séquence amplifié. Plusieurs stratégies basées sur différents principes sont disponibles. Il faut distinguer deux situations selon que la séquence entière du fragment est connue ou non (Ahmadian, 2001) :

- 1) si la séquence du fragment est connue et en particulier si le type et le site de la mutation sont déjà caractérisés, des méthodes dites de diagnostic sont souvent utilisées. Parmi ces méthodes, le génotypage SNP permet de déterminer le génotype à un site caractérisé par une mutation ponctuelle. De nombreuses technologies de génotypage SNP ont été développées à haut et moyen débit (Syvänen, 2001). Les deux approches (Victor et MegaBACE) que nous décrirons font partie des technologies à moyen débit. Les applications du génotypage SNP sont d'un grand intérêt. Chez les plantes, les SNP ont été utilisés comme marqueurs moléculaires, pour cartographier des EST, pour relier des cartes physiques et génétiques ou encore pour des études de déséquilibre de liaison et d'association (Rafalski, 2002).
- 2) Dans le cas d'une séquence non connue, les stratégies basées sur la migration différentielle de fragments d'ADN de séquence différente sur un gel de polyacrylamide, par exemple, sont souvent utilisées. Les techniques SSCP et DGGE font partie de ce type d'approche. Une autre stratégie possible est la digestion des

Figure 29a

10 EST fournis par **Ilga Porth**.



Quatre stratégies de cartographie des EST à tester :

1. **Victor** et 2. **MegaBACE** (géotypage SNP).

3. **SSCP** et 4. **DGGE**.



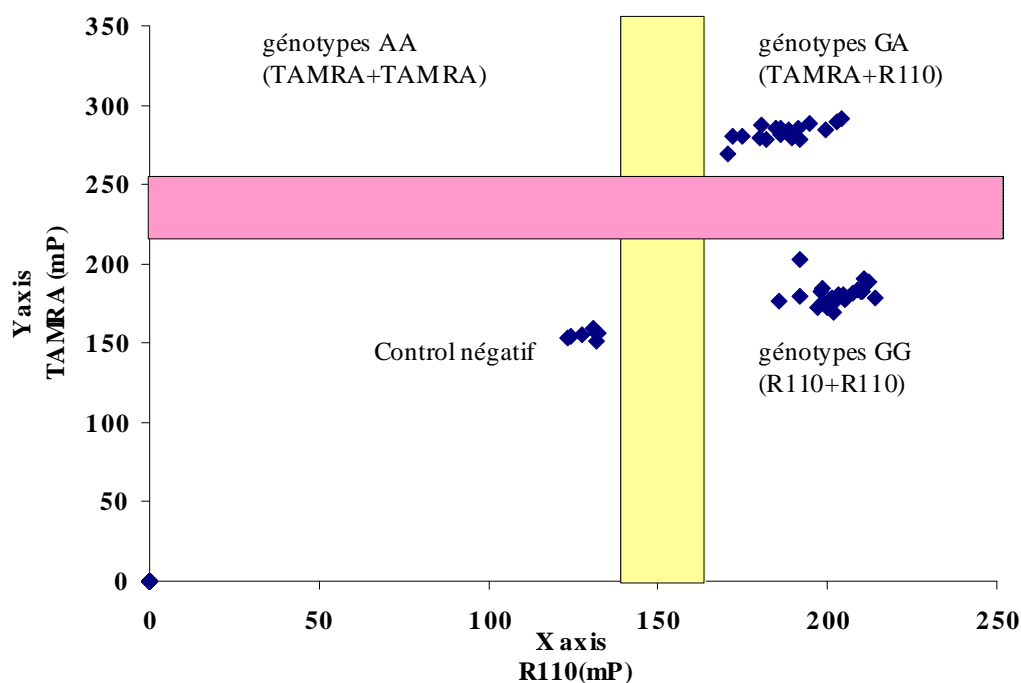
Choix de la méthode la plus efficace et la moins coûteuse pour la cartographie des EST fournis par **Jérémy Derory**.

	Coût (euros/génotype)
Victor	0,85
MegaBACE	2,13
SSCP	0,15
DGGE	0,05

EST (Porth et al., soumis, a et b) Resultat BLAST	Groupe de liaison chêne	Groupe de liaison châtaignier	Technique de cartographie
1T11 putative protein	10	10	Victor
1T12 putative oxyase protein	non lié (distordu)	3	SSCP
1T21 60S ribosomal protein L10	1	6	DGGE
1T25 Unknown protein	6	11	MegaBACE
1T57 Dc3 promoter-binding factor-2	4	2	MegaBACE
1T62 betaine aldehyde dehydrogenase-like protein	10	10	Victor
2T3 mtN19-like protein	10	10	agarose
2T11 putative glutathione transférase	non lié	8	SSCP
2T13 oxyase-like protein	11	3	SSCP
2T32 cytosolic ribosomal protein S11	2	1	MegaBACE

Figure 29b

Profil de génotypage obtenu grâce à la méthode basée sur Victor (Annexe V). L'EST 1T11 présente un site SNP polymorphe entre les deux parents. Le parent femelle est homozygote GG et le parent mâle hétérozygote GA, donc deux génotypes possibles seront détectés dans la descendance : GG et GA. En abscisse, la valeur de fluorescence polarisée du fluorochrome R110 et en ordonnée celle du fluorochrome TAMRA.



fragments amplifiés par des enzymes de restriction différentes suivie d'une séparation électrophorétique.

Il est évident qu'une troisième stratégie consiste à séquencer directement le fragment d'intérêt dans tous les individus de l'échantillon (Ahmadian, 2001).

Quatre méthodes de cartographie des EST ont été testées lors du génotypage des 10 EST fournies par Ilga Porth afin de choisir l'approche la plus adaptée à notre cas pour procéder à la cartographie des EST provenant du travail de Jérémy Derory (figure 29a).

D.3.2.1. Les techniques de génotypage SNP : Victor et MegaBACE.

De nombreuses technologies de génotypage des SNP ont été proposées, elles sont basées sur des principes différents (méthodes d'hybridation, amplification allèle spécifique, extension d'amorce, filtres à haute densité, restriction enzymatique) et elles sont toujours en évolution (Syvänen, 2001). Les deux méthodes que nous avons testées sont basées sur le principe 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 (termineur) marqué avec un fluorochrome. Une fois l'incorporation des didéoxynucléotides fluorescents réalisée, la réaction se termine et le mélange réactionnel est prêt pour la révélation. Dans le cas du séquenceur capillaire MegaBACE et du kit MegaBACE SnuPe Genotyping Kit (Amersham Pharmacia Biotech, Inc. 2001, USA), les amorces qui ont incorporé la base marquée au site polymorphe sont séparées par électrophorèse capillaire. Les 4 didéoxynucléotides fluorescents marqués sont révélés par quatre lasers spécifiques de leur longueur d'onde. Ensuite, le logiciel SNP profiler (Amersham Pharmacia Biotech, AB, Uppsala, Sweden) permet de visualiser les électrophorégrammes et donc de déduire les génotypes.

Le principe de révélation est différent dans le cas du Victor et du « Acyclo Prime TM- FP SNP detection kit » (Perkin Elmer Life Sciences, Boston, MA, USA). En effet dans ce cas le kit de réaction est spécifique pour chaque couple de bases possible à chaque site SNP. Les deux didéoxynucléotides sont marqués avec les deux fluorochromes TAMRA et R110 et le principe de révélation est basé sur la fluorescence polarisée (Chen et al., 1999 ; Hsu et al., 2001). L'appareil Victor mesure donc la fluorescence polarisée de l'amorce SNP avec la base marquée incorporée. Les valeurs de polarisation sont inversement proportionnelles à la vitesse de rotation de la molécule fluorescente, qui, à son tour, est inversement liée au volume de la molécule. Quand l'amorce SNP est liée à la base marquée, il augmente sa polarisation en

Figure 30a

Principe de la méthode SSCP et différents profils de ségrégation dans la descendance. Les deux premiers puits à gauche contiennent les parents femelle et mâle.

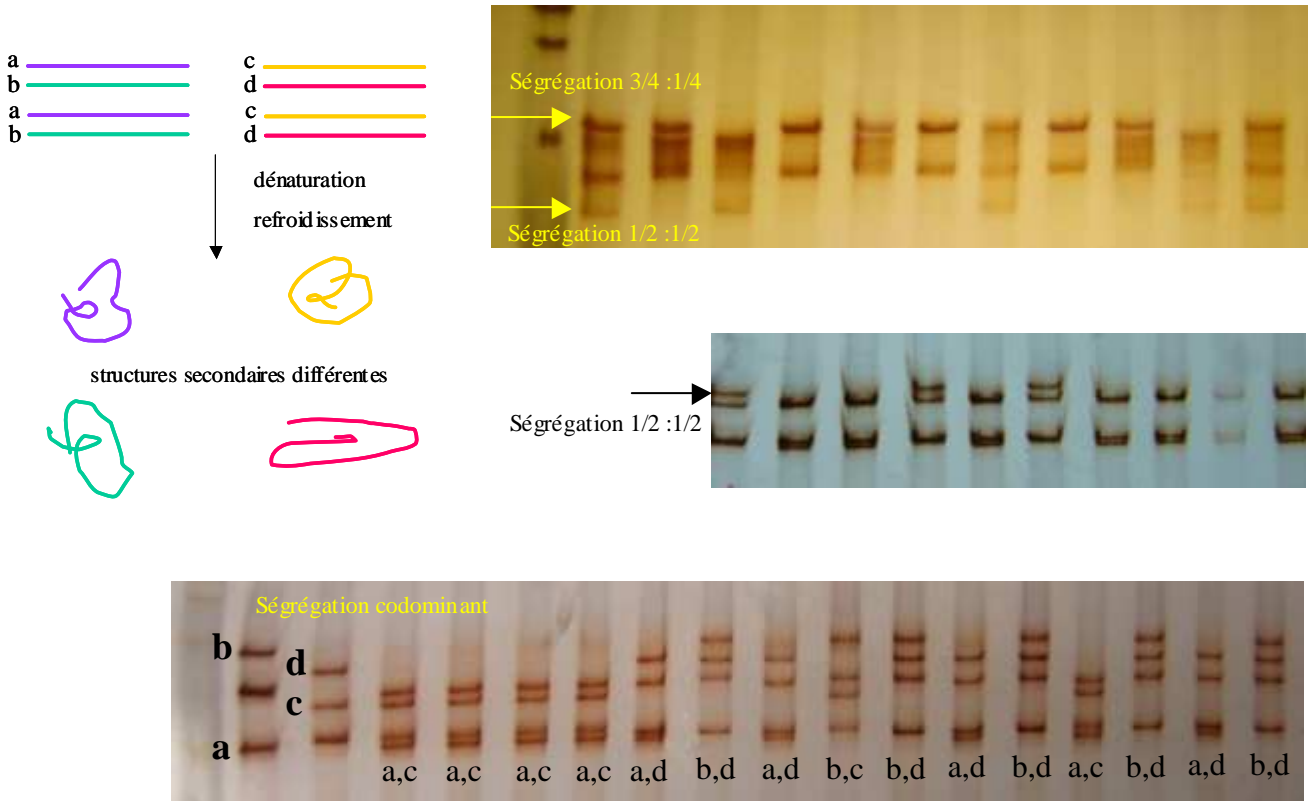
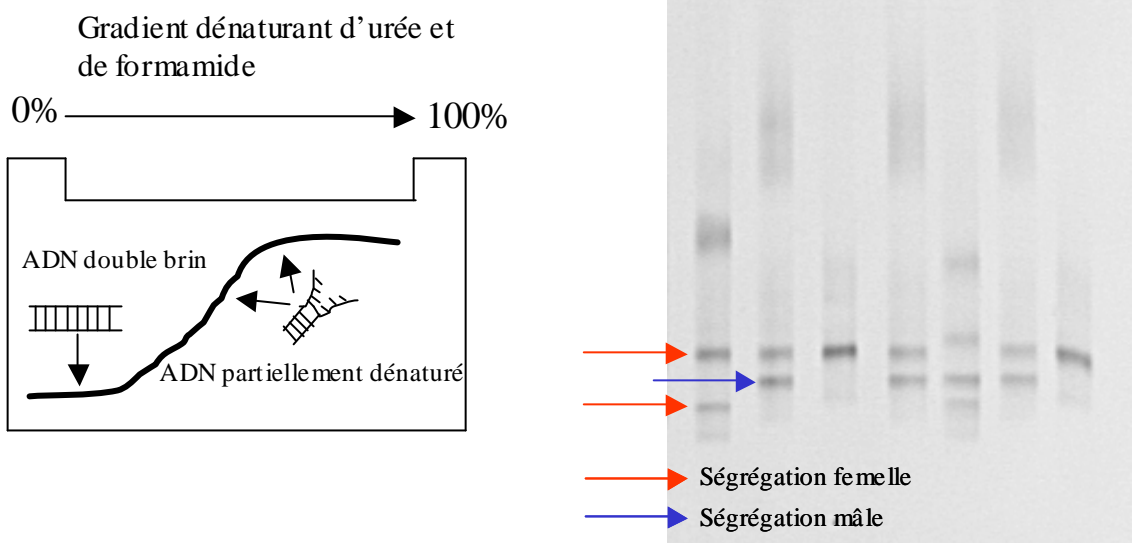


Figure 30b

Principe de la méthode DGGE. A gauche le résultat obtenu dans le cas d'un gradient dénaturant perpendiculaire au champ électrique. A droite un profil obtenu dans la descendance dans le cas d'un gradient dénaturant parallèle au champ électrique.



fonction du type de fluorochrome (soit TAMRA soit R110), ce qui permet donc de distinguer lequel des deux fluorochromes a été incorporé. Le protocole détaillé du génotypage utilisant la machine Victor est présenté en annexe V. Un profil de génotypage obtenu après lecture par la machine Victor est montré dans la figure 29b.

Pour les EST 1T25, 1T57, 1T11 et 1T62 (figure 29a), une dizaine d'individus ont été génotypés en utilisant l'approche Victor et MegaBACE, ce qui a permis de démontrer la cohérence entre les deux méthodes. Les EST 1T21 et 2T11 ont été également testées avec les deux approches, mais dans les deux cas un profil de génotypage clair n'a pas pu être obtenu. Cela suggère que, parfois, l'amorce juste à coté du site SNP n'est pas efficace dans la réaction d'extension de l'amorce. En effet, pour les amorces SNP, les mêmes contraintes que dans le cas des amorces PCR sont imposées : il faut éviter les structures secondaires, les zones riches en G/C et avoir une température de fusion entre 60 et 80°C. Mais, à chaque site SNP, deux amorces sont possibles ; si aucun des deux ne marche dans la réaction d'extension, le site SNP ne peut pas être utilisé avec ces deux méthodes. En ce qui concerne la manipulation, le génotypage avec le MegaBACE nécessite une étape de purification des produits après la réaction d'extension qui n'est pas nécessaire avec le système Victor. De plus, l'utilisation de la machine Victor est très conviviale, la lecture des échantillons très rapide, l'optimisation du protocole se fait avec peu d'individus et plusieurs sites SNP à la fois et le coût d'une réaction de génotypage est de 0,85 euros, moins chère qu'une réaction de génotypage au MegaBACE, estimée à 2,13 euros. Il est clair que les applications du système Victor par rapport à celles du séquenceur capillaire MegaBACE sont plutôt limitées. Néanmoins, en disposant des deux systèmes le système Victor est à préférer pour le génotypage SNP.

D.3.2.2. Les techniques SSCP et DGGE.

Les deux techniques SSCP et DGGE ont aussi été testées car les deux étaient disponibles au laboratoire (Bodénès et al., 1996 ; Plomion et al., 1999 ; Chagné, 2004). Il s'agit de deux techniques plutôt simples, pour lesquelles la connaissance de la séquence du fragment n'est pas nécessaire. La figure 30 (a et b) montre schématiquement le principe dans les deux cas. Dans le cas de la technique SSCP (Orita et al., 1989), le fragment d'ADN amplifié et dénaturé est chargé sur un gel non dénaturant d'acrylamide MDE³⁰. Ce type d'acrylamide est particulièrement sensible aux différences de conformation des molécules d'ADN et sa composition permet la séparation de ces molécules en fonction de la taille et de la conformation. La probabilité de révéler des mutations est donc beaucoup plus importante

³⁰ MDE : "Mutation Detection Enhancement".

Tableau 15. Liste des marqueurs STS dérivés des EST : séquence des amorces, résultats d'amplification et de cartographie (en vert, les STS cartographiés chez les deux espèces, en jaune chez le chêne et en bleu clair chez le châtaignier).

Nom du Locus ^a	Amorce « forward » (5' → 3')	Amorce « reverse » (5' → 3')	Taille attendue (pb) ^b	Protocole PCR (Q-C) ^c	Taille obtenue ^b	Technique de cartographie	Individus génotypés	Groupe de liaison Q ^d	Groupe de liaison C ^d
Projet « Chénédeb »									
CL4	CGGCAGGTACACAAATCTAA	AACGGCACCGTAAGGTTCTT	301	1-2	301	SSCP	57Q-56C	1	6
CL5	TTGTGTAACCAAGTTCGACTCT	AGTTGTTGGCGTGTCTTT	243	1-1	1200	SSCP	56Q-57C	5	4
CL7	TCAGACTCGAGCAAGGTAAGG	GCTCAGGATGGAATCTCAGG	180	1-1	180				
CL8	GAAGCACAAACAATGTTATCCCTAA	CCGGCAGGTACAATAGGA	178	1-2	300	SSCP-Q/DGGE-C	57Q-52C	9	2/4
CL9	CTGCACCGGAGAGGATCA	TACTCGGTGAAGGCTAGTGC	280	1-1	280				
CL10	GAATCGCCAAAGTGAAACCCAT	GGGGTTCTTTCAGAAACGATA	333	1-1	550	SSCP	55Q-56C	2	1
CL12	GCCGAGGTACAAAAATCCAT	ACCATGAGAGCCAAAGTGGAA	215	na-2	800C	SSCP	56C		9
CL15	CCGTCAAAATGATCCAAATCC	GCCCGAGGTAAGGCTAGCGAAG	601	na-na					
CL17	GGGATTCATGCGCACTTCT	CATACTTTGCATCGCCAAACA	424	1-1	1400Q-1500C	SSCP	51Q-55C	4	2
CL18	AATCGCCCTTAGCGTGGT	AACCACAACCATGCTAGTTTATCA	319	na-na					
CL20	GAGAAAAACACATAGAGATCAGATG	GCTCGGTAGCGAGAAAGAAG	153	3-3	200Q-250C	SSCP	57Q-52C	12 ^d	12
CL21	GGCAGTACCGTTCTGGAGAG	AACGGAGGGTGGGCTTCT	179	na-na					
CL24	CAAAATCTGATCCCGCAAAT	GCCATTGACACAAAGAACCAC	443	1-1	500	SSCP	57C		1
CL27	TCAACACCAACACCAACACC	CAGGTACCAAGCAAGTCCACA	191	4-4	191Q-400C	DGGE-Q	56Q	6	
CL29	CGGCAGGTACAAACTTACTATC	GCTGGCTAAATCGGAGCTT	266	3-3	266				
CL31	CCAAATCTCTTAGTGGCAAGG	GCTTTGGGATGATGTTTGG	255	1-1	500	SSCP	51Q-56C	5	4
CL34	ATTTCGAAGCTCGACTGCAT	CAGAACAAGCTGCTGTGGAA	428	na-na					
CL41	GAAGTTGCTTCTGGGATCA	GGCTCGTACGTGAACCTCCTC	244	5-na	500Q	SSCP	55Q	1	
CL48	CAGTTGCACCTTGAGCCATA	CAGGTACAAGCCGGAGAAA	344	6-na	1600Q				
S01A03	GGTATGGGAGCTGGTCGTAA	AGAGAGCCAACAGGGAGACA	382	1-1	500	SSCP	51Q-57C	7	5
S01E07	AAGCTTAGACCTCGCTTGC	GGCTGGCAGCTGTATTGT	145	1-1	400Q-145C	SSCP	56Q-57C	5 ^d	4
S02F02	ACTGAAAATCGCTCAGGACT	GCTGGATATCCTTGGGCATA	164	6-6	164	SSCP	57Q-56C	1	6
S02G03	ACGCTCCAAAACITGGTCAG	CGGGCAGTACTTAATGAGG	206	1-1	700	SSCP	57Q-49C	3	8
S03E05	AAGCTCCCATGTTTGGATG	CTGCACCTTGACGCTTGTGT	123	5-5	123	DGGE	57Q-49C	2	1 ^d
S05C04	AGCCCTTAGGACAGCCCTCAT	AGAGAGATGAGTGGTCTGGAG	170	na-na					
S05D10	GGCCGAGGTACTCAACATCA	GCCTCCGTTCAAGATCAATGT	268	na-na					
S06B07	GTCCCTCCAATTCGCCAATTA	ATGGGAACAGTTAGCCGATG	307	1-1	307	SSCP	57Q-56C	8	7
S06B09	AGCAAGTATGGCCCTGGTTA	CGGGCAGGTACATGATGCT	270	6-6	270				
S06C07	GGCCGAGGTACACTCTTTTC	GGTTTACCTCCCTGGGTTTC	185	na-na					

CL83	ATGGCTCGTACCAAGCAAAC	GGGCAGGTACTTTCGGATCT	171	1-1	171	SSCP	57Q-57C	5	4
CL89	TTGTAGCCGAGGATTGTTC	GTACGGGGGCTGAAAAG	107	na-na					
CL90	ACCAGGCAATGTGAATGTG	GTAGCTGTGACCCACCAAT	104	1-1	104				
CL93	ACCTTACCCTGGGACTTCT	TGCTCAAGAGGCTGTGAAGA	238	1-1	400	SSCP	55Q-57C	7	5
CL94	CAAACCTCCCATTCACATC	CCCAAGTCAGCACAAACTT	289	6-6	289	SSCP	57Q-55C	6	11
CL95	GCAAGTCTCTCCACAAACT	GGTGGCAAGCACCACTAAT	120	3-3	120	SSCP	57Q	6	
CL96	TCCTCTTCCAGCACCTCTA	TTATGGCAGGCTTGGGTTAG	210	1-1	500	SSCP	57Q-54C	9 ^d	9
CL97	CCATTGAGCTCCTTCTCAGC	CTTCAACAGTGCAGCAAAG	190	6-6	190	SSCP	57Q	2	
CL101	TGCCTTCTTACCAGCTCCTC	GAAAGAGGGAGACCACGTCA	115	1-1	200	SSCP	57C		1
Gènes candidates (Derory J.)									
S06G05									
S08C11			279		279				
PHY B/D			316		316	SSCP	57Q-57C	2	1
CL24			260		260	SSCP-Q/DGGE-C	57Q-49C	1/5	11
DET2			410		2000				
S08A01			214		1500				
S07B10			210		500	SSCP	46Q	3 ^d	informatif
CL11			360		360	SSCP	56Q	8	
CL16			327		327	SSCP	56Q	9	informatif
GA3			270		600Q	SSCP	51Q	8	informatif
GA20			347		600	SSCP	48Q	5	
S08G04			409		500Q				
CL1			393		1000	SSCP	57Q	11	
			235		1400				

a S : singleton (transcrit présent en une seule copie dans la banque d'ADNc) ; CL : cluster (séquence consensus dérivant de l'alignement de plusieurs EST) ; PHY, DET2, GA3 et GA20 : il s'agit de quatre gènes candidats fonctionnels (Derory et al., soumis)

b La taille attendue de la bande est celle calculée à partir de la séquence EST. Les estimations des tailles de bandes amplifiées sont calculées sur la base d'une migration sur gel d'agarose et donc elles sont très approximatives. Les STS, qui ont donné un produit d'amplification mais ne sont pas cartographiés, soit ils sont monomorphes soit les conditions de migration (SSCP et DGGE) n'ont pas pu être optimisées.

c Le protocole PCR 1 correspond à celui décrit au paragraphe D.3.4. ; les 10 cycles de « touchdown », pendant lesquels la température varie de 58 à 53°C, sont suivis de 25 cycles à 53°C. Protocole 2 : 55-50°C/25 cycles à 50°C ; protocole 3 : 55-50°C/30 cycles à 50°C ; protocole 4 : 65-60°C/30 cycles à 60°C ; protocole 5 : 58-53°C/30 cycles à 53°C ; protocole 6 : 60-55°C/30 cycles à 55°C ; protocole 7 : 68-63°C/25 cycles à 63°C ; protocole 8 : 67-62°C/30 cycles à 62°C ; na : non amplifié.

d Tous les STS ont été cartographiés en utilisant un seuil de LOD > 6.0, à l'exception des STS indiqués par la lettre d pour lesquels le seuil est de 4.0<LOD<6.0.

(jusqu'à 80% plus élevée) qu'en utilisant l'acrylamide normale. De plus, l'optimisation des conditions de migration est beaucoup plus rapide que dans le cas de l'acrylamide normale. La migration électrophorétique est faite entre 4 et 20°C et la séparation des molécules d'ADN se réalise principalement en fonction de la conformation, qui elle-même dépend de la séquence de la molécule. La révélation est réalisée par coloration au nitrate d'argent. La technique est très simple, l'équipement nécessaire n'est pas cher, et surtout 90% des mutations présentes sont détectées (Nataraj et al., 1999). L'estimation du coût par génotype est de 0,15 euros. L'inconvénient de cette technique réside dans l'absence d'une théorie qui puisse prévoir la conformation des molécules d'ADN dans les différentes conditions expérimentales. Cela empêche parfois l'optimisation des conditions de séparation des différentes conformations, et par conséquent les profils peuvent être difficiles à interpréter. La figure 30a montre quelques profils obtenus et les différents types de ségrégation possibles. Il est à noter que plusieurs mutations à la fois peuvent être révélées, ce qui permet une lecture codominante des profils de ségrégation.

Au contraire de la technique SSCP, la méthode DGGE est basée sur une théorie connue dans les moindres détails (Myers et al., 1987). Les molécules d'ADN amplifiées sont chargées sans être dénaturées (double brin) sur un gel qui contient un gradient linéaire de produits dénaturants (urée et formamide). Le principe de séparation est basé sur les propriétés de fusion (ou dénaturation) de l'ADN en solution (figure 30b). Chaque partie d'une molécule d'ADN (domaine de fusion) se dénature à une température et dans des conditions de concentration de produits dénaturants précises. La température et les concentrations de fusion dépendent de la séquence de la molécule d'ADN. Lors de la migration d'une molécule d'ADN double brin le long d'un gradient dénaturant, la proportion d'ADN dénaturé augmente en diminuant la vitesse de migration de la molécule, ce qui permet de séparer les différentes mutations. Normalement, le gradient dénaturant est parallèle au champ électrique (figure 30b) ; dans ce cas les différents fragments peuvent être séparés. En revanche, si le champ électrique est perpendiculaire au gradient, une courbe de dénaturation d'une molécule donnée d'ADN est obtenue, ce qui permet de déterminer avec précision les concentrations de dénaturation optimales pour chaque molécule. Lors d'une séparation difficile entre deux mutations données, un segment riche en GC peut être ajouté aux amorces afin de changer les propriétés de dénaturations. La technique nécessite un équipement particulier (BioRADTM, Dcode, « Universal Mutation Detection System ») et la migration est souvent faite à 58°C. La révélation est faite par coloration au bromure d'ethidium et le coût d'une réaction de génotypage est vraiment faible (0,05 euros). Néanmoins, l'optimisation des conditions de

séparation est plus lourde par rapport à la technique SSCP et surtout l'efficacité de révélation des mutations reste plus faible dans le cas de la technique DGGE (Sheffield et al., 1993).

Etant donné que :

- la machine Victor n'était pas disponible (elle a été testée au laboratoire pendant un mois en avril 2003 et elle n'a été achetée qu'en septembre 2004) ;
- le coût faible des réactions de génotypage était une priorité ;
- l'optimisation des conditions de génotypage pour un grand nombre de SNP est plutôt difficile, longue et coûteuse avec le MegaBACE ;

les techniques SSCP et DGGE ont été choisies pour cartographier les EST.

D.3.3. Les critères de choix des amorces pour l'amplification des EST.

Vu l'objectif prioritaire d'aligner les douze groupes de liaison entre le chêne et le châtaignier, les amorces utilisées pour l'amplification des EST, ont été choisies afin d'avoir une probabilité élevée d'amplification et donc de transfert entre les deux espèces. Pour cela, parmi les séquences de chêne une centaine a été choisie sur la base de l'annotation correspondant à des gènes de fonction connue. Une recherche d'homologues de la séquence du chêne a été réalisée dans les bases de données grâce à l'aide du logiciel BLAST³¹ (<http://www.ncbi.nlm.nih.gov/blast/>, Altschul et al., 1990). Les séquences homologues (deux ou trois à chaque fois) ont été alignées à l'aide du logiciel CLUSTALW (<http://www.ebi.ac.uk/clustalw/>). La région la plus conservée entre les séquences alignées a donc été choisie afin d'établir la paire d'amorces d'amplification (figure 31a), ce qui devrait permettre une bonne efficacité d'amplification chez le châtaignier. Les amorces ont été dessinées et caractérisées avec l'aide du logiciel Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) en respectant trois critères principaux : 1) les séquences des amorces devaient être situées dans la partie la plus conservée ; 2) la taille du fragment amplifié devait être environ de 100 à 450 paires de bases ; 3) la température de fusion devait être comprise entre 55 et 60 °C. La séquence des amorces correspond évidemment à celle de l'EST du chêne. Nous avons déjà expliqué la raison du premier critère de choix des amorces. En ce qui concerne la taille attendue du fragment, deux raisons ont guidé ce choix, d'une part les techniques SSCP et DGGE, utilisées pour la révélation du polymorphisme fournissent les meilleurs résultats avec des fragments de 200 à environ 600 paires de bases. Le respect de ces limites pour la longueur du fragment amplifié est beaucoup plus important en DGGE qu'en

³¹ BLAST : "Basic Local Alignment Search Tool".

Figure 31a

EST S02F02 (Histone H3.2) : séquence ESTchêne alignée avec celle de *B. vulgaris* et *V. vinifera*. En vert, les séquences des amorces utilisées dans notre étude.

```

Beta vulgaris
Vitis vinifera
EST Quercus
A CCAGAGAGC ACAGAGC TCCTG ATC CGAAAAGCTGC CATTTC CAAGAGGTG 101
A CCAGAGAGC ACTGAGC TTTTGG ATC AGGAAAC TCC CATTTC CAAGAGGTG 250
-----CATAGCCTT 9
** ** **

Beta vulgaris
Vitis vinifera
EST Quercus
GTTCTGTAANT TGCTC AGGACTT CAA GACCGAT CTC CGFTT CCAGAGCCA 151
GTCAGAGAANT TGCCG AGACTT CAA GACTGAT CTC CGFTT CCAGAGCCA 300
GTACTGGAAAT CGGTC AGGACTT CAA GACTGAT CTC CGFTT CCAGAGCCA 59
** ** **
*****
*****

Beta vulgaris
Vitis vinifera
EST Quercus
C GCTGTGTG TGG CCGTTC CAA GAGGCTG CTGAGGCTT A CCTTG TSGGTCCTT 201
T GCTGTGTG TGG CACTGCA GGAGG CCG CGGAGGC CTR CCTTG TSGGTCCTG 350
C GCTGTGTG TGG CACTGCA GGAGG CTG CTGAGGCTT A CCTTG TSGGTCCTG 109
*****
** ** **
*****
*****

Beta vulgaris
Vitis vinifera
EST Quercus
TTTGAGGACACT AACCTT TCGTCCG ATC CACGCCA AGC GTGTC ACATTA TG 251
TTTGAGGACACT AACCTT TCGTCCG ATC CACGCCA AGC GTGTC ACATTA TG 400
TTTGAGGACACT AACCTT TCGTCCG ATC CACGCCA AGC GTGTC ACATTA TG 159
** ** **
*****
*****

Beta vulgaris
Vitis vinifera
EST Quercus
CCTAAGGATATT CAGCTT TCGTCCG ATC CACGCCA AGC GTGTC ACATTA TG 301
CCCAAGGATATT CAGCTT TCGTCCG ATC CACGCCA AGC GTGTC ACATTA TG 450
CCCAAGGATATT CAGCTT TCGTCCG ATC CACGCCA AGC GTGTC ACATTA TG 209
** ** **
*****
*****

Beta vulgaris
Vitis vinifera
EST Quercus
TATTAATGTTGATG TCTGCTCCG ATC CACGCCA AGC GTGTC ACATTA TG 349
GG--AAGGTGATG GAG CCGCCG TA GTTGTGATGG--TGGT GHT CTT--TAT 495
TC---AAGTCA TATGA AT GAAAA ATTGTGGAGG GCAAGAAA TTGCGGC TG 256
* * *

Beta vulgaris
Vitis vinifera
EST Quercus
TAGTA GTTGTGATG TTT GTTTA GTTTAA TTCAGG -ACTA ATAGGGGTT 398
AAGTGTTT AGA TTTAGT CTTTA GGCAGAGGTA TG CAAAGG ATAGAA CT 545
C GCAAGGT GCA CCGAA TT GAATA TGA TGGGGTGT TTTTGGG AA GGGTGT 306
** ** **
*****
*****

Beta vulgaris
Vitis vinifera
EST Quercus
GT TTTTGG--TTGGGAT--GATG GTA TTC----- 424
ATAT GATGCTTTGATAT--GGTG GTGTG CA GTTAT CAA GGT--GGGTG TT 591
GTC TTTTGGAA TGTTA TCTGTTA GTT AGCAGCT AGT TAGGT CTAGAAGA 356
* * *
*****
*****

```

Dans le cas d'un "backcross", 50 individus permettent de détecter une liaison génétique avec un marqueur à environ 20 cM de distance, dans 90% des cas. La liaison génétique est toujours détectée si le marqueur se trouve à une distance de 10 cM ou moins.

Figure 31b

Quatre exemples d'amplification des STS chez les deux espèces. A chaque fois 4 échantillons, 2 de chêne (Q) et 2 de châtaignier (C) sont représentés.

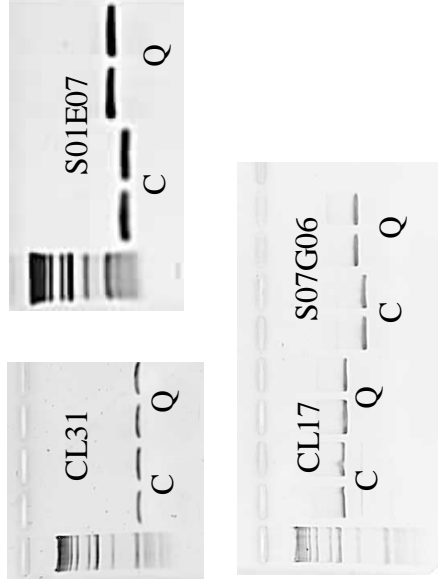


Figure 31c

Puissance statistique de détection de la liaison génétique (tiré de Liu, 1998).

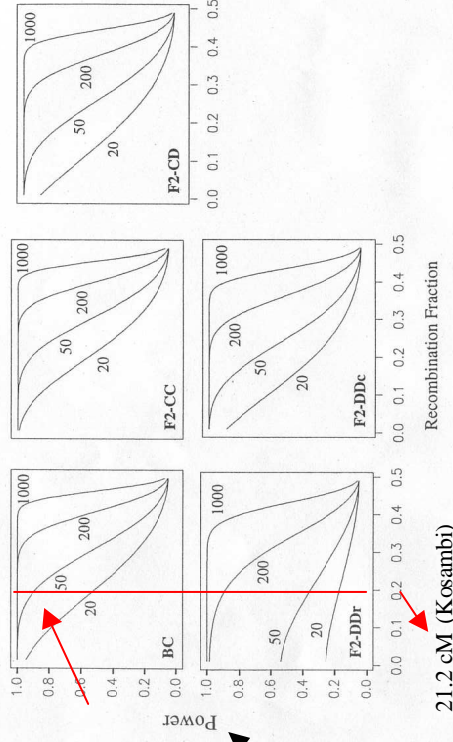


Figure 6.5 Statistical power for detecting linkage with sample sizes of 20, 50, 200 and 1000 using the backcross model (BC) and F2 models with different marker combinations. See Table 6.17 for notation.

21.2 cM (Kosambi)

SSCP. D'autre part, les amorces sont établies sur une séquence EST, par conséquent la position du fragment d'EST amplifié par rapport à la séquence du gène n'est pas connue. Cela veut dire qu'il y a la possibilité de se situer entre deux exons séparés par un intron dans la séquence du gène. Puisque l'ADN amplifié pour la cartographie est l'ADN génomique, il est possible que des introns soient inclus dans le fragment amplifié, ce qui augmente la taille du fragment amplifié. Il est donc prudent de prévoir des tailles plus petites au départ pour obtenir, même en présence d'introns, des fragments dont l'amplification ne perde pas en efficacité (les fragments longs sont plus difficiles à amplifier) et faciles à séparer en SSCP ou DGGE. Enfin, une température d'hybridation autour de 55-60 °C a été imposée pour tous les couples d'amorces afin de pouvoir amplifier toutes les EST dans les mêmes conditions. Il faut préciser que le produit amplifié ne constitue plus une EST. Les marqueurs obtenus par amplification d'un fragment de gène grâce aux amorces établies sur un fragment EST sont souvent appelés STS. Soixante-treize amorces ont été donc établies pour amplifier autant de STS (tableau 15). Treize gènes candidats, dont 9 EST et 4 gènes candidats fonctionnels ont été ajoutés à cette liste (tableau 15). Tous les STS ont été testés chez le chêne et le châtaignier.

D.3.4. L'amplification des EST chez les deux espèces.

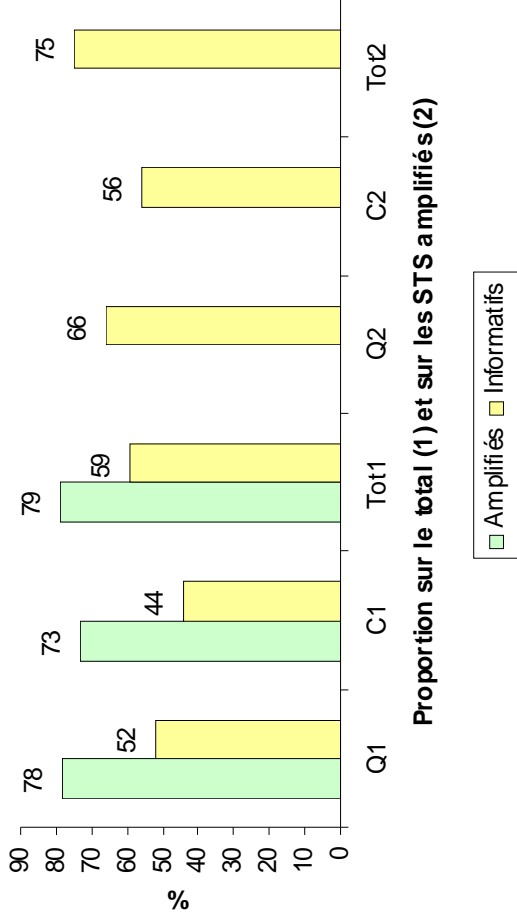
L'amplification des STS a été faite en utilisant un protocole de PCR « touchdown » (Don et al., 1991). Par rapport à une amplification normale, la température d'hybridation pendant une PCR « touchdown » n'est pas constante, mais elle diminue au fur et à mesure que les cycles de PCR se succèdent. Cela permet de commencer l'amplification à une température d'hybridation plus élevée, donc dans des conditions plus spécifiques afin d'éviter les mésappariements dans les premiers cycles de PCR. Ensuite, la diminution de la température d'hybridation et la disponibilité de plusieurs molécules d'ADN cible spécifiques, produites lors des premiers cycles, donnent lieu à une amplification en général plus efficace. Comme les séquences des amorces ont été choisies dans des régions conservées, le protocole devrait être efficace chez le chêne ainsi que chez le châtaignier et il devrait permettre d'obtenir des produits d'amplification spécifiques et surtout homologues chez les deux espèces.

La réaction de PCR a été faite dans un volume total de 12,5 µl contenant 1,25 µl de tampon 10X (Invitrogen, Carlsbad, Calif., USA), 2 mM de MgCl₂, 200 µM de chaque dNTP, 0,2 µM de chaque amorce, 100 µg/ml de BSA³², 5-10 ng d'ADN génomique et 0,4 unités de *Taq* Polymerase (Invitrogen, Carlsbad, Calif., USA). Le protocole d'amplification utilisé consiste

³² BSA : albumine de sérum bovin.

Figure 31d

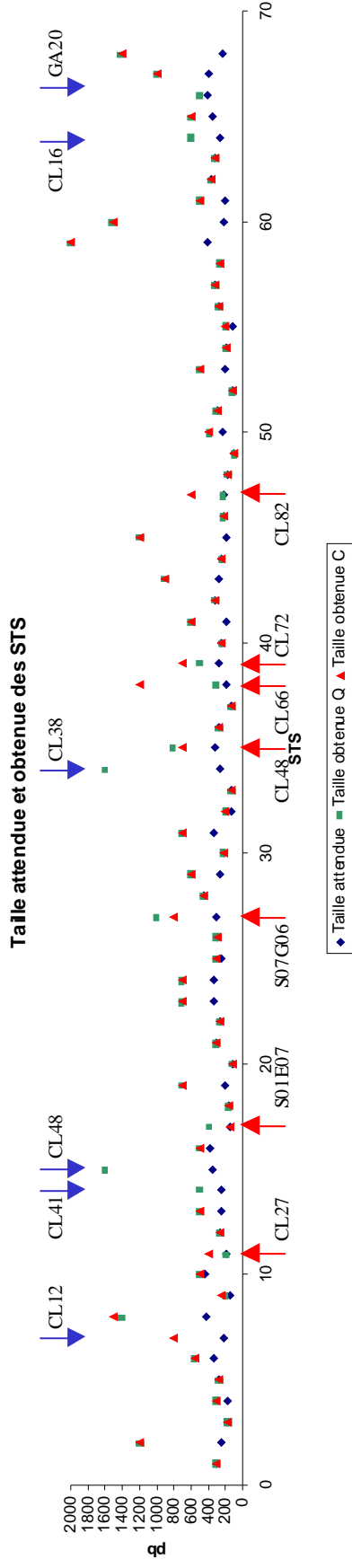
En vert, proportion d'amplification des STS chez le chêne (Q), chez le châtaignier (C) et totale (Tot). En jaune, proportion des STS informatifs en considérant le nombre total des STS testés (1) et le nombre des STS qui ont donné une bonne amplification (2).



Proportion sur le total (1) et sur les STS amplifiés (2)

Figure 31e

Taille attendue et obtenue de la bande amplifiée : en ordonnée la taille en pb et en abscisse les STS. Losange bleu : taille attendue sur la base de la séquence EST. Carré vert : taille obtenue chez le chêne. Triangle rouge : taille obtenue chez le châtaignier. Flèches bleues, STS amplifiés soit chez le chêne (Q) soit chez le châtaignier (C). Flèches rouges, STS qui donnent une bande amplifiée différente en taille entre les deux espèces.



en une première étape de 10 minutes à 94°C ; une deuxième étape de dix cycles de « touchdown » pendant lesquels la température d'hybridation diminue de 5°C, c'est à dire 0,5 °C chaque cycle de 45 secondes à 94°C, 45 secondes à la température d'hybridation et 1 minute à 72°C; une troisième étape de 25-30 cycles de 45 secondes à 94°C, 45 secondes à la température d'hybridation et 1 minute à 72°C ; et enfin 10 minutes à 72°C. Les différentes conditions d'amplification optimisées pour chaque STS sont présentées au tableau 15. Tous les STS ont été testés au départ avec le même protocole qui est caractérisé par une diminution de température de 58 à 53 °C pendant les 10 cycles de « touchdown » et 25 cycles à 53°C. Les conditions ont ensuite été modifiées selon les résultats de la première amplification et optimisées afin d'obtenir une seule bande amplifiée chez les deux espèces. La présence de plusieurs niveaux d'amplification est à éviter absolument car, dans ce cas, la séparation basée sur la conformation en SSCP et le point de fusion en DGGE peut donner des profils très difficiles à interpréter. De plus, la séquence des bandes révélées avec les deux techniques est nécessaire pour ensuite identifier les vrais orthologues entre les deux espèces.

Les résultats d'amplification des STS après optimisation montrent une cohérence considérable entre les deux espèces. Parmi les 86 STS testés, 78% ont amplifié une seule bande nette chez le chêne et 73% chez le châtaignier (figure 31d). Cinq STS, qui ont donné un produit chez le chêne, n'ont pas été amplifiés ou ont montré des profils multibande chez le châtaignier et un seul STS amplifié chez le châtaignier n'a pas pu être amplifié chez le chêne. Les amorces qui ont été écartées de l'étude amplifient soit plusieurs bandes, même après avoir augmenté la température d'hybridation de 4-6 degrés Celsius, soit ne donnent aucune amplification. Les différentes bandes amplifiées par un seul couple d'amorces peuvent correspondre à différents éléments de la même famille multigénique ou tout simplement à des produits non spécifiques. Comme la cartographie de ces produits pourrait être problématique pour l'étude de la synténie entre chêne et châtaignier nous avons choisi de ne pas faire d'efforts d'optimisation et de cartographie dans ce cas là.

La taille attendue des produits d'amplification sur la base de la séquence EST de départ, varie de 104 à 454 pb (le seul cas, dont la taille attendue était 601 pb, n'a pas été amplifié, tableau 15, figure 31e). Les tailles des bandes obtenues varient de 120 à 2000 pb. Différentes situations sont alors possibles : 1) la taille correspond à celle prévue dans les deux espèces ; 2) la taille correspond à celle prévue dans une seule de deux espèces ; 3) la taille ne correspond pas à celle prévue dans les deux espèces. La figure 31b montre 4 exemples d'amplification. Dans 49% des cas, la taille de la bande amplifiée correspond à celle attendue sur la base de la

séquence EST. En revanche, 41% des STS amplifient des fragments plus longs que prévu. Les résultats sont cohérents entre les deux espèces avec peu d'exceptions (7 (10%) STS donnent un fragment de taille différente chez le chêne et le châtaignier, figure 31e).

Trois conclusions intéressantes peuvent être tirées de ces premiers résultats :

- 1) le transfert de STS dérivés des EST du chêne chez le châtaignier semble très efficace. L'amplification de 70 à 80% des STS et la concordance des résultats d'amplification entre les deux espèces mettent en évidence que les EST sont une source optimale de marqueurs pour la cartographie génétique comparée.
- 2) Si ce résultat est comparé avec l'efficacité de transfert des SSR entre ces deux espèces, l'avantage de l'utilisation des EST est encore plus évident. En effet, entre 47 et 63% des SSR d'une espèce ont pu être amplifiés chez l'autre. Il ne faut pas oublier que les 86 STS du départ ne constituent pas encore des marqueurs optimisés dans l'espèce source, comme l'étaient les 53 SSR de chêne et les 30 de châtaignier. Si nous considérons les 67 STS qui amplifient une seule bande chez le chêne comme les marqueurs à transférer, alors 94% (63) ont été bien amplifiés chez le châtaignier.
- 3) Enfin, il y a concordance des résultats entre les deux espèces en ce qui concerne la taille de la bande amplifiée et donc la présence d'introns dans le fragment amplifié. Cela suggère que la structure des gènes soit conservée entre les deux espèces dans la plupart des cas.

D.3.5. La cartographie des EST et l'alignement des cartes génétiques du chêne et du châtaignier.

Les deux techniques SSCP et DGGE ont été utilisées pour la révélation du polymorphisme dans les fragments amplifiés. Tout d'abord, les produits d'amplification des deux parents et 6 individus F_1 de chaque pedigree ont été chargés sur un gel SSCP. Si aucun polymorphisme n'était révélé, les mêmes individus étaient chargés sur un gel DGGE. Si, dans ce dernier cas, aucun polymorphisme n'était détecté, le STS n'a pas pu être cartographié. Nous avons décidé d'utiliser en premier la technique SSCP pour sa plus grande efficacité à révéler des mutations et la simplicité d'optimisation des conditions de migration par rapport à la technique DGGE. La plupart des STS ont été cartographiés par SSCP, la technique DGGE a été utilisée pour 9 cas et 17 STS n'ont pas pu être cartographiés car ils montraient un profil monomorphe avec les deux techniques et dans les deux espèces (tableau 15), ou bien les conditions SSCP et DGGE n'ont pas pu être optimisées.

Cinquante et un individus supplémentaires ont été génotypés pour tous les STS polymorphes. Le nombre total d'individus génotypés est donc de 57. Ce nombre est un compromis entre une puissance satisfaisante de détection de la liaison génétique et une contrainte technique. La figure 31c montre que dans le cas d'un backcross, 50 individus sont suffisants pour détecter une liaison génétique avec un marqueur qui se trouve à 20 cM de distance dans 90% des cas. La liaison génétique est toujours détectée si le marqueur se trouve à 10 cM de distance. Les deux cartes génétiques utilisées sont saturées, les douze groupes de liaison sont bien définis et tout marqueur ajouté aux deux cartes sera presque toujours lié à moins de 20 cM de distance d'un autre marqueur de la carte. Il apparaît donc évident que 57 est un effectif suffisant pour détecter la liaison génétique. En revanche, comme nous le verrons plus loin, cet effectif faible peut poser des problèmes lors de la détermination de l'ordre des marqueurs proches. Dans la chambre électrophorétique utilisée pour la migration SSCP, un maximum de 56 échantillons à la fois (deux gels de 28) peut être chargé. Lors de la phase de test, 3 STS chez les deux espèces (48 échantillons et trois puits de marqueurs de migration) sont testés à la fois. Ensuite les deux gels sont chargés avec les 51 individus, les deux parents et 3 marqueurs de migration.

Parmi les 86 STS de départ, 45 (52%) se sont révélés polymorphes et informatifs chez le chêne et 38 (44%) chez le châtaignier. Si nous considérons les 68 STS qui ont donné une seule bande d'amplification et qui ont donc été utilisés par la suite, cela veut dire qu'au moins un polymorphisme informatif a été détecté dans 66% des cas chez le chêne et dans 56% des cas chez le châtaignier (figure 31d). Il est donc évident que les deux techniques utilisées permettent de détecter de façon efficace les mutations de séquence et, rappelons le ici, les EST s'avèrent plus utiles que les SSR pour la cartographie comparée. En effet, l'amplification, dans plusieurs cas, de séquences dérivant probablement des introns ou des extrémités 3'-UTR³³ augmente la probabilité de détecter des polymorphismes (jusqu'à trois fois d'après Zhu et al., http://www.intl-pag.org/pag/9/abstracts/P3e_10.html). En effet, ces séquences sont plus variables que celles des exons. D'ailleurs, sur les 17 STS qui n'ont pas pu être cartographiés, 11 ne semblent pas avoir amplifié d'introns.

Le seuil de LOD pour la détection de la liaison génétique a été fixé à 6.0. Dans 6 cas le LOD a été abaissé jusqu'à 4 (tableau 15), mais la fiabilité du résultat est confirmée par la liaison avec plusieurs marqueurs du même groupe et surtout par le fait que le même STS a été cartographié chez le groupe homologue de l'autre espèce. Comme il a été montré à la figure

³³ UTR : "Untranslated regions".

31c, le nombre de 57 individus s'est révélé suffisant pour la détection de la liaison dans les deux cartes génétiques.

La figure 32 représente les douze groupes de liaison associés entre le chêne et le châtaignier. La cartographie de 53 STS chez le chêne et 45 chez le châtaignier, dont 37 communs chez les deux espèces, a enfin permis de relier les douze groupes de liaison. La stratégie de construction de la carte consensus de chaque espèce sera décrite au paragraphe E.2.1. Cinquante-cinq marqueurs orthologues (18 SSR, 1 isozyme, 1 5SrDNA, 35 STS) sont distribués sur les douze groupes de liaison qui contiennent de 2 à 7 marqueurs orthologues. Il faut noter que 4 STS et 1 SSR ont été cartographiés chez une famille de chêne interspécifique qui ne correspond pas à celle utilisée dans ce travail. Cependant, la carte génétique dérivant du croisement interspécifique a été reliée en partie à celle du croisement 3PxA4 (Jauffres, 2003) et les quatre marqueurs ont pu être positionnés indirectement sur la carte 3PxA4. Deux locus SSR ont été cartographiés en plus chez le chêne dans cette version de la carte: CsCAT15 (Q3, Jauffres, 2003) et EMCs11 (Q7, Barreneche et al., en préparation). Une différence importante entre cette comparaison et celle montrée dans la figure 26a est l'association du groupe Q6 avec C11 et Q7 avec C5. Dans la première comparaison, le groupe Q7 avait été associé à C11 à l'aide du microsatellite EMCs1. L'identité de séquence à ce locus est élevée et donc le couple de groupes semblait être homologue. Les nouveaux résultats obtenus à l'aide des EST ne s'accordent pas avec la conclusion précédente.

Deux questions se posent :

- 1) le microsatellite EMCs1 est-il un orthologue ou bien un paralogue ? Le marqueur EMCs1 est un marqueur qui montre plusieurs niveaux d'amplification chez le châtaignier. Jauffres (2003) a cartographié ce locus dans le croisement interspécifique et il ne se trouve pas sur l'homologue du groupe Q7. De plus, Paul Sisco a également intégré ce locus dans la carte génétique du châtaignier américain et chinois et il ne se trouve pas sur l'homologue du groupe C11 (communication personnelle). Malgré l'identité de séquence élevée entre les deux espèces, due probablement à la comparaison après le clonage d'un fragment qui ne correspond pas au fragment cartographié, il semble que le locus cartographié chez le chêne pédonculé et le châtaignier européen soit un locus paralogue.
- 2) Le microsatellite QpZAG9 (Q7-C5), qui avait été considéré comme paralogue lors de la première comparaison, semble être un locus orthologue dans cette nouvelle association. En effet, le fragment amplifié chez le châtaignier contient une longue insertion par rapport au chêne à l'intérieur de laquelle se trouve un deuxième motif répété. L'alignement de

Tableau 16. Gènes correspondant aux STS cartographiés.

EST ^a	Résultat BLASTX EST Quercus (26/8/04) ^b	Identité Q/C ^c
1T11	gij28416671 gb AAO42866.1 At5g41980 [Arabidopsis thaliana] 2e-33, unknown protein	99,0
1T12	gij50428684 gb AAT77035.1 putative oxidoreductase [Oryza sativa (japonica cultivar-group)] 6e-24	93,5*
1T21	gij1902894 dbj BAA19462.1 QM family protein [Solanum melongena]60S RIBOSOMAL PROTEIN L10 1e-05	94,0
1T25	no hits	97,2
1T57	gij38324686 gb AY438652.1 Quercus petraea promoter-binding factor 2-like gene, 3' UTR (BLASTN, Porth et al.)	93,5
1T62	gij21592995 gb AAM64944.1 betaine aldehyde dehydrogenase, putative [Arabidopsis thaliana] 6e-15	91,9
2T11	no hits	89*
2T3	gij38324687 gb AY438653.1 Quercus petraea mtN19-like gene, 3' UTR (BLASTN, Porth et al.)	94,3
2T13	no hits	94,5
2T32	gij38324702 gb AY438661.1 Quercus petraea ribosomal protein S11 gene, partial cds and 3' UTR (BLASTN, Porth et al.)	93,0
CL4	gij70772 pir HSWT4 histone H4 - wheat 2e-39	89,5-93,1
CL5	gij13430184 gb AAK25759.1 ribosomal protein L18a [Castanea sativa] 2e-42	-
CL8	gij38093740 gb AAR10856.1 putative 60S ribosomal protein L13 [Oryza sativa (japonica cultivar-group)] 2e-27(cold-induced protein)	81,3
CL10	gij8388627 emb CAB94147.1 ribosomal protein S27 [Arabidopsis thaliana] 2e-35	89,3
CL12	gij407800 emb X75423.1 GHRP60S G.hirsutum mRNA for ribosomal protein 41, large subunit (RL41) 5e-17 (BLASTN)	-
CL17	gij6648215 gb AAF21213.1 unknown protein [Arabidopsis thaliana] 5e-95	93,1
CL20	gij20138097 sp Q9ZVWQ DAD1_CITUN Defender against cell death 1 (DAD-1) (CitDAD-1) 1e-9	95,3
CL24	gij1928991 gb AAC03416.1 heat shock protein 70 precursor [Citrullus lanatus] 4e-38 (dnaK-type molecular chaperone CSS1 precursor)	90,7*
CL27	gij38343924 emb CAE54590.1 dehydrin/response ABA [Fagus sylvatica] 1e-6	-
CL31	gij9795604 gb AAF98422.1 dormancy-associated protein [Arabidopsis thaliana] 5e-29 (auxin-repressed protein)	92,7
CL41	gij21592589 gb AAM64538.1 cinnamoyl-CoA reductase-like protein [Arabidopsis thaliana] 3e-42	92,9*
S01A03	gij1362041 pir S56673 ribosomal protein S23.e, cytosolic (clone RJ3) - garden strawberry 3e-76	94,3
S01E07	gij47496696 dbj BAD19063.1 auxin response factor 3 [Cucumis sativus] 2e-23	86,0
S02F02	gij829279 emb CAA31967.1 histone H3 (AA 1-120) [Medicago sativa] 4e-19	86,3
S02G03	gij18405061 ref NP_564667.1 thylakoid lumen 18.3 kDa protein [Arabidopsis thaliana] 4e-45 (unknown protein)	90,3
S03E05	gij132156 sp P13951 RBS_SINAL Ribulose biphosphate carboxylase small chain (RuBisCO small subunit) 3e-23	91,7
S06B07	gij34979117 gb AAQ83694.1 pseudo-response regulator protein [Oryza sativa (indica cultivar-group)] 5e-17	93,3
S06E10	gij1561774 gb AAB08874.1 malate dehydrogenase [Vitis vinifera] 8e-70	94,5
S07A08	gij30687712 ref NP_851044.1 raffinose synthase family protein / seed imbibition protein, putative (din10) [Arabidopsis thaliana] 3e-38	86,1*
S07A09	gij28144351 tpg DAA00870.1 TPA: PDR1 ABC transporter [Arabidopsis thaliana] 2e-51	95,7
S07C03	gij18394486 ref NP_564024.1 expressed protein [Arabidopsis thaliana] 2e-31	96,6*
S07G06	gij4586253 emb CAB40994.1 auxilin-like protein [Arabidopsis thaliana] 4e-25	90,9*
S08A03	gij5931655 emb CAB56581.1 squamosa promoter binding protein-like 1 [Arabidopsis thaliana] 3e-33	94,1*
S08A07	gij29569824 gb AAO84915.1 galactinol synthase [Cucumis sativus] 2e-29	88,0*
S08C05	gij38175819 tpg DAA01952.1 TPA: putative ABC (ATP binding cassette) transporter 1; AtABC1 [Arabidopsis thaliana] 6e-45	95,5*
S08D11	gij27542942 gb AAO16642.1 beta-1,3-glucanase [Fragaria x ananassa] 8e-28	88,4*
S08H11	gij10567808 gb AAG16758.1 putative glutathione S-transferase T3 [Lycopersicon esculentum] 2e-30	86,7
CL38	gij599726 emb CAA86851.1 Lea5 protein [Citrus sinensis] 5e-22	95,7*
CL66	gij2497906 sp Q96386 MT3_CARPA Metallothionein-like protein type 3 (MT-3) 1e-25	-
CL75	gij25354731 pir C86158 hypothetical protein F22D16.19 - Arabidopsis thaliana 8e-20	95,4
CL76	gij1172571 sp P42066 PPCK_CUCSA Phosphoenolpyruvate carboxykinase [ATP] (PEP carboxykinase) (PEPCK) 2e-47	95,4
CL77	gij130285 sp P00289 PLAS_SPIOL Plastocyanin, chloroplast precursor [Spinacia oleracea] 8e-53	91,2*
CL78	gij18015 emb CAA37898.1 hemoglobin II [Casuarina glauca] 5e-56	92,6
CL80	gij15149875 emb CAC39481.2 metallothionein-like protein [Quercus suber] 2e-25 class II	100*
CL81	gij25044847 gb AAM28295.1 PVR3-like protein [Ananas comosus] 3e-20	92,6
CL82	22326731 ref NP_196706.2 oxygen-evolving complex-related [Arabidopsis thaliana] putative protein 2e-17	89,5*
CL83	gij122070 sp P08860 H32_ORYSA Histone H3 1e-23	93,4*
CL93	gij10180025 gb AAG13986.1 60S ribosomal protein L24 [Prunus avium] 3e-17	94,7
CL94	gij28194659 gb AAO33591.1 putative early light induced protein [Arachis hypogaea] 3e-59	94,3
CL95	gij131390 sp P16059 PSBP_PEA Oxygen-evolving enhancer protein 2, chloroplast precursor (OEE2) (photosys.II) 6e-26	-
CL96	gij38347062 emb CAE05467.2 OSJNBa0006A01.3 [Oryza sativa (japonica cultivar-group)] 8e-27	79,6
CL97	gij10834738 gb AAG23799.1 chloroplast NADP-dependent glyceraldehyde 3-phosphate dehydrogenaseA subunit [Cucurbita pepo] 2e-46	91,9*
CL101	gij21593611 gb AAM65578.1 cytosolic ribosomal protein S11 [Arabidopsis thaliana] 5e-34	100*
S08C11	gij41352037 gb AAS00691.1 metal-nicotianamine transporter YSL1 [Arabidopsis thaliana] 2e-28	94,4
PHY B/D	gij475498 gb AAA17821.1 phytochrome B/D [Quercus turbinella] 2e-29	84,7-95,2

^a Les STS S08A01, S07B10, CL11, CL16, GA3 et S08G04 (tableau 15) n'ont pas été séquencés.

^b Résultat de l'alignement BLASTX avec les séquences EST. Dans 4 cas (1T57, 2T3, 2T32 et CL12), BLASTX n'a donné aucun résultat, donc il s'agit d'un alignement BLASTN.

^c A l'exception de S01E07 et CL96 (cf. paragraphe D.3.5.), tous les séquences des STS correspondent au même gène par alignement BLASTX. L'identité de séquence entre le produit amplifié chez le chêne et chez le châtaignier a été calculée pour tous les STS cartographiés chez les deux espèces. Si le STS a été cartographié chez une seule espèce ou une de deux séquence n'est pas disponible (CL80, châtaignier), l'alignement a été effectué avec la séquence EST de départ (*). Pour 5 cas (-) la réaction de séquence n'a pas fonctionné.

séquence des régions flanquantes entre les deux espèces avait été fait sans considérer le motif répété commun entre les deux mais avec le motif répété en plus révélé chez le châtaignier. Si ce motif est enlevé, l'identité de séquence monte à 85%.

Il est évident que seule la cartographie d'un grand nombre de locus permettra de manière définitive de vérifier s'il y a un réarrangement entre ces deux groupes de liaison et si les marqueurs EMCs1 et QpZAG9 sont des orthologues au bien des paralogues.

Trois conclusions peuvent être tirées de la figure 32 : 1) à ce stade de l'étude, les douze groupes de liaison du chêne pédonculé et du châtaignier européen semblent être homologues, ce qui nous laisse penser qu'aucun réarrangement chromosomique majeur ne s'est produit depuis la divergence des deux espèces ; 2) la synténie est bien conservée à l'exception de trois cas (CL66, CL8 et le phytochrome), qui seront discutés plus loin ; 3) la colinéarité est aussi globalement conservée, à l'exception de 5 inversions. Les cinq inversions peuvent être dues tout simplement à des erreurs de cartographie. En effet, même si 57 individus sont suffisants pour la détection de la liaison génétique, cet effectif faible peut poser des problèmes dans la détermination de l'ordre des marqueurs au sein du même groupe de liaison. Dans la plupart des cas, l'ordre obtenu coïncide entre les deux espèces. Même en utilisant 57 individus, la macrocolinéarité est presque toujours observée. Néanmoins, à courte distance (au-dessous de 5 cM), cet effectif n'est pas suffisant pour obtenir une bonne fiabilité de l'ordre des marqueurs. Toutes les inversions se produisent entre marqueurs étroitement liés (à moins de 5 cM de distance) et dans un cas, le marqueur impliqué (ESTS03E05, C1) a été lié avec un LOD = 5.0. Il est donc probable qu'il s'agisse de simples erreurs de cartographie.

Afin de valider l'amplification du gène correspondant à l'EST de départ et l'homologie de séquence entre les deux espèces, le produit amplifié dans chaque espèce a été directement séquencé sur MegaBACE (Amersham « DYEnamic™ ET terminator sequencing kit ») et les séquences obtenues ont été alignées entre elles et avec celle de l'EST du départ. Les résultats sont montrés dans le tableau 16. L'homologie de séquence avec l'EST de départ et entre les deux espèces a été confirmée. Néanmoins, il y a deux exceptions : les séquences des fragments amplifiés dans les cas de S01E07 et CL96 ne donnent aucun résultat après une recherche BLASTX et BLASTN³⁴ dans les bases des données. Un fragment très court

³⁴ Dans le cas d'un alignement BLASTX la séquence est traduite et comparée aux séquences protéiques. Pour des séquences codantes, ce type d'alignement permet plus efficacement de repérer les protéines homologues des espèces éloignées phylogénétiquement entre lesquelles la séquence nucléotidique peut être moins conservée.

(d'environ 100 paires de bases) a pu être aligné avec la séquence de l'EST de départ dans les deux cas. Le reste de la séquence correspond probablement à un intron. Cela peut expliquer l'absence de réponse dans les bases de données et la faible homologie de séquence entre les deux espèces pour le STS CL96 (79,6%). Il faut toutefois noter que, dans d'autres cas même en présence d'introns, l'homologie de séquence entre les deux espèces reste élevée (> 85%).

Trois STS (CL8, CL66 et PHY) ne semblent pas positionnés sur des groupes homologues entre les deux espèces (tableau 15, figure 32). Il est intéressant de noter que dans les trois cas il s'agit de familles multigéniques (tableau 16). Le STS CL8 correspond au gène de la protéine ribosomale L13. Chez *Arabidopsis*, 8 loci différents correspondant à ce gène ont été cartographiés sur trois chromosomes (Barakat et al., 2001). Malgré les duplications du génome d'*Arabidopsis*, il semble que les gènes des protéines ribosomales soient organisés en familles et que plusieurs copies soient distribuées sur les cinq chromosomes. Bien qu'une seule bande ait été amplifiée chez les deux espèces, deux locus indépendants ont été cartographiés par SSCP chez le châtaignier (groupes 2 et 4) et un troisième locus qui ne correspond à aucun des ceux du châtaignier a été identifié chez le chêne (groupe 9). Le STS PHY présente la même situation que CL8. Il s'agit du gène du phytochrome. Cinq gènes différents ont été caractérisés pour le phytochrome chez les angiospermes (PHYA-PHYE, Mathews et Sharrock, 1996). Deux locus indépendants ont été identifiés chez le chêne (groupes 1 et 5) et un locus chez le châtaignier (groupe11) lequel encore une fois ne correspond pas aux locus du chêne. Malgré l'amplification d'une seule bande, le profil SSCP observé chez le chêne est plutôt complexe, suggérant que plusieurs loci ont été amplifiés à la fois. L'optimisation des conditions de la migration du gel DGGE pour la révélation du polymorphisme du châtaignier a été très difficile à cause, probablement, de plusieurs bandes, ce qui confirme encore une fois l'amplification de plusieurs loci. Vu la conservation de l'homologie de séquence plutôt élevée entre ces gènes, il est tout a fait probable que différents membres de la même famille multigénique aient été cartographiés chez les deux espèces.

Enfin, le STS CL66 montre un fragment d'amplification très différent en taille entre les deux espèces et les deux locus ne sont pas localisés sur deux groupes homologues (Q2 et C7). Malheureusement, la séquence de ce fragment n'a pas été obtenue. Cependant, l'EST de départ correspond au gène de certaines protéines capables de lier les métaux (« metallothioneins »). Il s'agit également d'une famille multigénique très complexe dont 7 gènes ont été cartographiés chez *Arabidopsis* (Guo et al., 2003). La cartographie des différents gènes d'une famille multigénique est de grand intérêt afin d'étudier l'évolution des

gènes dupliqués, mais il est indispensable de connaître les séquences des différents membres pour pouvoir les distinguer.

D.3.6. Quelques conclusions sur la structure du génome des Fagacées.

En ce qui concerne la structure du génome du chêne et du châtaignier, les données à notre disposition indiquent que :

- la taille physique est conservée : 0,94 pg/C pour le chêne et 0,81 pg/C pour le châtaignier ;
- les estimations de la taille génétique ne sont pas significativement différentes : autour de 1200 cM pour le chêne et 1050 cM pour le châtaignier ;
- le nombre haploïde est conservé et égal à 12 dans les deux espèces. Les douze groupes de liaison, qui pourraient correspondre aux douze chromosomes, semblent homologues entre les deux espèces ;
- la conservation de la macrosynténie et de la macrocolinéarité a été démontrée et aucun réarrangement majeur n'a été détecté jusqu'à présent ;
- la structure des gènes semble être conservée entre les deux espèces.

Il est évident que le nombre encore très limité de marqueurs en commun entre les deux génomes nous empêche de tirer des conclusions définitives. Comme nous avons pu le voir dans l'introduction, l'évolution du génome des plantes semble être caractérisée par un plus grand nombre de microréarrangements que de macroréarrangements. Les microréarrangements peuvent être révélés dans le cadre de comparaisons physiques ou à l'aide de la cartographie génétique fine. Par conséquent, les microréarrangements qui peuvent différencier le chêne et le châtaignier, à ce stade de l'étude, ne peuvent être détectés. De plus, en utilisant peu de marqueurs, les résultats peuvent être biaisés. Il ne faut pas oublier que certaines duplications et réarrangements du génome d'*Arabidopsis* et du riz n'ont été détectés que lors du séquençage complet de leur génome (Bancroft, 2001 ; Paterson et al., 2004 ; Salse et al., 2004).

Toutefois, quelques hypothèses peuvent être avancées, à la lumière des résultats déjà acquis sur d'autres espèces de plantes. Tout d'abord, le rapport entre la taille physique et génétique de deux espèces est similaire et il n'est pas trop élevé : 0,77 Mb/cM chez le chêne et 0,81 Mb/cM chez le châtaignier (calculé comme décrit dans le paragraphe B.4.2.). Cela laisse penser que les deux génomes ne sont pas trop différents en composition et en contenu de séquences répétées, lesquelles ne devraient pas constituer une proportion majeure de ces génomes. La valeur de ce rapport est en effet similaire à celui de la tomate (0,7 Mb/cM), qui a

un génome avec une quantité plutôt faible de séquences répétées (autour de 20%). Le faible contenu en séquences répétées semblerait confirmé par la facilité de transférabilité de microsatellites. En effet, chez les gymnospermes et les Asteraceae, qui possèdent un génome grand et riche en séquences répétées, il est très difficile de transférer les locus microsatellites d'une espèce à l'autre.

Il semble que le modèle du « système génétique unique » des Poaceae peut s'appliquer également aux Fagacées. La conservation de la macrosynténie et de la macrocolinéarité indiquent que le transfert des connaissances moléculaires entre les deux espèces est possible et qu'apparemment il y a une bonne conservation de la position des gènes. Toutefois, la polyploïdie et l'expansion des rétrotransposons, qui ont marqué l'histoire évolutive du génome des Poaceae, ne semblent pas être les mécanismes impliqués dans la différenciation des génomes du chêne et du châtaignier. Il est plus probable que la différenciation et l'évolution de ces génomes, comme chez la plupart des eucotylédones, ait eu lieu par duplication de séquences ou de segments du génome et leur divergence. Les résultats décrits pour les EST CL8, CL66 et le gène PHY laissent penser que l'étude de l'organisation des familles multigéniques pourrait donner des éléments de réponse à l'évolution du génome des Fagacées. Une fois les groupes de liaison homologues identifiés grâce aux gènes en simple copie, les EST correspondant aux différents membres d'une famille multigénique peuvent être cartographiés pour étudier leur organisation au sein du génome et leur différenciation fonctionnelle.

En ce qui concerne la structure des gènes, nous savons que la structure en exons et introns d'un même gène orthologue est souvent conservée entre les espèces ; d'ailleurs les gènes d'*Arabidopsis* ont été utilisés pour l'annotation des gènes du riz (Liu et al., 2001). Il n'est donc pas étonnant que les gènes du chêne et du châtaignier aient une structure similaire. Il est, en revanche, possible que le même gène se trouve dans une position relative différente par rapport au centromère chez les deux espèces. Comme cela a été montré chez les Solanaceae (Tanksley et al., 1992) et chez le maïs (Gaut et al., 2000), la variation de fréquence de recombinaison tout au long du chromosome, et notamment au niveau des centromères, peut influencer sur la diversité génétique d'un gène donné. Le même gène peut donc avoir une histoire évolutive différente en fonction de sa position relative par rapport au centromère. Aucune inversion paracentrique n'a été encore détectée entre le chêne et le châtaignier, mais il est probable que ce mécanisme fréquent chez les espèces qui montrent une conservation importante de la synténie et de la colinéarité, comme par exemple les Solanaceae, ait eu un rôle majeur dans la différenciation du génome des Fagacées.

E. Cartographie comparée des QTL chez le chêne pédonculé et le châtaignier européen

QTL mapping is a statistical minefield from which few of the current generation of researchers will emerge unscathed.

(Carson M., Manager, NZFRI, Biotechnology Division, ~ 1997)

E.1. LA DÉTECTION DE QTL CHEZ LE CHÊNE PÉDONCULÉ.

E.1.1. Les caractères étudiés chez le chêne pédonculé.

Une étude de recherche de QTL a été menée chez le chêne pédonculé par Saintagne (2003). Son travail de thèse a consisté à étudier la distribution des régions génomiques qui différencient les deux espèces, chêne pédonculé (*Quercus robur*) et chêne sessile (*Q. petraea*). Pour atteindre cet objectif deux stratégies ont été adoptées : d'une part une recherche de QTL contrôlant les caractères phénotypiques qui différencient et qui ne différencient pas les deux espèces et d'autre part une stratégie basée sur l'estimation du coefficient de différenciation génétique interspécifique (G_{ST}) pour plusieurs marqueurs cartographiés. Parmi les caractères pris en compte dans le cadre de la première approche, la croissance en hauteur, la discrimination isotopique du carbone 13 et la date de débourrement (appelée de manière générique débourrement par la suite) ont été étudiés dans la famille de plein-frères 3PxA4. La carte génétique utilisée pour la détection des QTL est une actualisation de celle de Barreneche et al. (1998). Cette nouvelle carte a été construite en utilisant un effectif d'environ 220 individus et un sous-échantillon de marqueurs bien répartis sur le génome. Le degré de saturation des deux cartes parentales, calculée comme nous l'avons décrit au paragraphe B.4.2., est autour de 65%.

Les mesures phénotypiques ont été effectuées pendant trois ans et en utilisant des dispositifs expérimentaux différents. En effet, plusieurs copies des individus de la famille F_1 obtenues par bouturage sont disponibles. Comme décrit par Saintagne (2003), le débourrement a été mesuré trois fois : en 1999 dans le dispositif de Pierroton (dispositif 1, mesure appelée NJ1), en 2000 dans la tranche 1 du dispositif de Bourran (dispositif 2, mesure appelée NJ2) et en 2002 dans la tranche 2 du dispositif de Bourran (dispositif 3, mesure appelée NJ3). La détection des QTL a été réalisée en utilisant les trois mesures et en cumulant les mesures des dispositifs 2 et 3 (NJ2_3). La discrimination isotopique du carbone 13 a été également mesurée pendant trois années : en 2000 et 2001 dans le dispositif 2 ($\delta^{13}C_1$, $\delta^{13}C_2$), en 2002 dans le dispositif 3 ($\delta^{13}C_3$). Les mesures ont été effectuées par l'INRA de Nancy dans le cadre du projet OAKFLOW (<http://www.pierroton.inra.fr/Oakflow/>). En revanche, les mesures de croissance ont été effectuées une seule fois pendant l'hiver 2001 sur les boutures du dispositif 3 qui étaient à leur troisième saison de croissance depuis le bouturage. Les caractères mesurés correspondent : au nombre d'unité de croissance (N_{UC}), à la longueur moyenne d'une unité de croissance (L_{UC}), à la croissance totale de l'arbre durant la troisième

saison (Crois), qui est égale à la somme de toutes les unités de croissance et, enfin, à la hauteur totale de chaque individu à la fin de la troisième saison de végétation (Ht).

E.1.2. Les résultats obtenus chez le chêne sont-ils comparables à ceux du châtaignier ?

Grâce à la recherche de QTL chez le chêne pédonculé, nous disposons de données qui peuvent être comparées aux résultats obtenus chez le châtaignier. Le tableau 17 (a, b et c), tiré de Saintagne (2003), montre les QTL détectés chez le chêne contrôlant les trois caractères étudiés. Puisque chez le châtaignier la détection des QTL a été faite à un risque de 1^{ière} espèce de 20 % au niveau du génome, ce même seuil a été utilisé dans le cas du chêne. Le même logiciel (MultiQTL) et la même stratégie d'analyse ont été utilisés dans les deux cas. L'effectif de l'expérience est comparable entre les deux espèces. Entre 88 et 220 individus ont été utilisés pour l'analyse chez le chêne et entre 100 et 173 chez le châtaignier. Même si la famille F₁ du chêne est composée de 278 individus et celle du châtaignier de 152 individus (186, la première année de mesures), les données manquantes (génotypiques et phénotypiques) entraînent une réduction de l'effectif réel. Cet effectif réel correspond à 174 individus en moyenne chez le chêne et 153 chez le châtaignier pour le débourrement, 145 individus chez le chêne et 137 chez le châtaignier pour la discrimination isotopique du carbone 13 et enfin, 177 individus chez le chêne et 150 chez le châtaignier pour la croissance. Les QTL ont été détectés pendant plusieurs années chez les deux espèces. Globalement, les deux expériences de détection des QTL sont, donc, comparables. La seule différence importante est la présence de plusieurs copies végétatives par descendant chez le chêne qui ont permis de conduire la détection des QTL dans deux dispositifs indépendants et d'avoir des répétitions des mesures. Cette différence se traduit par une meilleure estimation des valeurs phénotypiques des plein frères du pedigree de cartographie. L'utilisation du clonage a également permis d'estimer l'héritabilité pour les trois caractères. Les estimations varient beaucoup entre les sites : de 0,15 à 0,52 pour le débourrement, de 0,32 à 0,80 pour la discrimination isotopique et de 0,04 à 0,23 pour la croissance en hauteur.

Le tableau 17 montre la totalité des QTL détectés chez les deux cartes parentales du chêne dans les différentes expériences décrites dans le paragraphe E.1.1. Trente-neuf QTL impliqués dans le débourrement ont été détectés et la variance phénotypique expliquée varie de 3,2 à 10,8%. Les 9 QTL impliqués dans la discrimination isotopique du carbone 13 expliquent de 3,8 à 34,4% de la variance phénotypique. Enfin, 13 QTL contrôlant la croissance en hauteur, qui expliquent de 5,1 à 18,7% de cette variance, ont été détectés en 2001. Les QTL impliqués dans le débourrement sont distribués sur les douze groupes de liaison. Six groupes différents

Tableau 17. a) QTL impliqués dans le débourrement détectés chez le chêne pédonculé (modifié de Saintagne, 2003).

Caractère	GL	N	Lod	α_{chr}	α_G	P1	P2	IC	d	P	$V_D=0$		$V_D=1/2 V_A$	
											S	PV	S	PV
NJ1	2f	191	3.6	0.001	5%	16.7	16.7	18.5	2.4	5.8				
NJ1	4f	189	3.3	0.002	5%	24.5	34.5	51.2	2.2	5.3				
NJ1	6m	216	4.8	0.001	5%	51.5	49.1	15.0	-3.3	10.7				
NJ1	7f	162	4.4	0.001	5%	72.5	69.8	27.6	2.6	7.1				
NJ1	9f	207	3.3	0.001	5%	48.1	33.7	36.3	2.2	5.0				
NJ1	10f	220	6.1	0.001	5%	19.4	24.3	31.9	-3.0	8.7				
NJ1	11f	207	4.2	0.001	5%	72.5	64.5	31.6	2.6	6.8				
NJ2	2f	148	3.6	0.001	5%	50.8	48	20.0	1.0	9.5	9.6	2.4	16.7	2.5
NJ2	3m	157	2.0	0.013	20%	22.1	21.3	10.9	0.8	5.7	5.7	1.4	10.0	1.5
NJ2	4f	156	2.5	0.010	10%	71.2	68.4	35.9	0.9	5.8	5.8	1.5	10.2	1.5
NJ2	7f	98	2.1	0.023	20%	79.8	64.2	50.2	1.2	8.2	8.3	2.1	14.4	2.2
NJ2	9m	129	2.2	0.010	15%	0	13.7	47.2	1.0	6.8	6.8	1.7	12.0	1.8
NJ2	10m	89	2.0	0.016	20%	0	4.1	24.1	1.1	9.7	9.8	2.5	17.1	2.6
NJ3	1m	156	8.3	0.001	5%	0	1.4	5.8	1.8	10.1	6.0	3.1	10.5	3.4
NJ3	2f	181	6.0	0.001	5%	43.7	42.8	20.6	1.9	10.0	5.9	3.1	10.4	3.3
NJ3	2m	154	2.4	0.011	10%	76.9	60.3	40.4	-1.0	3.2	1.9	1.0	3.3	1.1
NJ3	4m	168	5.7	0.001	5%	38.4	34.4	28.0	1.7	8.7	5.2	2.7	9.1	2.9
NJ3	4f	185	2.0	0.037	20%	27.7	36.7	51.0	1.2	3.9	2.3	1.2	4.1	1.3
NJ3	5m	174	4.0	0.001	5%	13.5	14.9	20.2	-1.3	4.8	2.9	1.5	5.0	1.6
NJ3	6m	192	5.0	0.001	5%	50.2	49.3	11.9	-1.3	5.2	3.1	1.6	5.4	1.7
NJ3	7f	122	3.5	0.001	5%	81.9	73.2	32.3	1.8	10.2	6.1	3.1	10.6	3.4
NJ3	9f	122	2.2	0.015	15%	15.9	20.6	28.8	1.5	6.1	3.6	1.9	6.3	2.0
NJ3	9m	185	5.0	0.001	5%	33.1	30.9	25.9	1.4	5.7	3.4	1.8	5.9	1.9
NJ3	10f	193	4.1	0.001	5%	27.7	22.9	20.2	-1.6	7.2	4.3	2.2	7.5	2.4
NJ3	10m	172	6.4	0.001	5%	34.8	28.4	19.0	-1.5	6.2	3.7	1.9	6.5	2.1
NJ3	11m	178	8.9	0.001	5%	47.7	49.4	16.2	1.9	9.6	5.7	3.0	10.0	3.2
NJ3	12f	179	2.7	0.003	5%	25.2	26.8	13.2	-1.4	5.3	3.2	1.6	5.5	1.8
NJ2+3	1m	172	6.2	0.001	5%	7.1	7.83	12.3	1.73	13.9	8.4	3.4	14.8	3.7
NJ2+3	2f	196	4.9	0.001	5%	43.6	36.2	31.0	1.33	7.8	4.7	1.9	8.3	2.0
NJ2+3	4m	200	3.8	0.001	5%	8.7	27.4	40.6	1.04	4.8	2.9	1.2	5.1	1.3
NJ2+3	5m	189	3.4	0.002	5%	18.3	20.9	32.3	-1.07	5.0	3.0	1.2	5.3	1.3
NJ2+3	6m	209	3.3	0.002	5%	55.9	52.9	16.4	-0.98	4.2	2.5	1.0	4.5	1.1
NJ2+3	7f	132	4.4	0.001	5%	79.8	73.9	27.6	1.58	10.8	6.5	2.7	11.5	2.8
NJ2+3	8f	163	2.011	0.009	15%	27.3	18.8	16.3	-0.94	3.8	2.3	0.9	4	1
NJ2+3	9m	201	4.4	0.001	5%	27.9	22.1	24.3	1.35	8.0	4.9	2.0	8.5	2.1
NJ2+3	10f	213	5.4	0.001	5%	24.8	20.5	15.1	-1.40	8.5	5.2	2.1	9.0	2.2
NJ2+3	10m	186	3.3	0.002	5%	34.8	28.7	21.4	-1.04	4.5	2.7	1.1	4.8	1.2
NJ2+3	11f	210	2.6	0.004	5%	63	58.5	28.4	0.96	3.9	2.4	1.0	4.1	1.0
NJ2+3	11m	207	5.1	0.001	5%	65.5	53.9	26.3	1.20	6.1	4	1.5	6.4	1.6
Moyenne										6.5	4.5	1.8	7.9	1.9

Tableau 17. b) QTL impliqués dans la discrimination isotopique du carbone 13 détectés chez le chêne pédonculé (modifié de Saintagne, 2003).

Caractère	GL	N	Lod	α_{chr}	α_G	P1	P2	IC	d	P	$V_D=0$		$V_D=1/2V_A$	
											S	PV	S	PV
$\delta^{13}C_1$	1f	144	2.1	0.015	20%	68	47.4	57.8	0.27	6.3	4.2	2.2	7.4	2.4
$\delta^{13}C_2$	2m	88	3.6	0.001	5%	97.6	90.8	32.5	0.52	12.7	8.0	6.4	14.0	7.2
$\delta^{13}C_3$	6m	178	2.5	0.004	5%	14.6	15	11.5	0.20	4.4	2.7	1.7	4.7	1.9
$\delta^{13}C_1$	8m	130	5.6	0.001	5%	47.9	44.8	17.0	0.42	15.8	9.7	6.1	17.0	6.8
$\delta^{13}C_1$	9m	153	2.2	0.016	20%	15.7	15.6	25.1	0.21	3.8	2.6	1.3	4.5	1.4
$\delta^{13}C_3$	9m	156	2.8	0.003	5%	0	6.67	31.0	0.21	4.9	3.0	1.9	5.3	2.1
$\delta^{13}C_1$	11m	160	10.53	0.001	5%	57.1	57.7	9.0	-0.51	22.0	14.8	7.6	25.9	8.2
$\delta^{13}C_3$	11m	190	11.46	0.001	5%	64.3	64.6	7.9	-0.44	21.7	13.4	8.4	23.4	9.3
$\delta^{13}C_2$	11m	106	11.13	0.001	5%	64.2	63.8	4.2	-0.89	34.4	21.6	17.2	37.9	19.5
Moyenne										10.5	6.7	4.5	11.6	4.9

Tableau 17. c) QTL impliqués dans la croissance en hauteur détectés chez le chêne pédonculé (modifié de Saintagne, 2003).

Caractère	GL	N	LoD	α_{chr}	α_G	P1	P2	IC	d	P	$V_D=0$		$V_D=1/2 V_A$	
											S	PV	S	PV
Ht	5m	164	2.5	0.004	5%	53.4	48.9	28.4	-5.8	6.7	5.8	1.3	10.1	1.4
Ht	6m	188	3.2	0.001	5%	28.3	29	13.4	-5.7	6.4	5.5	1.3	9.7	1.3
Ht	10m	170	7.5	0.001	5%	18.4	19.1	12.8	-9.6	17.3	14.9	3.5	26.1	3.6
Ht	11f	172	3.5	0.001	5%	40.0	36.3	30.8	6.9	9.1	7.8	1.8	13.7	1.9
Crois	3m	191	5.1	0.001	5%	22.1	27	19.3	-0.5	9.8	9.5	1.8	16.7	1.8
Crois	10f	178	3.9	0.001	5%	12.2	16.1	26.5	-0.5	9.5	9.2	1.7	16.2	1.8
Crois	10m	170	8.2	0.001	5%	15.5	16.5	9.6	-0.7	18.7	18.2	3.4	31.8	3.5
L _{UC}	3m	191	5.5	0.001	5%	21.3	24.6	18.2	-0.3	10.7	11.6	1.8	20.3	1.9
L _{UC}	10m	170	7.0	0.001	5%	15.8	17.3	11.3	-0.3	16.4	17.8	2.8	31.2	2.9
L _{UC}	10f	170	2.0	0.014	15%	5.1	18.3	42.1	-0.2	5.9	6.4	1.0	11.2	1.0
N _{UC}	10f	178	6.1	0.001	5%	12.2	16.1	17.0	-0.3	13.7	25.9	1.9	45.3	1.9
N _{UC}	10m	170	3.1	0.001	5%	17.9	23.7	22.9	-0.2	8.9	16.8	1.2	29.5	1.2
N _{UC}	11f	193	2.5	0.008	5%	51.1	44.4	25.1	0.2	5.1	9.6	0.7	16.9	0.7
Moyenne										10.2	11.9	1.8	20.9	1.8

Avec

GL : le groupe de liaison (m = carte mâle et f = carte femelle)

N : le nombre d'individus pris en compte dans l'analyse

LOD : la valeur de Lod maximale obtenue sur le chromosome après 1000 permutations

α_{chr} : L'erreur de type I au niveau du chromosome obtenue après 1000 permutations

α_G : L'erreur de type I en pourcentage au niveau du génome

P1 : La position la plus probable du QTL sur le chromosome d'après la valeur de LOD maximale

P2 : La position du QTL la plus probable après 1000 ré-échantillonnages réalisés par bootstrap

IC : intervalle de confiance (longueur en cM) du QTL obtenu après 1000 ré-échantillonnages réalisés par bootstrap

d : effet de substitution allélique au QTL

P : le pourcentage de variance des moyennes de clones expliquée par le QTL dans la descendance

S : le pourcentage de variance additive expliquée par le QTL dans la population d'où sont issus les parents pour deux valeurs contrastées de variance de dominance ($V_D=0$ et $V_D=1/2V_A$)

PV : le pourcentage de variance phénotypique expliquée par le QTL dans la population d'où sont issus les parents pour deux valeurs contrastées de variance de dominance ($V_D=0$ et $V_D=1/2V_A$)

sont impliqués dans le contrôle de la discrimination isotopique du carbone treize et 5 groupes contiennent au moins un QTL impliqué dans la croissance en hauteur. En ce qui concerne le débourrement, les groupes 2f, 4f, 6m, 7f, 9f, 9m, 10f et 10m contiennent des QTL stables, détectés lors des différentes expériences. Les groupes 9m et 11m contiennent des QTL stables impliqués dans la discrimination isotopique du carbone 13. Enfin, la croissance en hauteur a été mesurée une seule fois, néanmoins toutes les mesures permettent de détecter au moins un QTL sur le groupe 10f et/ou 10m, suggérant un QTL majeur dans cette région du génome.

E.2. LA CONSTRUCTION D'UNE CARTE CONSENSUS CHEZ LE CHÊNE ET CHEZ LE CHÂTAIGNIER ET LA COMPARAISON DES QTL.

Les questions auxquelles nous essaierons de répondre dans la partie qui suit sont les suivantes : le contrôle génétique des caractères adaptatifs est-il similaire chez le chêne et le châtaignier ? Est-il possible d'identifier des régions génomiques conservées qui contrôlent ces caractères ?

Pour que la comparaison des QTL soit possible il faut que leur position par rapport à des marqueurs communs aux deux espèces soit connue. Il faut rappeler que : 1) les deux cartes génétiques avec lesquelles la détection des QTL a été effectuée ne contiennent pas beaucoup de marqueurs communs entre le chêne et le châtaignier ; 2) de plus, sous la stratégie de double pseudo test-cross la détection des QTL est faite sur les deux cartes parentales séparément. Si le nombre de marqueurs codominants informatifs chez les deux parents est limité, comme c'est le cas du chêne et du châtaignier, même la comparaison de la position des QTL entre les groupes homologues d'une même espèce peut poser des problèmes.

Cependant, la carte utilisée pour la détection des QTL dérive, dans chacune des deux espèces, de la carte sur laquelle tous les marqueurs orthologues ont été cartographiés. Il est donc possible intégrer toutes les informations (marqueurs orthologues et position des QTL) sur une seule carte consensus dans chacune des deux espèces. Cela nous permettra enfin de comparer la position des QTL entre les deux espèces grâce aux marqueurs orthologues. La stratégie d'intégration des informations et de comparaison a donc été la suivante :

- dans chacune des deux espèces, une carte consensus a été construite. Cette dernière contient certains des marqueurs des deux cartes parentales utilisées pour la détection des QTL et tous les marqueurs orthologues entre les deux génomes ;
- les QTL ont été projetés sur cette carte consensus grâce aux marqueurs communs ;
- la position des QTL chez les deux espèces a été comparée grâce aux intervalles communs identifiés par les marqueurs orthologues.

E.2.1. La stratégie de construction de la carte consensus.

La carte consensus a été construite à l'aide des deux logiciels « MapMaker » et « JoinMap » décrits au paragraphe B.1.8. La stratégie a consisté en deux étapes :

- 1) dans un premier temps, le logiciel « MapMaker » a été utilisé afin de construire une carte jalon de chaque parent dans chacune des deux espèces. L'option « ripple » a été

utilisée afin de vérifier la fiabilité statistique de l'ordre des marqueurs (cf. paragraphe B.4.2., $LOD > 2.0$). Les données des deux cartes originales ont été utilisées (Barreneche et al., 1998 ; Annexe I). L'effectif de ces deux cartes est donc de 94 individus chez le chêne et 96 chez le châtaignier. Cinquante-sept de ces individus ont été génotypés avec les STS. Un sous échantillon de marqueurs de ces deux cartes, génotypés sur environ 220 individus chez le chêne et 186 individus chez le châtaignier, a été utilisé pour la détection des QTL. L'ordre optimisé des marqueurs dans ces deux nouvelles cartes jalon coïncide, en général, avec celui de la carte de détection des QTL. Les différences observées concernent toujours des marqueurs étroitement liés (l'ordonnement de marqueurs très proches est plus sensible à l'effectif de la famille). Les STS (23 chez le chêne et 26 chez le châtaignier), pour lesquels le seuil de « ripple » était de $LOD > 2.0$, ont aussi été intégrés dans les deux cartes jalon. La nouvelle carte génétique du chêne couvre une distance de 852,6 cM (femelle) et 876,0 cM (mâle). Selon la nouvelle estimation de la longueur génétique du génome du chêne (1360,7 cM femelle et 1237,8 cM mâle), la saturation est de 63% (femelle) et de 69% (mâle). La carte femelle du châtaignier couvre 863,4 cM. L'estimation de la taille du génome est de 1073 cM, donc la saturation est de 80%. La carte mâle couvre 866,6 cM, qui correspond à une saturation de 69%, l'estimation de la taille du génome étant de 1262 cM. Les résultats sont cohérents avec ceux obtenus précédemment.

- 2) La deuxième étape de construction de la carte consensus a été effectuée à l'aide du logiciel « JoinMap ». C'est grâce à ce logiciel que les données de ségrégation des deux cartes parentales peuvent être intégrées dans une carte consensus. La liaison des deux cartes parentales est faite par les marqueurs ségrégeant 3:1, 1:2:1 ou bien 1:1:1:1. En général, dans une carte consensus à cause d'une plus grande densité des marqueurs et du faible contenu informatif des couples de marqueurs 1:1/3:1 et 1:1/1:2:1 (cf. paragraphe B.1.5.), l'optimisation de l'ordre des marqueurs est beaucoup plus difficile et plusieurs ordres également probables sont souvent possibles. Le logiciel « JoinMap » offre la possibilité grâce à l'option « fixed order » de fixer l'ordre de certains marqueurs et positionner les autres par rapport à cet ordre fixé. Nous avons donc choisi d'utiliser cette option pour la construction des cartes consensus. Tout d'abord, si l'ordre donné par « JoinMap » correspondait à celui obtenu en utilisant le logiciel « MapMaker » et à l'ordre sur la carte de détection des QTL, cette option n'a pas été utilisée. Si, en revanche, l'introduction des marqueurs 3:1, 1:2:1 et de certains STS entraînait des incohérences importantes (inversions de marqueurs à une distance

supérieure à 5 cM) alors l'ordre de certains marqueurs a été fixé afin de garder l'ordre optimisé par le logiciel « MapMaker » et dans la plupart des cas l'ordre de la carte jalon de détection des QTL. Une fois cet ordre fixé, les autres marqueurs ont été intégrés. Tous les marqueurs orthologues ont été intégrés dans les deux cartes consensus. De cette manière la carte consensus contient, d'une part, un nombre suffisant de marqueurs communs avec la carte de détection des QTL pour pouvoir intégrer par la suite les informations de la détection des QTL et, d'autre part, les marqueurs orthologues entre chêne et châtaignier pour la comparaison de la position des QTL chez les deux espèces. Il est évident que pour tous les marqueurs très proches l'incertitude sur l'ordre des marqueurs sur la carte consensus est élevée et quelques inversions entre la carte consensus et la carte de détection des QTL sont apparues (groupes C9, Q1, Q2, Q9, Q11).

La nouvelle carte consensus du chêne et du châtaignier est présentée dans la figure 34.

E.2.2. Le logiciel BioMercator et la projection des QTL sur la carte consensus de chaque espèce.

La nouvelle carte consensus de chaque espèce contient maintenant plusieurs marqueurs communs avec la carte utilisée lors de la détection des QTL. Grâce à ces marqueurs nous pouvons positionner les QTL sur la carte consensus de chaque espèce. Le logiciel « BioMercator » a été utilisé pour atteindre cet objectif (Arcade et al., 2004). Ce logiciel a été développé afin d'intégrer les résultats de cartographie provenant de cartes génétiques différentes qui possèdent des marqueurs communs et afin de synthétiser les données de plusieurs expériences indépendantes de détection de QTL. Il est donc possible de construire une carte consensus qui contient tous les gènes, les marqueurs génétiques et la position des QTL dérivant de plusieurs expériences. Le logiciel offre aussi la possibilité d'effectuer une méta-analyse des QTL basée sur l'algorithme de Goffinet et Gerber (2000). Cette analyse permet de vérifier si plusieurs QTL liés au même caractère et positionnés dans la même région génomique sont issus du même locus. N'étant pas dans les conditions de pouvoir faire une méta-analyse (la détection des QTL chez le chêne et chez le châtaignier a été faite sur des individus issus du même croisement contrôlé, donc il ne s'agit pas d'expériences indépendantes), le logiciel a été utilisé pour la projection de la position et de l'intervalle de confiance des QTL de la carte de détection des QTL sur la carte consensus des deux espèces. Les marqueurs communs entre les deux cartes sont utilisés dans le logiciel pour calculer un rapport homothétique de projection. Si m_1 , m_2 , et m_3 sont les distances en cM de l'extrémité

du groupe de liaison de trois marqueurs liés (dans l'ordre m_1 - m_2 - m_3) sur la carte A et m_1' , m_2' , et m_3' les orthologues dans le même ordre sur la carte B, le rapport de projection sera : $R_1=(m_2-m_1)/(m_2'-m_1')$ et $R_2=(m_3-m_2)/(m_3'-m_2')$. Chaque intervalle commun a son rapport R de projection. L'homothétie est appliquée pour la position du QTL et l'intervalle de confiance. La projection faite de cette manière implique que : 1) si la distance relative entre les couples de marqueurs sur la carte consensus est différente de la distance sur la carte « projetée » (dans notre cas la carte de détection des QTL), alors l'intervalle de confiance projeté peut apparaître de taille différente par rapport à la taille sur la carte de détection des QTL ; 2) les extrémités de l'intervalle de confiance sont toujours limitées aux marqueurs communs entre la carte consensus et la carte de détection des QTL (voir le groupe C10 consensus comme exemple, figure 34). Cela implique que les intervalles de confiance sur la position des QTL peuvent apparaître non symétriques par rapport à la position du QTL, si le QTL est proche d'une des extrémités du groupe de liaison dans la carte de détection des QTL.

Quatre-vingt QTL détectés chez le châtaignier et 61 chez le chêne ont donc été projetés sur la carte consensus de chaque espèce. La figure 34 montre la position des ces QTL par rapport aux marqueurs orthologues entre les deux espèces.

E.2.3. La comparaison des QTL.

Grâce aux données disponibles, les QTL impliqués dans trois caractères adaptatifs différents peuvent être comparés entre le chêne pédonculé et le châtaignier européen. La comparaison a été réalisée en trois étapes : 1) tout d'abord, le nombre et les effets des QTL détectés ont été comparés entre les deux espèces ; 2) ensuite, nous avons vérifié d'une part si des QTL stables contrôlant des caractères phénotypiques identiques ont été détectés sur des groupes de liaison homologues entre les deux espèces, et d'autre part si la position des QTL est conservée entre les deux espèces (« colocalisation »); 3) enfin, nous avons essayé de vérifier si les « colocalisations » ne sont pas dues au simple fait du hasard. Puisque les QTL sont distribués sur les douze groupes de liaison, la probabilité d'avoir une « colocalisation » entre les deux espèces simplement due au hasard est élevée. Lin et al. (1995) ont proposé une formule qui permet de calculer, sous l'hypothèse de distribution aléatoire des QTL, la probabilité que la « colocalisation » entre QTL dans une région donnée du génome soit due au hasard (encadré 5). Cette probabilité a donc été calculée afin de valider du point de vue statistique les « colocalisations » observées.

Il est évident que cette comparaison présente des limites importantes :

- les QTL détectés ne correspondent pas à la totalité des QTL contrôlant ces caractères chez les deux espèces et les QTL à effet très faible ne sont pas détectés;
- l'imprécision de l'estimation de la position du QTL est très élevée. Les intervalles de confiance de la position des QTL sont souvent plus grandes que la moitié de la taille du groupe de liaison ;
- les QTL ont été projetés sur une carte consensus contenant plusieurs marqueurs qui n'ont pas été directement utilisés pour la détection des QTL.

Ces limites seront prises en compte dans la discussion des résultats de la comparaison des QTL entre les deux espèces.

E.2.3.1. Comparaison du nombre et des effets des QTL.

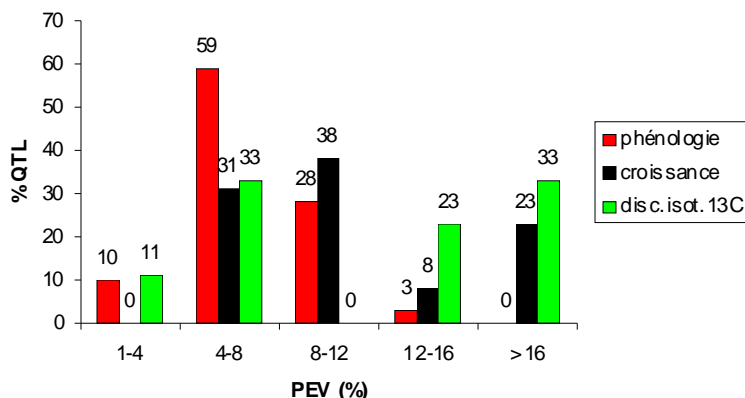
Au cours des trois années de mesures, 39 QTL (27 sans considérer l'analyse combinée des sites 2 et 3) contrôlant le débourrement ont été détectés chez le chêne et 31 (21 sans considérer l'analyse des données ajustées à l'effet année) chez le châtaignier. Chez le chêne pendant les trois années, 9 QTL impliqués dans la discrimination isotopique du carbone 13 ont été détectés. Chez le châtaignier, dans les mêmes conditions (donc sans tenir compte des QTL dérivant de l'analyse sur les données ajustées à l'effet année), 12 QTL contrôlant ce caractère ont été détectés. La croissance en hauteur a été mesurée une seule année chez le chêne grâce à quatre types de mesures, la moyenne de QTL par mesure est de 3,25 plus élevée que chez le châtaignier. Comme nous avons pu le voir dans le paragraphe C.3.3., chez le châtaignier dans le cas de la croissance en hauteur peu de QTL par caractère ont été détectés (1,8). De cette simple comparaison, nous pouvons conclure que :

- En ce qui concerne le débourrement et la discrimination isotopique du carbone 13, le nombre total de QTL détectés sur les trois années est comparable.
- Comme cela a été montré chez d'autres espèces (tableau 8) également dans le cas du chêne et du châtaignier, le débourrement est le caractère pour lequel le nombre le plus élevé de QTL a été détecté.
- Chez le chêne la décomposition du caractère hauteur en plusieurs composantes de la croissance a été efficace pour la détection des QTL. Cela peut expliquer la différence significative de QTL détectés sur une année entre les deux espèces.

La distribution de la proportion de variance phénotypique expliquée par les QTL détectés sur chêne est montrée dans la figure 33a. La comparaison avec la figure 24c met en évidence quelques différences. Tout d'abord, chez le chêne plus de QTL à effet très faible ($PEV <$

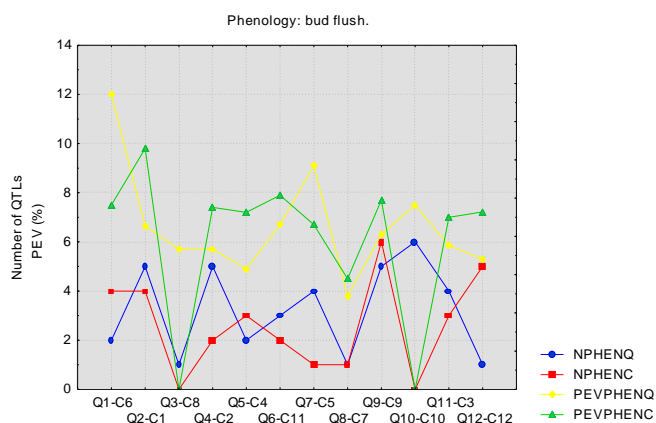
Figure 33

a) Distribution des valeurs de variance phénotypique expliquée par les QTL détectés chez le chêne.

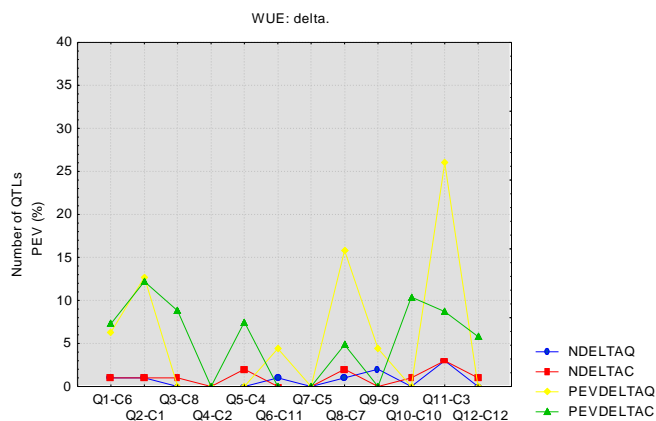


Pour chaque couple d'homologues le nombre de QTL détectés (en bleu chez le chêne et en rouge chez le châtaignier) et la moyenne des valeurs de variance phénotypique expliquée par les QTL (en jaune chez le chêne et en vert chez le châtaignier) sont représentés.

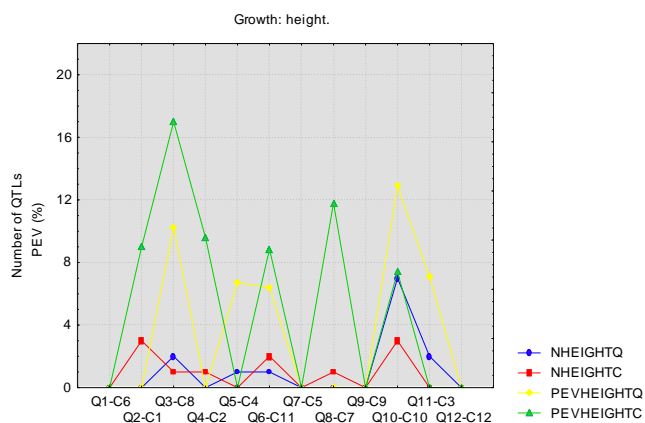
b) Débourrement. Tous les QTL sont pris en compte.



c) Discrimination isotopique du carbone 13. Les QTL détectés chez le châtaignier grâce à l'analyse des données ajustées à l'effet année n'ont pas été considérés.



d) Croissance en hauteur. Les QTL détectés chez le châtaignier grâce à l'analyse des données ajustées à l'effet année n'ont pas été considérés.

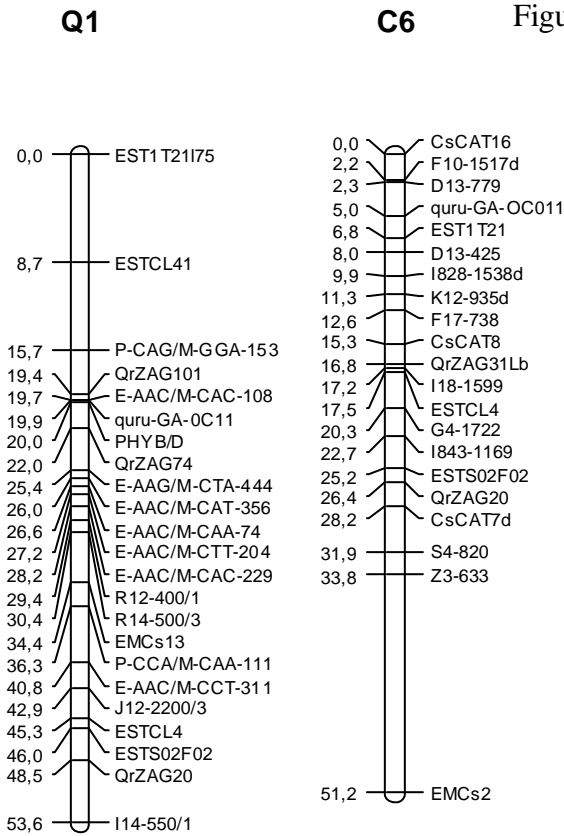


4,0%) ont pu être détectés. Puisque l'effectif est en général comparable entre les deux espèces, il est probable que les mesures effectuées sur plusieurs copies du même individu ont contribué à mettre en évidence des effets plus petits grâce à une meilleure précision de l'estimation des valeurs phénotypiques.

En ce qui concerne la distribution des valeurs de PEV, il semble que le débourrement est contrôlé chez le chêne ainsi que chez le châtaignier par plus de QTL à effet faible et modéré que de QTL à effet fort (nous considérons les QTL à effet faible ou modéré si la proportion de la variance phénotypique expliquée est comprise entre 4 et 12%. En revanche, les QTL à effet fort montrent, dans notre cas, une valeur de PEV > 12,0%). En revanche, en ce qui concerne la croissance en hauteur et la discrimination isotopique du carbone 13, il semblerait qu'il existe une différence significative entre les deux espèces car la distribution en forme de L des effets des QTL n'est apparemment pas respectée chez le chêne. Néanmoins, cette différence entre les deux espèces dérive du fait que chez le chêne 3 QTL contrôlant la discrimination isotopique du carbone 13 et 3 QTL impliqués dans la croissance en hauteur ayant en effet supérieur à 16% correspondent, en effet, au même QTL. Le faible nombre de QTL détectés pour ces deux caractères et cette redondance donnent l'impression d'une distribution qui ne correspond pas à celle du châtaignier. En réalité chez les deux espèces il semble que les QTL à effet faible ou modéré sont toujours plus nombreux que les QTL à effet fort. Cependant, il est à noter que la discrimination isotopique du carbone 13 et la croissance en hauteur semblent être caractérisées par une proportion plus élevée de QTL à effet fort (PEV > 12,0%) que le débourrement, mais les données ne nous permettent pas de tirer une conclusion définitive.

La figure 33 (b, c et d) montre le nombre de QTL détectés et la variance phénotypique expliquée pour les trois caractères étudiés. Pour chaque couple de groupes homologues entre le chêne et le châtaignier, la figure montre le nombre total et la valeur moyenne de PEV des QTL détectés. Certains QTL stables ont été détectés sur des groupes homologues. Il s'agit des groupes Q2-C1, Q9-C9 pour la phénologie, Q10-C10 pour la croissance en hauteur et Q11-C3 pour la discrimination isotopique du carbone 13. Il est également possible de mettre en évidence des différences importantes entre les deux espèces. Par exemple, 6 QTL impliqués dans le débourrement ont été détectés sur le groupe Q10 du chêne, alors que sur l'homologue de châtaignier aucun QTL n'a été détecté. De la même manière, le groupe C1 du châtaignier contient 3 QTL contrôlant la croissance en hauteur mais aucun QTL n'a été détecté sur l'homologue de chêne (Q2). Les QTL impliqués dans la discrimination isotopique du carbone

Figure 34



Les deux cartes consensus du chêne (Q, le nombre du groupe correspond à celui reporté par Barreneche et al., 1998) et du châtaignier (C, le nombre du groupe correspond à celui reporté par Casasoli et al., 2001) sont montrées à gauche avec tous les marqueurs utilisés pour la construction de ces cartes. En bas, les mêmes deux groupes sont représentés avec peu de marqueurs pour mieux identifier les intervalles communs entre les deux cartes. Les marqueurs orthologues sont représentés en rouge. Les QTL projetés sont représentés sur la droite des groupes de liaison (en rouge les QTL impliqués dans le débourrement, en vert dans la discrimination isotopique du carbone 13 et en noir dans la croissance). Les intervalles communs ont été identifiés comme décrit dans le texte.

Intervalles comparés : 3.

QTL Q : 1 débourrement,

1 delta.

QTL C : 2 delta,

1 diamètre,

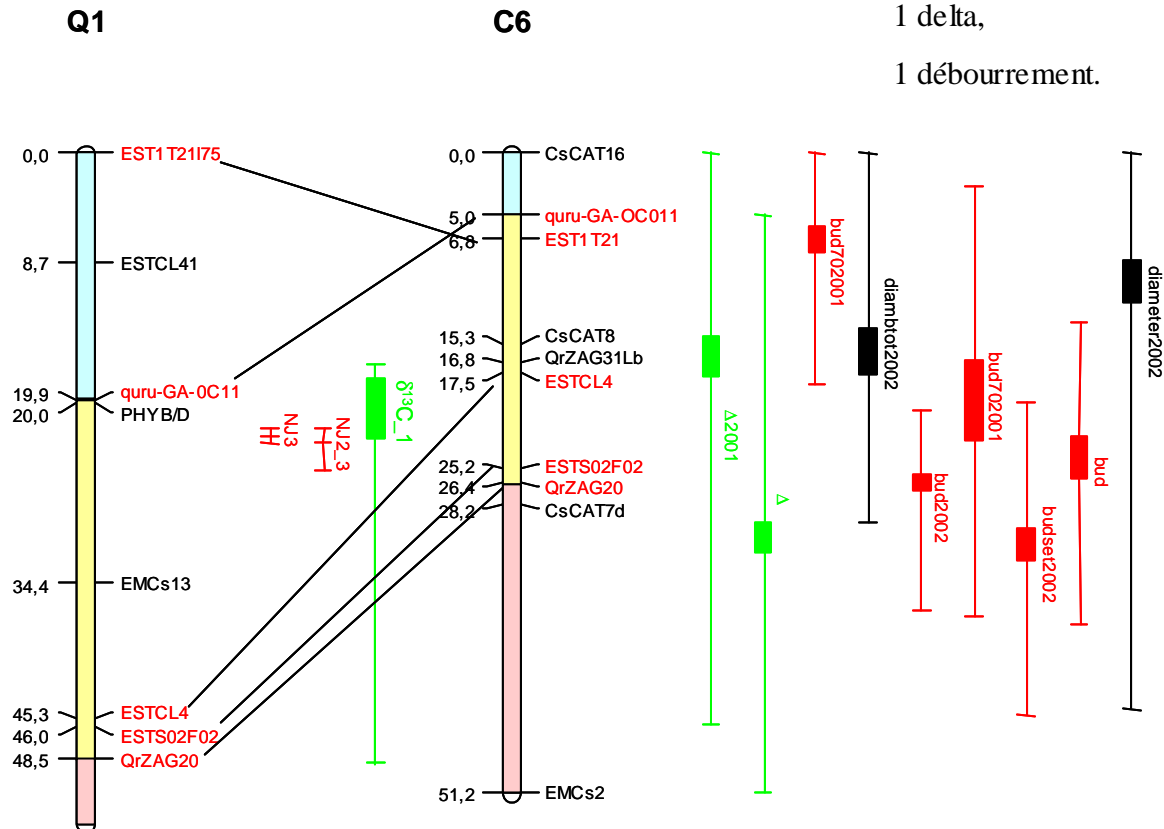
1 débourrement

1 arrêt de croissance.

“Colocalisations” :

1 delta,

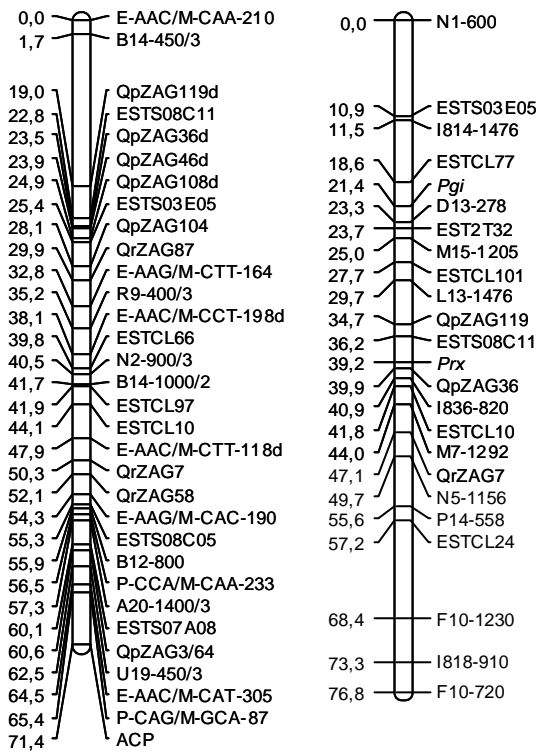
1 débourrement.



Q2

C1

Figure 34



Intervalles comparés : 4.

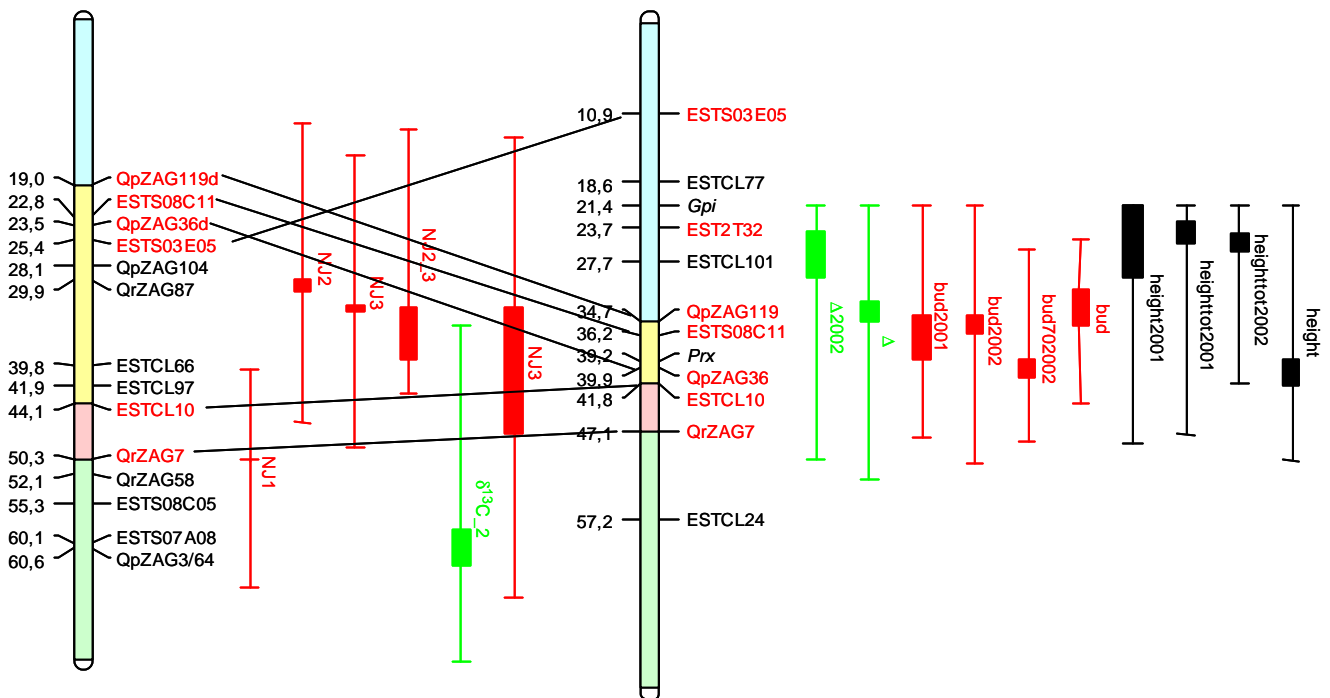
QTL Q : 2 débourrement,
1 delta.

QTL C : 1 delta,
1 débourrement,
2 hauteur.

“Colocalisations” :
1 débourrement.

Q2

C1

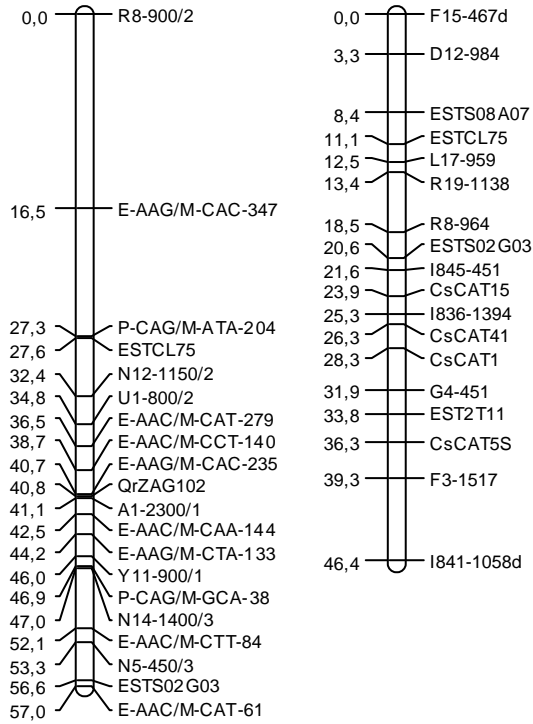


Le marqueur EST2T32 se trouve sur le groupe Q2 mais il a été cartographié dans le croisement interspécifique (voir texte).

Q3

C8

Figure 34



Intervalles comparés : 2.

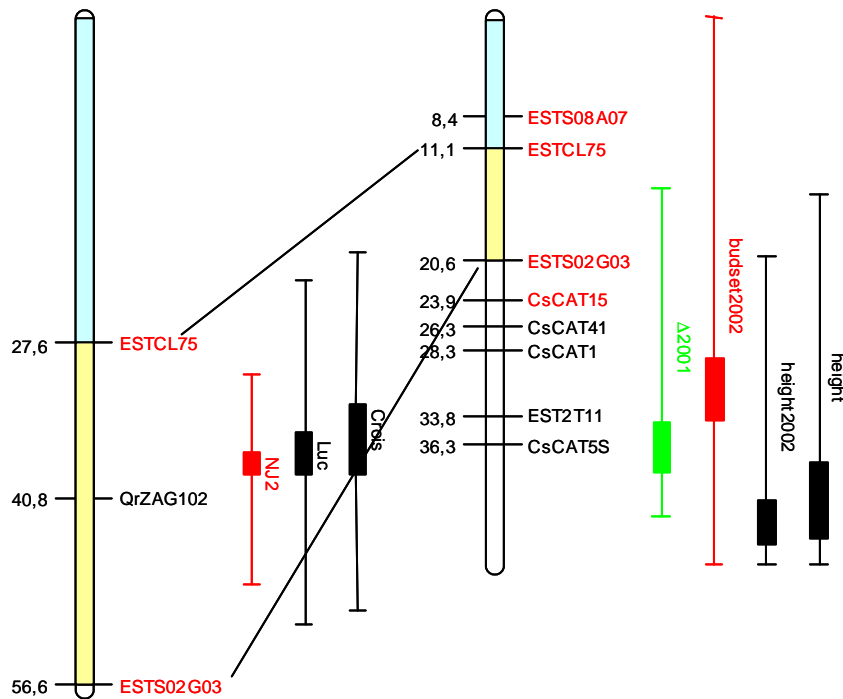
QTL Q : 1 débourrement,
1 hauteur.

QTL C : 1 delta,
1 arrêt de croissance,
1 hauteur.

“Colocalisations” : aucune.

Q3

C8

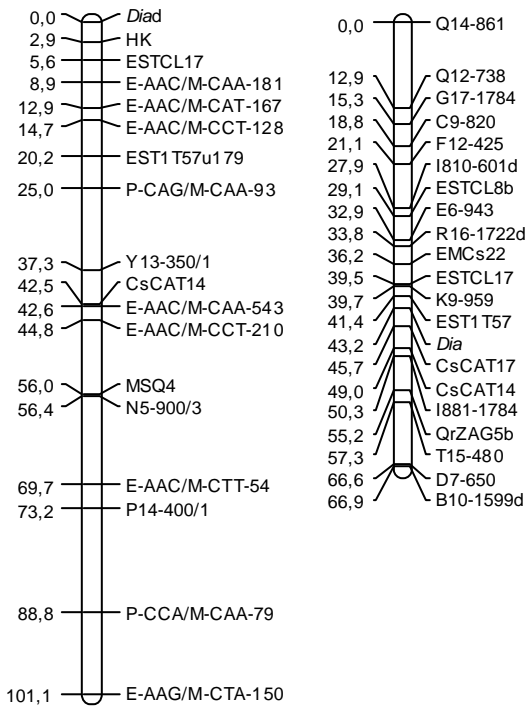


S08A01 se trouve dans le groupe Q3, mais la position n'a pas pu être définie à cause de l'effet faible, tableau 15. Il correspond à l'STS S08A07. CsCAT15 se trouve également sur le groupe Q3 mais il a été cartographié dans le croisement interspécifique (voir texte).

Figure 34

Q4

C2



Intervalles comparés : 3.

QTL Q : 2 débourrement.

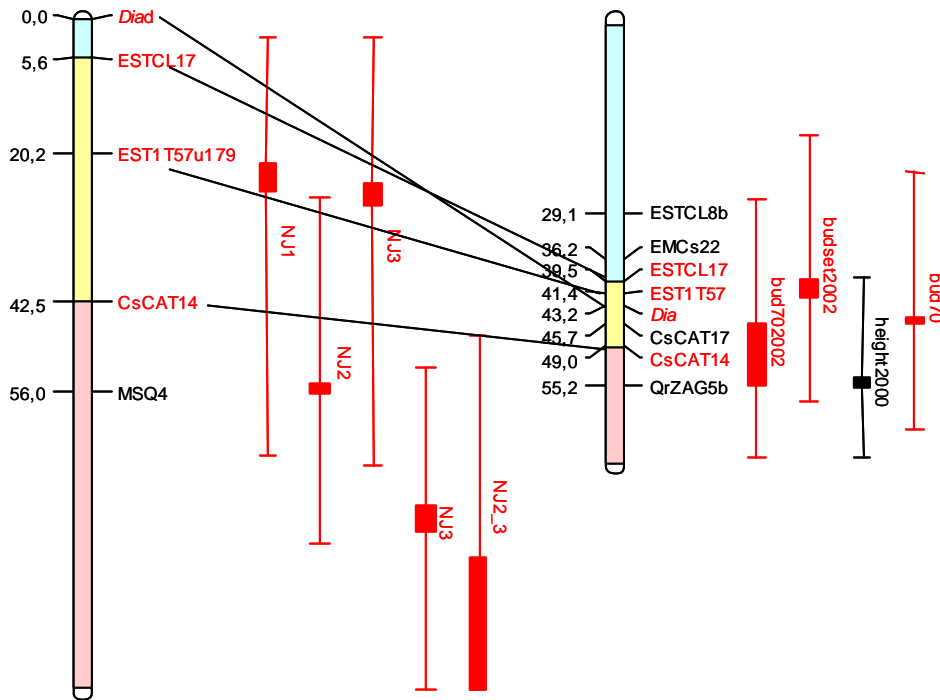
QTL C : 1 débourrement,
1 arrêt de croissance,
1 hauteur.

“Colocalisations” :

1 débourrement.

Q4

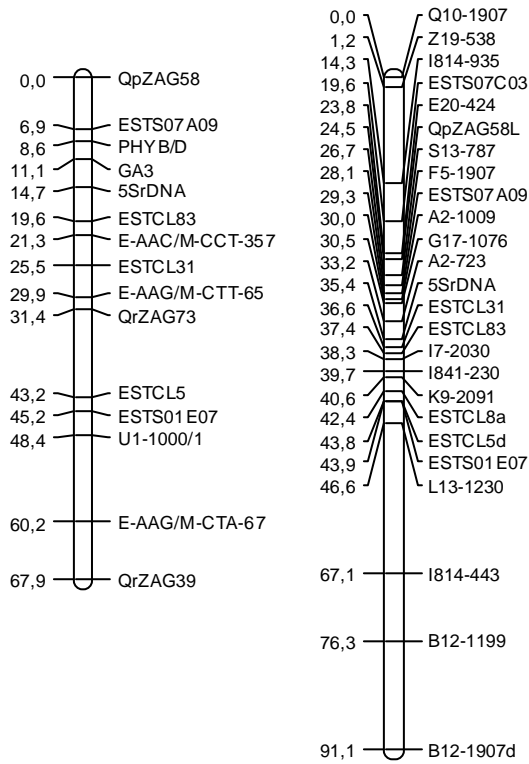
C2



Q5

C4

Figure 34



Intervalles comparés : 3.

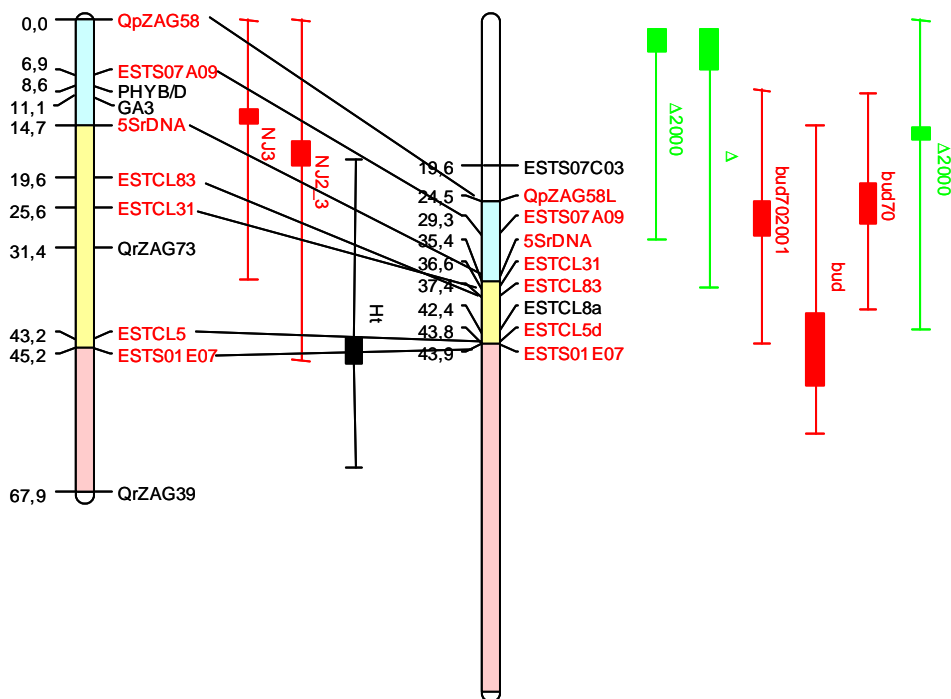
QTL Q : 1 débourrement,
1 hauteur.

QTL C : 1 delta
2 débourrement.

“Colocalisations” :
1 débourrement.

Q5

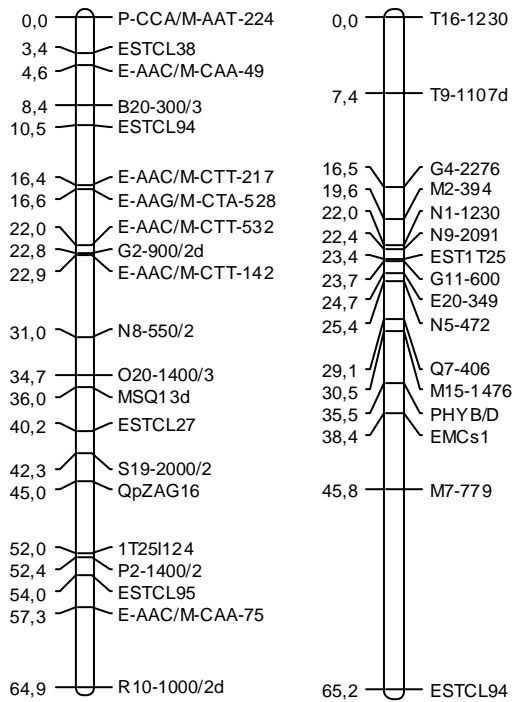
C4



Q6

C11

Figure 34



Intervalles comparés : 2.

QTL Q : 1 delta,
1 débourrement,
1 hauteur.

QTL C : 1 débourrement,
2 hauteur,
1 diamètre.

“Colocalisations” :

1 débourrement,
1 hauteur.

Q6

C11

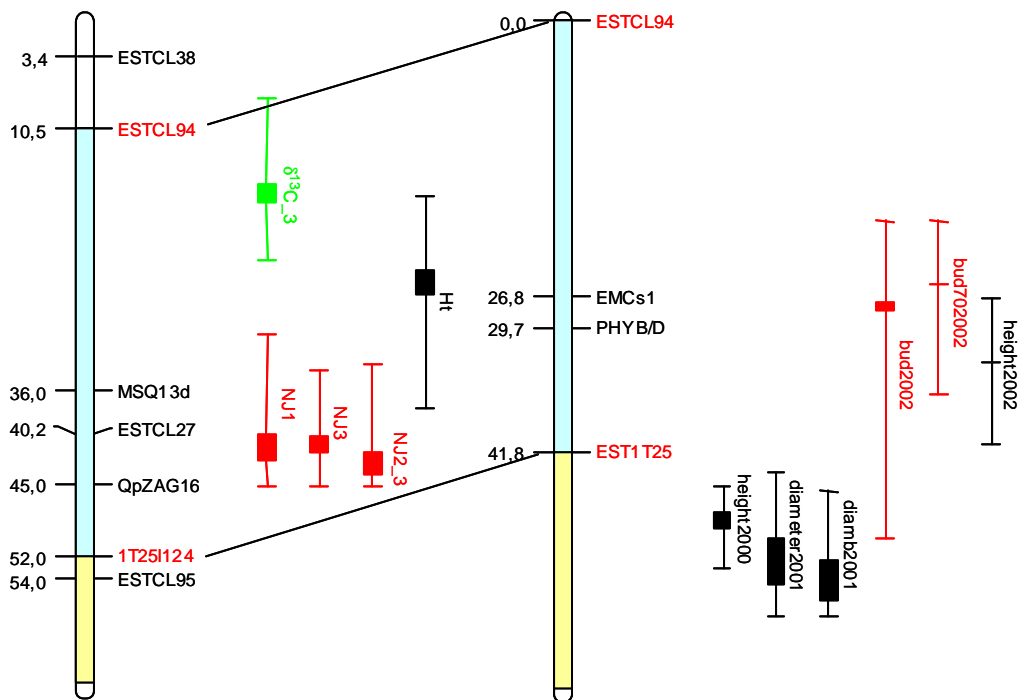
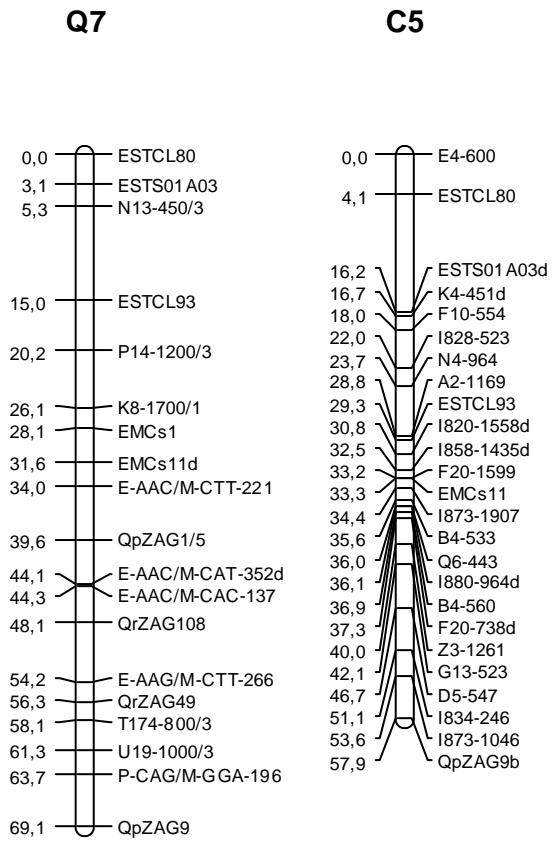


Figure 34



Intervalles comparés : 2.

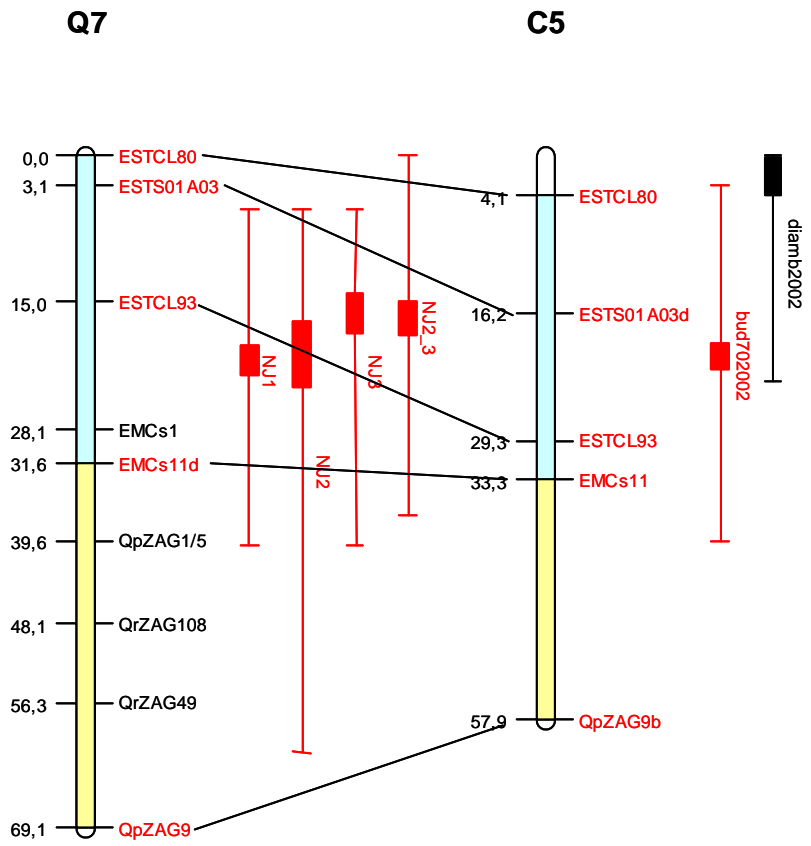
QTL Q : 1 débourrement.

QTL C : 1 débourrement,

1 diamètre.

“Colocalisations” :

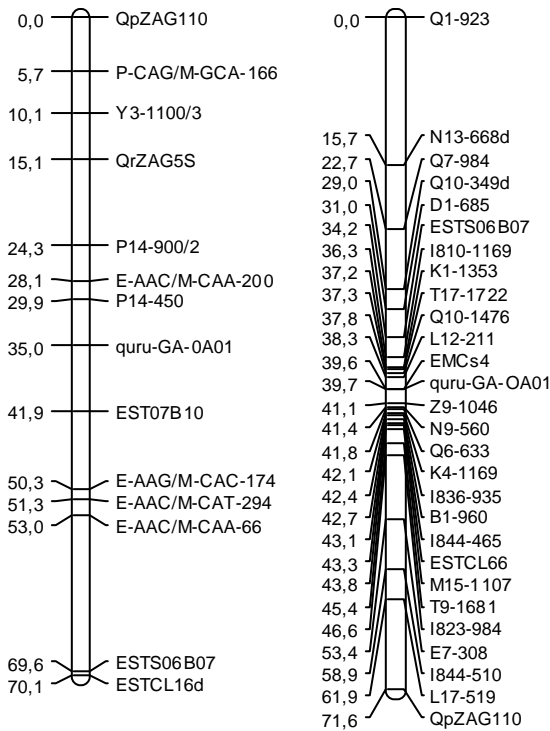
1 débourrement.



Q8

C7

Figure 34



Intervalles comparés : 2.

QTL Q : 1 delta,
1 débourrement.

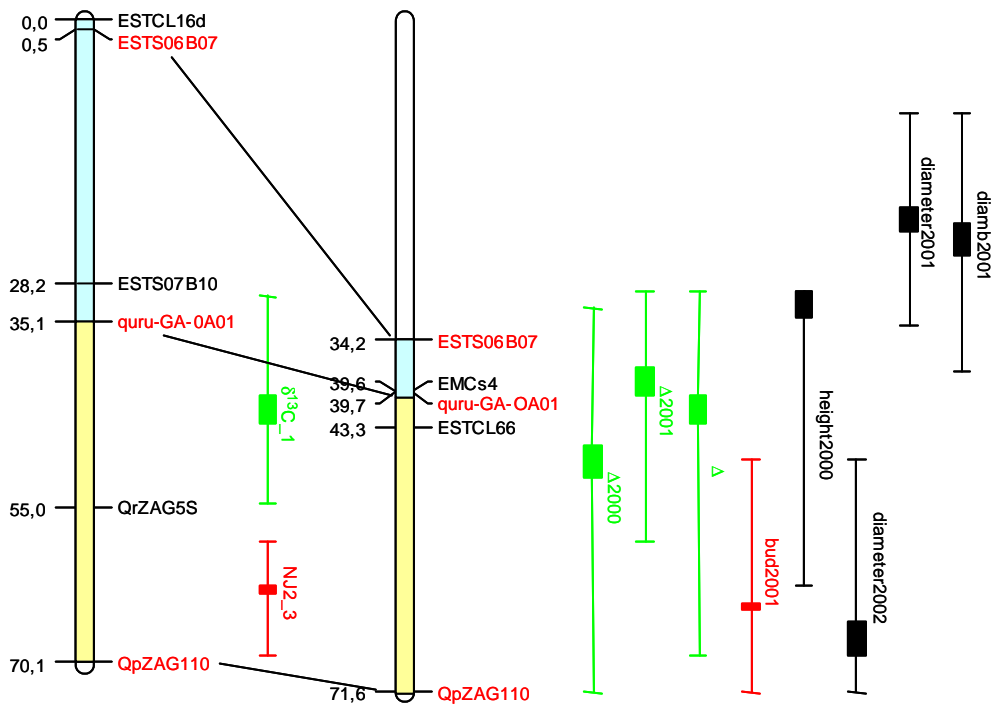
QTL C : 1 delta,
1 débourrement,
1 hauteur,
2 diamètre.

“Colocalisations” :

1 delta,
1 débourrement.

Q8

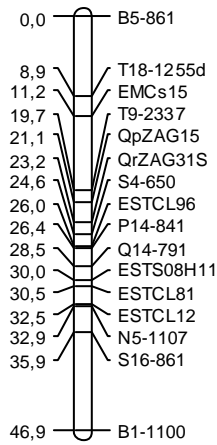
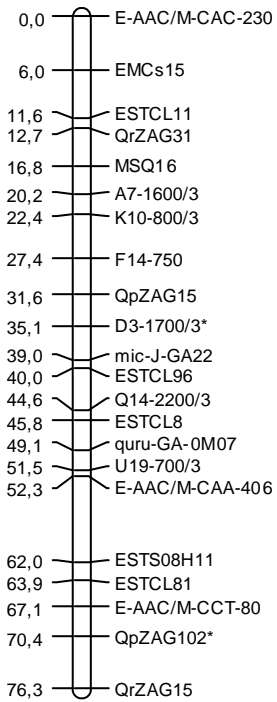
C7



Q9

C9

Figure 34



Intervalles comparés : 4.

QTL Q : 1 delta,
2 débourrement.

QTL C : 2 débourrement,
1 arrêt de croissance,
1 diamètre.

“Colocalisations” :

2 débourrement.

Q9

C9

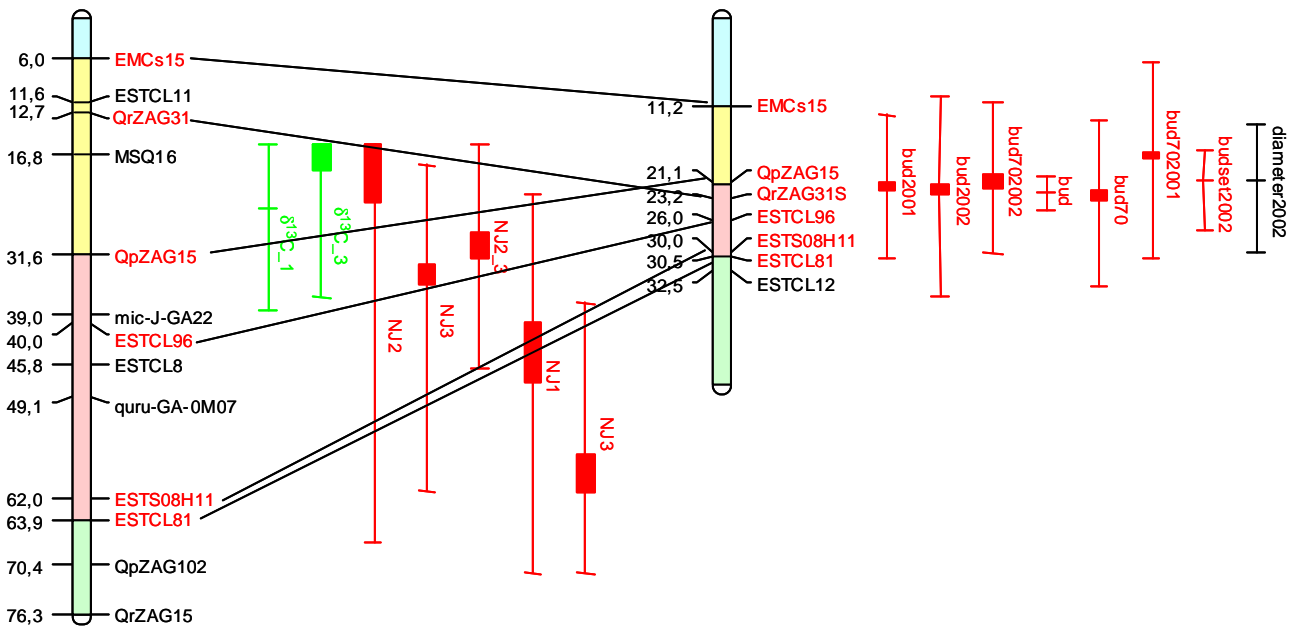
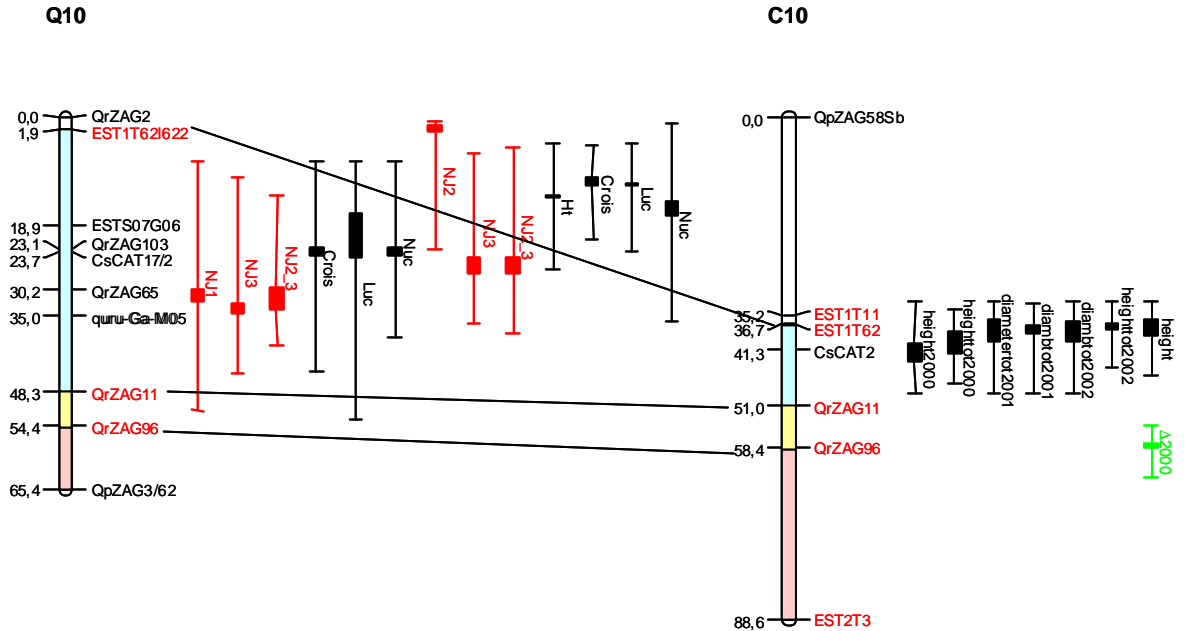
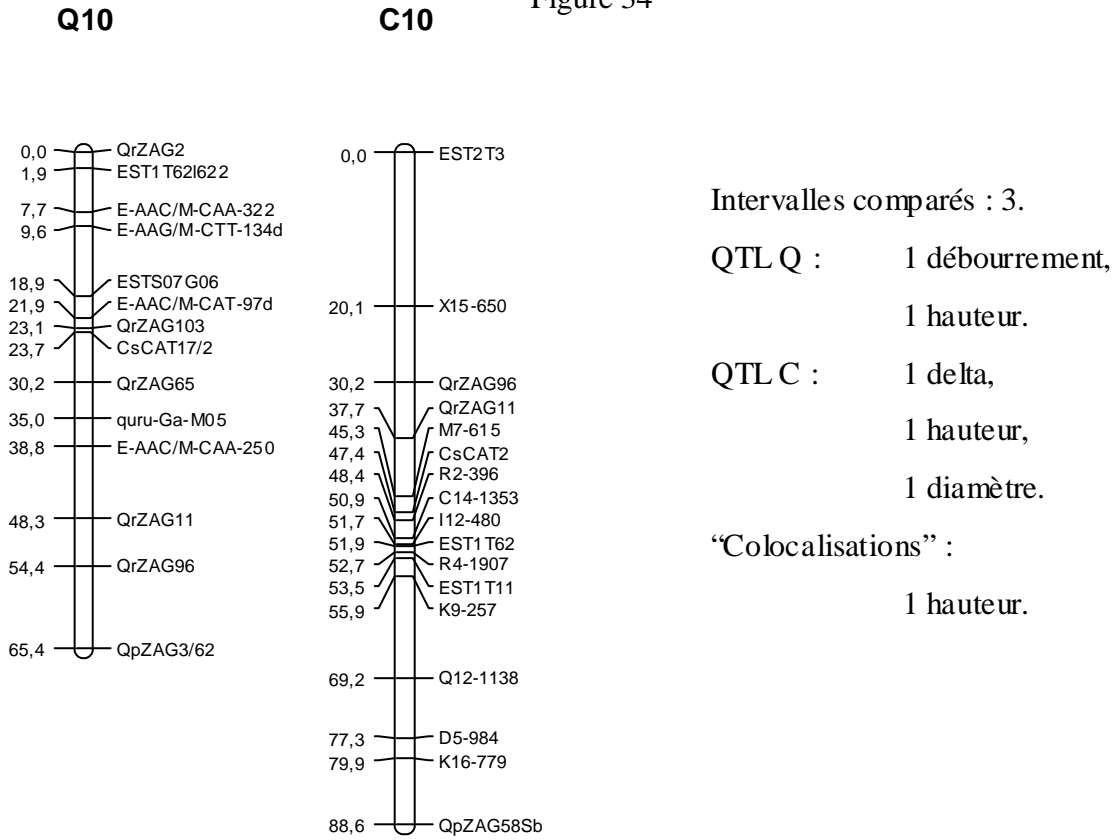
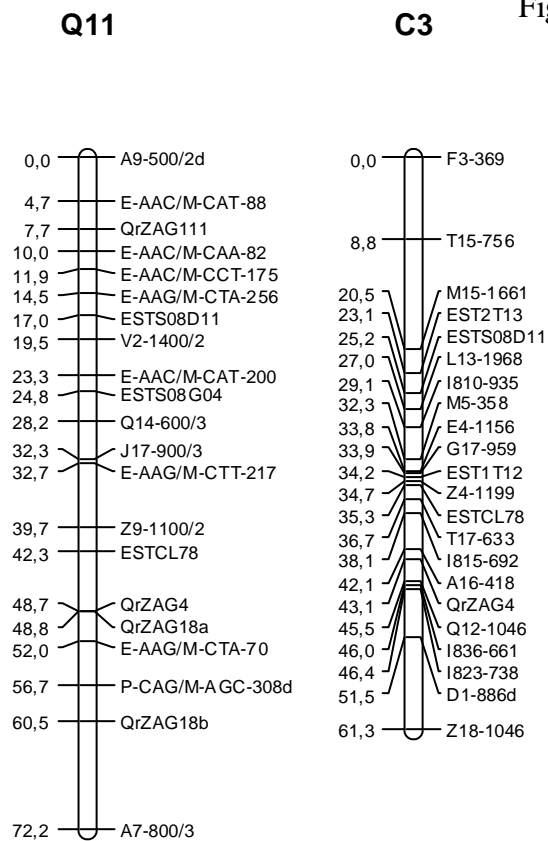


Figure 34



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

Figure 34

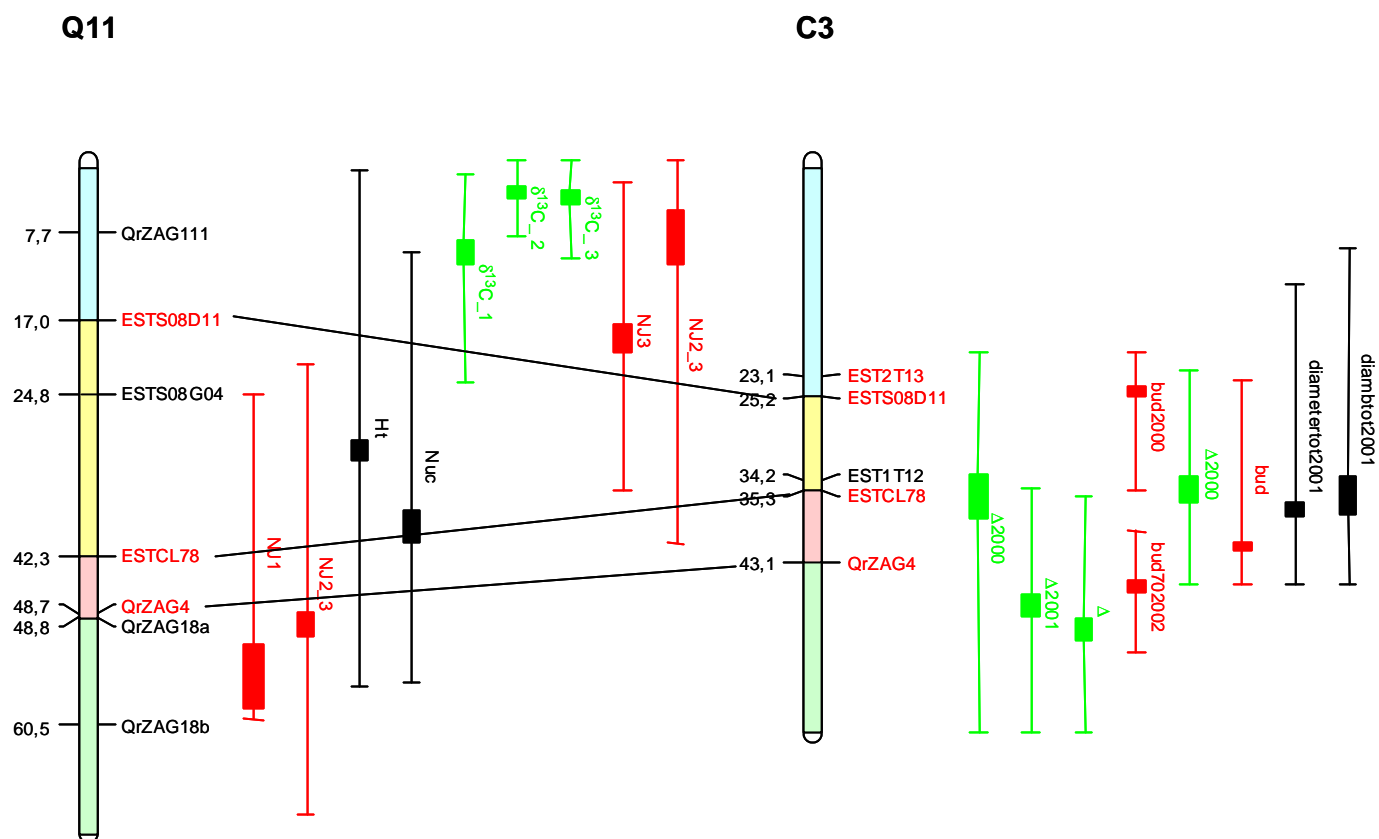


Intervalles comparés : 4.

QTL Q : 1 delta
2 débourrement,
1 hauteur.

QTL C : 2 delta,
2 débourrement,
1 diamètre.

“Colocalisations” :
2 débourrement.



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

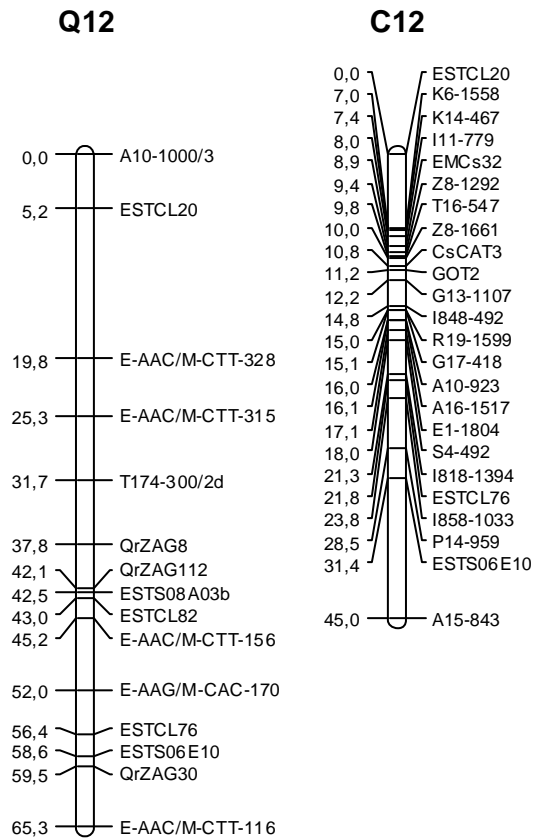


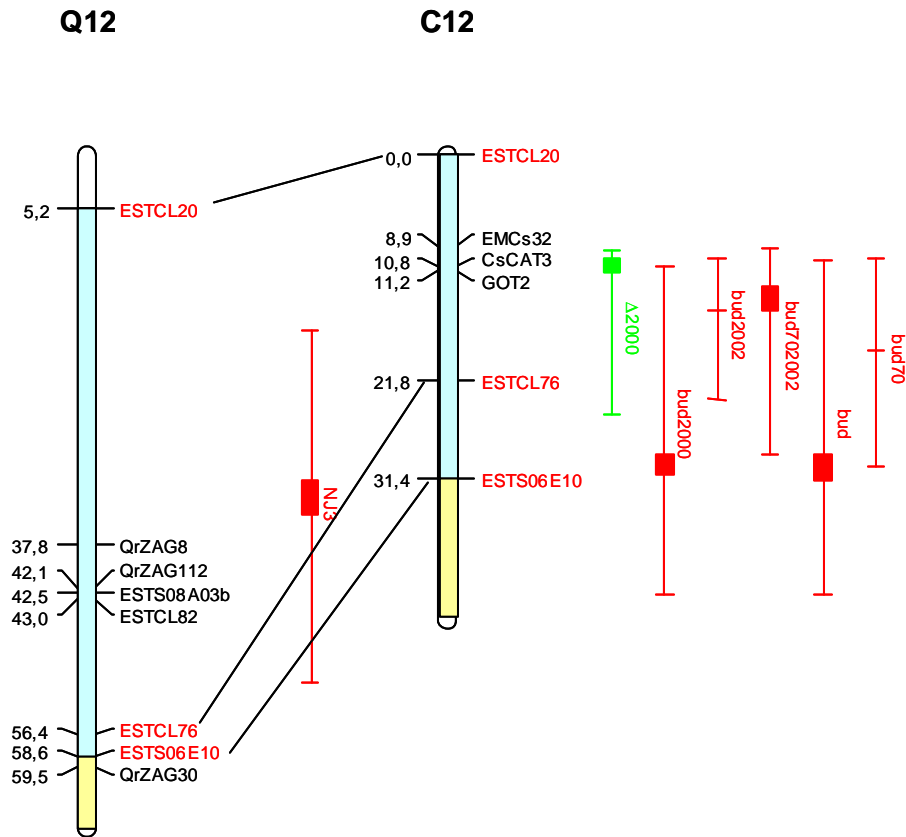
Figure 34

Intervalles comparés : 2.

QTL Q : 1 débourrement.

QTL C : 1 delta,
1 débourrement.

“Colocalisations” :
1 débourrement.



13 détectés sur les couples d'homologues Q5-C4 et Q9-C9 ne correspondent pas entre les deux espèces.

Les résultats de cette comparaison globale sont, dans son ensemble, peu surprenants. Les QTL stables pourraient correspondre à des gènes impliqués dans les voies biochimiques principales qui contrôlent un processus physiologique donné. Ces QTL seraient presque toujours détectés et probablement correspondent entre espèces puisqu'il s'agit de processus de base impliqués dans le contrôle des caractères étudiés et qui auraient été conservés au cours de l'évolution. Les différences observées entre les deux espèces peuvent elles aussi être expliquées par le dispositif de détection des QTL utilisé et par la biologie des deux espèces. D'une part, les QTL détectés chez les deux espèces sont un sous-ensemble du nombre total des QTL contrôlant les caractères étudiés et d'autre part, comme les conditions expérimentales des deux essais sont différentes et indépendantes, il est possible que des mécanismes différents soient impliqués dans les deux situations. Enfin, il est aussi possible que certains des mécanismes contrôlant ces caractères adaptatifs soient différents entre les deux espèces. Comme nous le verrons dans le paragraphe E.2.3.3., plusieurs explications sont possibles. Après l'analyse globale du nombre et de l'effet des QTL impliqués dans les trois caractères, nous avons comparé la position de tous les QTL détectés chez les deux espèces à l'aide des intervalles communs identifiés par les marqueurs orthologues.

E.2.3.2. Comparaison de la position des QTL.

Les QTL détectés sur les deux cartes parentales de chaque espèce ont été projetés, comme décrit au paragraphe E.2.2., sur la carte consensus du chêne et du châtaignier. La figure 34 montre en haut les deux cartes consensus qui contiennent tous les marqueurs orthologues cartographiés chez les deux espèces et certains marqueurs de la carte de détection des QTL qui ont permis la projection des QTL sur cette nouvelle carte. En bas de la même figure, les deux cartes consensus sont représentés avec peu de marqueurs et les QTL projetés afin d'identifier plus facilement les intervalles communs entre les cartes des deux espèces. Sur la base de ces données, la comparaison de la position des QTL a été réalisée en trois étapes :

- 1) **Identification des intervalles communs.** Les intervalles communs sont définis par les marqueurs orthologues. Pour que l'intervalle soit comparable il faut qu'il soit au moins de 5 cM chez les deux espèces. Comme les STS ont été cartographiés en utilisant 57 individus, l'ordre des marqueurs n'est pas fiable au-dessous de cette distance. Les marqueurs qui montrent des inversions entre les deux espèces n'ont pas

Encadré 5

Sous l'hypothèse de distribution aléatoire des QTL, la probabilité que la "colocalisation" entre les QTL dans une région donnée du génome soit due au hasard est égale à :

$$p = \frac{\binom{l}{m} \binom{n-l}{s-m}}{\binom{n}{s}} \quad \text{Lin et al. (1995).}$$

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.

Débourrement :
n = 34
l = 16 (Q)
s = 13 (C)
m = 12
p = 3,53x10⁻⁵.

Delta :
n = 34
l = 10 (C)
s = 6 (Q)
m = 2
p = 0,35554.

Hauteur :
n = 34
l = 8 (C)
s = 5 (Q)
m = 2
p = 0,26163.

été pris en compte. Quand plusieurs marqueurs orthologues sont proche de l'extrémité d'un intervalle et la distance entre eux est inférieure à 5 cM, le marqueur génotypé sur plus d'individus (par exemple, EMCs11, génotypé sur 96 individus, dans le groupe C5 par rapport à ESTCL93, génotypé sur 57 individus, ou QrZAG20 par rapport à ESTCL4 dans le couple d'homologues Q1-C6) ou le plus informatif (par exemple, ESTS06E10, ségrégeant 1:1, par rapport à ESTCL76, ségrégeant 3:1 dans le couple d'homologues C12-Q12) est considéré comme extrémité de l'intervalle.

- 2) **Définition de QTL uniques.** Comme le montre la figure 34, plusieurs QTL sont positionnés dans la même région génomique. Il s'agit de QTL détectés, par exemple, sur deux années différentes ou bien sur la carte femelle et sur la carte mâle. Il peut s'agir évidemment du même QTL détecté plusieurs fois. A chaque fois que la position la plus probable d'un ensemble de QTL se trouve dans le même intervalle, nous avons décidé de considérer ces QTL comme un QTL unique localisé dans cet intervalle. Si la position la plus probable est située entre deux intervalles, deux situations sont possibles : a) d'autres QTL sont positionnés dans un des deux intervalles. Dans ce cas là, si les intervalles de confiance des QTL sont superposés (plus de 50% de leur longueur), alors nous considérons qu'il s'agit du même QTL localisé dans l'intervalle contenant la plus part des QTL ; b) en revanche, s'il n'y a aucun QTL dans les deux intervalles alors le QTL est positionné dans l'intervalle qui contient la plus grande partie de l'intervalle de confiance. Cette définition de QTL unique est une définition pratique pour pouvoir comparer la position des QTL chez les deux espèces.
- 3) **« Colocalisation » de QTL entre le chêne et le châtaignier.** Les QTL sont colocalisés entre les deux espèces si la position la plus probable se trouve dans le même intervalle. La figure 34 montre le résultat obtenu dans chaque couple de groupes de liaison homologues.

En suivant cette logique, nous avons identifié 34 intervalles communs entre les deux espèces (encadré 5). Seize QTL uniques impliqués dans le débourrement ont été identifiés chez le chêne et 13 chez le châtaignier, dont 12 « colocalisent » entre les deux espèces. Parmi les 10 QTL contrôlant la discrimination isotopique identifiés chez le châtaignier et le 6 chez le chêne, 2 sont colocalisés. Enfin, en ce qui concerne la croissance en hauteur, 8 QTL uniques ont été identifiés chez le châtaignier, 5 chez le chêne et deux d'entre eux « colocalisent ». Comme le montre l'encadré 5, vu le nombre de QTL détectés sur les douze groupes de liaison, les « colocalisations » dans le cas des QTL impliqués dans la discrimination

isotopique du carbone 13 et la croissance en hauteur semblent dues au hasard. En revanche, les 12 « colocalisations » des QTL contrôlant le débourrement semblent significatives.

Les résultats obtenus mettent en évidence que :

- le contrôle génétique des trois caractères adaptatifs étudiés, en terme de nombre de QTL détectés et distribution globale de la variance phénotypique expliquée, est plutôt conservé entre les deux espèces. Sur les 16 QTL conservés entre les deux espèces, 11 contribuent pour des proportions comparables à la variance phénotypique des caractères. En revanche, une différence importante en ce qui concerne la conservation de la position des QTL a été mise en évidence. La position des QTL impliqués dans le débourrement est remarquablement conservée par rapport aux QTL contrôlant la discrimination isotopique du carbone 13 et la croissance en hauteur. Nous tenterons de trouver une explication à cette différence dans le paragraphe suivant.
- Nous ne pouvons pas exclure le fait que les deux « colocalisations » des QTL impliqués dans la croissance en hauteur soient due au hasard. Cependant, la coïncidence dans le même intervalle du couple Q10-C10 (figure 34) de 7 QTL impliqués dans la croissance en hauteur chez le chêne et 7 QTL impliqués dans la croissance en hauteur et en diamètre chez le châtaignier, laisse penser que nous avons identifié une région génomique conservée contrôlant la croissance dans les deux espèces.
- Les QTL stables contrôlant le débourrement détectés indépendamment chez les deux espèces sont colocalisés sur deux groupes homologues. Un seul QTL impliqué dans la croissance et un autre dans la discrimination isotopique du carbone 13 tous les deux stables chez le châtaignier « colocalisent » avec un QTL détecté dans une région homologue du chêne.

E.2.3.3. Quelles conclusions peuvent être tirées de cette comparaison ?

Malgré les limites de cette comparaison (il ne faut pas oublier que nous ne disposons que d'un sous échantillon des locus qui contrôlent les caractères adaptatifs étudiés chez les deux espèces, que la localisation de ces locus sur la carte génétique est très imprécise et que les intervalles communs identifiés entre les cartes génétiques du chêne et du châtaignier sont plutôt grossiers), il y a, néanmoins, trois résultats qui méritent d'être relevés du point de vue biologique :

- 1) le contrôle génétique des trois caractères adaptatifs semble être similaire entre les deux espèces (au vu du déterminisme génétique des caractères : nombre de QTL, proportion de la variabilité génétique expliquée), mais les QTL stables identifiés chez les deux espèces ne sont pas toujours conservés.
- 2) Il existe des différences importantes entre les trois caractères adaptatifs étudiés en ce qui concerne la conservation des QTL entre les deux espèces.
- 3) La conservation des QTL impliqués dans le débourrement est remarquable, vu le temps de divergence entre les deux espèces.

Le déterminisme génétique des caractères adaptatifs semble donc conservé au cours de l'évolution en ce qui concerne le nombre de QTL impliqués et la distribution de la variance expliquée. Cette conservation entre les deux espèces montre que ces caractères sont contrôlés par plus de QTL à effet faible et modéré que des QTL à effet fort. Néanmoins, comme prévu dans la théorie de Orr (1998), des QTL à effet fort sont toujours présents. Notamment, en ce qui concerne la discrimination isotopique du carbone 13 et la croissance en hauteur moins de QTL ont été détectés par rapport au débourrement et la proportion de QTL à effet fort semble être plus élevée dans les deux premiers cas. De plus, probablement grâce aussi au nombre plus élevé de QTL détectés chaque année, le débourrement semble être contrôlé par plus de QTL stables que les autres caractères. Ce dernier résultat est en accord avec celui obtenu par Jermstad (2001a, 2003) et il semble confirmer l'héritabilité élevée de ce caractère. Ces résultats semblent indiquer qu'il existe des différences selon le type de caractère, mais globalement pour chaque caractère les résultats sont cohérents entre les deux espèces.

La différence entre discrimination isotopique du carbone 13 et croissance en hauteur d'une part et débourrement de l'autre est d'autant plus évidente si nous considérons la comparaison de la position des QTL. Malgré l'exception du groupe Q10 du chêne sur lequel un QTL stable impliqué dans le débourrement n'a pas été détecté chez le châtaignier, tous les QTL stables contrôlant le débourrement sont conservés entre les deux espèces. En ce qui concerne la croissance en hauteur le couple d'homologues Q10-C10 semble contenir un QTL stable chez les deux espèces. Enfin, le QTL stable contrôlant la discrimination isotopique du carbone 13 identifié sur le couple Q11-C3 ne semble pas localisé dans une région homologue. Parmi les deux « colocalisations » concernant la discrimination isotopique (Q1-C6 et Q8-C7), seul le groupe C7 contient des QTL stables. Bien que le nombre plus élevé de QTL impliqués dans le débourrement puisse expliquer ce résultat simplement par chance, les 12 « colocalisations »

dans le cas du débourrement sont significatives. Il est à noter qu'une différence similaire entre types de caractères a été mise en évidence chez les Brassicaceae. Comme nous avons pu le voir dans l'introduction, il s'agit d'une famille dont le génome a subi plusieurs réarrangements. Mais, la plupart des QTL contrôlant la hauteur ne sont pas conservés entre différentes populations de *Brassica oleracea* (Lan et Paterson, 2001), tandis que certains QTL impliqués dans le temps de floraison semblent correspondre entre les genres *Brassica* et *Arabidopsis*, malgré les duplications de ces génomes (Osborn et al., 1997). Comment pouvons-nous expliquer cette différence entre les caractères ? Plusieurs hypothèses peuvent être faites :

- tout d'abord, ce résultat peut être une conséquence de la différence d'héritabilité des caractères étudiés. Comme nous avons pu le voir les estimations d'héritabilité de ces caractères varient beaucoup. Cependant, le débourrement est considéré comme étant un caractère à héritabilité modérée ou élevée, alors que la croissance en hauteur et la discrimination isotopique du carbone 13 sont des caractères à héritabilité faible ou modérée (cf. paragraphe C.3.3.5.). Des valeurs élevées d'héritabilité peuvent parfois être obtenues pour la discrimination isotopique et la croissance lorsque dans un environnement donné un gène majeur détermine la variabilité des individus utilisés pour l'estimation. Chez le chêne la présence d'un QTL impliqué dans la discrimination isotopique du carbone 13 sur le groupe 11 à effet majeur peut expliquer l'estimation élevée de son héritabilité. Nous savons que la puissance de détection des QTL dépend de façon importante de l'héritabilité du caractère (cf. paragraphe C.2.3.). Il est donc probable que pour un caractère comme le débourrement l'héritabilité plus élevée et, par conséquent, la stabilité des locus impliqués d'un environnement à l'autre pourraient expliquer une efficacité plus élevée de détection des QTL par rapport à des caractères moins héréditaires comme la croissance en hauteur et la discrimination isotopique du carbone 13. Dans le cas du débourrement grâce à une détection des QTL plus efficace, il serait plus facile d'identifier les QTL conservés entre les espèces.
- Le débourrement semble aussi un caractère « plus simple » que la discrimination isotopique du carbone 13 et la croissance en hauteur. Le stimulus environnemental principal qui déclenche la reprise de l'activité des bourgeons est la température (Howe et al., 2003). Il est donc probable que les mécanismes moléculaires, qui se mettent en place lors de la reprise végétative, soient dépendants d'une voie de transduction du signal qui a été conservée entre les espèces. Cela pourrait expliquer aussi la stabilité de certains QTL détectés pendant plusieurs années. Au contraire, la croissance ou

l'efficacité d'utilisation de l'eau d'une plante peuvent résulter d'un nombre plus élevé de composantes ou de mécanismes physiologiques. Ces derniers caractères sont des caractères synthétiques dont l'expression phénotypique est la réponse à un nombre important de stimuli environnementaux. La disponibilité de nutriments dans le sol, la disponibilité en eau, la concentration de CO₂, la température, la présence de stress abiotiques et biotiques sont quelques-uns des facteurs majeurs qui peuvent avoir un effet sur la croissance d'un arbre et auxquelles la plante doit répondre au cours de sa saison de croissance (Kirschbaum, 2000). D'ailleurs, tout facteur qui influence, par exemple, la photosynthèse, la conductance stomatique, la transpiration a des effets sur l'efficacité d'utilisation de l'eau et par conséquent la croissance d'une plante (Brugnoli et Farquhar, 2000). La plante répond ainsi différemment selon le type de stimulus en activant des processus physiologiques et moléculaires qui dépendent fortement de l'environnement. Il est donc probable que la discrimination isotopique du carbone 13 et la croissance soient des caractères plus complexes et composés que le débourrement. Cela pourrait expliquer l'héritabilité souvent plus faible, l'instabilité des QTL d'un environnement à l'autre et la moindre conservation des QTL entre espèces. Il ne faut pas oublier, par exemple, qu'en 2000 et 2001 les individus de châtaignier étaient placés dans des pots. Il est à noter que dans cette situation les conditions de croissance ne sont pas comparables à celles subies par les plantes sur le terrain. Afin de bien comparer les QTL contrôlant ces caractères entre les deux espèces, il faudrait d'une part utiliser des conditions environnementales homogènes et d'autre part décomposer ces caractères complexes en leurs composantes fonctionnelles élémentaires.

- Le travail de Wu (1998b), enfin, nous suggère une séduisante explication à la différence de conservation des QTL entre les trois caractères. Wu a utilisé l'approche de détection des QTL afin d'identifier des locus contrôlant la plasticité phénotypique de la croissance chez le peuplier. Il existe deux hypothèses quant à la nature des gènes responsables de la plasticité phénotypique : d'une part, l'hypothèse de la « sensibilité allélique » et d'autre part, celle de la régulation génique (Via et al., 1995). Selon la première hypothèse les mêmes locus seraient impliqués dans la détermination de la variabilité phénotypique d'un caractère donné dans des environnements différents. Mais les allèles à ces locus auraient des effets différents selon le type d'environnement ou bien des allèles différents seraient exprimés dans des environnements différents. L'hypothèse de régulation génique, au contraire, prévoit que les gènes de régulation

pourraient déterminer des phénotypes différents selon l'environnement grâce à l'activation ou bien la répression de gènes structuraux. Dans son article, Wu (1998b) fait l'hypothèse que les QTL détectés dans un environnement particulier et qui ne sont pas stables d'un environnement à l'autre constituent probablement les gènes responsables de la plasticité phénotypique déterminée par la régulation génique. Au contraire, les QTL stables d'un environnement à l'autre seraient les gènes contrôlant le caractère dont la plasticité phénotypique se réalise surtout par « sensibilité allélique ». D'après Wu (1998b) la régulation génique serait le phénomène principal à la base de la plasticité phénotypique de la croissance chez le peuplier. La plasticité phénotypique est un véritable caractère, lui-même en évolution (Schlichting, 1986). Les bases génétiques et moléculaires de la plasticité phénotypique ne sont pas encore bien connues, toutefois il est probable que les deux mécanismes cités auparavant soient tous les deux responsables de la variation phénotypique du même génotype dans des environnements différents. La régulation génique, les différents effets des différents allèles et probablement l'épistasie joueraient un rôle dans la détermination génétique de la plasticité phénotypique (Wu, 1998b). En considérant les trois caractères adaptatifs étudiés, nous pouvons alors faire l'hypothèse d'une plasticité phénotypique basée sur des mécanismes différents dans les trois cas. Imaginons, en effet, que la plasticité phénotypique de la croissance et de la discrimination isotopique du carbone ¹³ soit surtout déterminée par des mécanismes de régulation génique. Ces mécanismes répondraient aux différents stimuli environnementaux en déclenchant la réponse physiologique la plus adaptée dans l'environnement en question. Il est probable que les gènes structuraux activés ou réprimés soient différents dans les différentes situations. Cela pourrait expliquer l'instabilité des QTL d'un environnement à l'autre et l'absence de conservation de la plupart des QTL entre les espèces. Au contraire, la plasticité phénotypique du débourrement pourrait être principalement déterminée par le mécanisme de « sensibilité allélique ». Comme le principal stimulus du débourrement est la température, la réponse moléculaire sera caractérisée principalement par la voie de transduction du signal activé par ce facteur environnemental. Il est probable que les mécanismes moléculaires impliqués soient donc les mêmes, ce qui expliquerait encore la stabilité et la conservation de la plupart des QTL entre les deux espèces étudiées. Toutefois, si cette hypothèse est vraie, alors il faut en conclure que pour ce caractère la diversité adaptative entre les individus, les populations et les espèces dérive des mêmes locus.

Le troisième résultat d'intérêt est, donc, la conservation de la plupart des QTL impliqués dans le débourrement entre le chêne pédonculé et le châtaignier européen. Comment expliquer cette conservation et le maintien du polymorphisme à ces locus en dépit de 60 millions d'années de divergence ? Il est à noter que chez les familles de plantes qui montrent une grande conservation de la synténie et de la colinéarité, la « colocalisation » de certains QTL, contrôlant le même caractère, a déjà été mise en évidence. Par exemple, chez les Solanaceae il a été montré que les QTL impliqués dans des caractères du fruit et de la feuille, sélectionnés indépendamment chez des espèces différentes au cours de la domestication, sont conservés (Paterson et al., 1991 ; Doganlar et al., 2002b ; Frary et al., 2003 ; Frary et al., 2004). Néanmoins, la conservation des QTL n'implique pas forcément que les mêmes gènes soient impliqués de la même manière dans le processus étudié. Par exemple, en ce qui concerne les Solanaceae et les caractères de domestication, il semble que :

- les QTL conservés n'expliquent pas la même proportion de variance phénotypique chez les différentes espèces (Doganlar et al., 2002b) ;
- le même gène impliqué dans un processus commun peut être régulé de façon différente chez les différentes espèces (Thorup et al., 2000) ;
- il existe des QTL qui ne correspondent pas entre les espèces qui pourraient expliquer la diversité interspécifique (Doganlar et al., 2002b).

Une situation similaire a été mise en évidence chez les Poaceae. Malgré la conservation étonnante de la position des QTL contrôlant les caractères liés à la domestication (Paterson et al., 1995 ; Hu et al., 2003), récemment Doust et al. (2004) ont démontré que le gène *tb1*, qui correspond au candidat positionnel majeur contrôlant le tallage chez le maïs, a un effet mineur chez le millet et que dans la même région génomique d'autres gènes pourraient représenter de meilleurs candidats dans le cas du millet. Ces exemples montrent clairement que la correspondance de QTL entre les espèces n'implique pas forcément que les gènes impliqués soient les mêmes. De plus, les études avancées chez la drosophile ont montré que plusieurs mutations différentes du même locus, souvent dans la partie non codante du gène et donc responsables probablement de variations au niveau de la régulation du gène, semblent impliquées dans la variation des caractères complexes (Phillips, 1999). Cela suggère que même quand le locus impliqué correspond entre espèces, le type de mutation et par conséquent l'effet du locus sur le caractère quantitatif peuvent être différents.

Cela dit, il faut noter que chez le chêne et le châtaignier la conservation des QTL contrôlant le débourrement reste exceptionnelle. Même si cela n'implique pas que les gènes impliqués soient les mêmes, il est néanmoins probable que quelque-uns parmi les QTL conservés soient déterminés par des gènes orthologues. Chez les arbres forestiers, le débourrement est un caractère qui est soumis à une forte sélection naturelle, comme le montre la différenciation élevée entre les populations provenant, par exemple, de latitudes ou altitudes différentes. Cependant, la variabilité pour ce caractère à l'intérieur des populations reste élevée, suggérant que les populations ne sont pas forcément fixées à des allèles divergents (Howe et al., 2003). Dans l'hypothèse de la coïncidence au moins de certains des gènes impliqués dans le débourrement chez les deux espèces, nous devons alors expliquer comment le polymorphisme a été maintenu à ces locus, malgré la forte pression sélective et 60 millions d'années de divergence entre les deux espèces. Une explication possible peut être trouvée dans l'hypothèse que nous avons faite auparavant concernant la plasticité phénotypique. Si les mécanismes moléculaires impliqués dans le débourrement ont été conservés au cours de l'évolution et la plasticité phénotypique de ce caractère est principalement déterminée par la « sensibilité allélique », alors il est probable qu'une hétérozygotie élevée soit maintenue à ces locus pour que la plasticité phénotypique du caractère reste également élevée. Une autre hypothèse est une probable sélection balancée³⁵ ou disruptive³⁶ pour ce caractère déterminée par un phénomène de coévolution entre l'arbre et des espèces d'insectes dont le cycle vital est lié au débourrement. Il a été montré que différentes espèces d'insectes herbivores et défoliateurs adaptent leur phénologie d'éclosion à la date de débourrement de leur espèce hôte, par exemple le chêne pédonculé (Crawley et Akhteruzzaman, 1988). Tikkanen et Lyytikäinen-Saarenmaa (2002) ont mis en évidence que le temps d'éclosion d'une de ces espèces d'insectes (*Operophtera brumata*) est contrôlé génétiquement et que le synchronisme entre la date d'éclosion des insectes et la date de débourrement de son hôte (le chêne pédonculé) est de grande importance pour la survie des larves. Ce synchronisme serait soumis à une forte pression sélective. Chez le chêne pédonculé, la pression de sélection représentée par les insectes défoliateurs conduit à un étalement des dates de débourrement au sein des populations : la fitness moyenne de la population sera d'autant plus élevée que cet étalement couvre une large période. Au plan moléculaire cela doit se traduire au sein des populations par une forte variabilité allélique résultant soit de la sélection balancée, soit de la sélection

³⁵ Sélection balancée : situation dans laquelle la fitness des hétérozygotes est plus grande que celle des homozygotes. Nous disons alors qu'il y a un avantage des hétérozygotes.

³⁶ Sélection disruptive : situation dans laquelle les phénotypes extrêmes sont favorisés par rapport aux phénotypes intermédiaires.

disruptive. Le maintien d'une variabilité élevée pour la date de débourrement serait ainsi un mécanisme adaptatif des populations de chêne afin de se protéger de l'attaque de ces insectes (Tikkanen et Tiitto, 2003). Ce phénomène concerne plusieurs espèces y compris le châtaignier et il pourrait expliquer l'hétérozygotie élevée aux locus contrôlant le débourrement. Si cette forme de sélection était présente chez les deux espèces, nous pourrions comprendre qu'il n'y ait pas eu fixation des allèles aux différents QTL. Néanmoins, cette hypothèse suggère aussi que ces forces évolutives (sélection balancée ou disruptive) se soient maintenues durant d'aussi grandes périodes de temps (plusieurs dizaines de millions d'années). A défaut, nous pourrions suggérer également l'hypothèse que le polymorphisme soit maintenu aussi par des taux de mutation plus élevés à ces QTL.

E.3. IDENTIFICATION DE GÈNES CANDIDATS POUR LE CARACTÈRE DE DÉBOURREMENT.

E.3.1. Le concept de gène candidat.

Le concept de gène candidat s'est développé dans le domaine des études sur les maladies humaines. La définition que nous trouvons dans le « National Human Genome Research Institute » est la suivante : il s'agit d'un gène, qui se trouve dans une région génomique impliquée dans une maladie donnée, dont le produit protéique suggère qu'il soit le déterminant génétique de la maladie. Cette définition contient les différents aspects du concept de gène candidat, qui est maintenant largement utilisé dans tout domaine de la génomique animale et végétale. En règle générale, un gène candidat est un gène, pour lequel il existe un certain nombre de présomptions conduisant à le rendre responsable de la variation phénotypique au caractère observé. Ces présomptions sont de différentes natures (Pflieger et al., 2001) :

- 1) présomption relative à la position du gène sur la carte génétique : il s'agit d'un gène cartographié dans une région génomique elle-même impliquée dans le contrôle génétique d'un caractère. Par exemple, il peut s'agir d'un gène qui « colocalise » avec un QTL ;
- 2) présomption relative à la fonction du gène chez d'autres espèces : il s'agit d'un gène dont la fonction, connue grâce à des études de biologie moléculaire, biochimie et physiologie, semble être liée à un caractère donné ;
- 3) présomption relative à l'expression du gène dans l'espèce étudiée : les gènes différenciellement exprimés en réponse à un stress quelconque, au cours d'un processus de développement ou encore entre deux états physiologiques différents, peuvent être impliqués dans le phénomène en question.

Traditionnellement, en biologie humaine, l'approche dénommée « des gènes candidats » correspond à l'utilisation des ces gènes au sein d'études d'association³⁷ (Kwon et Goate, 2000). En biologie végétale, un gène candidat correspond d'une manière générale à un gène pour lequel tout un faisceau de présomptions suggère qu'il est impliqué dans le caractère étudié. Dans ce cadre, un gène dont la fonction serait potentiellement impliquée dans un caractère et, par ailleurs, localisé dans une région génomique contrôlant le processus en

³⁷ Etudes d'association : recherche d'une association statistique entre la variabilité allélique à un locus donné et la variabilité phénotypique d'un caractère.

question, constitue un candidat pertinent à considérer pour des études d'association et/ou de transformation de mutants (donc de complémentation fonctionnelle) afin de valider son implication dans le processus étudié. Notre étude comparative entre les deux genres, a indéniablement fourni certaines présomptions de nature « positionnelle ».

Comme les STS, cartographiés dans le cadre de cette thèse, dérivent d'une étude d'identification de gènes différentiellement exprimés au cours du débourrement et les QTL contrôlant ce caractère ont été positionnés sur la même carte, nous pouvons utiliser les « colocalisations » entre STS et QTL afin d'identifier des gènes candidats pour le caractère de débourrement. Les gènes candidats identifiés de cette manière constituent des candidats positionnels, fonctionnels et dans certains cas expressionnels mais dont le rôle dans le processus de débourrement reste à démontrer. Afin de valider l'implication de ces gènes dans le débourrement du chêne et du châtaignier, des études d'association en populations naturelles seront nécessaires. Il est, en effet, à noter que la cartographie des STS avec 57 individus ne permet pas une localisation précise de ces gènes et, étant donné l'intervalle de confiance sur la position des QTL, les « colocalisations » peuvent se produire simplement par chance. De plus, comme nous le verrons plus loin, plusieurs STS, correspondant à des gènes différentiellement exprimés au cours du débourrement (candidats expressionnels) peuvent « colocaliser » avec le même QTL. Il est aussi probable qu'aucun des STS colocalisés ne soit responsable des QTL détectés. En effet, la région couverte par l'intervalle de confiance contient des centaines de gènes et nous n'avons cartographié qu'un très faible nombre de ces gènes. Même dans l'hypothèse où l'STS cartographié serait responsable du QTL détecté, il est aussi probable que la mutation à la base de la variation du QTL dans la descendance du croisement contrôlé soit différente de celle responsable de la variation en populations naturelles. Malgré les limites de la « colocalisation » entre QTL et gènes candidats illustrées par le travail de Liu et al. (2003), ce type d'analyse constitue une approche pertinente pour sélectionner les gènes candidats à tester dans des études d'association ultérieures.

E.3.2. Quelques gènes candidats positionnels : « colocalisation » entre QTL et EST.

Nous décrirons deux exemples concernant les couples d'homologues Q1-C6 et Q5-C4 afin de démontrer comment les résultats obtenus dans le cadre de cette thèse peuvent être utilisés pour l'identification de gènes candidats potentiels. Aucune analyse détaillée de la fonction des ces gènes ne sera présentée, cela n'étant pas le but de cette thèse, mais simplement un exemple de démarche d'identification de gènes candidats sera fourni. Dans les deux cas que nous allons décrire, un QTL contrôlant le débourrement a été détecté et il est conservé entre le

chêne et le châtaignier. Nous avons donc considéré tous les STS qui ont été cartographiés dans l'intervalle contenant ces QTL afin de vérifier la présence de fonctions potentiellement impliqués dans le débourrement.

Groupes Q1 et C6. Les STS cartographiés dans l'intervalle d'intérêt sont : PHYB/D, ESTCL4 et ESTS02F02, correspondant aux gènes du phytochrome, de l'histone H4 et de l'histone H3 (tableau 16).

Les gènes du phytochrome, impliqué chez les plantes dans la perception de la lumière, ont été déjà colocalisés avec des QTL impliqués dans le débourrement et l'arrêt de croissance chez le peuplier (Frewen et al., 2000). Ce gène semble impliqué dans l'arrêt de croissance et le début de la période de dormance, car la perception de la diminution de la longueur de la journée est très importante dans ce processus (Horvath et al., 2003). En effet, un QTL impliqué dans l'arrêt de croissance a été positionné chez le châtaignier sur ce groupe, même si la position ne semble pas correspondre à celle du gène du phytochrome qui n'a été cartographié que chez le chêne.

La « colocalisation » des gènes des histones semble être très intéressante. Ces gènes interviennent au cours de la phase S du cycle cellulaire. Derory et al. (soumis) ont démontré que les deux histone H4 et H3 sont différentiellement exprimés entre l'état quiescent du bourgeon et le début du débourrement. Devitt and Stafstrom (1995) ont également montré que l'ARN messager de l'histone H4, en faible quantité dans le bourgeon dormant, augmente au cours du débourrement et diminue quand le bourgeon arrête sa croissance pour rentrer dans la phase de dormance. Il est intéressant de noter que chez le châtaignier deux QTL impliqués dans le débourrement et l'arrêt de croissance « colocalisent » dans cette région du groupe de liaison. Les QTL impliqués dans ces deux caractères localisés sur le groupe 6M du châtaignier ont un effet de substitution allélique concordant (positif pour les deux QTL, tableau 10) ce qui se traduit dans notre système de mesure par une association entre un débourrement tardif et un arrêt de croissance précoce ou vice versa. Dans le paragraphe C.3.3.7 nous avons fait l'hypothèse que des gènes impliqués dans la régulation (activation et/ou inhibition) de l'activité mitotique du bourgeon sont responsables de ces QTL. La « colocalisation » avec les gènes des histones renforce cette hypothèse.

Groupes Q5 et C4. Ces groupes contiennent plusieurs STS fonctionnellement intéressants. Un QTL contrôlant le débourrement est conservé entre les deux espèces, deux QTL semblent présents chez le châtaignier. Chez les deux espèces, l'intervalle de confiance des QTL

contient tous les STS d'intérêt : ESTS07A09 (transporteur ABC « ATP Binding Cassette »), PHY B/D (phytochrome), GA3 (voie de synthèse des gibberellines), ESTCL31 (protéine associée à la dormance, réprimée par l'auxine), ESTCL83 (Histone H3), ESTCL8 (protéine ribosomale L13), ESTCL5 (protéine ribosomale L18a) et ESTS01E07 (facteur 3 de réponse à l'auxine).

Les transporteurs ABC sont des canaux ioniques impliqués dans la perception du signal lumineux (Spalding, 2000). Il est intéressant de noter que ce gène est positionné près d'un des gènes du phytochrome (PHY B/D), une autre protéine importante pour la perception du signal lumineux. Vu les interactions entre le phytochrome B et les hormones gibbérellines dans la perception de la lumière, notamment dans la régulation de la croissance de l'hypocotyle chez *Arabidopsis* (Borevitz et al., 2002), il est également intéressant de noter la position proche à ces gènes de celui codant pour la protéine GA3, qui intervient dans la biosynthèse des hormones gibbérellines.

Les EST CL31, CL83, CL8, CL5 et S01E07 sont tous différemment exprimés au cours du débourrement chez le chêne (Derory et al., soumis). Particulièrement, les hormones auxine et gibbérellines sont impliqués dans le processus de dormance (Rohde et al., 2000). La présence dans cette région d'un gène impliqué dans la voie de biosynthèse des hormones gibbérellines et de deux gènes codant pour des protéines qui répondent à l'auxine pourrait donc expliquer la détection de plusieurs QTL contrôlant le débourrement dans le même intervalle.

Les deux exemples montrent clairement que plusieurs gènes candidats intéressants peuvent être colocalisés dans la même région contenant un QTL. Il n'est pas possible, à ce stade de l'étude, de savoir lequel de ces gènes est responsable des QTL, s'il s'agit du même gène impliqué chez le chêne et le châtaignier et, étant donné le faible nombre de gènes cartographiés et la probabilité élevée de « colocalisations » par chance, il est aussi probable qu'aucun de ces gènes ne soit vraiment responsables des QTL détectés. La validation de l'implication dans le débourrement des gènes candidats identifiés en suivant cette démarche est donc nécessaire et elle peut se réaliser dans le cadre d'études d'association en populations naturelles.

F. Conclusion et perspectives.

E quindi uscimmo a riveder le stelle.

(Dante Alighieri, Divina Commedia, Inferno)

F.1. LA DÉTECTION DE QTL CONTRÔLANT LES CARACTÈRES ADAPTATIFS : EST-ELLE UNE STRATÉGIE INTÉRESSANTE POUR L'ÉTUDE DES BASES GÉNÉTIQUES DE L'ADAPTATION ?

- La construction de la carte génétique du châtaignier a fourni un outil précieux pour l'identification des régions génomiques qui contrôlent les caractères adaptatifs. La famille de plein-frères, issue d'un croisement contrôlé entre deux individus appartenant à deux populations adaptées à des environnements contrastés, a montré une grande variabilité pour les caractères étudiés. Cela dérive en partie du niveau élevé d'hétérozygotie des espèces forestières qui permet une détection efficace des QTL dans une descendance de type F₁. Comme l'ont souligné Orr et Coyne (1992), la détection des QTL, contrôlant les caractères adaptatifs, appliquée à des croisements entre individus appartenant à des populations naturelles est une stratégie efficace pour la dissection de ces caractères complexes en composantes mendéliennes.
- Les résultats obtenus chez le châtaignier mettent en évidence la distribution en forme de L des effets des QTL prévue par la théorie de Orr (1998). Hurme et al. (2000) ont démontré chez le *Pinus sylvestris* que des allèles à effet fort contrôlant des caractères d'adaptation sont parfois détectés dans les populations naturelles. Les caractères adaptatifs semblent donc être contrôlés par beaucoup de locus à effet faible ou modéré et peu de locus à effet fort.
- Des QTL stables d'une année à l'autre ainsi que des QTL spécifiques d'une année ont pu être détectés dans notre expérience. Jermstad et al. (2003) ont mis en évidence que chez *Pseudotsuga menziesii* certains QTL contrôlant le début et l'arrêt de la saison de croissance sont stables d'un environnement à l'autre. En revanche, d'autres QTL semblent spécifiques d'un environnement. La détection des QTL permet donc de révéler à la fois des locus qui interviennent toujours dans la variabilité phénotypique des caractères adaptatifs et des locus qui ne sont impliqués que dans certaines conditions environnementales. Comme le montre Wu (1998b), la détection des QTL pourrait permettre une étude des locus impliqués dans la plasticité phénotypique des caractères adaptatifs.
- La détection de la majorité des QTL sur la carte femelle du châtaignier nous a mené à formuler une hypothèse sur l'adaptation locale de ce génotype à son environnement d'origine. Chez l'orge, une approche QTL a été récemment utilisée pour étudier les

bases génétiques de la différenciation adaptative entre deux populations naturelles dérivant d'environnements contrastés (Verhoeven et al., 2004).

L'approche QTL peut donc fournir des éléments de réponse aux questions encore ouvertes de certains aspects de la génétique des caractères complexes, notamment des caractères adaptatifs. Il faut à cet égard rappeler l'intérêt d'utiliser des populations naturelles et d'étudier des caractères qui aient subi l'action de la sélection naturelle. Les espèces forestières nous offrent des exemples intéressants de ce point de vue. Toutefois, la détection des QTL chez ces espèces reste limitée par les effectifs faibles, le temps de génération long de ces organismes et les expériences souvent très lourdes à réaliser (essais au champ, délai nécessaire à l'évaluation phénotypique, etc.).

F.2. LA COMPARAISON DES QTL ENTRE LE CHÊNE ET LE CHÂTAIGNIER : POUVONS-NOUS TIRER QUELQUES CONCLUSIONS CONCERNANT L'ÉVOLUTION DES CARACTÈRES ADAPTATIFS ?

La cartographie de marqueurs orthologues chez le chêne et le châtaignier et l'alignement des 12 groupes de ces deux espèces, nous a fourni une base pour la comparaison des QTL contrôlant les mêmes caractères adaptatifs chez les deux espèces. La raison principale pour laquelle cette comparaison a été réalisée est de vérifier si les locus contrôlant les trois caractères adaptatifs étudiés sont conservés entre les deux espèces. Nous avons montré que :

- le déterminisme génétique du débourrement, de la croissance en hauteur et de la discrimination isotopique du carbone 13 est plutôt conservé en termes de nombre de QTL détectés et de variance phénotypique expliquée par les QTL. Malgré certaines limites de la détection des QTL que nous avons discutées tout au long du manuscrit, il semble que ces caractéristiques (nombre et effets) restent conservés au cours de l'évolution.
- Les QTL contrôlant le débourrement sont conservés d'une espèce à l'autre. La stabilité des QTL contrôlant ce caractère a été vérifiée au niveau intra spécifique et elle semble respectée au niveau interspécifique.
- En revanche, peu de QTL contrôlant la croissance en hauteur et la discrimination isotopique sont conservés entre les deux espèces. Au niveau intra spécifique une instabilité des QTL contrôlant ces caractères a également été observée. La différence d'héritabilité et la complexité des trois caractères, ainsi que peut-être une plasticité

phénotypique basée sur des mécanismes génétiques différents pourraient expliquer cette différence entre caractères.

- Concernant le débourrement une différence importante a été mise en évidence sur le groupe Q10-C10. Un QTL stable chez le chêne n'a pas été détecté chez le châtaignier. En revanche, même si la plupart des QTL pour la croissance en hauteur ne sont pas conservés, néanmoins sur ce même couple d'homologue beaucoup de QTL ont été détectés chez les deux espèces, suggérant la présence d'une région d'intérêt pour ce caractère conservée au cours de l'évolution.

Le résultat le plus significatif de cette comparaison est la conservation des QTL impliqués dans le débourrement. Apparemment, les mêmes régions génomiques sont impliquées dans le contrôle de ce caractère, malgré les 60 millions d'années de divergence entre les deux espèces. Ce résultat n'implique pas que les mêmes gènes soient impliqués dans les deux espèces, toutefois la stabilité de ces QTL au cours de l'évolution est une évidence de la stabilité du génome des Fagacées suggérant que la diversification interspécifique ne se soit pas produite par le biais de réarrangements chromosomiques majeurs. La différence en termes de conservation des QTL entre d'une part le débourrement et d'autre part la croissance et la discrimination isotopique du carbone 13 suggère une différence au niveau des mécanismes adaptatifs impliqués dans les trois caractères. Nous avons proposé quelques hypothèses qui restent, toutefois, purement spéculatives. Cette comparaison représente un point de départ pour l'étude de l'évolution de ces caractères dans ces espèces. La cartographie de gènes candidats pour les trois caractères semble une stratégie pertinente pour la suite de cette étude. La « colocalisation » de ces gènes avec les QTL détectés chez les deux espèces fournit des candidats pertinents à valider dans des études d'association comme nous l'avons signalé dans la partie E.3. La validation de l'implication de ces gènes candidats dans les caractères pourra éclaircir certaines des mécanismes moléculaires sous-jacents aux QTL.

F.3. LA CARTOGRAPHIE GÉNÉTIQUE COMPARÉE CHEZ LE CHÊNE ET LE CHÂTAIGNIER : VERS LE « SYSTÈME GÉNÉTIQUE UNIQUE » DES FAGACÉES.

La cartographie génétique de marqueurs orthologues (SSR et STS dérivés des EST) nous a permis d'associer les 12 groupes de liaison du chêne pédonculé et du châtaignier européen. La similarité du karyotype, de la taille génétique et physique des deux génomes, la conservation de la macro synténie, de la macro colinéarité et des QTL constituent un ensemble

d'indications en faveur d'un génome plutôt stable. Trois perspectives intéressantes peuvent être envisagées :

- 1) nous avons démontré que le transfert de connaissances moléculaires et génétiques est possible entre ces deux espèces. Cela constitue un résultat de grande importance tout particulièrement pour le châtaignier. Chez le châtaignier la disponibilité de séquences EST est encore très limitée. Etant donné les résultats de cette thèse, les EST disponibles chez le chêne pourront être facilement exploitées chez le châtaignier. Les marqueurs orthologues cartographiés chez les deux espèces devraient permettre d'associer les groupes de liaison de tout autre carte génétique développée dans d'autres espèces de chêne et de châtaignier. Les résultats obtenus par Paul Sisco le confirment : la carte génétique dérivée d'un croisement interspécifique du châtaignier américain et chinois a été alignée avec la carte du châtaignier européen grâce aux SSR orthologues. Ces mêmes marqueurs orthologues pourront être utilisés au sein de la famille des Fagacées pour des études de diversité génétique, phylogénie, flux de gènes.
- 2) La résolution de notre étude de cartographie comparée reste très faible. Afin de comprendre les différences et donc les mécanismes de la différenciation évolutive des génomes du chêne et du châtaignier, il faudrait augmenter le nombre des marqueurs orthologues communs entre les deux espèces. Comme nous avons pu le voir dans l'introduction de ce manuscrit, les micro réarrangements et les inversions paracentriques sont souvent impliqués dans l'évolution des génomes des angiospermes. De nombreux marqueurs et un effectif de cartographie plus élevé seront nécessaires afin de révéler ce type de réarrangements. Deux perspectives d'études semblent envisageables afin de mieux comprendre la structure du génome des Fagacées. D'une part, certains gènes très conservés au cours de l'évolution des angiospermes ont été identifiés et cartographiés chez plusieurs familles d'eucotylédones (Fulton et al., 2000 ; Dominguez et al., 2003). Il a été ainsi démontré que certains blocs de gènes restent plutôt conservés au cours de l'évolution. Ces gènes ont été amplifiés chez *Beta vulgaris*, *Helianthus annuus*, *Prunus* spp., et *Arabidopsis thaliana* et les séquences des amorces développées pour l'amplification de ces gènes sont disponibles (Dominguez et al., 2003). Il semble donc possible de cartographier les mêmes gènes chez le chêne et le châtaignier et vérifier d'abord si ces blocs sont également conservés chez les Fagacées et où ils sont situés dans le génome. D'autre part, la stratégie dénommée en anglais « mapping in silico » pourrait également

apporter des résultats intéressants à la structure du génome. Il s'agit tout simplement de positionner par homologie de séquences sur le génome d'une espèce modèle les EST, par exemple, cartographiés chez une autre espèce afin d'identifier les régions conservées et les réarrangements entre les deux génomes (Zhu et al., 2003 ; Gebhardt et al., 2003). La cartographie d'un nombre important d'EST chez le chêne et le châtaignier pourrait permettre de réaliser ce type d'approche et de comparer la carte génétique du chêne et du châtaignier avec le génome du peuplier qui sera bientôt complètement séquencé (Brunner et al., 2004). Afin de cartographier un très grand nombre d'EST sur un effectif plus élevé, les techniques de « transcriptome mapping » et « ecotilling » semblent très intéressantes (Li et al., 2003 ; Comai et al., 2004).

- 3) Les deux cartes génétiques du chêne et du châtaignier contiennent désormais un ensemble d'informations génétiques (QTL) et moléculaires (EST) qui peuvent être exploitées dans une vision plus fonctionnelle. Il nous semble intéressant, par exemple, d'engager la cartographie des différents membres d'une même famille multigénique, comme celle du phytochrome. Une fois les gènes orthologues identifiés et caractérisés entre les deux espèces, des perspectives très intéressantes peuvent être envisagées. Les gènes orthologues et paralogues de la même famille multigénique pourraient être utilisés à la fois chez le chêne et chez le châtaignier afin d'étudier d'une part leur organisation dans le génome et d'autre part, grâce à des études de diversité nucléotidique, leur différenciation éventuelle au niveau fonctionnel. L'identification de gènes candidats pour les trois caractères étudiés, leur « colocalisation » avec les QTL conservés entre les deux espèces et, enfin, leur validation au sein d'études d'association en populations naturelles fourniront des outils de grand intérêt pour l'étude de la diversité adaptative intra et interspécifique dans la famille des Fagacées.

Références bibliographiques

Une vie entière à lire aurait comblé mes vœux.

(Michel Houellebecq, Extension du domaine de la lutte)

A

Acarkan A., Rossberg M., Koch M., Schmidt R. (2000) Comparative genome analysis reveals extensive conservation of genome organisation for *Arabidopsis thaliana* and *Capsella rubella*. *Plant J.* 23 : 55-62.

Adams K.L., Cronn R., Percifield R., Wendel J.F. (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc. Natl. Acad. Sci. USA* 15 : 4369-4371.

Ahmadian A. (2001) Approaches for analysis of mutations and genetic variations. Departement of Biotechnology, Royal Institute of Technology, KTH, Stockholm, Sweden, ISBN 91-7283-062-X.

Ahn S., Tanksley S.D. (1993) Comparative linkage maps of the rice and maize genomes. *Proc. Natl. Acad. Sci. USA* 90 : 7980-7984.

Ahuja M.R. (2001) Recent advances in molecular genetics of forest trees. *Euphytica* 121 : 173-195.

Akhunov E.D., Akhunova A.R., Linkiewicz A.M., Dubcovsky J., Hummel D., Lazo G.R., Chao S., Anderson O.D., David J., Qi L., Echalié B., Gill B.S., Miftahudin, Gustafson J.P., La Rota M., Sorrells M.E., Zhang D., Nguyen H.T., Kalavacharla V., Hossain K., Kianian S.F., Peng J., Lapitan N.L., Wennerlind E.J., Ndauti V., Anderson J.A., Sidhu D., Gill K.S., McGuire P.E., Qualset C.O., Dvorak J. (2003a) Synteny perturbations between wheat homologous chromosomes caused by locus duplications and deletions correlate with recombination rates. *Proc. Natl. Acad. Sci. USA* 100 : 10836-10841.

Akhunov E.D., Goodyear A.W., Geng S., Qi L.L., Echalié B., Gill B.S., Miftahudin, Gustafson J.P., Lazo G.R., Chao S., Anderson O.D., Linkiewicz A.M., Dubcovsky J., La Rota M., Sorrells M.E., Zhang D., Nguyen H.T., Kalavacharla V., Hossain K., Kianian S.F., Peng J., Lapitan N.L., Gonzalez-Hernandez J.L., Anderson J.A., Choi D.W., Close T.J., Dilbirligi M., Gill K.S., Walker-Simmons M.K., Steber C., McGuire P.E., Qualset C.O., Dvorak J. (2003b) The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. *Genome Res.* 13 : 753-763.

Aldrich P.R., Michler C.H., Sun W., Romero-Severson J. (2002) Microsatellite markers for Northern red oak (Fagaceae: *Quercus rubra*). *Mol. Ecol. Notes* 2 : 472-474.

Allard R.W. (1956) Formulas and tables to facilitate the calculation of recombination values in heredity. *Hilgardia* 24 : 235-278.

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215 : 403-410.

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796-815.

Arcade A., Labourdette A., Falque M., Mangin B., Chardon F., Charcosset A., Joets J. (2004) BioMercator: integrating genetic maps and QTL towards discovery of candidate genes. *Bioinformatics* doi:10.1093/bioinformatics/bth230.

Axeisson T., Shavorskaya O., Lagercrantz U. (2001) Multiple flowering time QTLs within several Brassica species could be the result of duplicated copies of one ancestral gene. *Genome* 44 : 856-864.

B

Babula D., Kaczmarek M., Barakat A., Delseny M., Quiros C.F., Sadowski J. (2003) Chromosomal mapping of *Brassica oleracea* based on ESTs from *Arabidopsis thaliana*: complexity of the comparative map. *Mol. Gen. Genomics* 268 : 656-665.

Ball C.A. et Cherry J.M. (2001) Genome comparisons highlight similarity and diversity within the eukaryotic kingdoms. *Curr. Opin. Chem. Biol.* 5 : 86-89.

Bancroft I. (2001) Duplicate and divergence: the evolution of plant genome microstructure. *Trends Genet.* 17 : 89-93.

Barakat A., Matassi G., Bernardi G. (1998) Distribution of genes in the genome of *Arabidopsis thaliana* and its implications for the genome organization of plants. *Proc. Natl. Acad. Sci. U S A.* 95 :10044-10049.

Barakat A., Szick-Miranda K., Chang I.F., Guyot R., Blanc G., Cooke R., Delseny M., Bailey-Serres J. (2001) The organization of cytoplasmic ribosomal protein genes in the *Arabidopsis* genome. *Plant Physiol.* 127 : 398-415.

Barber V.A., Juday G.P., Finney B.P. (2000) Reduced growth of Alaskan white spruce in the twentieth century from temperature-induced drought stress. *Nature* 40 : 668-673.

Barnes S. (2002) Comparing *Arabidopsis* to other flowering plants. *Curr. Opin. Plant Biol.* 5 : 128-133.

Barreneche T., Bodenes C., Lexer C., Trontin J.F., Fluch S., Streiff R., Plomion C., Roussel G., Steinkellner H., Burg K., Favre J.M., Glössl J., Kremer A. (1998) A genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5SrDNA markers. *Theor. Appl. Genet.* 97 : 1090-1103.

Barton N.H., Keightley P.D. (2002) Understanding quantitative genetic variation. *Nature Rev. Genet.* 3 : 11-21.

Bateson W. (1913) Mendel's principles of heredity. Cambridge University Press, Cambridge.

Beavis W.D. (1995) The power and deceit of QTL experiments: lessons from comparative QTL studies. In Proceedings of the 49th Annual Corn and Sorghum Industry Research Conference, ASTA, Washington DC, pp 250-266.

Beckmann J.S., Soller M. (1983) Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. *Theor. Appl. Genet.* 67 : 35-43.

Ben Chaim A., Paran I., Grube R.C., Jahn M., van Wijk R., Peleman J. (2001) QTL mapping of fruit-related traits in pepper (*Capsicum annuum*). *Theor. Appl. Genet.* 102 : 1016-1028.

- Bennetzen J.L. et Freeling M. (1993) Grasses as a single genetic system: genome composition, colinearity and compatibility. *Trends Genet.* 9 : 259-261.
- Bennetzen J.L. (1998) The structure and evolution of angiosperm nuclear genomes. *Curr. Opin. Plant Biol.* 1 : 103-108.
- Bennetzen J.L. (2000a) Transposable element contributions to plant gene and genome evolution. *Plant Mol. Biol.* 42 : 251-269.
- Bennetzen J.L. (2000b) Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. *Plant Cell* 12 : 1021-1029.
- Bennetzen J.L., Ramakrishna W. (2002) Numerous small rearrangements of gene content, order and orientation differentiate grass genomes. *Plant Mol. Biol.* 48 : 821-827.
- Bennetzen J.L., Ma J. (2003) The genetic colinearity of rice and other cereals on the basis of genomic sequences analysis. *Curr. Opin. Plant Biol.* 6 : 128-133.
- Bevan M., Murphy G. (1999) The small, the large and the wild: the value of comparison in plant genomics. *Trends Genet.* 15 : 211-214.
- Blanc G., Barakat A., Guyot R., Cooke R., Delseny M. (2000) Extensive duplication and reshuffling in the Arabidopsis genome. *Plant Cell* 12 : 1093-1101.
- Bodénès C., Laigret F., Kremer A. (1996) Inheritance and molecular variations of PCR-SSCP fragments in pedunculate oak (*Quercus robur* L.). *Theor. Appl. Genet.* 93 : 348-354.
- Bonierbale M.W., Plaisted R.L., Tanksley S.D. (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* 120 : 1095-1103.
- Borevitz J.O., Maloof J.N., Lutes J., Dabi T., Redfern L.J., Trainer G.T., Werner J.D., Asami T., Berry C.C., Weigel D., Chory J. (2002) Quantitative trait loci controlling light and hormone response in two accessions of *Arabidopsis thaliana*. *Genetics* 160 : 683-696.
- Borevitz J.O., Chory J. (2004) Genomics tools for QTL analysis and gene discovery. *Curr. Opin. Plant Biol.* 7 : 132-136.
- Botta R., Akkac A., Marinoni D., Bounous G., Kampfer S., Steinkellner H., Lexer C. (1999) Evaluation of microsatellite markers for characterizing chestnut cultivars. *Acta Hort.* 494 : 277-282.
- Boutin S.R., Young N.D., Olson T.C., Yu Z.-H., Shoemaker R.C., Vallejos C.E. (1995) Genome conservation among three legume genera detected with DNA markers. *Genome* 38 : 928-937.
- Bowers J.E., Abbey C., Anderson S., Chang C., Draye X., Hoppe A.H., Jessup R., Lemke C., Lenington J., Li Z., Lin Y.R., Liu S.C., Luo L., Marler B.S., Ming R., Mitchell S.E., Qiang D., Reischmann K., Schulze S.R., Skinner D.N., Wang Y.W., Kresovich S., Schertz K.F., Paterson A.H. (2003a) A high-density genetic recombination map of sequence-tagged sites

for *Sorghum*, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. *Genetics* 165 : 367-386.

Bowers J.E., Chapman B.A., Rong J., Paterson A.H. (2003b) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422 : 433-438.

Bradshaw H.D., 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 : 963-973.

Brendel O., Pot D., Plomion C., Rozenberg P., Guehl J.M. (2002) Genetic parameters and QTL analysis of $\delta^{13}\text{C}$ and ring width in maritime pine. *Plant Cell Env.* 25 : 945-953.

Brendel V., Kurtz S., Walbot V. (2002) Comparative genomics of *Arabidopsis* and maize: prospects and limitations. *Genome Biol.* 3 : 1005.1-1005.6.

Britvin E., Minkov D., Glikson L., Ronin Y., Korol A. (2001) MultiQTL, an interactive package for genetic mapping of correlated quantitative trait complexes in multiple environments, version 2.0 (Demo). Plant and Animal Genome Conference IX, San Diego, CA.

Brondani C., Pereira R., Brondani V., Rangel P.H.N., Ferreira M.E. (2001) Development and mapping of *Oryza glumaepatula*-derived microsatellite markers in the interspecific cross *Oryza glumaepatula* x *O. sativa*. *Hereditas* 134 : 59-71.

Brown G.R., Kadel III E.E, Bassoni D.L., Kiehne K.L., Temesgen B., van Buijtenen J.P., Sewell M.M., Marshall K.A., Neale D.B. (2001) Anchored reference loci in loblolly pine (*Pinus taeda* L.) for integrating pine genomics. *Genetics* 159 : 799-809.

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., 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 : 1537-1546.

Brown S.M., Hopkins M.S., Mitchell S.E., Senior M.L., Wang T.Y., Duncan R.R., Gonzalez-Candelas F., Kresovich S. (1996) Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum (*Sorghum bicolor* L. Moench). *Theor. Appl. Genet.* 93 : 190-198.

Brubaker C.L., Paterson A.H., Wendel J.F. (1999) Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome* 42 : 184-203.

Brueggeman R., Rostoks N., Kudrna D., Kilian A., Han F., Chen J., Druka A., Steffenson B., Kleinhofs A. (2002) The barley stem rust-resistance gene *Rpg1* is a novel disease-resistance gene with homology to receptor kinases. *Proc. Natl. Acad. Sci. USA* 99 : 9328-9333.

Brugnoli E., Farquhar G.D. (2000) Photosynthetic fractionation of carbon isotopes. In *Photosynthesis: Physiology and Metabolism. Advances in Photosynthesis* (eds R.C. Leegood, T.D. Sharkey & S. von Caemmerer) Vol. 9 Chapter 17: 399-434. Kluwer Academic Publishers, Netherlands.

Brunner A.M., Busov V.B., Strauss S.H. (2004) Poplar genome sequence: functional genomics in an ecologically dominant plant species. *Trends Plant Sci.* 9 : 49-56.

Buck E., Hadonou M., James C., Blakesley D., Russell K. (2003) Isolation and characterisation of polymorphic microsatellites in European chestnut (*Castanea sativa* Mill.). *Mol. Ecol. Notes* 3 : 239-241.

Burke J.M., Lai Z., Salmaso M., Nakazato T., Tang S., Heesacker A., Knapp S.J., Rieseberg L.H. (2004) Comparative mapping and rapid karyotypic evolution in the genus *Helianthus*. *Genetics* 167 : 449-457.

Byrne M., Marquez-Garcia M.I., Uren T., Smith D.S., Moran G.F. (1996) Conservation and genetic diversity of microsatellite loci in the genus *Eucalyptus*. *Aust. J. Bot.* 44 : 331-341.

Byrne M., Murrell L.C., Owen J.V., Kriedemann P., Williams E.R., Moran G.F. (1997) Identification and mode of action of quantitative trait affecting seedling height and leaf area in *Eucalyptus nitens*. *Theor. Appl. Genet.* 94: 674-681.

C

Camus A. (1936-1954) Les chênes, monographie du genre *Quercus* et monographie du genre *Lithocarpus*. Encyclopédie Economique de Sylviculture. Vol. VI, VII, VIII. Académie des Sciences. Paris.

Cervera M.T., Storme V., Ivens B., Gusmao J., Liu B.H., Hostyn V., Van Slycken J., Van Montagu M., Boerjan W. (2001) Dense genetic linkage maps of three *Populus* species (*Populus deltoides*, *P. nigra* and *P. trichocarpa*) based on AFLP and microsatellite markers. *Genetics* 158 : 787-809.

Chagné D., Lalanne C., Madur D., Kumar S., Frigerio J.-M., Krier C., Decroocq S., Savouré A., Bou-Dagher-Kharrat M., Bertocchi E., Brach J., Plomion C. (2002) A high density genetic map of maritime pine based on AFLPs. *Ann. For. Sci.* 59 : 627-636.

Chagné D., Brown G.R., Lalanne C., Madur D., Pot D., Neale D.B., Plomion C. (2003) Comparative genome and QTL mapping between maritime and loblolly pines. *Mol. Breed.* 12 : 185-195.

Chagné D. (2004) Développement de marqueurs moléculaires chez le pin maritime (*Pinus pinaster* Ait.) et cartographie génétique comparée des conifères. Thèse Université Henry Poincaré, Nancy I.

Chagné D., Chaumeil P., Ramboer A., Collada C., Guevara A., Cervera M.T., Vendramin G.G., Garcia Virginie, Frigerio J.-M., Echt C., Richardson T., Plomion C. Cross species transferability and mapping of genomic and cDNA SSRs in pines. *Theor. Appl. Genet.*, sous presse.

Chakravarti A., Lasher L.A., Reefer J.E. (1991) A maximum likelihood method for estimating genome length using genetic linkage data. *Genetics* 128: 175-182.

- Chao S., Sharp P.J., Worland A.J., Warham E.J., Koeber R.M.D., Gale M.D. (1989) RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. *Theor. Appl. Genet.* 78 : 495-504.
- Chen M., SanMiguel P., De Oliveira A.C., Woo S.S., Zhang H., Wing R.A., Bennetzen J.L. (1997) Microcollinearity in *sh2*-homologous regions of the maize, rice and sorghum genomes. *Proc. Natl. Acad. Sci. USA* 94 : 3431-3435.
- Chen X., Levine L., Kwok P.-Y. (1999) Fluorescence polarization in homogeneous nucleic acid analysis. *Genome Res.* 9 : 492-498.
- Chen X., Cho Y.G., McCouch S.R. (2002) Sequence divergence of rice microsatellites in *Oryza* and other plant species. *Mol. Genet. Genomics* 268 : 331-343.
- Cheung W.Y., Landry B.S. (1996) Current status of genome mapping in the Cruciferae. In *Genome mapping in Plants*, A.H. Paterson ed. R.G. Landes Co. Georgetown, TX. pp 179-191.
- Cheung W.Y., Champagne G., Hubert N., Landry B.S. (1997) Comparison of the genetic maps of *Brassica napus* and *Brassica oleracea*. *Theor. Appl. Genet.* 94 : 569-582.
- Choi H.-K., Kim D., Uhm T., Limpens E., Lim H., Mun J.-H., Kalo P., Penmetsa R.V., Seres A., Kulikova O., Roe B. A., Bisseling T., Kiss G.B., Cook D.R. (2004) A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. *Genetics* 166 : 1463-1502.
- Churchill G.A., Doerge R.W. (1994) Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963-971.
- Comai L., Young K., Till B.J., Reynolds S.H., Greene E.A., Codomo C.A., Enns L.C., Johnson J.E., Burtner C., Odden A.R., Henikoff S. (2004) Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. *Plant J.* 37 : 778-786.
- The Complex Trait Consortium (2003) The nature and identification of quantitative trait loci: a community's view. *Nature Genet. Rev.* 4 : 911-916.
- Conedera M., Krebs P., Tinner W., Pradella M., Torriani D. (2004) The cultivation of *Castanea sativa* (Mill.) in Europe: from its origin to its diffusion on a continental scale. *Veget. Hist. Archaeobot.* 13 : 161-179.
- Conner P.J., Brown S.K., Weeden N.F. (1998) Molecular-marker analysis of quantitative traits for growth and development in juvenile apple trees. *Theor. Appl. Genet.* 96 : 1027-1035.
- Costa P. (1999) Réponses moléculaire, physiologique et génétique du pin maritime à une contrainte hydrique. Thèse Université Henri Poincaré, Nancy I.
- Crawley M.J., Akhteruzzaman M. (1988) Individual variation in the phenology of oak trees and its consequences for herbivorous insects. *Func. Ecol.* 2 : 409-415.

Cronn R.C., Small R.L., Wendel J.F. (1999) Duplicated genes evolve independently after polyploid formation in cotton. *Proc. Natl. Acad. Sci. USA* 96 : 14406-14411.

Crow J.F. (1990) Anecdotal, historical and critical commentaries on genetics. Mapping functions. *Genetics* 125 : 669-671.

D

Danin-Poleg Y., Reis N., Baudracco-Arnas S., Pitrat M., Staub J.E., Oliver M., Arus P., deVicente C.M., Katzir N. (2000) Simple sequence repeats in *Cucumis* mapping and map merging. *Genome* 43 : 963-974.

Darvasi A., Weinreb A., Minke V., Weller J.I., Soller M. (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* 134 : 943-951.

Darvasi A., Soller M. (1994) Optimum spacing of genetic markers for determining linkage between marker loci and quantitative trait loci. *Theor. Appl. Genet.* (1994) 89 : 351-357.

Darwin C., (1859) *On the origin of species*. John Murray, London.

Debener T., Mattiesch L. (1999) Construction of a genetic linkage map for roses using RAPD and AFLP markers. *Theor. Appl. Genet.* 99 : 891-899.

Delseny M. (2004) Re-evaluating the relevance of ancestral shared synteny as a tool for crop improvement. *Curr. Opin. Plant Biol.* 7 : 126-131.

Dempster A.P., Laird N.M., Rubin D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc.* 39 : 1-38.

Derory J., Léger P., Garcia V., Hauser M.-T., Plomion C., Glössl J., Kremer A. (2004) Transcriptome analysis of bud burst in sessile oak (*Quercus petraea* Liebl.). *Plant Mol. Biol., soumis*.

Devey M.E., Sewell M.M., Uren T.L., Neale D.B. (1999) Comparative mapping in loblolly and radiata pine using RFLP and microsatellite markers. *Theor. Appl. Genet.* 99 : 656-662.

De Vienne D., Causse M. (1997) La cartographie et la caractérisation des locus contrôlant la variation des caractères quantitatifs. In de Vienne D. *Les marqueurs moléculaires en génétique et biotechnologie végétales*. INRA éditions. pp. 89-118.

Devitt M.L., Stafstrom J.P. (1995) Cell cycle regulation during growth-dormancy cycles in pea axillary buds. *Plant Mol. Biol.* 29 : 255-265.

Devos K.M., Atkinson M.D., Chinoy C.N., Liu C.J., Gale M.D. (1993) RFLP-based genetic map of the homoeologous group 2 chromosomes of wheat, rye, and barley. *Theor. Appl. Genet.* 85 : 784-792.

Devos K.M., Beales J., Nagamura Y., Sasaki T. (1999) *Arabidopsis*-rice: will colinearity allow gene prediction across the eudicot-monocot divide? *Genome Res.* 9 : 825-829.

- Devos K.M., Gale M.D. (2000) Genome relationships : the grass model in current research. *Plant Cell* 12 : 637-646.
- Devos K.M., Pittaway T.S., Reynolds A., Gale M.D. (2000) Comparative mapping reveals a complex relationship between the pearl millet genome and those of foxtail millet and rice. *Theor. Appl. Genet.* 100 : 190-198.
- Devos K.M., Brown J.K.M., Bennetzen J.L. (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res.* 12 : 1075-1079.
- Di Gaspero G., Peterlunger E., Testolin R., Edwards K.J., Cipriani G. (2000) Conservation of microsatellite loci within the genus *Vitis*. *Theor. Appl. Genet.* 101 : 301-308.
- Dirlewanger E., Cosson P., Tavaud M., Aranzana M.J., Poizat C., Zanetto A., Arus P., Laigret F. (2002) Development of microsatellite markers in peach (*Prunus persica* L Batsch) and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). *Theor. Appl. Genet.* 105 : 127-138.
- Dirlewanger E., Graziano E., Joobeur T., Garriga-Calderé F., Cosson P., Howad W., Arús P. (2004) Comparative mapping and marker-assisted selection in Rosaceae fruit crops. *Proc. Natl. Acad. Sci. USA* www.pnas.org/cgi/doi/10.1073/pnas.0307937101
- Distelfeld A., Uauy C., Olmos S., Schlatter A.R., Dubcovsky J., Fahima T. (2004) Microcolinearity between a 2-cM region encompassing the grain protein content locus Gpc-6B1 on wheat chromosome 6B and a 350-kb region on rice chromosome 2. *Funct. Integr. Genomics* 4 : 59-66.
- Doerge R.W. (2002) Mapping and analysis of quantitative trait loci in experimental populations. *Nature Rev. Genet.* 3 : 43-52.
- Doganlar S., Frary A., Daunay M.C., Lester R.N., Tanksley S.D. (2002a) A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the solanaceae. *Genetics* 161 : 1697-1711.
- Doganlar S., Frary A., Daunay M.C., Lester R.N., Tanksley S.D. (2002b) Conservation of gene function in the solanaceae as revealed by comparative mapping of domestication traits in eggplant. *Genetics* 161 : 1713-1726.
- Dominguez I., Graziano E., Gebhardt C., Barakat A., Berry S., Arús P., Delseny M., Barnes S. (2003) Plant genome archaeology: evidence for conserved ancestral chromosome segments in dicotyledonous plant species. *Plant Biotech. J.* 1 : 91-99.
- Don R. H., Cox P.T., Wainwright B. J., Baker K., Mattick J. S. (1991) Touchdown PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* 19 : 4008.
- Doust A.N., Devos K.M., Gadberry M.D., Gale M.D., Kellogg E.A. (2004) Genetic control of branching in foxtail millet. *Proc. Natl. Acad. Sci. USA* 15 : 9045-9050.

Dow B.D., Ashley M.V., Howe H.F. (1995) Characterization of highly variable (GA/CT)_n microsatellites in the bur oak, *Quercus macrocarpa*. *Theor. Appl. Genet.* 91 : 137-141.

Dow B.D., Ashley M.V. (1996) Microsatellite analysis of seed dispersal and saplings parentage in bur oak, *Quercus macrocarpa*. *Mol. Ecol.* 5 : 615-627.

Doyle J.J., Doyle J.L. (1987) Isolation of plant material from fresh tissue. *Focus* 12: 13-15.

Dubcovsky J., Ramakrishna W., SanMiguel P.J., Busso C.S., Yan L., Shiloff B.A., Bennetzen J.L. (2001) Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. *Plant Physiol.* 125 :1342-1353.

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

E

Echt C.S., Vendramin G.G., Nelson C.D., Marquardt P. (1999) Microsatellite DNA as shared genetic markers among conifer species. *Can. J. For. Res.* 29 : 365-371.

Edwards M.D., Stuber C.W., Wendel J.F. (1987) Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. *Genetics* 116 : 113-125.

Ehrlich J., Sankoff D., Nadeau J.H. (1997) Synteny conservation and chromosome rearrangements during mammalian evolution. *Genetics* 147 : 289-296.

Eichler E.E. et Sankoff D. (2003) Structural dynamics of eukaryotic chromosome evolution. *Science* 301 : 793-797.

Elsik C.G., Williams C.G. (2000) Retroelements contribute to the excess low-copy number DNA in pine. *Mol. Gen. Genet.* 264 : 47-55.

Elsik C.G., Williams C.G. (2001) Families of clustered microsatellites in a conifer genome. *Mol. Genet. Genomics* 265 : 535-542.

Emebiri L.C., Devey M.E., Matheson A.C., Slee M.U. (1998) Age-related changes in the expression of QTLs for growth in radiata pine seedlings. *Theor. Appl. Genet.* 97 : 1053-1061.

F

Fang D.Q., Roose M.L., Krueger R.R., Federici C.T. (1997) Fingerprinting trifoliolate orange germ plasm accessions with isozyme, RFLPs, and inter-simple sequence repeat markers. *Theor. Appl. Genet.* 95 : 211-219.

Farquhar G.D., Ehleringer J.R., Hubick K.T. (1989) Carbon isotope discrimination and photosynthesis. *Ann. Rev. Plant Phys. Mol. Biol.* 40 : 503-537.

Feuillet C., Keller B. (2002) Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution. *Ann. Botany* 89 : 3-10.

Fineschi S., Turchini D., Villani F., Vendramin G.G. (2000) Chloroplast DNA polymorphism reveals little geographical structure in *Castanea sativa* Mill. (Fagaceae) throughout southern European countries. *Mol. Ecol.* 9 : 1495-1503.

Fisher P.J., Richardson T.E., Gardner R.C. (1998) Characteristics of single- and multi-copy microsatellites from *Pinus radiata*. *Theor. Appl. Genet.* 96 : 969-979.

Fisher R.A. (1930) *The genetical theory of natural selection*. Oxford Univ Press, Oxford.

Frary A., Nesbitt T.C., Grandillo S., Knaap E., Cong B., Liu J., Meller J., Elber R., Alpert K.B., Tanksley S.D. (2000) fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science* 289 : 85-88.

Frary A., Doganlar S., Daunay M.C., Tanksley S.D. (2003) QTL analysis of morphological traits in eggplant and implications for conservation of gene function during evolution of solanaceous species. *Theor. Appl. Genet.* 107 : 359-370.

Frary A., Fulton T.M., Zamir D., Tanksley S.D. (2004) Advanced backcross QTL analysis of a *Lycopersicon esculentum* x *L. pennellii* cross and identification of possible orthologs in the Solanaceae. *Theor. Appl. Genet.* 108 : 485-496.

Freeling M. (2001) Grasses as a single genetic system: reassessment 2001. *Plant Physiol.* 125 : 1191-1197.

Frewen B.E., Chen T.H., Howe G.T., Davis J., Rohde A., Boerjan W., Bradshaw H.D. Jr. (2000) Quantitative trait loci and candidate gene mapping of bud set and bud flush in *Populus*. *Genetics* 154 : 837-845.

Fulton T.M., Van der Hoeven R., Eannetta N.T., Tanksley S.D. (2002) Identification, analysis, and utilization of conserved ortholog set markers for comparative mapping genomics in higher plants. *Plant Cell* 14 : 1457-1467.

G

Gale M.D., Devos K.M. (1998) Comparative genetics in the grasses. *Proc. Natl. Acad. Sci. USA* 95 : 1971-1974.

Gaut B.S., Le Thierry d'Ennequin M., Peek A.S., Sawkins M.C. (2000) Maize as a model for the evolution of plant nuclear genomes. *Proc. Natl. Acad. Sci. USA* 97 : 7008-7015.

Gebhardt C., Walkemeier B., Henselewski H., Barakat A., Delseny M., Stüber K. (2003) Comparative mapping between potato (*Solanum tuberosum*) and *Arabidopsis thaliana* reveals structurally conserved domains and ancient duplications in the potato genome. *Plant J.* 34 : 529-541.

Gerber S., Mariette S., Streiff R., Bodénès C., Kremer A. (2000) Comparison of microsatellites and AFLP markers for parentage analysis. *Mol. Ecol.* 9 : 1037-1048.

Glazier A.M., Nadeau J.H., Aitman T.J. (2002) Finding genes that underlie complex traits. *Science* 298 : 2345-2349.

Goff S.A. (1999) Rice as a model for cereals genomics. *Curr. Opin. Plant Biol.* 2 : 86-89.

Goff S.A., Ricke D., Lan T.-H., Presting G., Wang R., Dunn M., Glazebrook J., Sessions A., Oeller P., Varma H., Hadley D., Hutchison D., Martin C., Katagiri F., Lange B.M., Moughamer T., Xia Y., Budworth P., Zhong J., Miguel T., Paszkowski U., Zhang S., Colbert M., Sun W.-L., Chen L., Cooper B., Park S., Wood T.C., Mao L., Quail P., Wing R., Dean R., Yu Y., Zharkikh A., Shen R., Sahasrabudhe S., Thomas A., Cannings R., Gutin A., Pruss D., Reid J., Tavtigian S., Mitchell J., Eldredge G., Scholl T., Miller R.M., Bhatnagar S., Adey N., Rubano T., Tusneem N., Robinson R., Feldhaus J., Macalma T., Oliphant A., Briggs S. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296 : 92–100.

Goffinet B., Gerber S. (2000) Quantitative trait loci: a meta-analysis. *Genetics* 155 : 463-473.

Grant D., Cregan P., Shoemaker R.C. (2000) Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 97 : 4168-4173.

Grattapaglia D., Sederoff R. (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* 137 : 1121-1137.

Grattapaglia D., Bertolucci F.L., 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. *Theor. Appl. Genet.* 90 : 933-947.

Groover A., Devey M., Fiddler T., Lee J., Megraw R., Mitchel-Olds T., Sherman B., Vujcic S., Williams C., Neale D. (1994) Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. *Genetics* 138 : 1293-1300.

Grube R.C., Radwanski E.R., Jahn M. (2000) Comparative genetics of disease resistance within the Solanaceae. *Genetics* 155 : 873-887.

Gualtieri G., Kulikova O., Limpens E., Kim D.J., Cook D.R., Bisselin T., Geurts R. (2002) Microsynteny between pea and *Medicago truncatula* in the SYM2 region. *Plant Mol. Biol.* 5 : 225-235.

Guimarães C.T., Sills G.R., Sobral B.W.S. (1997) Comparative mapping of Andropogoneae: *Saccharum* L. (sugarcane) and its relation to sorghum and maize. *Proc. Natl. Acad. Sci. USA* 94 : 14261-14266.

Guo H., Moose S.P. (2003) Conserved noncoding sequences among cultivated cereal genomes identify candidate regulatory sequence elements and patterns of promoter evolution. *Plant Cell* 15 : 1143-1158.

Guo W.-J., Bundithya W., Goldsbrough P.B. (2003) Characterization of the *Arabidopsis* metallothionein gene family: tissue-specific expression and induction during senescence and in response to copper. *New Phyt.* 159 : 369-381.

Guyomarc'h H., Sourdille P., Charmet G., Edwards K.J., Bernard M. (2002) Characterisation of polymorphic microsatellite markers from *Aegilops tauschii* and transferability to the D-genome of bread wheat. *Theor. Appl. Genet.* (2002) 104 : 1164-1172.

H

Haldane J.B.S. (1919) The combination of linkage values, and the calculation of distance between the loci of linked factors. *J. Genet.* 8 : 299-309.

Heslop-Harrison J.S. (2000) Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. *Plant Cell* 12 : 617-635.

Hodgetts R.B., Aleksiuik M.A., Brown A., Clarke C., Macdonald E., Nadeem S., Khasa D. (2001) Development of microsatellite markers for white spruce (*Picea glauca*) and related species. *Theor. Appl. Genet.* 102 : 1252-1258.

Horvath D.P., Anderson J.V., Chao W.S., Foley M.E. (2003) Knowing when to grow: signals regulating bud dormancy. *Trends Plant Sci.* 8 : 534-540.

Howe G.T., Saruul P., Davis J., Chen T.H.H. (2000) Quantitative genetics of bud phenology, frost damage, and winter survival in an F₂ family of hybrid poplars. *Theor. Appl. Genet.* 101 : 632-642.

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

Hsu T.M., Chen X., Duan S., Miller R.D., Kwok, P.-Y. (2001) Universal SNP genotyping assay with fluorescence polarization detection. *Biotechniques* 31 : 560-570.

Hulbert S.H., Richter T.E., Axtell J.D., Bennetzen J.L. (1990) Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. *Proc. Natl. Acad. Sci. USA* 87 : 4251-4255.

Hu F.Y., Tao D.Y., Sacks E., Fu B.Y., Xu P., Li J., Yang Y., McNally K., Khush G.S., Paterson A.H., Li Z.K. (2003) Convergent evolution of perenniality in rice and sorghum. *Proc. Natl. Acad. Sci. USA* 100 : 4050-4054.

Huntley B., Birks H.J.B. (1983) An atlas of past and present pollen maps for Europe: 0-13000 years ago. Cambridge University Press, Cambridge.

Hurme P., Sillanpää M.J., Arjas E., Repo T., Savolainen O. (2000) Genetic basis of climatic adaptation in Scot pine by Bayesian quantitative trait locus analysis. *Genetics* 156 : 1309-1322.

I

Ilic K., SanMiguel P.J., Bennetzen J.L. (2003) A complex history of rearrangement in an orthologous region of the maize, sorghum and rice genomes. *Proc. Natl. Acad. Sci. USA* 100 : 12265-12270.

IPCC, Climate change 2001. Third assessment report of the Intergovernmental Panel on Climate Change ICCP (WG I & II) Cambridge University Press, Cambridge, 2001.

J

Jackson S.A., Cheng Z., Wang M.L., Goodman H.M., Jiang J. (2000) Comparative fluorescence in situ hybridization mapping of a 431-kb *Arabidopsis thaliana* bacterial artificial chromosome contig reveals the role of chromosomal duplications in the expansion of the *Brassica rapa* genome. *Genetics* 156 : 833-838.

Jansen R.C. (1993) Interval mapping of multiple quantitative trait loci. *Genetics* 135 : 205-211.

Jauffres A. (2003) Cartographie comparée de Fagacées à l'aide de microsatellites. Maîtrise de biologie des populations et des écosystèmes, UFR sciences et techniques Côte Basque, Anglet.

Jermstad K.D., Bassoni D.L., Jech K.S., Wheeler N.C., Neale D.B. (2001a) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. I. Timing of vegetative bud flush. *Theor. Appl. Genet.* 102 : 1142-1151.

Jermstad K.D., Bassoni D.L., Wheeler N.C., Anekonda T.S., Aitken S.N., Adams W.T., Neale D.B. (2001b) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. II. Spring and fall cold-hardiness. *Theor. Appl. Genet.* 102 : 1152-1158.

Jermstad K.D., Bassoni D.L., Jech K.S., Ritchie G.A., Wheeler N.C., 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 : 1489-1506.

Jiang C., Wright R.J., El-Zik K.M., Paterson A.H. (1998) Polyploid formation created unique avenues for response to selection in *Gossypium*. *Proc. Natl. Acad. Sci. USA* 95 : 4419-4424.

K

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

Kao C.H., Zeng Z.B., Teasdale R.D. (1999) Multiple interval mapping for quantitative trait loci. *Genetics* 152 : 1203-1216.

Kaplinsky N. J., Braun D.M., Penterman J., Goff S.A., Freeling M. (2002) Utility and distribution of conserved noncoding sequences in the grasses. *Proc. Natl. Acad. Sci. USA* 99 : 6147-6151.

Karhu A., Dieterich J.H., Savolainen O. (2000) Rapid expansion of microsatellite sequences in Pines. *Mol. Biol. Evol.* 17 : 259-265.

Kathryn A., VandenBosch A, Stacey G. (2003) Summaries of legumes genomics projects from around the globe. Community resources for crops and models. *Plant Physiol.* 131 : 840-865.

Kaya Z., Sewell M.M., Neale D.B. (1999) Identification of quantitative trait loci influencing annual height- and diameter-increment growth in loblolly pine (*Pinus taeda* L.). *Theor. Appl. Genet.* 98 : 586-592.

- Kearsey M.J., Farquhar A.G. (1998) QTL analysis in plants: where are we now? *Heredity* 80 :137-142.
- Keller B., Feuillet C. (2000) Colinearity and gene density in grass genomes. *Trends Genet.* 5 : 246-251.
- Kellogg E.A. (1998) Relationships of cereals crops and other grasses. *Proc. Natl. Acad. Sci. USA* 95 : 2005-2010.
- Kesseli R.V., Paran I., Michelmore R.W. (1994) Analysis of a detailed genetic linkage map of *Lactuca sativa* (lettuce) constructed from RFLP and RAPD markers. *Genetics* 136 : 1435-1446.
- Kesseli R., Witsenboer H., Babaloukhan H., Chase M., Hill M., Zhang Y.-H., Michelmore R. (1995) The abundance of simple sequence repeats and prospects for comparative mapping with SSRs, RFLPs and SCARs in the Compositae. Plant and Animal Genome III Conference, San Diego, CA, January 1995.
- Kesseli R.V., Michelmore R.W. (1996) The Compositae: systematically fascinating but specifically neglected. In *Genome mapping in Plants*, A.H. Paterson ed. R.G. Landes Co. Georgetown, TX. pp 179-191.
- Kianian S.F., Quiros C.F. (1992) Generation of a Brassica oleracea composite RFLP map: linkage arrangements among various populations and evolutionary implications. *Theor. Appl. Genet.* 84 : 544-554.
- Kijas J.M.H., Fowler J.C.S., Thomas M.R. (1995) An evaluation of sequence tagged microsatellite site markers for genetic analysis within *Citrus* and related species. *Genome* 38 : 349-355.
- Kimura M. (1983) *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge.
- Kinlaw C.G., Neale D.B. (1997) Complex gene families in pine genomes. *Trends Plant. Sci.* 2 : 356-359.
- Kirschbaum M.U.F. (2000) Forest growth and species distribution in a changing climate. *Tree Physiol.* 20 : 309-322.
- Komulainen P., Brown G.R., Mikkonen M., Karhu A., García-Gil M.R., O'Malley D., Lee B., Neale D.B., Savolainen O. (2003) Comparing EST-based genetic maps between *Pinus sylvestris* and *Pinus taeda*. *Theor. Appl. Genet.* 107 : 667-678.
- Kondrashov A.S. (1999) Comparative genomics and evolutionary biology. *Curr. Opin. Genet. Dev.* 9 : 624-629.
- Korol A.B., Ronin Y.I., Nevo E., Hayes P.M. (1998a) Multi-interval mapping of correlated trait complexes. *Genetics* 80 : 273-284.

- Korol A.B., Ronin Y.I., Nevo E. (1998b) Approximate analysis of QTL-environment interactions with no limits on the number of environments. *Genetics* 148 : 2015-2028.
- Korol A.B., Ronin Y.I., Itskovich A.M., Peng J., Nevo E. (2001) Enhanced efficiency of quantitative trait loci mapping analysis based on multivariate complexes of quantitative traits. *Genetics* 157 : 1789-1803.
- Kosambi D.D. (1944) The estimation of map distance from recombination values. *Ann. Eugen.* 12 : 172-175.
- Koski V., Skroppa T., Paule L., Wolf H., Turok J. (1997) Technical guidelines for genetic conservation of Norway spruce (*Picea abies* (L.) Karst). European Forest Genetic Resources Programme (EUFORGEN). International Plant Genetic Resources Institute (IPGRI).
- Kowalski S.P., Lan T.H., Feldmann K.A., Paterson A.H. (1994) Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization. *Genetics* 138 : 499-510.
- Krebs P., Conedera M., Pradella M., Torriani D., Felber M., Tinner W. (2004) Quaternary refugia of the sweet chestnut (*Castanea sativa* Mill.): an extended palynological approach. *Veget. Hist. Archaeobot.* 13 : 145-160.
- Kremer A. (1992) Prediction of age-age correlations of total height based on serial correlations between height increments in maritime pine (*Pinus pinaster* Ait.). *Theor. Appl. Genet.* 85 : 152-158.
- Kresovich S., Szewc-McFadden A.K., Bliet S.M., McFerson J.R. (1995) Abundance and characterization of simple-sequence repeats (SSRs) isolated from a size-fractionated genomic library of *Brassica napus* L. (rapeseed) *Theor. Appl. Genet.* 91 : 206-211.
- Kriebel H.B. (1985) DNA sequence components of *Pinus strobus* nuclear genome. *Can. J. For. Res.* 15 : 1-4.
- Kruglyak S., Durrett R.T., Schug M.D., Aquadro C.F. (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proc. Natl. Acad. Sci. USA* 95 : 10774-10778.
- Ku H.-M., Vision T., Liu J., Tanksley S.D. (2000) Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny. *Proc. Natl. Acad. Sci. USA* 97 : 9121-9126.
- Kubisiak T.L., Hebard F.V., Nelson C.D., Zhang J., Bernatzky R., Huang H., Anagnostakis S.L., Doudrick R.L. (1997) Molecular mapping of resistance to blight in an interspecific cross in the genus *Castanea*. *Phytopathology* 87 : 751-759.
- Kurata N., Moore G., Nagamura Y., Foote T., Yano M., Minobe Y., Gale M. (1994) Conservation of genome structure between rice and wheat. *Bio/Technology* 12 : 276-278.
- Kutil B.L., Williams C.G. (2001) Triplet-repeat microsatellite shared among hard and soft pines. *J. Hered.* 92 : 327-332.

Kwon J.M., Goate A.M. (2000) The candidate gene approach. *Alcohol Res. Health* 24 : 164-168.

L

Lagercrantz U., Lydiate D.J. (1996) Comparative genome mapping in Brassica. *Genetics* 144 : 1903-1910.

Lagercrantz U. (1998) Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that Brassica genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics* 150 : 1217-1228.

Lagercrantz U., Axelsson T. (2000) Rapid evolution of the family of CONSTANS LIKE genes in plants. *Mol. Biol. Evol.* 17 : 1499-1507.

Lan T.H., DelMonte T.A., Reischmann K.P., Hyman J., Kowalski S.P., McFerson J., Kresovich S., Paterson A.H. (2000) An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. *Genome Res.* 10 : 776-788.

Lan T.H., Paterson A.H. (2001) Comparative mapping of QTLs determining the plant size of *Brassica oleracea*. *Theor. Appl. Genet.* 103 : 383-397.

Lander E.S., Botstein D. (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121 : 185-199.

Lander E.S., Green P., Abrahamson J., Barlow A., Daly M.J., Lincoln S.E., Newburg L. (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1 : 174-181.

Lander E., Kruglyak L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet.* 11 : 241-247.

Landry B.S., Kesseli R.V., Farrara B., Michelmore R.W. (1987) A genetic map of lettuce (*Lactuca sativa*) with restriction fragment length polymorphism, isozyme, disease resistance, and morphological markers. *Genetics* 116 : 331-337.

Lang G.A. (1987) Dormancy: a new universal terminology. *Hort. Science* 22 : 817-820.

Lange K., Boehnke M. (1982) How many polymorphic marker genes will it take to span the human genome? *Am. J. Hum. Genet.* 34 : 842-845.

La Rota M., Sorrells M.E. (2004) Comparative DNA sequence analysis of mapped wheat ESTs reveals the complexity of genome relationships between rice and wheat. *Funct. Integr. Genomics* 4 : 34-46.

Laurie D.A., Devos K.M. (2002) Trends in comparative genetics and their potential impacts on wheat and barley research. *Plant. Mol. Biol.* 48 : 729-740.

- Lauteri M., Scartazza A., Guido M.C., Brugnoli E. (1997) Genetic variation in photosynthetic capacity, carbon isotope discrimination and mesophyll conductance in provenances of *Castanea sativa* adapted to different environments. *Funct. Ecol.* 11: 675-683.
- Lauteri M., Monteverdi M.C., Sansotta A., Cherubini M., Spaccino L., Villani F. (1999) Adaptation to drought in European chestnut. Evidences from a hybrid zone and from controlled crosses between drought and wet adapted populations. *Acta Hort.* 494 : 345-353.
- Lauteri M., Pliura A., Monteverdi M.C., Brugnoli E., Villani F., Eriksson G. (2004) Genetic variation in carbon isotope discrimination in six European populations of *Castanea sativa* Mill. originating from contrasting localities. *J. Evol. Biol.* doi: 10.1111/j.1420-9101.2004.00765.x.
- Lawton-Rauh A., Robichaux R.H., Purugganan M.D. (2003) Patterns of nucleotide variation in homeologous regulatory genes in the allotetraploid Hawaiian silversword alliance (Asteraceae). *Mol. Ecol.* 12 : 1301-1313.
- Lee J.-M., Grant D., Vallejos C.E., Shoemaker R.C. (2001) Genome organization in dicots. II Arabidopsis as a “bridging species” to resolve genome evolution events among legumes. *Theor. Appl. Genet.* 103 : 765-773.
- Lerceteau E., Plomion C., Andersson B. (2000) AFLP mapping and detection of quantitative trait loci (QTLs) for economically important traits in *Pinus sylvestris*: a preliminary study. *Mol. Breed.* 6 : 451-458.
- Lexer C. (1999) Application of microsatellite markers to genetic map construction and genetic analysis of forest seed material in oaks (*Quercus* spp). PhD Thesis. Universität für Bodenkultur. Vienna. Austria. pp 73.
- Li G., Gao M., Yang B., Quiros C.F. (2003) Gene for gene alignment between the Brassica and Arabidopsis genomes by direct transcriptome mapping. *Theor. Appl. Genet.* 107 : 168-180.
- Li Y.-C., Korol A.B., Fahima T., Beiles A., Nevo E. (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol. Ecol.* 11 : 2453-2465.
- Li Y.-C., Korol A.B., Fahima T., Nevo E. (2004) Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.* 21 : 991-1007.
- Lian C., Zhou Z., Hogetsu T. (2001) A simple method for developing microsatellite makers using amplified fragments of Inter-Simple Sequence Repeat (ISSR). *J. Plant Res.* 114, 381-385.
- Lin Y.R., Schertz K.F., 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 : 391-411.
- Lincoln S.E., Lander E.S. (1992) Systematic detection of errors in genetic linkage data. *Genomics* 14 : 604-610.

Liu B, Brubaker C.L., Mergeai G., Cronn R.C., Wendel J.F. (2001) Polyploid formation in cotton is not accompanied by rapid genomic changes. *Genome* 44 : 321-330.

Liu B.-H. (1998) *Statistical Genomics. Linkage, mapping, and QTL analysis.* CRC Press, Boca Raton, New York.

Liu H., Sachidanandam R., Stein L. (2001) Comparative genomics between rice and *Arabidopsis* shows scant collinearity in gene order. *Genome Res.* 11 : 2020-2026.

Liu Y.-S., Gur A., Ronen G., Causse M., Damidaux, Buret M., Hirschberg J., Zamir D. (2003) There is more to tomato fruit colour than candidate carotenoid genes. *Plant Biotech. J.* 1 : 195-207.

Livingstone K.D., Lackney V.K., Blauth J.R., van Wijk R., Jahn M.K. (1999) Genome mapping in capsicum and the evolution of genome structure in the Solanaceae. *Genetics* 152 : 1183-1202.

Lukens L., Zou F., Lydiate D., Parkin I., Osborn T. (2003) Comparison of a *Brassica oleracea* genetic map with the genome of *Arabidopsis thaliana*. *Genetics* 164 : 359-372.

Lynch M., Conery J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290 : 1151-1155.

M

Ma J., Devos K.M., Bennetzen J.L. (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res.* 14 : 860-869.

Machon N., Burel L., Lefranc M., Frascaria-Lacoste N. (1996) Evidence of genetic drift in chestnut populations. *Can. J. For. Res.* 26 : 905-908.

Manos P.S., Steele K.P. (1997) Phylogenetic analyses of “higher” Hamamelidiae based on plastid sequence data. *Am. J. Bot.* 84 : 1407-1419.

Manos P.S., Stanford A.M. (2001) The historical biogeography of Fagaceae: tracking the tertiary history of temperate and subtropical forests of the northern hemisphere. *Int. J. Plant Sci.* 162 : s77-s93.

Manos P.S., Zhou Z.-K., Cannon C.H. (2001) Systematics of Fagaceae: phylogenetic tests of reproductive trait evolution. *Int. J. Plant Sci.* 162 : 1361-1379.

Marinoni D., Akkak A., Bounous G., Edwards K.J., Botta R. (2003) Development and characterization of microsatellite markers in *Castanea sativa* (Mill). *Mol. Breed.* 11 : 127-136.

Marques C.M., Vasquez-Kool J., Carocha V.J., Ferreirs J.G., O'Malley D.M., Liu B.H., Sderoff R. (1999) Genetic dissection of vegetative propagation traits in *Eucalyptus tereticornis* and *E. globulus*. 99 : 936-946.

- Marques C.M., Brondani R.P.V., Grattapaglia D., Sederoff R. (2002) Conservation and synteny of SSR loci and QTLs for vegetative propagation in four *Eucalyptus* species. *Theor. Appl. Genet.* 105 : 474-478.
- Masterson J. (1994) Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* 264 : 421-423.
- Mathews S., Sharrock R.A. (1996) The phytochrome gene family in grasses (Poaceae): a phylogeny and evidence that grasses have a subset of the loci found in dicot angiosperms. *Mol. Biol. Evol.* 13 : 1141-1150.
- Matsuoka Y., Mitchell S.E., Kresovich S., Goodman M., Doebley J. (2002) Microsatellites in *Zea*-variability, patterns of mutations, and use for evolutionary studies. *Theor. Appl. Genet.* 104 : 436-450.
- Mauricio R. (2001) Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. *Nature Rev. Genet.* 2 : 370-381.
- Mayer K., Murphy G., Tarchini R., Wambutt R., Volckaert G., Pohl T, Dusterhoft A., Stiekema W., Entian K.D., Terry N., Lemcke K., Haase D., Hall C.R., van Dodeweerd A.M., Tingey S.V., Mewes H.W., Bevan M.W., Bancroft I. (2001) Conservation of microstructure between a sequenced region of the genome of rice and multiple segments of the genome of *Arabidopsis thaliana*. *Genome Res.* 11 : 1167-1174.
- McClintock B. (1984) The significance of responses of the genome to challenge. *Science* 226 : 792-801.
- Meyers B.C., Chin D.B., Shen K.A., Sivaramakrishnan S., Lavelle D.O., Zhang Z., Michelmore R.W. (1998) The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. *Plant Cell* 10 : 1817-1832.
- Michelmore R.W., Knapp S.J., Rieseberg L., Kesseli R.V., Bradford K., Jackson L., Sudarshana P. (2001) CompositDB and comparative genomics of domesticated traits in lettuce and sunflower. *Plant and Animal Genome IX Conference San Diego, CA, January 2001.*
- Ming R., Liu S.C., Lin Y.R., da Silva J., Wilson W., Braga D., van Deynze A., Wenslaff T.F., Wu K.K., Moore P.H., Burnquist W., Sorrells M.E., Irvine J.E., Paterson A.H. (1998) Detailed alignment of saccharum and sorghum chromosomes: comparative organization of closely related diploid and polyploid genomes. *Genetics* 150 : 1663-1682.
- Ming R., Del Monte T.A., Hernandez E., Moore P.H., Irvine J.E., Paterson A.H. (2002) Comparative analysis of QTLs affecting plant height and flowering among closely-related diploid and polyploid genomes. *Genome* 45 : 794-803.
- Mitchell-Olds T. et Clauss M.J. (2002) Plant evolutionary genomics. *Curr. Opin. Plant Biol.* 5 : 74-79.
- Monteverdi M.C. (2004) Meccanismi adattativi intra-specifici a condizioni contrastanti per disponibilità idrica in *Castanea sativa* Mill. Università della Tuscia. Viterbo.

Moore G., Roberts M., Aragon-Alcaide L., Foote T. (1997) Centromeric sites and cereal chromosome evolution. *Chromosoma* 105 : 321-323.

Morgan T.H. (1932) *The scientific basis of evolution*. Norton, New York.

Morgante M., Olivieri A.M. (1993) PCR amplified microsatellites as markers in plant genetics. *Plant J.* 3 : 175-182.

Morgante M., Hanafey M., Powell W. (2002) Microsatellites are preferentially associated with non-repetitive DNA in plant genomes. *Nature Genet.* 30 : 194-200.

Morton N.E. (1955) Sequential tests for the detection of linkage. *Am. J. Hum. Genet.* 7 : 277-318.

Murphy W.J., Stanyon R., O'Brien S.J. (2001) Evolution of mammalian genome organization inferred from comparative gene mapping. *Genome Biol.* 2 : 0005.1-0005.8.

Myers R.M., Maniatis T., Lerman L.S. (1987) Detection and localization of single base changes by denaturing gradient gel-electrophoresis. *Meth. Enzymol.* 155 : 501-527.

N

Nadeau J.H. et Sankoff D. (1998) Counting on comparative maps. *Trends Genet.* 14 : 495-501.

Nagaoka T., Ogihara Y. (1997) Applicability of inter-simple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP et RAPD markers. *Theor. Appl. Genet.* 94 : 597-602.

Nataraj A.J., Olivos-Glander I., Kusukawa N., Highsmith W.E.Jr (1999) Single-strand conformation polymorphism and heteroduplex analysis for gel-based mutation detection. *Electrophoresis* 20 : 1177-1185.

Nixon K.C. (1989) Origins of Fagaceae. P.R. Crane, S. Blackmore, eds. *Evolution systematics and fossils history of the Hamamelidae*. Vol. 2 "Higher" Hamamelidae. Pp.23-44. Clarendon, Oxford.

O

Oh K., Hardeman K., Ivanchenko M.G., Ellard-Ivey M., Nebenführ A., White T.J., Lomax T.L. (2002) Fine mapping in tomato using microsynteny with the *Arabidopsis* genome: the *Diageotropica* (*Dgt*) locus. *Genome Biol.* 3 : 0049.1-0049.11.

O'Neill C.M., Bancroft I. (2000) Comparative physical mapping of segments of the genome of *Brassica oleracea* var. *alboglabra* that are homoeologous to sequenced regions of chromosomes 4 and 5 of *Arabidopsis thaliana*. *Plant J.* 23 : 233-243.

Orita M., Iwahana H., Kanazawa H., Hayashi K., Sekiya T. (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. USA* 86 : 2766-2770.

Orr A.H., Coyne J.A. (1992) The genetics of adaptation: a reassessment. *Amer. Nat.* 140 : 725-742.

Orr A.H. (1998) The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52, 935-949.

Orr A.H. (2000) Adaptation and the cost of complexity. *Evolution* 54 : 13-20.

Orti G., Pearse D.E., Avise J.C. (1997) Phylogenetic assessment of length variation at a microsatellite locus *Proc. Natl. Acad. Sci. USA* 94 : 10745-10749.

Osborn T.C., Kole C., Parkin I.A., Sharpe A.G., Kuiper M., Lydiate D.J., Trick M. (1997) Comparison of flowering time genes in *Brassica rapa*, *B. napus* and *Arabidopsis thaliana*. *Genetics* 146 : 1123-1129.

P

Paran I., Zamir D. (2003) Quantitative traits in plants: beyond the QTL. *Trends Genet.* 19 : 303-306.

Parmesan C., Yohe G. (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature* 421 : 37-42.

Paterson A.H., Damon S., Hewitt J.D., Zamir D., Rabinowitch H.D., Lincoln S.E., Lander E.S., Tanksley S.D. (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. *Genetics* 127 : 181-197.

Paterson A.H., Lin Y.-R., Li Z., Schertz K.F., Doebley J.F., Pinson S.R.M., Liu S.-C., Stansel J.W., Irvine J.E. (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* 269 : 1714-1718.

Paterson A.H., Lan T.-H., Reischmann K.P., Chang C., Lin Y.-R., Liu S.-C., Burow M.D., Kowalski S.P., Katsar C.S., DelMonte T.A., Feldmann K.A., Schertz K.F., Wendel J.F. (1996) Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence. *Nature Genet.* 14 : 380-382.

Paterson A.H., Bowers J.E., Burow M.D., Draye X., Elsik C.G., Jiang C.-X., Katsar C.S., Lan T.-H., Lin Y.-R., Ming R., and Wright R.J. (2000) Comparative genomics of plant chromosomes. *Plant Cell* 12 : 1523-1539.

Paterson A.H., Bowers J.E., Peterson D.G., Estill J.C., Chapman B.A. (2003) Structure and evolution of cereals genomes. *Curr. Opin. Genet. Dev.* 13 : 644-650.

Paterson A.H., Bowers J.E., Chapman B.A. (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc. Natl. Acad. Sci. USA* 101 : 9903-9908.

Peakall R., Gilmore S., Keys W., Morgante M., Rafalski A. (1998) Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. *Mol. Biol. Evol.* 15 : 1275-1287.

- Peng J., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M., Flintham J.E., Beales J., Fish L.J., Worland A.J., Pelica F., Sudhakar D., Christou P., Snape J.W., Gale M.D., Harberd N.P. (1999) "Green revolution" genes encode mutant gibberellin response modulators. *Nature* 400 : 256-261.
- Pereira M.G., Lee M., Bramel-Cox P., Woodman W., Doebley J., Whitkus R. (1994) Construction of an RFLP map in sorghum and comparative mapping in maize. *Genome* 37:236-243.
- Perry D.J. et Furnier G.R. (1996) *Pinus banksiana* has at least seven alcohol dehydrogenase genes in two linked groups. *Proc. Natl. Acad. Sci. USA* 93 : 13020-13023.
- Pflieger S., Lefebvre V., Causse M. (2001) The candidate gene approach in plant genetics: a review. *Mol. Breed.* 7 : 275-291.
- Phillips P.C. (1999) From complex traits to complex alleles. *Trends Genet.* 15 : 6-8.
- Pigliucci M., Villani F., Benedettelli S. (1990) Geographic and climatic factors associated with the spatial structure of gene frequencies in *Castanea sativa* Mill. from Turkey. *J. Genet.* 69: 141-149.
- Pillen K., Pineda O., Lewis C.B., Tanksley S.D. (1996) Status of genome mapping tools in the taxon Solanaceae. In *Genome mapping in Plants*, A.H. Paterson ed. R.G. Landes Co. Georgetown, TX. pp 281-308.
- Plieske J., Struss D. (2001) Microsatellite markers for genome analysis in *Brassica*. I. Development in *Brassica napus* and abundance in *Brassicaceae* species. *Theor. Appl. Genet.* 102 : 689-694.
- Pliura A. Erikson G. (2002) Genetic variation in juvenile height and biomass of open-pollinated families of six *Castanea sativa* Mill. populations in a 2 x 2 factorial temperature x watering experiment. *Silvae Genet.* 51 : 152-160.
- Plocik A., Layden J., Kesseli R. (2004) Comparative analysis of NBS domain sequences of NBS-LRR disease resistance genes from sunflower, lettuce, and chicory. *Mol. Phyl. Evol.* 31 : 153-163.
- Plomion C., Durel C.E., O'Malley D.M. (1996) Genetic dissection of height in maritime pine seedlings raised under accelerated growth conditions. *Theor. Appl. Genet.* 93 : 849-858.
- Plomion C., Hurme P., Frigerio J.-M., Ridolphi M., Pot D., Pionneau C., Avila C., Gallardo F., David H., Neutlings G., Campbell M., Canovas F.M., Savolainen O., Bodénès C., Kremer A. (1999) Developing SSCP markers in two *Pinus* species. *Mol. Breed.* 5 : 21-31.
- Poncet V., Martel E., Allouis S., Devos K.M., Lamy F., Sarr A., Robert T. (2002) Comparative analysis of QTLs affecting domestication traits between two domesticated x wild pearl millet (*Pennisetum glaucum* L., Poaceae) crosses. *Theor. Appl. Genet.* 104 : 965-975.

Porth I., Koch M., Berenyi M., Burg A., Burg K. (2004a) Identification of adaptation specific differences in the mRNA expression of sessile and pedunculate oak based on osmotic stress induced genes. *Tree Physiol.*, soumis.

Porth I., Berenyi M., Scotti-Saintagne C., Barreneche T., Kremer A., Burg K. (2004b) Linkage mapping of osmotic stress induced genes of oak. *Theor. Appl. Genet.*, soumis.

Powell W., Machray G.C., Provan J. (1996) Polymorphism revealed by simple sequence repeats. *Trends Plant Sci.* 1 : 215-222.

Q

Quiros C.F., Grellet F., Sadowski J., Suzuki T., Li G., Wroblewski T. (2001) Arabidopsis and Brassica comparative genomics: sequence, structure and gene content in the ABI-Rps2-Ck1 chromosomal segment and related regions. *Genetics* 157 : 1321-1330.

R

Rafalski A. (2002) Applications of single nucleotide polymorphisms in crop genetics. *Curr. Opin. Plant Biol.* 5 : 94-100.

Ramakrishna W., Dubcovsky J., Park Y.J., Busso C., Emberton J., SanMiguel P., Bennetzen J.L. (2002) Different types and rates of genome evolution detected by comparative sequence analysis of orthologous segments from four cereal genomes. *Genetics* 162:1389-1400.

Rajora O.P., Rahman M.H., Dayanandan S., Mosseler A. (2001) Isolation, characterization, inheritance and linkage of microsatellite DNA markers in white spruce (*Picea glauca*) and their usefulness in other spruce species. *Mol. Gen. Genet.* 264 : 871-882.

Rebai A., Goffinet B., Mangin B. (1995) Comparing power of different methods for QTL detection. *Biometrics* 51 : 87-99.

Reinisch A.J., Dong J.M., Brubaker C.L., Stelly D.M., Wendel J.F., Paterson A.H. (1994) A detailed RFLP map of cotton, *Gossypium hirsutum* x *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. *Genetics* 138 : 829-847.

Riechmann J.L., Heard J., Martin G., Reuber L., Jiang C.-Z., Keddie J., Adam L., Pineda O., Ratcliffe O.J., Samaha R.R., Creelman R., Pilgrim M., Broun P., Zhang J.Z., Ghandehari D., Sherman B. K., Yu G.-L. (2000) *Arabidopsis* transcription factors : genome-wide comparative analysis among eukaryotes. *Science* 290 : 2105-2110.

Rieseberg L.H., Baird S.J.E., Gardner K.A. (2000) Hybridization, introgression, and linkage evolution. *Plant Mol. Biol.* 42 : 205-224.

Rieseberg L.H. (2001) Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* 16 : 351-358.

Rieseberg L.H., Raymond O., Rosenthal D.M., Lai Z., Livingstone K., Nakazato T., Durphy J.L., Schwarzbach A.E., Donovan L.A., Lexer C. (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301 : 1211-1216.

Röder M.S., Korzun V., Wendehake K., Planschke J., Tixier M., Leroy P., Galal M.W. (1998) A microsatellite map of wheat. *Genetics* 149 : 2007-2023.

Rohde A., Howe G.T., Olsen J.E., Moritz T., van Montagu M., Junttila O., Boerjan W. (2000) Molecular aspects of bud dormancy in trees. In *Molecular Biology of Woody Plants* (eds. Jain S.M. and Minocha S.C.) Vol. 1 : 89-134. Kluwer Academic Publishers, Netherlands.

Rohde A., Prinsen E., De Rycke R., Engler G., Van Montagu M., Boerjan W. (2002) PtABI3 Impinges on the growth and differentiation of embryonic leaves during bud set in poplar. *Plant Cell* 14 : 1885-1901.

Rong J., Abbey C., Bowers J.E., Brubaker C.L., Chang C., Chee P.W., Delmonte T.A., Ding X., Garza J.J., Marler B.S., Park C., Pierce G.J., Rainey K.M., Rastogi V.K., Schulze S.R., Trolinder N.L., Wendel J.F., Wilkins T.A., Williams-Coplin T.D., Wing R.A., Wright R.J., Zhao X., Zhu L., Paterson A.H. (2004) A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics* 166 : 389-417.

Ronin Y.I., Korol A.B., Nevo E. (1999) Single- and multiple-trait mapping analysis of linked quantitative trait loci: some asymptotic analytical approximations. *Genetics* 151 : 387-396.

Root T.L., Price J.T., Hall K.R., Schneider S.H., Rosenzweig C., Pounds J.A. (2003) Fingerprints of global warming on wild animals and plants. *Nature* 421 : 57-60.

Rosberg M., Theres K., Acarkan A., Herrero R., Schmitt T., Schumacher K., Schmitz G., Schmidt R. (2001) Comparative sequence analysis reveals extensive microcolinearity in the *Lateral Suppressor* regions of the tomato, *Arabidopsis*, and *Capsella* genomes. *Plant Cell* 13 : 979-988.

Rossetto M., McNally J., Henry R.J. (2002) Evaluating the potential of SSR flanking regions for examining taxonomic relationships in the Vitaceae. *Theor. Appl. Genet.* 104 : 61-66.

Rutter P.A., Payne J.A., Miller G. (1990) Chestnut. Dans J.N. Moore et J.R. Ballington, Jr. (Eds.), *Genetic resource for temperate fruit and nut crops*, 761-788. The international Society for Horticultural Science, Wageningen, The Netherlands.

Ryder C.D., Smith L.B., Teakle G.R., King G.J. (2001) Contrasting genome organisation: two regions of the *Brassica oleracea* genome compared with collinear regions of the *Arabidopsis thaliana* genome. *Genome* 44 : 808-817.

S

Saal B., Plieske J., Hu J., Quiros C.F., Struss D. (2001) Microsatellite markers for genome analysis in Brassica. II Assignment of rapeseed microsatellites to the A and C genomes and genetic mapping in *Brassica oleracea* L. *Theor. Appl. Genet.* 102 : 695-699.

Saintagne C. (2003) Distribution des régions génomiques différenciant deux espèces proches : le chêne sessile (*Quercus petraea*) et le chêne pédonculé (*Q. robur*). Thèse Université Henry Poincaré, Nancy I.

Saintagne C., Bodenes C., Barreneche T., Pot D., Plomion C., Kremer A. (2004) Distribution of genomic regions differentiating oak species assessed by QTL detection. *Heredity* 92 : 20-30.

Salse J., Piégu B., Cooke R., Delseny M. (2002) Synteny between *Arabidopsis thaliana* and rice at the genome level: a tool to identify conservation in the ongoing rice genome sequencing project. *Nucleic Acids Res.* 30 : 2316-2328.

Salse J., Piegu B., Cooke R., 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 J.* 38 : 396-409.

Sambrook J., Fritsch E.F., Maniatis T. (1989) *Molecular cloning: a laboratory manual* (Ford N., Nolan C., Ferguson M., eds), 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sankoff D. et Nadeau J.H. (2003) Chromosome rearrangements in evolution : from gene order to genome sequence and back. *Proc. Natl. Acad. Sci. USA* 100 : 11188-11189.

Sax K., (1923) the association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 8 : 552-560.

Saxe H., Cannell M.G.R., Johnsen O., Ryan M.G., Vourlitis G. (2001) Tree and forest functioning in response to global warming. *New Phytol.* 149 : 369-400.

Scalfi M., Troglio M., Piovani P., Leonardi S., Magnaschi G., Vendramin G.G., Menozzi P. (2004) A RAPD, AFLP, and SSR linkage map, and QTL analysis in European beech (*Fagus sylvatica* L.). *Theor. Appl. Genet.* 180 : 443-441.

Schachter B. (2001) A tree grows in Manhattan. *HMS Beagle, The BioMedNet Magazine*, September 14, Issue 110.

Schlichting C.D. (1986) The evolution of phenotypic plasticity in plants. *Ann. Rev. Ecol. Syst.* 17 : 667-693.

Schmidt R. (2000) Synteny : recent advances and future prospects. *Curr. Opin. Plant Biol.* 3 : 97-102.

Scotti I., Magni F., Paglia G.P., Morgante M. (2002) Trinucleotide microsatellites in Norway spruce (*Picea abies*): their features and the development of molecular markers. *Theor. Appl. Genet.* 106 : 40-50.

Sewell M.M., Bassoni D.L., Megraw R.A., Wheeler N.C., Neale D.B. (2000) Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.).I. Physical wood properties. *Theor. Appl. Genet.* 101 : 1273-1281.

Sewell M.M., Neale D.B. (2000) Mapping quantitative traits in forest trees. In *Molecular Biology of woody plants* (eds Jain S.M. et Minocha S.C.), pp. 407-423. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Sheffield V.C., Beck J.S., Kwitek A.E., Sandstrom D.W., Stone E.M. (1993) The sensitivity of single-strand conformation polymorphism analysis for detection of single base substitutions. *Genomics* 16 : 325-332.

- Sheperd M., Cross M., Maguire T.L., Dieters M.J., Williams C.G., Henry R.J. (2002) Transpecific microsatellites for hard pines. *Theor. Appl. Genet.* 104 : 819-827.
- Shoemaker R.C., Polzin K., Labate J., Specht J., Brummer E.C., Olson T., Young N., Concibido V., Wilcox J., Tamulonis J.P., Kochert G., Boerma H.R. (1996) Genome duplication in soybean (*Glycine* subgenus *soja*). *Genetics* 144 : 329-338.
- Skov E. (1998) Are RAPD-markers reproducible between different laboratories? A case study of *Picea abies* (L.) KARST. *Silvae Genetica* 1998 47 : 283-287.
- Soltis P.S., Soltis D.E., Chase M.W. (1999) Angiosperm phylogeny inferred from multiple genes as tool for comparative biology. *Nature* 402 : 402-404.
- Soltis D.E. et Soltis P.S. (2003) The role of phylogenetics in comparative genetics. *Plant Physiol.* 132 : 1790-1800.
- Song R., Llaca V., Messing J. (2002) Mosaic organization of orthologous sequences in grass genomes. *Genome Res.* 12 : 1549-1555.
- Spalding E.P. (2000) Ion channels and the transduction of light signals. *Plant Cell Env.* 23 : 665-674.
- Stam P. (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *Plant J.* 3 : 739-744.
- Stam P., Van Ooijen J.W. (1995) JoinMap™ version 2.0: software for the calculation of genetic linkage maps. CPRO-DLO, Wageningen.
- Steinkellner H., Lexer C., Turetschek E., Glössl J. (1997a) Conservation of (GA)_n microsatellite loci between *Quercus* species. *Mol. Ecol.* 6: 1189-1194.
- Steinkellner H., Fluch S., Turetschek E., Lexer C., Streiff R., Kremer A., Burg K., Glössl J. (1997b) Identification and characterization of (GA/CT)_n microsatellite loci from *Quercus petraea*. *Plant Mol. Biol.* 33 : 1093-1096
- Stracke S., Sato S., Sandal N., Koyama M., Kaneko T., Tabata S., Parniske M. (2004) Exploitation of colinear relationships between the genomes of *Lotus japonicus*, *Pisum sativum* and *Arabidopsis thaliana*, for positional cloning of a legume symbiosis gene. *Theor. Appl. Genet.* 108 : 442-449.
- Streiff R., Labbé T., Bacilieri R., Steinkellner H., Glössl J., Kremer A. (1998) Within population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl assessed with isozymes and microsatellites. *Mol. Eco.l* 7 : 317-328.
- Sturtevant A.H. (1913) The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J. Exp. Zool.* 14 : 43-59.
- Sutton T., Whitford R., Baumann U., Dong C., Able J.A., Langridge P. (2003) The Ph2 pairing homoeologous locus of wheat (*Triticum aestivum*): identification of candidate meiotic genes using a comparative genetics approach. *Plant J.* 36 : 443-456.

Syvänen A.-C. (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Rev. Genet.* 2 : 930-942.

T

Tanksley S. D., Ganai M. W., Prince J. P., de-Vicente M. C., Bonierbale M. W., Broun P., Fulton T. M., Giovannoni J. J., Grandillo S., Martin G. B., Messeguer R., Miller J. C., Miller L., Paterson A. H., Pineda O., Roder M. S., Wing R. A., Wu W. and Young N. D. (1992) High density molecular linkage maps of the tomato and potato genomes. *Genetics* 132 : 1141-1160.

Tanksley S.D. (1993) Mapping polygenes. *Annu. Rev. Genet.* 27 : 205-233.

Taylor J.S., Durkin J.M.H., Breden F. (1999) The death of a microsatellite: a phylogenetic perspective on microsatellite interruptions. *Mol. Biol. Evol.* 16 : 567-572.

Tekaia F., Lazcano A., Dujon B. (1999) The genomic tree as revealed from whole proteome comparisons. *Genome Res.* 9 : 550-557.

Thabuis A., Palloix A., Pflieger S., Daubeze A.M., Caranta C., Lefebvre V. (2003) Comparative mapping of *Phytophthora* resistance loci in pepper germplasm: evidence for conserved resistance loci across Solanaceae and for a large genetic diversity. *Theor. Appl. Genet.* 106 : 1473-1485.

Thoday J.M. (1961) Location of polygenes. *Nature* 19 : 368-370.

Thorup T.A., Tanyolac B., Livingstone K.D., Popovsky S., Paran I., Jahn M. (2000) Candidate gene analysis of organ pigmentation loci in the Solanaceae. *Proc. Natl. Acad. Sci. USA* 97 : 11192-11197.

Thuriaux P. (1977) Is recombination confined to structural genes on the eucaryotic genome? *Nature* 268 : 460-462.

Tikhonov A.P., SanMiguel P.J., Nakajima Y., Gorenstein N.M., Bennetzen J.F., Avramova Z. (1999). Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. *Proc. Natl. Acad. Sci. USA* 96: 7409–7414.

Tikkanen O.-P., Lyytikäinen-Saarenmaa P. (2002) Adaptation of a generalist moth, *Operophtera brumata*, to variable budburst phenology of host plants. *Ent. Exp. Appl.* 103 : 123-133.

Tikkanen O.-P., Tiitto R.J. (2003) Phenological variation as protection against defoliating insects: the case of *Quercus robur* and *Operophtera brumata*. *Oecologia* 136 : 244-251.

Tóth G., Gáspári Z., Jurka J. (2000) Microsatellites in different eukaryotic genomes : survey and analysis. *Genome Res.* 10 : 967-981.

Tsarouhas V., Gullberg U., Lagercrantz U. (2002) An AFLP and RFLP linkage map and quantitative trait locus (QTL) analysis of growth traits in *Salix*. *Theor. Appl. Genet.* 105 : 277-288.

Tsarouhas V., Gullberg U., Lagercrantz U. (2003) Mapping of quantitative trait loci controlling timing of bud flush in *Salix*. *Hereditas* 138 : 172-178.

Tsumura Y., Ohba K., Strauss S.H. (1996) Diversity and inheritance of inter-simple sequence repeat polymorphisms in Douglas-fir (*Pseudotsuga menziesii*) and sugi (*Cryptomeria japonica*). *Theor. Appl. Genet.* 92 : 40-45.

V

van Buijtenen J.P. (2001) Genomics and quantitative genetics. *Can. J. Forest Res.* 31, 617-622.

Van Buuren M.L., Salvi S., Morgante M., Serhani B., Tuberosa R. (2002) Comparative genomic mapping between a 754 kb region flanking DREB1A in *Arabidopsis thaliana* and maize. *Plant Mol. Biol.* 48 : 741-750.

Vandepoele K., Simillion C., Van de Peer Y. (2003) Evidence that rice and other cereals are ancient aneuploids. *Plant Cell* 15 : 2192-2202.

Van der Hoeven R., Ronning C., Giovannoni J., Martin G., Tanksley S. (2002) Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. *Plant Cell* 14 : 1441-1456.

Van Ooijen J.W. (1992) Accuracy of mapping quantitative trait loci in autogamous species. *Theor. Appl. Genet.* 84 : 803-811.

Van Ooijen J.W. (1999) LOD significance thresholds for QTL analysis in experimental populations of diploid species. *Heredity* 83 : 613-624.

van Treuren R., Kuittinen H., Kärkkäinen K., Baena-Gonzalez E., Savolainen O. (1997) Evolution of microsatellites in *Arabis petraea* and *Arabis lyrata*, outcrossing relatives of *Arabidopsis thaliana*. *Mol. Biol. Evol.* 14 : 220-229.

Ventelon M., Deu M., Garsmeur O., Doligez A., Ghesquière A., Lorieux M., Rami J.F., Glaszmann J.C., Grivet L. (2001) A direct comparison between the genetic maps of sorghum and rice. *Theor. Appl. Genet.* 102 : 379-386.

Verhaegen D., Plomion C., Gion J.M., Poitel M., Costa P., Kremer A. (1997) Quantitative trait dissection analysis in Eucalyptus using RAPD markers: 1 Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different ages. *Theor. Appl. Genet.* 95 : 597-608.

Verhoeven K.J.F., Vanhala T.K., Biere A., Nevo E., van Damme J.M.M. (2004) The genetic basis of adaptive population differentiation: a quantitative trait locus analysis of fitness traits in two wild barley populations from contrasting habitats. *Evolution* 58 : 270-283.

Via S., Gomulkiewicz R., De Jong G., Scheiner S.M., Schlichting C.D., van Tienderen P.H. (1995) Adaptive phenotypic plasticity: consensus and controversy. *Trends Ecol. Evol.* 10 : 212-217.

Villani F., Pigliucci M., Benedettelli S., Cherubini M. (1991) Genetic differentiation among Turkish chestnut (*Castanea sativa* Mill.) populations. *Heredity* 66: 131-136.

Villani F., Pigliucci M., Lauteri M., Cherubini M., Sun O. (1992) Congruence between genetic, morphometric, and physiological data on differentiation of Turkish chestnut (*Castanea sativa*). *Genome* 35: 251-256.

Villani F., Pigliucci M., Cherubini M. (1994) Evolution of *Castanea sativa* Mill. in Turkey and Europe. *Genet. Res. Camb.* 63: 109-116.

Vision T.J., Brown D.G., Tanksley S.D. (2000) The origins of genomic duplications in *Arabidopsis*. *Science* 290 : 2114-2116.

Visscher P.M., Thompson R., Haley C.S. (1996) Confidence intervals in QTL mapping by bootstrapping. *Genetics* 143, 1013-1020.

Voorrips R.E. (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93 : 77-78.

Vos P., Hogers R., Bleeker M., Reijans M., van de Lee T., Hornes M., Frijters A., Pot J., Peleman J., Kuiper M., Zabeau M. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23 : 4407-4414.

W

Wakamiya I., Newton R.J., Johnston J.S., Price H.J. (1993) Genome size and environmental factors in the genus *Pinus*. *Amer. J. Bot.* 80 : 1235-1241.

Wakamiya I., Price H.J., Messina M.G., Newton R.J. (1996) Pine genome size diversity and water relations. *Physiol. Plant.* 96 : 13-20.

Walther G.-R., Post E., Convey P., Menzel A., Parmesan C., Beebee T.J.C., Fromentin J.-M., Hoegh-Guldberg O., Bairlein F. (2002) Ecological responses to recent climate change. *Nature* 416 : 389-395.

Wang Z., Weber J.L., Zhong G., Tanksley S.D. (1994) Survey of plant short tandem DNA repeats. *Theor. Appl. Genet.* 88 : 1-6.

Wendel J.F., Cronn R.C., (2003) Polyploidy and the evolutionary history of cotton. *Adv. Agron.* 78 : 139-186.

Weng C., Kubisiak T.L., Nelson C.D., Stine M. (2002) Mapping quantitative trait loci controlling early growth in a (logleaf pine x slash pine) x slash pine BC₁ family. *Theor. Appl. Genet.* 104 : 852-852.

Whitton J., Rieseberg L.H., Ungerer M.C. (1997) Microsatellite loci are not conserved across the Asteraceae. *Mol. Biol. Evol.* 14 : 204-209.

Williams J.G., Kubelik A.R., Livak K.J., Rafalski J.A., Tingey S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 18 : 6531-6535.

Wu R.L. (1998a) Genetic mapping of QTLs affecting tree growth and architecture in *Populus*: implication for ideotype breeding. *Theor. Appl. Genet.* 96 : 447-457.

Wu R.L. (1998b) The detection of plasticity genes in heterogeneous environments. *Evolution* 52 : 967-977.

Wu R.L., Bradshaw H.D., Stettler R.F. (1998) Developmental quantitative genetics of growth in *Populus*. *Theor. Appl. Genet.* 97 : 1110-1119.

Wu R., Ma C.X., Zhu J., Casella G. (2002) Mapping epigenetic quantitative trait loci (QTL) altering a developmental trajectory. *Genome* 45 : 28-33.

X

Xu L.-A. (2004) Diversité de l'ADN chloroplastique et relations phylogénétiques au sein des Fagacées et du genre *Quercus*. Thèse Université Henry Poincaré, Nancy I.

Y

Yamamoto T., Kimura T., Sawamura Y., Kotobuki K., Ban Y., Hayashi T., Matsuta N. (2001) SSRs isolated from apple can identify polymorphism and genetic diversity in pear. *Theor. Appl. Genet.* 102 : 865-870.

Yan H.H., Mudge J., Kim D.-J., Larsen D., Shoemaker R.C., Cook D.R., Young N.D. (2003) Estimates of conserved microsynteny among the genomes of *Glycine max*, *Medicago truncatula* and *Arabidopsis thaliana*. *Theor. Appl. Genet.* 106 : 1256-1265.

Yan H.H., Mudge J., Kim D.-J., Shoemaker R.C., Cook D.R., Young N.D. (2004) Comparative physical mapping reveals features of microsynteny between *Glycine max*, *Medicago truncatula* and *Arabidopsis thaliana*. *Genome* 47 : 141-155.

Yan L., Loukoianov A., Tranquilli G., Helguera M., Fahima T., Dubcovsky J. (2003) Positional cloning of the wheat vernalization gene VRN1. *Proc. Natl. Acad. Sci. U S A* 100 : 6263-6268.

Yan L., Loukoianov A., Blechl A., Tranquilli G., Ramakrishna W., SanMiguel P., Bennetzen J.L., Echenique V., Dubcovsky J. (2004) The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. *Science* 303 : 1640-1644.

Young E.T., Sloan J.S., van Riper K. (2000) Trinucleotide repeats are clustered in regulatory genes in *Saccharomyces cerevisiae*. *Genetics* 154 : 1053-1068.

Young N.D., Weeden N.F., Kochert G. (1996) Genome mapping in Legumes (Fam. Fabaceae). In *Genome mapping in Plants*, A.H. Paterson ed. R.G. Landes Co. Georgetown, TX. pp 211-227.

Yu J., Hu S., Wang J., Wong G.K., Li S., Liu B., Deng Y., Dai L., Zhou Y., Zhang X., Cao M., Liu J., Sun J., Tang J., Chen Y., Huang X., Lin W., Ye C., Tong W., Cong L., Geng J., Han Y., Li L., Li W., Hu G., Huang X., Li W., Li J., Liu Z., Li L., Liu J., Qi Q., Liu J., Li L., Li T., Wang X., Lu H., Wu T., Zhu M., Ni P., Han H., Dong W., Ren X., Feng X., Cui P., Li X., Wang H., Xu X., Zhai W., Xu Z., Zhang J., He S., Zhang J., Xu J., Zhang K., Zheng X.,

Dong J., Zeng W., Tao L., Ye J., Tan J., Ren X., Chen X., He J., Liu D., Tian W., Tian C., Xia H., Bao Q., Li G., Gao H., Cao T., Wang J., Zhao W., Li P., Chen W., Wang X., Zhang Y., Hu J., Wang J., Liu S., Yang J., Zhang G., Xiong Y., Li Z., Mao L., Zhou C., Zhu Z., Chen R., Hao B., Zheng W., Chen S., Guo W., Li G., Liu S., Tao M., Wang J., Zhu L., Yuan L., Yang H. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp indica). *Science* 296 : 79-92.

Z

Zeng Z.B. (1994) Precision mapping of quantitative trait loci. *Genetics* 136: 1457-1468.

Zeng Z.B., Kao C.H., Basten C.J. (1999) Estimating the genetic architecture of quantitative traits. *Genet. Res. Camb.* 74 : 279-289.

Zhang J., Zheng H. G., Aarti A., Pantuwan G., Nguyen T. T., Tripathy J. N., Sarial A. K., Robin S., Babu R. C., Nguyen Bay D., Sarkarung S., Blum A., Nguyen H. T. (2001) Locating genomic regions associated with components of drought resistance in rice: comparative mapping within and across species. *Theor. Appl. Genet.* 103 : 19-29.

Zhu H., Kim D.-J., Baek J.-M., Choi H.-K., Ellis L.C., Küester H., McCombie W.R., Peng H.-M., Cook D.R. (2003) Syntenic relationships between *Medicago truncatula* and Arabidopsis reveal extensive divergence of genome organization. *Plant Physiol.* 131 : 1018-1026.

Zietkiewicz E., Rafalski A., Labuda D. (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* 20 : 176-183.

Zohary D., Hopf M. (1988) Domestication of plants in the old world. Clarendon Press, Oxford.

Zoldos V., Papes D., Brown S.C., Panaud O., and Siljak-Yakovlev S. (1998) Genome size and base composition of seven *Quercus* species: inter- and intra-population variation. *Genome* 41 : 162-168.

Annexe I : A genetic linkage map of European chestnut (*Castanea sativa* Mill.) based on RAPD, ISSR and isozyme markers.

M.Casasoli, C. Mattioni, M. Cherubini, F.Villani (2001) Theor. Appl. Genet. 102: 1190-1199.

Annexe II : Identification of QTLs affecting adaptive traits in *Castanea sativa* Mill.

M. Casasoli, D. Pot, C. Plomion, M.C. Monteverdi, T. Barreneche, M. Lauteri, Villani F.
(2004) Plant Cell Envir. 27: 1088-1101.

Annexe III : Comparative mapping between *Quercus* and *Castanea* using simple-sequence repeats (SSRs).

T. Barreneche, M. Casasoli, K. Russell, A. Akkak, H. Meddour, C. Plomion, F. Villani, A. Kremer (2004) *Theor. Appl. Genet.* 108: 558-566.

Annexe IV : Comparison of QTLs for adaptive traits between oak and chestnut based on an EST consensus map.

M. Casasoli, J. Derory, C. Morera-Dutrey, O. Brendel, J. M. Guehl, F. Villani, A. Kremer
(manuscrit en préparation).

Annexe V : Protocole de génotypage SNP (VICTOR).