# Contribution à l'étude de la structure et du polymorphisme du génome du basidiomycète ectomycorhizien "Laccaria bicolor" (Maire) Orton et identification de QTLs de mycorhization chez les peupliers, "Populus trichocarpa Torr. & A. Gray ex Hook. et "Populus deltoides (Bartr.) Marsh

Jessy Labbé

U.F.R. Sciences et Techniques Biologiques
Ecole Doctorale Ressources Procédés Produits Environnement

UMR 1136 INRA/UHP
Interactions arbres/Micro-organismes
Centre INRA de Nancy

Thèse

Présentée pour l'obtention du titre de docteur de l'Université Henri Poincaré, Nancy 1

Biologie Végétale et Forestière

Par

Jessy Labbé

# Contribution à l'étude de la structure et du polymorphisme du génome du basidiomycète ectomycorhizien *Laccaria bicolor* (Maire) Orton et identification de QTLs de mycorhization chez les Peupliers, *Populus trichocarpa* Torr. & A. Gray ex Hook. et *Populus deltoides* (Bartr.) Marsh.

Soutenue en publique le 04 Novembre 2009

Membres du jury

| | |
|---|---|
| *Président:* | |
| Pierre Leblond | Professeur, Université Henri Poincaré, Nancy |
| *Rapporteurs :* | |
| Roland Marmeisse | Chargé de Recherches (HDR), CNRS, Lyon |
| Sylvain Jeandroz | Maître de conférences (HDR), ENESAD, Dijon |
| *Examinateurs* : | |
| Véronique Jorge | Chargé de recherches, INRA, Orléans |
| François Le Tacon | Directeur de recherches émérite, INRA, Nancy (directeur de thèse) |
| Francis Martin | Directeur de recherches, INRA, Nancy (co-directeur de thèse) |
| *Invité* : | |
| Patricia Faivre-Rampant | Chargé de Recherches, INRA, Evry-Val d'Essone |

Co-Financement :

*Labor omnia vincit improbus…*

*(Géorgiques de Virgile* I, 145-146)

A mes parents, à ma femme et à mes Tontons…..

# Table des matières

## Remerciements

Il est vain d'imaginer que quelques mots suffisent à remercier ou à exprimer une quelconque gratitude à tous ceux qui ont permis ce travail. Mais ces quelques mots seront des plus sincères.

Je tiens tout d'abord à remercier Roland Marmeisse, Sylvain Jeandroz, Pierre Leblond, Véronique Jorge et Patricia Faivre-Rampant qui me font l'honneur de juger ce travail.

J'adresse mes plus sincères remerciements à mes deux pères spirituels François Le Tacon et Francis Martin. Merci pour votre confiance, pour vos conseils, pour vos encouragements et pour la grande liberté que vous m'avez laissée. Je ne sais comment exprimer le plaisir que j'ai de travailler avec vous. J'espère avoir été à la hauteur en étant votre "Padawan" : « *La règle du Code Jedi stipule qu'un maître ne peut avoir qu'un seul Padawan en même temps…… »* quelle chance, moi j'ai eu deux maîtres, en même temps!

Je souhaite ensuite remercier Véronique Jorge et Catherine Bastien pour m'avoir initié à l'art de la cartographie génétique et de la détection de QTL ; merci de m'avoir accueilli à Orléans. Je remercie également l'ensemble du personnel du centre INRA d'Orléans.

Je remercie Tongming Yin, Jerry Tuskan et l'ensemble des collègues américains, les "T Vols", également pour leur accueil et tout ce qu'ils m'ont appris, je vous dis à bientôt!

Je souhaite également remercier tous les collègues avec qui j'ai pu interagir au travers des divers projets qui m'ont été confiés notamment, Hélène Hirzel-Niculita, René Smulder, Hadi Quesneville, Kathrin Donges… et bien d'autres, merci.

Egalement un grand merci à Pierre Leblond, Nathalie Leblond, Virginie Libante, Gérard Guédon, Florence Charron-Bourgoin et à l'ensemble de l'équipe pédagogique du Laboratoire de Génétique et Microbiologie avec qui j'ai pu prendre plaisir à l'enseignement dans d'agréables conditions.

Merci à Marc-André Selosse pour les discussions et les échanges que nous avons eus,…merci MAS!

Je voudrais remercier tous les membres du laboratoire IaM (les microbes !) et l'ensemble de l'UMR 1136 pour leurs expertises, leurs aides, leurs sourires et pour l'agréable ambiance de travail.

Je tiens à exprimer ma vive reconnaissance à mes compagnons de route Pierre-Emmanuel Courty, François Rineau, Aurélie Deveau, Veronica Pereda, Marlis Reich, Lucie Vincenot (youpi girl!), Judith Richter (ma belle blonde!), Sara Hortal (super de travailler avec toi copine!), Claude Murat, Adeline Rigal, Stéphane Hacquart, Angela Cusano, Jonathan Plett et Carine Commun, merci de vos coups de mains, de votre gentillesse et de votre bonne humeur.

Un p'tit clin d'œil et toute ma reconnaissance à mon p'tit staff médical, Béatrice Brembilla-Perrot, Simona Kolopp-Trusca, à leurs collègues et à Gérard Frey, pour votre compétence et votre soutien, merci de tout mon Coeur!

Enfin, j'adresse un GRAND MERCI à mes parents qui m'ont permis d'être ce que je suis et à ma femme, vous avez subi ces années de travail tout en m'apportant votre soutien inébranlable et l'équilibre qui m'est nécessaire, merci.

# 1. Introduction

Les associations étroites et durables entre espèces différentes sont un des facteurs clé de l'évolution qui est orientée par ces phénomènes de coopération, d'interaction et de dépendance mutuelle (Sapp, 1994). Parmi ces associations, les symbioses mycorhiziennes (associations entre les racines d'une plante et un champignon) sont universellement rencontrées dans les écosystèmes terrestres et pourraient avoir joué un rôle fondamental dans la colonisation de ce milieu par les plantes (Smith & Read, 1996 ; Selosse & Le Tacon, 1998 ; Read & Perez-Moreno, 2003).

En effet, le mutualisme mycorhizien implique près de 95% des végétaux (Malloch *et al.*, 1980) et seules quelques familles ne sont pas mycorhizées (Crucifères, Chénopodiacées, Saxifragacées, Caryophyllacées et Joncacées). L'absence de mycorhizes pourrait être due à la croissance dans les milieux riches ou aquatiques et/ou à la production de composés racinaires ayant une activité antifongique (Schreiner et Koide, 1993). Les mycorhizes sont communément classées en trois principaux groupes définis selon des critères anatomiques: les ectomycorhizes, les endomycorhizes et les ectendomycorhizes. Le groupe des endomycorhizes est actuellement le plus répandu (Nicolson, 1967). Les ectomycorhizes constituent le deuxième groupe le plus rencontré et sont principalement établies avec les grandes espèces sociales des forêts boréales et tempérées caducifoliées et des régions tropicales à saisons contrastées (Read *et al.*, 2003).

La symbiose ectomycorhizienne évolue depuis 130 à 180 millions d'années (LePage *et al.*, 1997) et joue un rôle central dans tous les grands processus biogéochimiques (biodisponibilité en azote minérale, altération

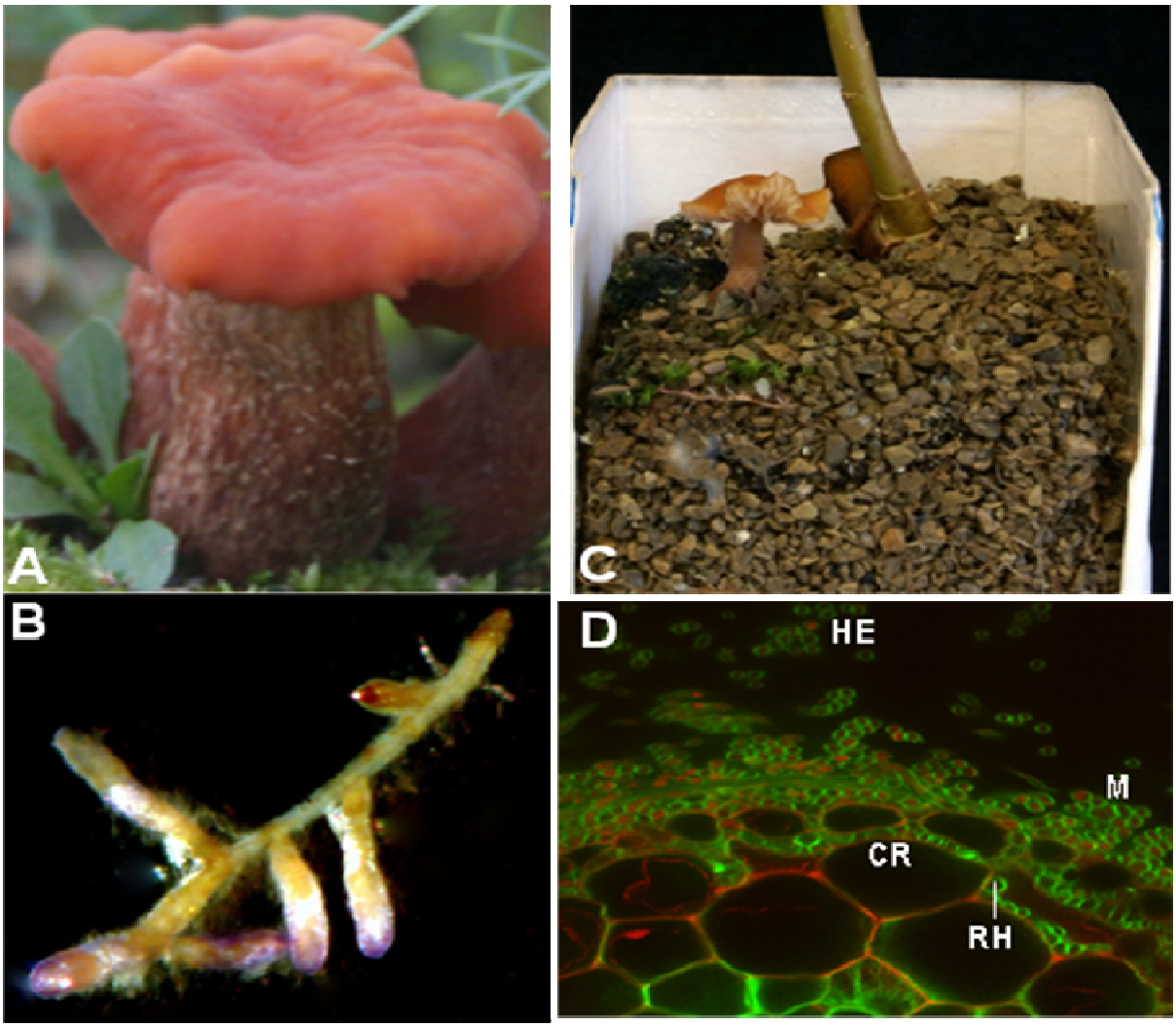



**Figure 1 : La symbiose ectomycorhizienne**

(A) Le basidiomycète *Laccaria bicolor* S238N (photographie D. Vairelles). (B) Plant de peuplier (*Populus trichocarpa*) colonisé par *Laccaria bicolor* (J. Labbé). (C) De courtes racines de peuplier colonisées par *L. bicolor* S238N. Le manteau d'aspect nacré et violacé entoure les racines courtes de *P. trichocarpa* (J. Labbé). (D) Coupe transversale (30 µm) de mycorhize de *Laccaria/Populus* vue au microscope confocal (600x) à balayage laser (l'iodure de propidium colore en rouge les parois des cellules végétales ainsi que les noyaux de la plante et du champignon ; l'UVitex colore en vert la chitine contenue dans les parois des cellules fongiques : les hyphes extramatriciels (HE) explorent le milieu tandis qu'un manteau est déjà formé autour de la racine (M) ; l'hyphe fongique a commencé à pénétrer entre les cellules épithéliales et les cellules du cortex de la racine (CR) pour former le réseau de Hartig (RH ; photographie de J. Felten).

des minéraux primaires, recyclage du carbone et minéralisation de la matière organique). Dans cette symbiose, le mycélium fongique colonise la surface racinaire sur laquelle il prolifère et forme un manchon pseudo-parenchymateux, puis il s'insinue dans la lamelle moyenne des cellules corticales de la racine de la plante-hôte formant un réseau mycélien strictement extracellulaire appelé réseau de Hartig (Figure 1 ; Mankel *et al.*, 2000). Ce dernier constitue l'interface symbiotique. Ce type d'association implique des champignons supérieurs variés, Ascomycètes (*Elaphomycetaceae*, *Tuber terfeziaceae*) ou Basidiomycètes (*Agaricaceae*, *Amanitaceae*, *Boletaceae*, *Cortinaceae*) appartenant à plus de 5000 espèces (Molina *et al.*, 1992, Barker *et al.*, 1998a) et des plantes-hôtes ligneuses (Gymnospermes ou Angiospermes). Malgré une relative unité structurale, cette diversité des partenaires suggère des émergences répétées et une grande diversité fonctionnelle (Hibbet & Donoghue, 1996).

Ce système très performant pour mobiliser et favoriser l'absorption par la racine des éléments minéraux du sol grâce au réseau mycélien rhizosphérique et tellurique très dense, (1000 m de mycélium/m de racine), améliore la nutrition de la plante-hôte (Chalot & Brun, 1998 ; Martin, 2007). L'ectomycorhize protège également la racine de l'hôte contre divers pathogènes telluriques. Cette protection est assurée mécaniquement par la barrière que constitue le manteau, mais aussi par accumulation de composés antibiotiques en réponse à la colonisation fongique (Frey *et al.*, 1997, Weiss *et al.*, 1997). En retour, le champignon bénéficie d'un apport carboné (10 à 40% des photosynthétats de la plante) (Högberg *et al.*, 2001).

Les effets bénéfiques sur les plantes-hôtes peuvent être mis à profit en foresterie et ont permis en sylviculture la mycorhization contrôlée de plants en pépinière (Olivier *et al.*, 1997 ; Le Tacon *et al.*, 1997) et en

plantations (Dell & Malajczuk, 1997). Ces micro-organismes ont donc une fonction essentielle dans le fonctionnement des écosystèmes forestiers et influent fortement sur la diversité et la productivité des forêts en ayant un effet sur les communautés (Gehring & Whitman, 1994).

La symbiose induit des modifications morphogénétiques et physiologiques qui se traduisent par des différences d'expression génique ; ces dernières sont observées au niveau du transcriptome et du protéome et impliquent vraisemblablement des signaux contrôlant l'organisation de réseaux de gènes (Duplessis *et al.*, 2005 ; Martin et *al.*, 2001 ; Johansson *et al.*, 2004, Martin & Nehls, 2009). Des travaux ayant pour objectif d'identifier les déterminants génétiques de cette symbiose ont permis le séquençage de milliers d'ESTs (Peter *et al.*, 2003 ; Voiblet *et al.*, 2001). L'analyse de micro-réseaux d'ADNc (Duplessis *et al.*, 2005 ; Le Quéré *et al.*, 2005) a ainsi permis la mise en évidence d'ensembles de gènes dont l'expression est coordonnée. Beaucoup de fonctions cellulaires sont surexprimées pendant le développement de la symbiose : notons la multiplication des cellules fongiques, et chez la plante-hôte la différenciation cellulaire, l'émission de signaux et la mise en place des mécanismes de défense (Figure 2 ; Martin *et al.*, 2007). Cependant, ces études sont limitées à l'analyse d'environ 10 à 20% des gènes présents chez les deux partenaires, d'où la nécessité de développer des analyses « génome entier ». Le récent séquençage des génomes du peuplier et du clitocybe laqué bicolore (ou laccaire bicolore) permet d'envisager l'approfondissement de nos connaissances des processus moléculaires qui régissent le développement et le fonctionnement de la symbiose.

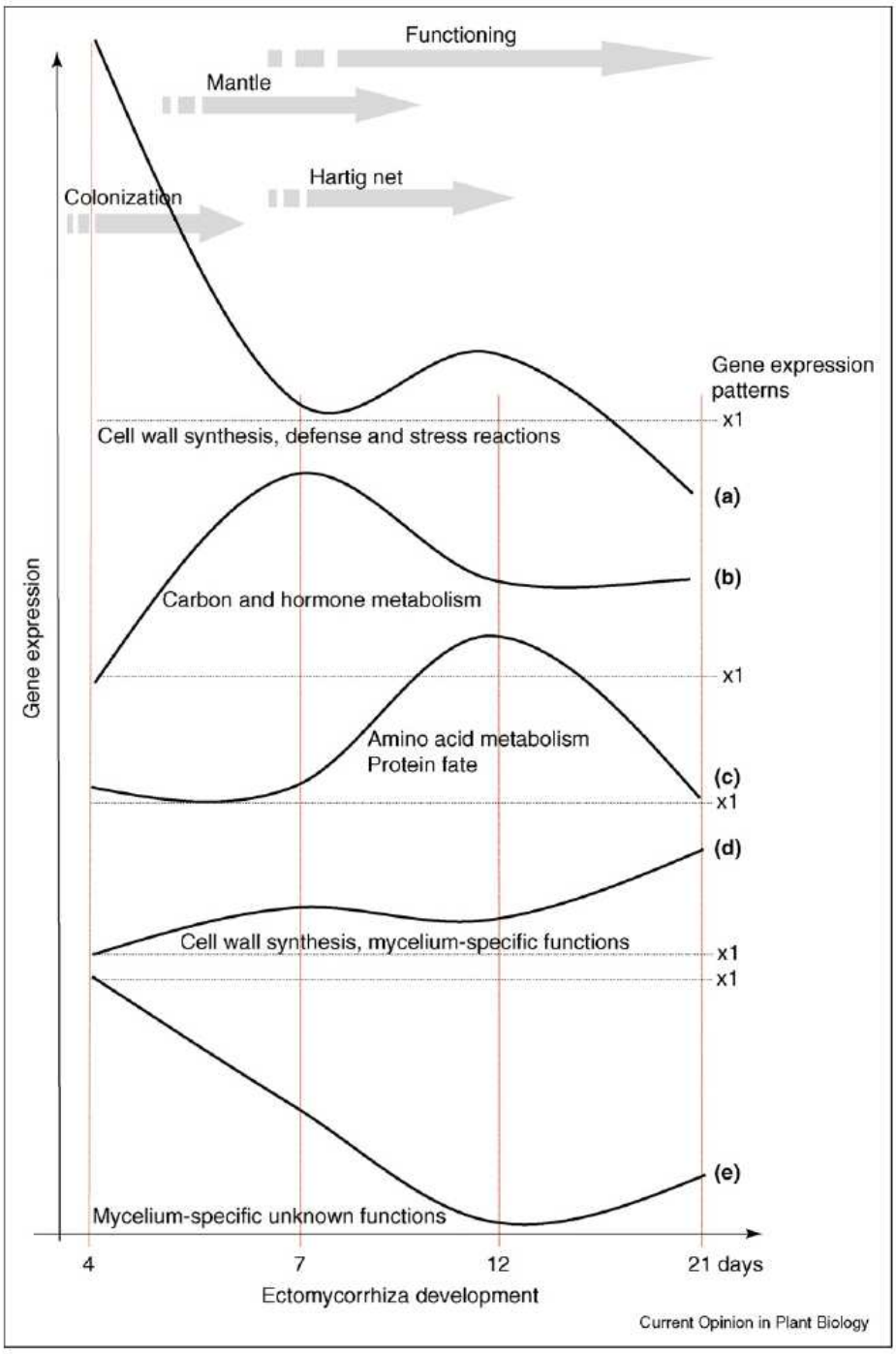



**Figure 2 : Profil d'expression des gènes de la plante et du champignon au cours du développement de l'ectomycorhize (d'après Martin *et al.,* 2007).** Les transcriptomes du symbiote fongique et de la plante-hôte ont été examinés dans les systèmes *Betula-Paxillus* et *Eucalyptus-Pisolithus* en utilisant les résultats des études de Duplessis *et al,* 2005 et Le Quéré *et al.*, 2005. Les groupes des gènes dont l'expression est coordonnée sont obtenus en utilisant un ensemble d'algorithmes hiérarchiques et non-hiérarchiques. Les modèles temporels majeurs d'induction ou de répression sont observés avec des groupes de gènes distincts à réponse précoce, moyenne ou tardive lors de la colonisation fongique.

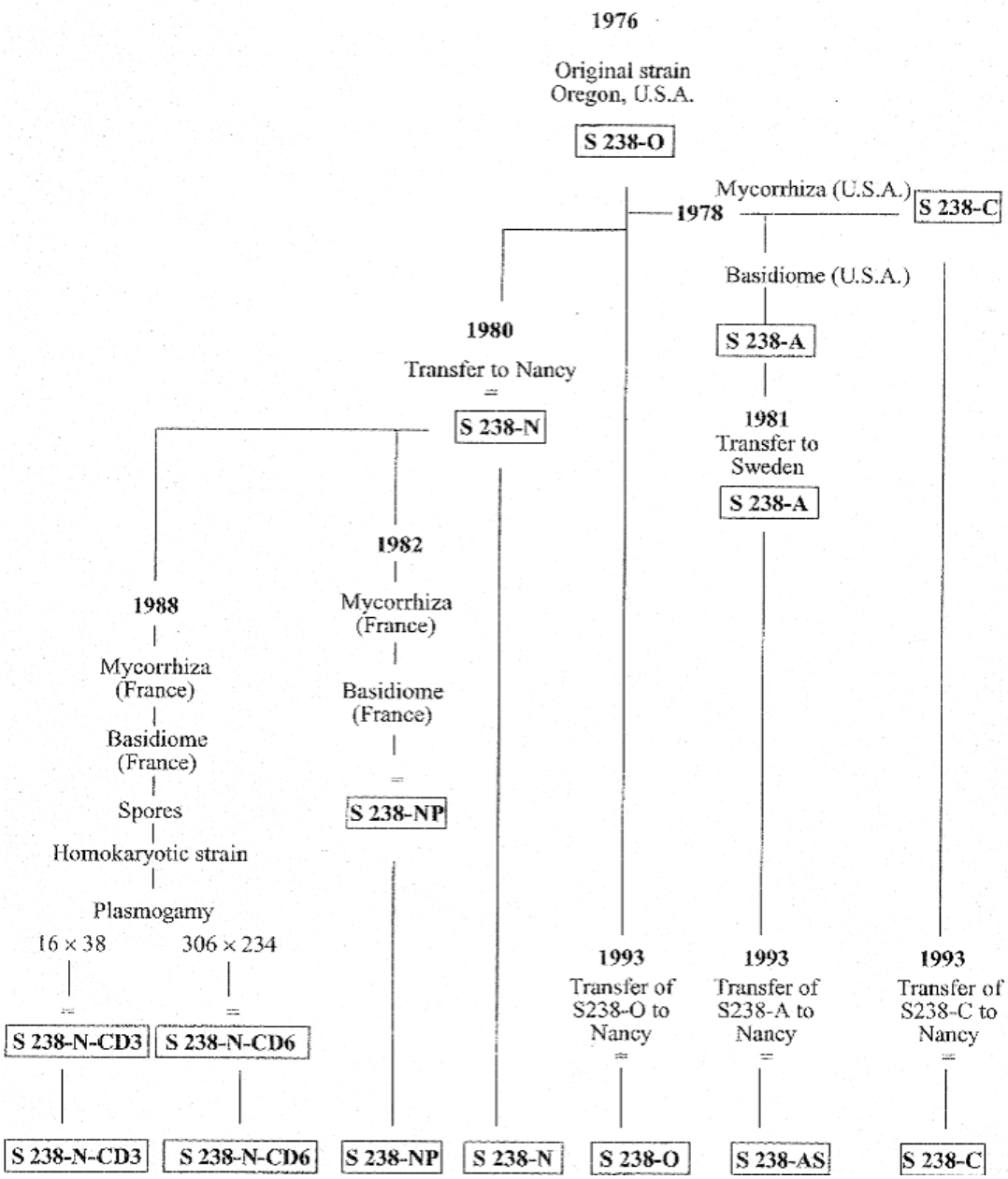



**Figure 3 : Origine et historique des souches de *L. bicolor* S238N (d'après Di Battista *et al.*, 1996).**

Ainsi notre projet de recherche s'inscrit d'une part dans le programme de séquençage et d'analyse du génome du champignon ectomycorhizien *Laccaria bicolor* et d'autre part dans la recherche des gènes impliqués dans l'établissement de la symbiose entre le peuplier et le clitocybe laqué bicolore. Nous présenterons d'abord quelques données sur la biologie, la génétique et

les ressources génomiques de ces espèces. Puis, j'exposerai les objectifs de ma thèse ainsi que la stratégie, le matériel et les méthodes utilisées.

## 1.1. Le basidiomycète *Laccaria bicolor* S238N

Le champignon ectomycorhizien *Laccaria bicolor* (clitocybe laqué bicolore) (Maire) P.D. Orton (*Basidiomycota, Agaricales, Hydnangiaceae* ; Matheny *et al*., 2007) forme des associations symbiotiques avec un grand nombre d'espèces d'arbres de l'hémisphère Nord (Mueller, 1991, 1992). D'un point du vue descriptif, le clitocybe laqué bicolore, épigé, possède un chapeau convexe et mamelonné de couleur bun-rose à ocre qui porte des lames échancrées doublées de lamelles d'une couleur claire rose-lilas (Figure 1). La base du pied est violette. En Europe comme en Amérique du Nord, *L. bicolor* est principalement associé à des Gymnospermes de la famille des *Pinaceae* bien qu'il soit parfois retrouvé associé à des *Caducifoliae* tels que le chêne et le hêtre. Comme d'autres champignons ectomycorhiziens, *L. bicolor* est donc ubiquiste (Raffle *et al*., 1985). La souche américaine S238N de *L. bicolor*, objet de cette étude, a été isolée en 1976 à partir d'un carpophore prélevé sous *Tsuga mertensiana* dans l'Oregon (USA) par Randy Molina et James Trappe, puis transférée à l'INRA de Nancy en 1980 (Di Battista et *al*., 1996). Sur la base de l'analyse RFLP de l'ADN ribosomal et du séquençage de la région intergénique IGS1 séparant l'ADN ribosomal 28s et 5s (Henrion et *al*., 1994), cette souche, originellement identifiée comme *L. laccata,* a été reclassée comme *L. bicolor* S238N (Armstrong *et al*., 1989). De plus l'étude de l'IGS1 ribosomique a révélé la variabilité génétique des monocaryons issus de la germination de spores de cette souche (Selosse & Le Tacon, 1998 ; Martin et *al*., 1999).

### 1.1.1. *Cycle du Clitocybe laqué bicolore*

Le cycle de développement de *L. bicolor* est du même type que celui de beaucoup de basidiomycètes supérieurs avec alternance de deux phases (Figure 4). La phase haploïde débute à la germination des spores qui produisent des hyphes haploïdes (pouvant contenir plusieurs noyaux identiques par article) que l'on nomme monocaryons. Ces monocaryons peuvent acquérir un noyau haploïde différent par plasmogamie avec un autre monocaryon. Il s'agit d'une espèce tétrapolaire hétérothallique (Kropp *et al.*, 1988). La compatibilité entre monocaryons est déterminée génétiquement par un système d'incompatibilité génétique. Deux gènes A et B commandent les phénomènes de compatibilités. Ils sont en effet déterminés par deux gènes ayant chacun deux allèles ; deux monocaryons sont alors compatibles si les allèles de leurs gènes A et B sont tous les deux différents.

Après plasmogamie entre deux monocaryons compatibles, commence alors la phase dicaryotique pendant laquelle deux noyaux différents par cellule coexistent (Esser & Kuehnen, 1967). La majeure partie de la phase végétative correspond au dicaryon obtenu ; les dicaryons forment les ectomycorhizes avec les racines des plantes-hôtes se trouvant à proximité. En conditions expérimentales, il est parfois possible d'obtenir des mycorhizes avec les monocaryons (Debaud *et al.*, 1988 ; Gardes *et al.*, 1990, Lamhamedi *et al.*, 1990). En conditions favorables, les hyphes dicaryotiques forment à partir des hyphes extra-matriciels des mycorhizes un amas cellulaire qui évolue en primordium puis en un organe charnu appelé carpophore (Figure 5).

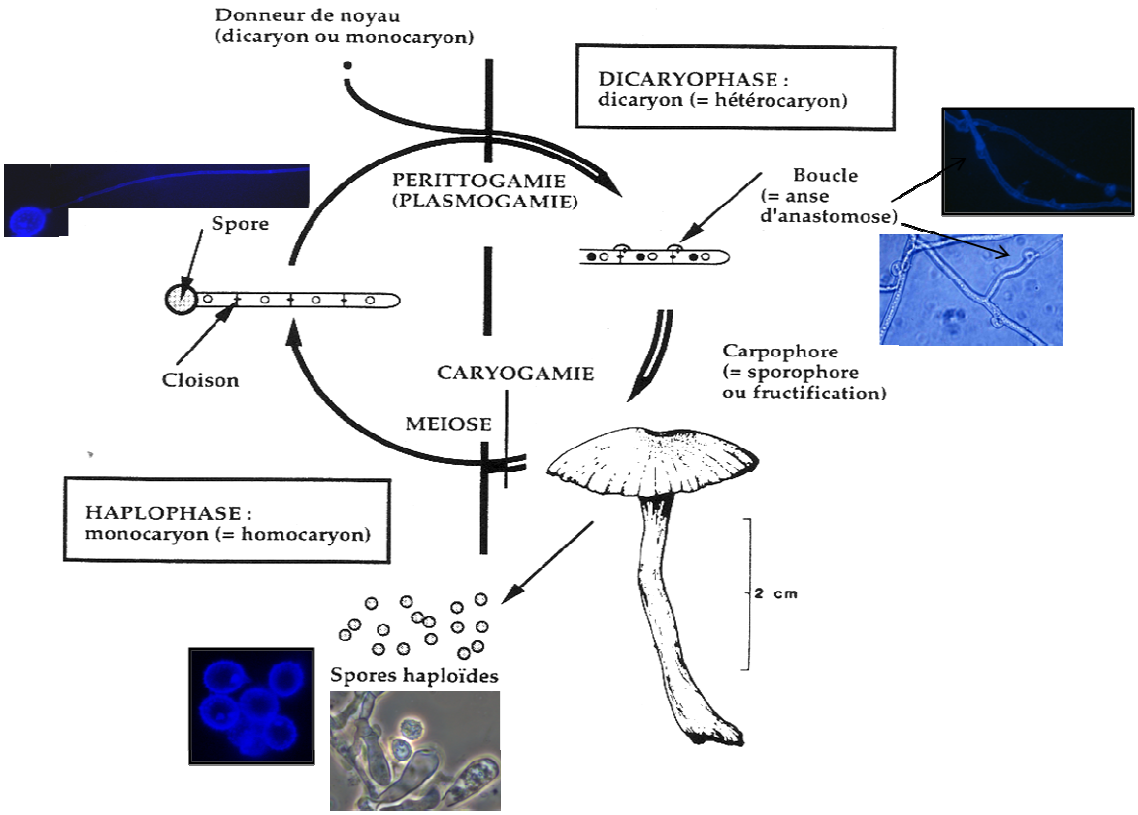



**Figure 4 : Cycle du clitocybe laqué bicolore** (photographies de microscopie confocale 40x, V. Legué ; photographies de microscopie optique 40x, J. Labbé).

Sur la face inférieure de cette structure aérienne se forment des lamelles où se différencient les spores méiotiques. Ces dernières sont formées sur des cellules en massue appelées basides dans lesquelles a eu lieu la caryogamie suivie de la méiose, bouclant ainsi le cycle de vie de cette espèce. La fructification s'observe en condition contrôlées en serre, quelques mois après l'inoculation de jeunes plants de Douglas ou de peuplier (Figure 5).

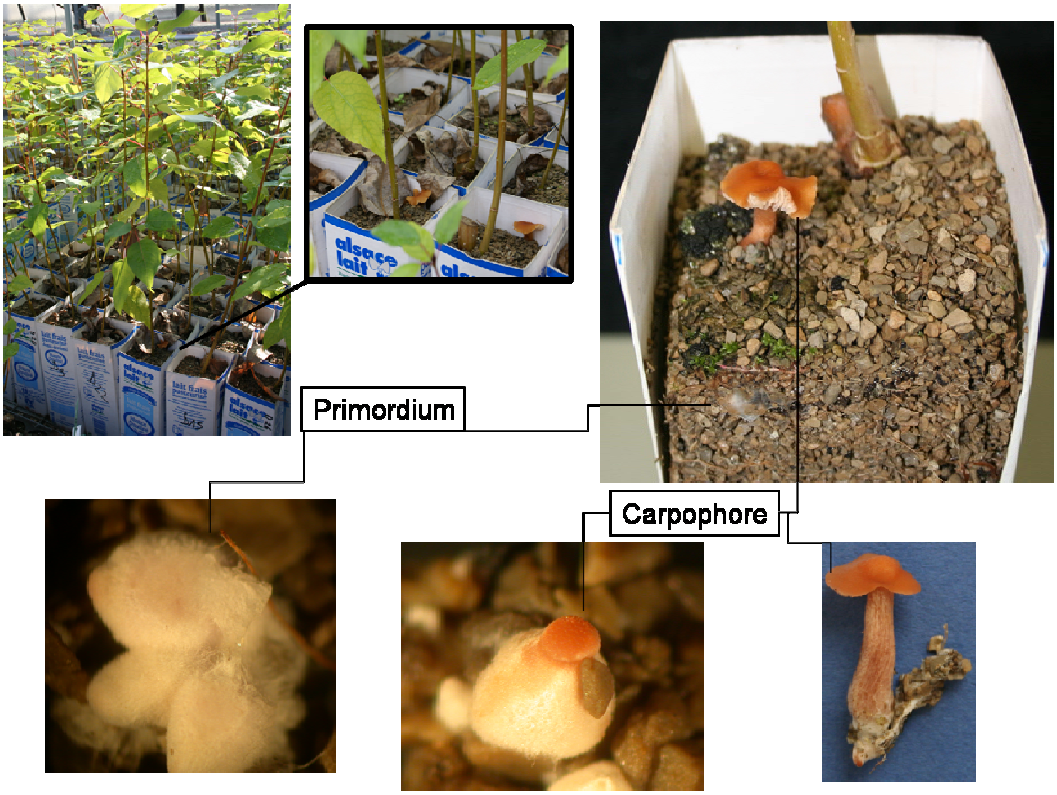



**Figure 5 : Fructifications de *L. bicolor* S238N observées sur des jeunes plants de peuplier (2mois après inoculation).**

### 1.1.2. *Un champignon utilisé en foresterie*

Les effets bénéfiques des mycorhizes sur la nutrition de leur hôte ont été mis à profit en foresterie par la mycorhization contrôlée de plants en pépinières. La production d'arbres forestiers à grande échelle se développe aux Etats-Unis dans les années 80 (Marx *et al.*, 1982). Son développement est initié en Europe dans les années 90, notamment sur un conifère d'origine

américaine, le Douglas (*Pseudotsuga menziesii*) qui est la première essence de reboisement en France (AFOCEL, 1997). Dix ans d'essais en pépinières et en plantations (Le Tacon *et al.*, 1992) ont permis la constitution d'un Groupement d'Intérêt Economique (regroupant à l'origine 4 pépiniéristes forestiers et l'INRA) qui utilise *L. bicolor* S238N pour l'inoculation en pépinière de plants de Douglas certifiés. L'inoculation par cette souche induit un gain de croissance important de l'hôte en pépinière puis plusieurs année après la transplantation en forêt (Le Tacon *et al.*, 1992 ; Selosse *et al.*, 2000).

## 1.2. Le peuplier

D'origine fort ancienne parmi les Angiospermes, cet arbre de grande taille (20 à 30 mètres) de la famille des *Salicaceae* a une croissance rapide et une longévité réduite (70 à 80 ans). Caduques, alternes, simples et palmatopennées, les feuilles de peuplier sont grandes à long pétiole. Les fleurs, presque toujours dioïques, (portées par des individus mâles ou femelles distincts), sont groupées en châtons pendants et apparaissent au printemps, avant les feuilles (Ying & Bagley, 1976). Les fruits sont des capsules arrondies laissant échapper à maturité des petites graines couvertes de poils blancs à l'aspect cotonneux. C'est une essence de pleine lumière des régions froides et tempérées de l'hémisphère Nord. Le peuplier a une nette préférence pour les sols frais, profonds et bien drainés. Il croît sur les terrains humides voire temporairement inondés.

Le genre *Populus* comprend plusieurs sections botaniques, dont le nombre, comme celui des espèces qui le composent, varie en fonction des époques et des classificateurs (Eckenwalder, 1996). Cinq sections sont le plus souvent reconnues : Leuce (peupliers blancs et peupliers trembles), Tacamahaca (dont *P. trichocarpa*), Leucoïde, Turanga et Aigeros ou peupliers noirs (dont *P. deltoides*). La majorité des espèces sont originaires d'Amérique

du Nord où les peupliers sont nommés *poplars*, mais aussi et plus spécifiquement *aspens* pour les peupliers trembles et *cottonwoods* pour les baumiers et les liards (termes désignant respectivement les peupliers producteurs de résines aromatiques utilisées dans les baumes et ceux producteurs de jeunes tiges flexibles utilisées en vannerie) en raison de leurs graines cotonneuses. D'autres espèces sont originaires d'Asie et d'Afrique du Nord. Le peuplier ne pousse pas à l'état spontané dans l'hémisphère Sud mais il peut y être cultivé aux latitudes symétriques de l'hémisphère Nord (ex. Chili). On distingue trois espèces à l'état spontané en France: le peuplier noir (*Populus nigra*), le peuplier tremble (*Populus tremula*) et le peuplier blanc (*Populus alba*) (Figure 6). Mais de nombreuses variétés issues de l'hybridation entre les peupliers européens et américains (euraméricains) et entre les peupliers américains (interaméricains) sont cultivées.


**Figure 6 : (A) *P. nigra* en conditions naturelles dans la vallée de la Durance (France) ; (B) peupleraie artificielle de Beaupré à Mirabeau (France) (photographies P. Frey).**

Cette essence produit un bois tendre assez résistant, adhérant et peu fissile. Se sciant difficilement, il est très valorisé en déroulage.

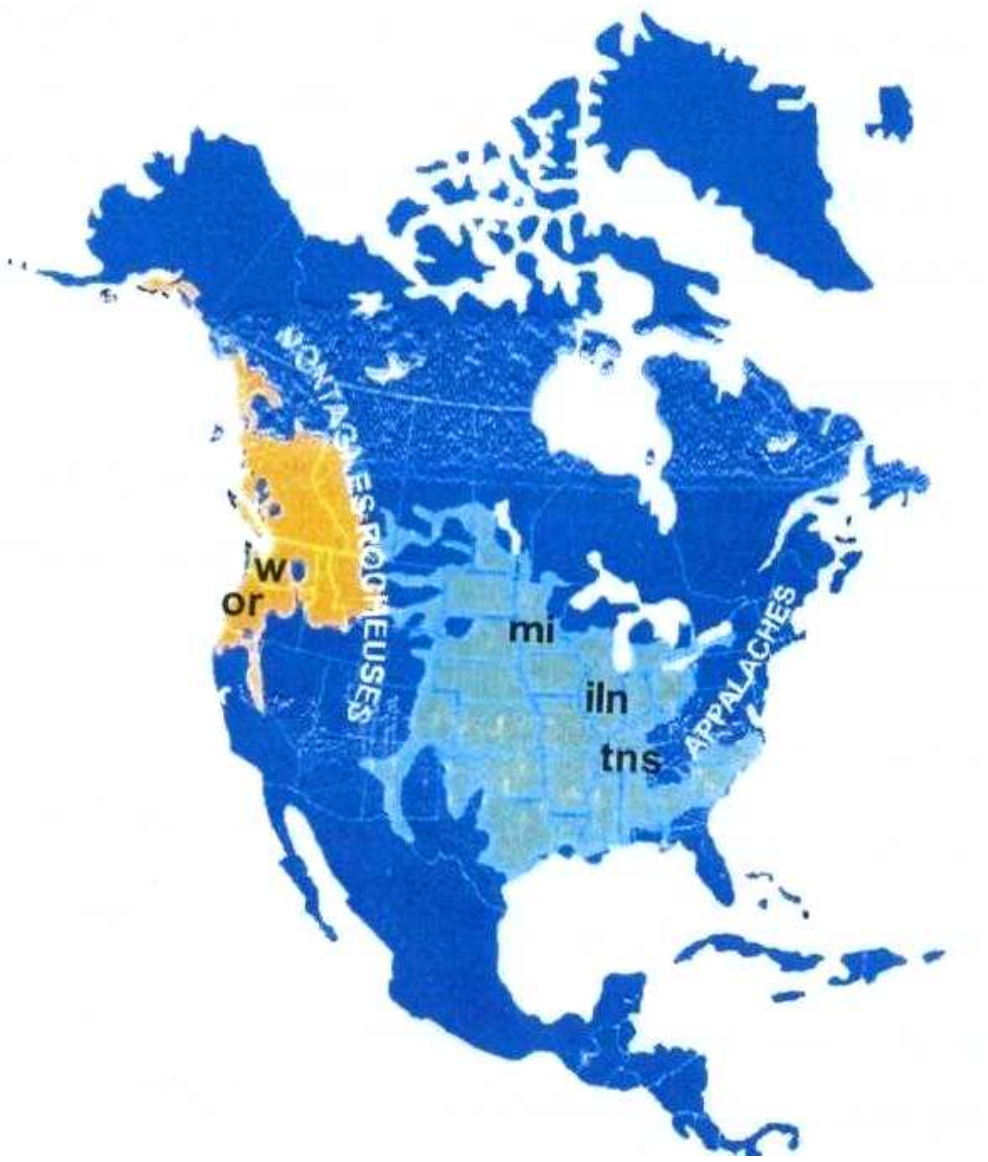



**Figure 7 : Aires de distribution naturelles de *Populus trichocarpa* (orange) et *Populus deltoides* (vert) (d'après Burn *et al.*, 1990).**

Il est surtout employé pour confectionner des emballages légers, des allumettes, en papeterie et pour la menuiserie courante (Dulbecco *et al.*, 1995 ; Viart, 1999). Il peut aussi être utilisé aux Etats-Unis pour dépolluer, valoriser et recoloniser d'anciens sites miniers en association avec des champignons mycorhiziens (Williams & Johnston, 1984 ; Shepperd *et al.*, 2001). Une nouvelle utilisation est envisagée depuis quelque temps : la production de biomasse par des taillis à courte révolution. Cette biomasse pourrait être valorisée directement à des fins énergétiques. La cellulose pourrait aussi être utilisée pour la production de carburant de deuxième génération (bioéthanol). Ainsi, basé sur les grandes possibilités de rendement en biomasse du peuplier, le récent projet ENERGYPOPLAR a pour objectif, la création de clones de peuplier à faible teneur en lignines. Une approche pluridisciplinaire devrait permettre de créer des peupliers à haut rendement et à bois à forte teneur en cellulose.

### 1.2.1. *Les symbioses chez le peuplier*

Le peuplier forme des associations « tripartite » tant avec des champignons ectomycorhiziens (ECMF) qu'avec des champignons endomycorhiziens à arbuscules (AMF) (Lodge, 2000 ; Debellis *et al.,* 2006 ; Krpata *et al.*, 2008 ; Newcombe *et al.*, 2009). Le peuplier peut être associé à plus d'une soixantaine de champignons etcomycorhiziens (Cripps *et al.*, 1993). La communauté fongique du peuplier est en effet très diversifiée, riche en basidiomycètes (plus de 43 espèces) et en ascomycètes. Elle est dominée par un ascomycète asexué *Cenococcum geophilum.* Les *Thelephoraceae* (*Tomentella spp.*) et les clitocybe laqués (*Laccaria spp*.) sont aussi bien représentés (Kaldorf *et al.,* 2004 ; Krpata *et al.*, 2008). Cependant les conditions environnementales, particulièrement l'humidité du sol, ont une grande influence sur le type de colonisation par les AMF et les ECMF. Une

modification d'équilibre entre AM et ECM a été observée lors d'inondation (Piotrowski *et al.*, 2008).

### 1.2.2. *Deux espèces américaines objet de cette étude : Populus deltoides & Populus trichocarpa*

Toutes deux originaires d'Amérique du Nord, l'une de l'Est (*Populus deltoides* Bartr. Ex Marsh.), l'autre de l'Ouest (*Populus trichocarpa* Torr. & A. Gray), elles occupent des aires naturelles très vastes allant des forêts humides côtières à des steppes arides continentales. Ainsi une très grande diversité en réponse à des conditions climatiques très variées peut être exprimée, permettant d'observer une importante variabilité à une faible échelle géographique (Figure 7 ; Burns *et al.*, 1990).

#### 1.2.2.1. *Populus deltoides*

Cette espèce est communément appelée *Southern Cottonwood*. Elle possède des feuilles en forme de delta majuscule au limbe denté et au pétiole aplati des deux côtés. Ses bourgeons sont légèrement résineux et saillants. *P. deltoides* s'étend de la côte Est des Etats-Unis aux pieds des Rocheuses et du Golfe du Mexique jusqu'au Canada. Ces peupliers des basses plaines de l'Est américain figurent parmi les plus grandes espèces qui poussent à l'Est des Rocheuses ; on rapporte des hauteurs allant jusqu'à 58 m et des diamètres jusqu'à 180 cm (Burdon, 2001). Cependant cette espèce exprime une forte sensibilité aux rouilles fongiques foliaires à *Melampsora medusae* (Nelson & Tauer, 1987).

#### 1.2.2.2. *Populus trichocarpa*

Cette espèce aussi appelée *Black Cottonwood* pousse dans les sites humides à l'Ouest des Rocheuses. Son aire de répartition naturelle s'étend au Nord jusqu'à l'Ile de Kodak dans le Golf de l'Alaska et jusqu'au Sud des

montagnes californiennes (Burdon, 2001). C'est un peuplier baumier caractérisé par de gros bourgeons très odorants et des feuilles ovoïdes. Cette espèce est beaucoup moins sensible que *P. deltoides* aux rouilles foliaires à *Melampsora medusae* (Ying & Bagley, 1976 ; Nelson & Tauer, 1987).

#### 1.2.2.3 *Les hybrides interaméricains entre P. deltoides et P. trichocarpa*

La domestication du peuplier repose sur deux processus naturels et fréquents chez cette essence que sont l'hybridation interspécifique et la multiplication végétative. Des copies conformes ou clones de plants de peuplier sont produits naturellement par des phénomènes de drageonnage ou de dispersion de fragments de tiges s'enracinant facilement. De même, l'hybridation interspécifique peut se produire naturellement entre espèces d'une même section botanique mais également entre certaines espèces de sections différentes (Eckenwalder, 1996). Les espèces du genre *Populus* étant diploïdes (2n=38), tous leurs hybrides sont fertiles.

Cependant, l'hybridation interspécifique maîtrisée peut permettre de combiner au sein d'un unique clone des caractères d'intérêt présents chez les deux espèces parentes. Les premiers hybrides euraméricains ont été introduits en Europe au dix-huitième siècle (1752 au Jardin Botanique de Nancy, Pourtet, 1951). Les premiers travaux d'hybridation interspécifique contrôlée sont attribués à Henry (Thielges, 1985), qui était convaincu des gains en croissance et en résistance aux maladies que l'on pouvait obtenir chez les arbres forestiers par cette voie. En 1912, deux descendances hybrides étaient obtenues dont une était le fruit d'une hybridation *P. deltoides* x *P. trichocarpa*. Dès cette époque, des individus intéressants ont été sélectionnés et clonés à partir de la génération F1.

Les hybrides interaméricains entre *P. deltoides* et *P. trichocarpa* possèdent d'exceptionnelles qualités de croissance dont on a démontré qu'elles sont issues de l'association de caractères hérités des deux espèces parentales (Stettler & Fenn, 1988 ; Stettler *et al.*, 1996).

Malheureusement, l'hybridation interspécifique n'a pas pour unique conséquence de combiner au sein d'une même plante des caractères d'intérêt hérités des deux espèces parentales. En effet, elle peut impliquer la rupture d'associations allèliques parentales et conduire à la perte de caractères d'intérêt :

- Soit par la faible probabilité de transmission ou de reconstitution chez les hybrides de caractères gouvernés par de nombreux gènes ;
- Soit par la ségrégation chez les hybrides de caractères gouvernés par un faible nombre de gènes à l'état hétérozygote.

Ce risque a été ignoré et de mauvaises stratégies de sélection et de modes de cultures intensifs l'ont amplifié, conduisant ces champions de la croissance qu'étaient les hybrides interaméricains entre *P. deltoides* et *P. trichocarpa* à subir aujourd'hui les attaques de rouille foliaire à *Melampsora larici-populina* jamais connues par leurs parents (Lefèvre *et al.*, 1998).

Nous avons utilisés pour ce travail, les hybrides interaméricains issus de la famille 54B détaillée dans le chapitre 2. Matériel et méthodes.

## 1.3. Les ressources génomiques et transcriptomiques

Des lacunes nombreuses subsistent dans nos connaissances de la biologie et de l'écologie des mycorhizes et des mécanismes contrôlant les interactions plantes-champignons, ECM, en particulier (Smith & Read, 2008). Pendant la dernière décennie, les recherches sur ces interactions ont été particulièrement

focalisées sur les échanges d'éléments nutritifs et de signaux de pré-mycorhization, utilisant des approches de physiologie moléculaire (Buscot *et al.*, 2000 ; Martin, 2006 ; Nehls *et al.*, 2007). Depuis 2005, l'essor des moyens techniques et la formation de consortia internationaux ont permis de développer des approches de génomique et de transcriptomique afin d'acquérir des bases de données permettant des études organisationnelle, structurale et fonctionnelle des génomes, clés de la compréhension des interactions plantes-champignons. Ainsi en 2009, la séquence de plus de 40 génomes d'espèces fongiques, représentatives des grands groupes taxonomiques et écologiques (saprophytes, symbiotes, endophytes, pathogènes de plantes ou d'animaux) (Xu *et al.*, 2006 ; Martin et Selosse, 2008) est disponible. Il en est de même en ce qui concerne les génomes de plantes (Jackson *et al.*, 2006).

### 1.3.1. *Laccaria bicolor*

Dans le cadre du projet « Populus Community Sequencing » (Martin et *al.*, 2004), le Joint Genome Institute (JGI) a réalisé le séquençage shotgun 10 X de la souche monosporale monocaryotique H82 de *Laccaria bicolor* S238N fournie par l'UMR 1136 IaM de l'INRA-Nancy. En février 2005, 878000 séquences étaient disponibles et déposées au NCBI Trace Archive : http://www.ncbi.nlm.gov/dbGSS). L'assemblage initial a été effectué à l'aide de l'algorithme assembleur JAZZ (Chapman *et al.*, données non publiées) qui a généré environ 600 scaffolds (portion de séquence génomique construite à partir d'un ensemble de plus petites séquences) pour une taille globale de 64,69 Mb dont 56,07 Mb appartenant à l'assemblage et 8,6 Mb de séquences correspondant vraisemblablement aux discontinuités. Un deuxième assemblage généré à l'aide de l'algorithme assembleur ARACHNE (Batzoglou *et al.*, 2002) fût ensuite disponible, générant des fragments

génomiques de plus grandes tailles (scaffolds) et de composition différentes au premier. Cette amélioration de l'assemblage génomique s'explique par une moindre sensibilité de l'algorithme ARACHNE aux séquences répétées, telles que les transposons. L'annotation automatique réalisée à l'aide d'algorithmes différents et complémentaires (FGENESH, EUGENE, GENEWISE, TWINSCAN) a conduit à l'identification de 20400 gènes codant des protéines. L'assemblage et l'annotation structurale automatique (Gènes, éléments transposables,…) sont les premières phases du décryptage, ils sont suivis d'une phase de curation et d'annotation manuelle. Les ressources disponibles sont actuellement les séquences assemblées des génomes nucléaire et mitochondrial (http://shake.jgi-psf.org/Lacbi1/Lacbi1.download.html) ; plusieurs banques d'ADNc et environ >150000 ESTs (Expressed Sequence Tags ou fragments de séquence exprimée: http://mycor.nancy.inra.fr/LaccariaDB/index.html).

Cependant dans de nombreuses régions, l'assemblage est difficile (régions télomèriques, transposons, absence de séquences chevauchantes) et comme nous l'avons vu ci-dessus, des fragments génomiques de tailles et de compositions différentes ont été générés par les algorithmes JAZZ et ARACHNE (Figure 8 et Figure 9). Ceci justifie dans un premier temps la construction de la carte génétique de *L. bicolor* S238N afin de faciliter et de corriger l'agencement des différents segments de séquence grâce à des marqueurs moléculaires (microsatellites et single nucleotide polymorphism SNP).

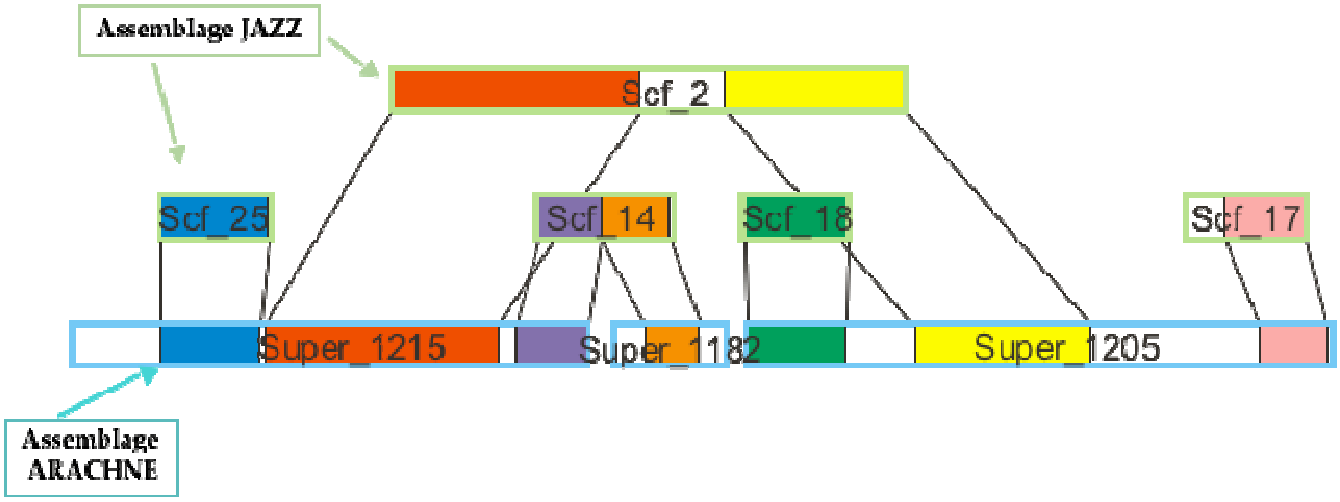



**Figure 8 : Comparaison des assemblages JAZZ et ARACHNE de la séquence génomique v1 de *L. bicolor* S238N.** Cette comparaison a été réalisée par Jan Wuyst (VIB - Université de Gand) à l'aide du logiciel ADHoRe qui détecte les régions homologues. On observe une différence de taille et de composition entre les scaffolds (Scf) générés par JAZZ et les supercontigs (Super) généré par ARACHNE.

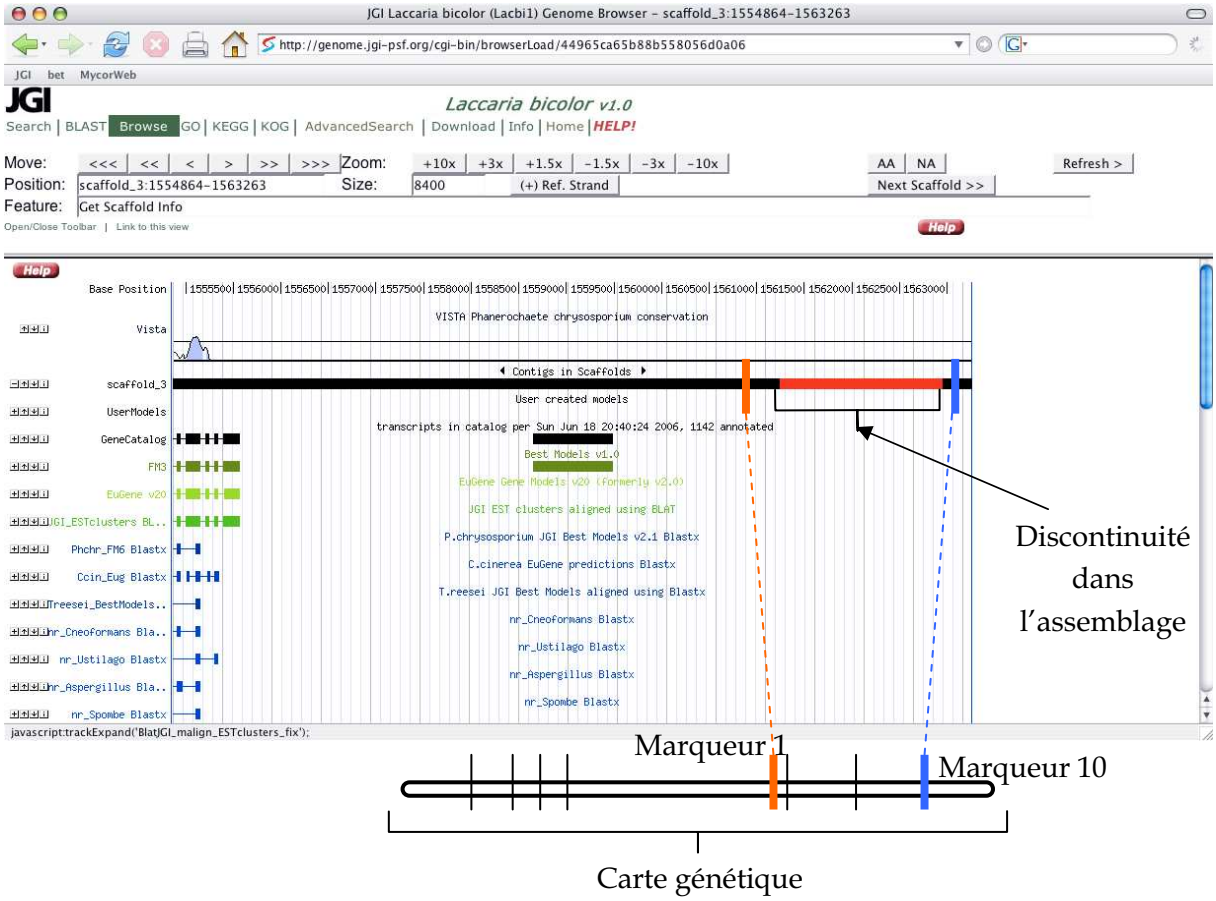



**Figure 9 : Exemple de discontinuité dans l'assemblage du scaffold 3 de la séquence génomique v1 de *L. bicolor* S238N et utilité de la carte génétique.**

### 1.3.2. *Le génome du peuplier*

Après l'arabette des dames (*Arabidospis thaliana*) et le riz (*Oryza sativa*), le peuplier est la troisième plante et le premier arbre séquencé (Tuskan *et al.*, 2006). Le Joint Genome Institute (JGI) a réalisé le séquençage shotgun 7,5 X du génome nucléaire du génotype femelle *P. trichocarpa* Nisqually-1 ainsi que des génomes chloroplastique et mitochondrial (Tuskan *et al.*, 2006). D'une taille de 485 Mb, le génome nucléaire est assemblé en 2447 scaffolds comprenant 420 Mb, 2,1 Mb correspondant aux discontinuités. Les génomes chloroplastique et mitochondrial sont assemblés circulairement avec une taille respective de 157 et 803 kb. De récents événements de duplication ont été mis en évidence dans 92 % génome (Figure 9). Une carte génétique d'une résolution de 5 cM (centiMorgan) permet d'ancrer physiquement l'assemblage génomique des 19 chromosomes à l'aide de 356 marqueurs microsatellites (Yin *et al.* 2004, Kelleher *et al.*, 2007). La plupart des 55793 gènes codant des protéines identifiés ont été ancrés à la carte génétique à l'aide de marqueurs microsatellites (Drost *et al.*, 2009). Des puces à ADN « génome entier » conçues par la société NimbleGen Systems on été développées à partir des modèles de gènes prédits (Tuskan *et al.*, 2006). A ceci, s'ajoutent diverses cartes génétiques plus ou moins saturées réalisées sur d'autres espèces de peupliers et permettant de cartographier une douzaine de QTL (Quantitative Trait Loci: loci impliqués dans la variabilité d'un caractère) ; (Cervera *et al.,* 2001 ; Jorge *et al.*, 2005 ; Street *et al.*, 2006) rassemblés dans la base de données PopGenIE (Populus Genome Integrative Explorer ; Sjödin *et al.*, 2009).

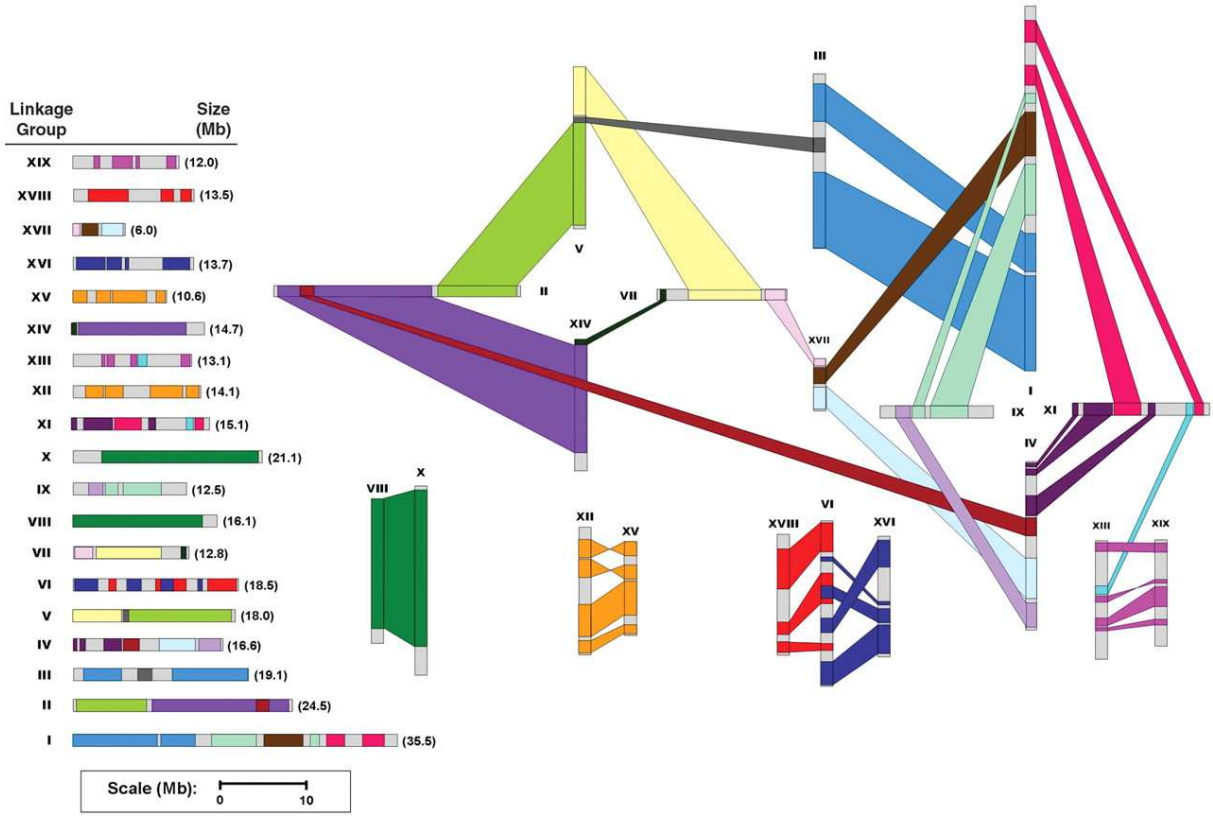



**Figure 9 : Illustration des évènements de duplication mis en évidence dans le génome du peuplier (d'après Tuskan *et al.*, 2006).** L'utilisation de couleurs communes sur des chromosomes différents référence les bocs de séquence homologue résultant de duplication.

## 1.4. Le développement de la symbiose ectomycorhizienne

La reconnaissance des deux partenaires est indispensable à la mise en place de la symbiose. Côté plante, d'après Horan & Chilvers (1990) la formation des ectomycorhizes serait dépendante d'une stimulation émanant de l'apex racinaire. Ainsi, les propagules des champignons mycorhiziens répondraient à une stimulation par les exsudats racinaires diversifiés en composés : flavonoïdes, diterpènes, hormones et nutriments variés (rutine, zéatine) ; (thèse de doctorat, Hubert Lagrange). Parmi ces composés actifs des exsudats, l'acide abiétique identifiée chez *Pinus sylvestis*, stimule la germination de spores de plusieurs espèces mais seulement du genre *Suillus*,

ce qui suggère une spécificité d'hôte (Fries, 1987). D'autre part, les exsudats racinaires de l'eucalyptus (dont la zéatine) provoquent une accumulation d'hypaphorine fongique chez *Pisolitus tinctorius* (Beguiristain *et al.*, 1995). Cette hypaphorine (composé indolique à activité anti-auxinique) a comme effet spécifique de diminuer le taux d'élongation des poils absorbants sans affecter le développement des racines (Ditengou *et al.*, 2000 ; Kawano *et al.*, 2001 ; Ditengou *et al.*, 2003). Côté fongique, d'après les travaux de Gay *et al.* (1992) et de Debaud *et al.* (1997), les champignons sont aussi capables de synthétiser de l'auxine ou acide indole-3-acétique (AIA), stimulant ainsi la croissance des racines latérales de la plante-hôte. L'AIA intervient aussi dans les processus d'adhésion du mycélium à la racine (Karabaghli-Degron *et al.*, 1998 ; Rincon *et al.*, 2001, 2003) et dans le développement du réseau de Hartig (Gay *et al.*, 1992). De plus, lors de ces premières étapes de reconnaissances, les champignons mycorhiziens provoquent des réactions semblables à la réponse hypersensible aux pathogènes, impliquant alors des éliciteurs (Salzer *et al.*, 1997). Cependant, durant la formation des ectomycorhizes, la plante produirait des composés diminuant la concentration de ces éliciteurs, aboutissant à la suppression des réponses de défense et à l'établissement de la mycorhize (Salzer *et al.*, 1997). Ainsi, l'adhésion à la surface la racine de l'hôte et la formation rapide de l'interface peuvent débuter.

L'adhésion des hyphes à la surface racinaire est une étape clé permettant la différenciation des structures symbiotiques. Lors des interactions compatibles, le champignon ectomycorhizien s'attache aux cellules épidermiques grâce à un mucigel polysaccharidique et à la sécrétion de matériel fibrillaire contenant des glycoprotéines (Martin *et al.*, 2001). Au point de contact avec la plante, la sécrétion des polymères fibrillaires extracellulaires est augmentée. Leurs mécanismes de synthèse et leur fonction sont encore inconnus (Tagu & Martin, 1996). Après attachement, les

hyphes se ramifient et forment le manteau ectomycorhizien autour de la racine. Des protéines spécifiques de la symbiose ectomycorhizienne, appelées protéines SR (Symbiosis-Related, par analogie avec PR : Pathogenesis-Related) sont alors surexprimées :

- chez la plante, on trouve des enzymes surexprimées sous l'effet de l'auxine et des hydrophobines (Mankel *et al.*, 2002)

- chez le champignon, des petites protéines pariétales fongiques, les SRAPs (Symbiosis-Regulated Acidic Polypeptides) et les hydrophobines sont aussi surexprimée (Duplessis *et al.*, 2001 ; Tagu *et al.*, 2001). Les premières analyses de la séquence génomique de *L. bicolor* ont mis en évidence un ensemble d'effecteurs de type petites protéines sécrétées (SSPs) dont la fonction est encore inconnue et dont certaines sont seulement exprimées dans les tissus symbiotiques. Les transcrits les plus surexprimés dans les mycorhizes codent pour des petites protéines riches en cystéines, les MiSSPs (mycorrrhiza-induced small secreted proteins). Parmi ces MiSSPs, MiSSP 7 est sécrétée à l'interface entre l'hyphe fongique et le réseau de Hartig (Figure 10 ; Martin *et al.*, 2008 ; Martin & Nehls 2009).

### 1.4.1. *Un contrôle génétique des deux partenaires*

Bien que l'environnement joue un rôle prépondérant, le développement de la symbiose ectomycorhizienne est également sous le contrôle génétique des deux partenaires (Tagu *et al.*, 2001 ; Voiblet *et al.*, 2001). Les données acquises sur le transcriptome symbiotique suggère que ce contrôle s'exerce sur les processus de reconnaissance plante-hôte-champignon, la pénétration et l'infection des tissus racinaires, la morphologie de la racine et des tissus fongiques et la réorganisation du métabolisme des symbiotes (Martin *et al.*, 2001 ; Tagu *et al.*, 2002 ; Peter *et*

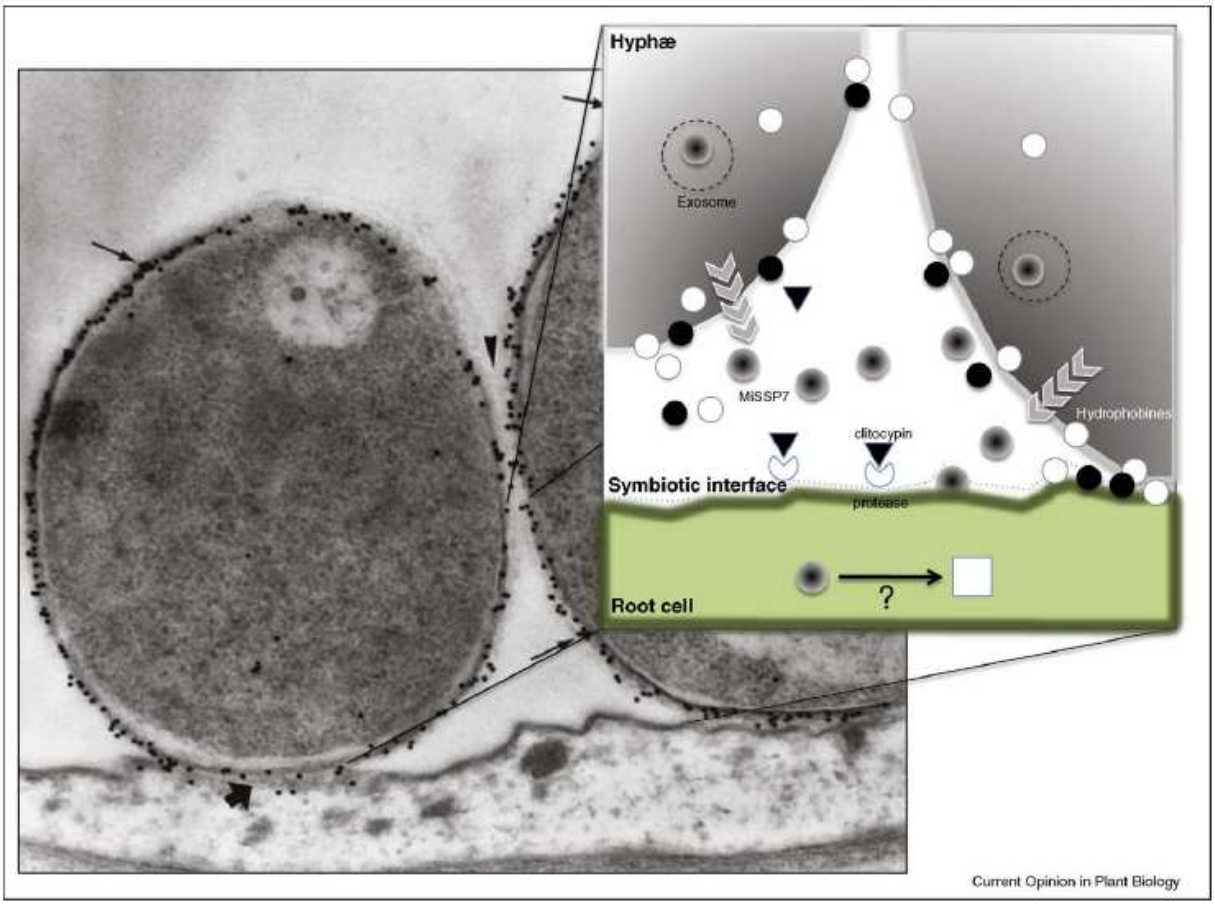



**Figure 10 : Sécrétion de protéines dans l'interface symbiotique apoplastique par le champignon ectomycorhizien (d'après Martin & Nehls, 2009).** La photo de gauche représente l'immuno-localisation (30000x) de la protéine SRAP32 dans l'apoplaste de *Pisolithus microcarpus* en contact avec la surface de racines d'*Eucalyptus globulus*. Cette protéine est régulée par la symbiose et contient un motif d'adhésion cellulaire RGD, et est localisée dans les parois cellulaires (flèches minces). L'image de droite représente le modèle hypothétique indiquant la présence de petites protéines sécrétées, induites par la formation de la mycorhize (cercles gris), comme MiSSP7 (Martin *et al.*, 2008). Il est suggéré que ces MiSSPs agissent réciproquement avec des protéines-cibles de la plante-hôte par des mécanismes inconnus. L'expression d'hydrophobines (cercles blancs) et de SRAP32 (cercles noirs) est induite par le développement ECM et est probablement impliquée dans l'attachement de l'hyphe à la surface de la racine. Les inhibiteurs de protéases et de cyclines sont fortement exprimés dans les tissus symbiotiques et permettraient d'inhiber les réactions de défense de la plante-hôte.

*al.*, 2003 ; Tagu *et al.*, 2003 ; Duplessis *et al.*, 2005 ; Martin *et al.*, 2008). Ces dernières années, la plupart des études conduites sur l'établissement de la symbiose ectomycorhizienne ont porté sur la caractérisation du transcriptome de plusieurs associations symbiotiques impliquant les champignons *Amanita muscaria, Laccaria bicolor, Paxillus involutus*, *Pisolithus tinctorius* et *Tuber borchii* (Voiblet *et al.*, 2000 ; Voiblet *et al.*, 2001 ; Lacourt *et al.*, 2002 ; Peter *et al.*, 2003 ; Sedda *et al.*, 2009 ; Martin & Tunlid 2009). Récapitulons brièvement les derniers acquis pour chaque partenaire. Le rôle de l'auxine dans la formation de la symbiose a suscité un intérêt particulier pendant plusieurs années. L'inoculation de peupliers par *L. bicolor* a confirmé le rôle de l'auxine fongique dans l'augmentation du nombre de racines chez la plante-hôte (Richter *et al.*, 2009). Cela a induit chez la plante, une augmentation du nombre de transcrits codant pour des transporteurs d'auxine et des gènes de réponse auxine/éthylène (Richter *et al.*, 2009). D'autre part, le développement d'ectomycorhizes d'*Amanita muscaria* n'est pas affecté par la transformation de peupliers trembles (*P. tremula* x *P. tremuloides*), exprimant dans les racines, des gènes de synthèse d'auxine (Hampp *et al.*, 1996). L'expression Coordonnée de malate synthase et d'autres gènes de métabolisme des lipides (Acétyl-CoA acétyltransférase), suggère que des changements dans le métabolisme des lipides puisse être important dans le processus de pré-infection ainsi que dans le transfert et l'utilisation du carbone chez le champignon (Hiremath *et al.*, 2006). Il a aussi été suggéré que les gènes d'hémoglobine peuvent jouer un rôle pendant le développement de l'ectomycorhize chez des hybrides de peuplier tremble (*P. tremula* x *P. tremuloides*) (Jokipii *et al.*, 2008).

L'expression différentielle de gènes dans des ectomycorhizes de peuplier a révélé des surexpressions de gènes codant aussi bien pour des protéines impliquées dans des réponses au stress ou de défense que pour des protéines impliquées dans la modification de la paroi cellulaire (Kohler *et al.*, données

non publiées). Les premiers profils de transcription de *L. bicolor* en interaction avec des racines de peuplier ont révélé une douzaine de gènes surexprimés dans les tissus symbiotiques (Martin *et al.,* 2008). Comme nous le disions ci-avant, les principales surexpressions détectées dans les mycorhizes correspondent à des gènes codant pour des petites protéines sécrétées riches en cystéine, dont MiSSP 7 (Figure 10 ; Martin *et al.,* 2008 ; Martin & Nehls, 2009).

## 1.5. Objectifs et organisation

Comme nous l'avons montré dans les parties précédentes, la compréhension des mécanismes moléculaires induisant et accompagnant l'établissement de la symbiose ectomycorhizienne constitue un objectif majeur des programmes de recherche de l'UMR 1136 sur les interactions plantes-champignons. L'essor technologique et la mutualisation des compétences de ces dernières années permettent d'acquérir de plus en plus de données essentielles à cette compréhension. C'est justement à la confluence des projets de séquençage des génomes de deux partenaires symbiotiques que s'inscrit cette thèse. Son objectif est de contribuer au développement et à l'exploitation des données génomiques et transcriptomiques du symbiote fongique *L. bicolor* et à la connaissance des interactions entre *L. bicolor* et le peuplier, en identifiant les déterminants génétiques contrôlant l'interaction.

Ce travail est constitué de quatre parties : 1. Introduction, 2. Matériel et méthodes, 3. Résultats, 4. Conclusions et perspectives. La première partie du chapitre 3 (Résultats), composée de cinq articles, est consacrée **à la caractérisation et au décryptage du génome de *L. bicolor* S238N**. Le premier article, collectif, fait la synthèse des connaissances acquises en 2008 sur le

génome de *L. bicolor* S*238*N. Les trois articles suivants ont trait au travail d'annotation auquel j'ai participé :

L'article 2 **Pattern of simple sequence repeats distribution in the ectomycorrhizal fungus *Laccaria bicolor* genome (Jessy Labbé, Claude Murat, Emmanuelle Morin, Francis Martin and François Le Tacon),** s'attache à caractériser les microsatellites ;

L'article 3 **Characterization of Transposable Elements in the *Laccaria bicolor* Genome (Jessy Labbé, Claude Murat, Benoit Hiselberger, Emmanuelle Morin, Marie-Pierre Oudot-Le Secq, Jan Wuyst, Igor Grigoriev, Pierre Rouzé, Hadi Quesneville, François Le Tacon and Francis Martin),** se propose de caractériser les éléments transposables ;

L'article 4 **Gene organization of the mating type regions in the ectomycorrhizal fungus *Laccaria bicolor* reveals distinct evolution between the two mating type loci (Hélène Niculita-Hirzel, Jessy Labbé, Annegret Kohler, François le Tacon, Francis Martin, Ian R. Sanders and Ursula Kües),** caractérise les régions de compatibilité sexuelle ;

Enfin l'article 5 **A genetic linkage map for the ectomycorrhizal fungus *Laccaria bicolor* and its alignment to the whole-genome sequence assemblies (J. Labbé, X. Zhang, T. Yin, J. Schmutz, J. Grimwood, F. Martin, G. A. Tuskan and F. Le Tacon)** est consacré à la construction de la carte génétique de *L. bicolor*.

La seconde partie du chapitre 3 (Résultats), composée des deux derniers articles, est consacrée **à la recherche des gènes impliqués chez les deux partenaires *L. bicolor* S238N et le peuplier.**

L'article 6 **Genetic variability of ectomycorrhizal development and functionning in an interspecific F1 poplar cross inoculated with *Laccaria bicolor* (Courty PE, Labbé J, Marçais B, Bastien C, Churin JL, Garbaye J, Martin F, Le Tacon F)** est consacré à l'étude de la variabilité génétique de la mycorhization dans une descendance interspécifique F1 de *P. deltoides* et de *P. trichocarpa* ;

Dans l'article 7, **Identification of quantitative trait loci affecting ectomycorrhizal symbiosis in an interspecific F1 poplar cross and differential expression of genes in ectomycorrhizas of the two parents*: Populus deltoides* and *Populus trichocarpa* (Jessy Labbé, Véronique Jorge, Annegret Kohler, Patrice Vion, Benoît Marçais, Catherine Bastien, Gerald A. Tuskan, Francis Martin and François Le Tacon),** nous avons tenté d'identifier les régions du génome du peuplier impliquées dans l'établissement de l'ectomycorhize et à la quantification de leurs effets par une stratégie de détection de QTL dans la même descendance et une stratégie d'expression différentielle de gènes chez les deux parents.

# 2. Matériel et méthodes

## 2.1. *Matériel fongique et conditions de culture*

La souche dicaryotique (diploïde) de *L. bicolor* S238N est utilisée pour réaliser ce travail, ainsi que 111 souches monosporales homocaryotiques (haploïdes). Ces souches monocaryotiques sont isolées à partir des basidiospores de carpophores de *Laccaria bicolor* S238N, d'abord en 1988 (49 carpophores) puis de 2004 à 2006 (62 carpophores). Ces carpophores ont été cultivés et récoltés en serre ou en pépinière sur substrat artificiel au pied de semis de Douglas préalablement inoculé avec la souche dicaryotique S238N. Les spores sont mises à germer, en conditions stériles, suivant la méthode décrite par Fries (1983).

Le mycélium végétatif haploïde est cultivé en boîte de pétri (∅ 90 mm) sur du milieu de Pachlewski (P5 ; Pachlewski and Pachlewska 1974). Le milieu de culture est recouvert d'une feuille de cellophane préalablement lavée avec une solution d'EDTA 1‰ portée à ébullition, puis rincée à l'eau déionisée (MilliQ, Millipore) et autoclavée. La maintenance des thalles fongiques est assurée par repiquages successifs tous les six mois. La croissance est alors initiée à partir de trois implants mycéliens par boîte et se déroule à 25°C à l'obscurité. On observe une variabilité phénotypique entre souche homocaryotique (Figure 11).

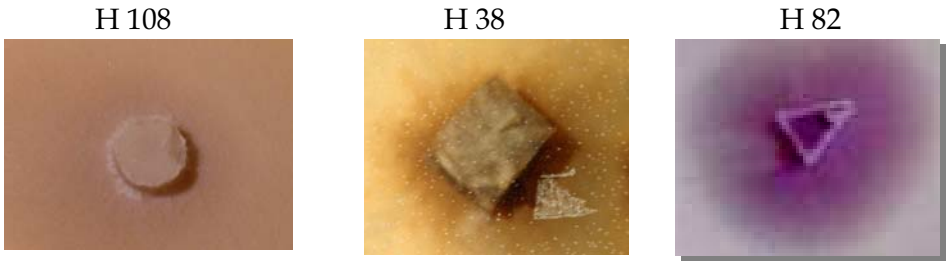



**Figure 11 : Mycélium végétatif de trois homocaryons issus de *L. bicolor* S238N.**
Ces photographies illustrent la diversité phénotypique entre les homocaryons issus de la souche dicaryotique *L. bicolor* S238N (photographies F. Rineau).

### 2.2. *Matériel végétal*

Les peupliers utilisés dans ce travail comme plante-hôte, appartiennent à la famille 54B (*P. deltoides* femelle 7302862 x *P. trichocarpa* mâle 101-74) fournie par Catherine Bastien et ses collègues de l'UAGPF (centre INRA d'Orléans). Cette famille constitue un modèle d'étude, non pas du fait de sa représentativité vis-à-vis des hybrides interaméricains ou d'un comportement particulier, mais surtout du fait du grand nombre de génotypes obtenus et conservés (≥ 1300). Cet effectif a permis le développement de marquage moléculaire, de cartographie génétique et de détection de QTL (Jorge *et al.*, 2005).

Cette famille est utilisée dans notre recherche des QTL de mycorhization du fait de son utilisation dans les travaux de Tagu *et al.* (2005) qui ont mis en évidence la co-localisation d'un QTL de mycorhization avec ceux de la résistance à la rouille *Melampsora larici-populina* sur le chromosome XIX du peuplier.

L'ensemble des génotypes constituant le matériel d'étude est conservé en pépinière au centre INRA d'Orléans sous forme de pieds-mères recépés chaque hiver pour fournir des boutures ligneuses. Ces boutures nous sont

fournies par Catherine Bastien et sont ensuite stockées en chambre froide (+4°C) jusqu'à leur mise en culture. Les conditions d'utilisation lors des expérimentations sont précisées dans le paragraphe 2.4. Protocoles expérimentaux.

### 2.3. *Généralités et définitions des techniques employées*

#### 2.3.1. *Les étapes principales du séquençage d'un génome*

Les étapes principales du séquençage d'un génome sont : le séquençage *sensu stricto*, l'assemblage des séquences, l'annotation *ab initio* de la séquence assemblée à l'aide d'algorithmes, suivie de l'annotation fonctionnelle et de l'annotation expertisée.

Entre 2001 et 2008, une des principales techniques utilisées pour séquencer le génome d'un organisme était le séquençage en shotgun ou séquençage massif aléatoire (Figure 12 ; Linton *et al.*, 2001 ; Venter *et al.*, 2001). Le principe de ce séquençage est de couper aléatoirement, grâce à des enzymes de restriction ou par nébulisation, 8 à 10 copies du génome de l'organisme en fragments d'environ 2000, 10 000 et 40 000 paires de bases. Ces fragments de restriction sont clonés dans des vecteurs réplicatifs (plasmides ou cosmides). Pour un génome de 50 à 60 Mb, des centaines de milliers de ces fragments clonés sont séquencés sur 500 à 1000 pb à chaque extrémité à l'aide de séquenceurs automatiques multicapillaires. A titre d'exemple, le Joint Genome Institute (DOE) génère 1 Gb par mois grâce à plusieurs séquenceurs multicapillaires, tels le MegaBACE 4000 (http://www4.amershambiosciences.com/aptrix/upp01077.nsf/Content/autodna_megabace4000). Vient ensuite l'assemblage des centaines de milliers de séquences en clusters, puis en contigs (séquences chevauchantes). Ces contigs sont ensuite ancrés *in silico* sur des fragments génomiques de grande taille (BAC >100kb ou cosmides < 40kb). Dans de nombreuses régions,

l'assemblage est impossible ou de mauvaise qualité spécialement dans le cas de régions hautement répétitives, (télomères, transposons, absence de séquences chevauchantes). Le séquençage spécifique et ciblé de ces discontinuités fait partie du processus de finition qui prend actuellement plus de temps que la détermination de l'ébauche de séquence initiale du génome. C'est alors que la carte génétique de l'organisme à séquencer facilite cet assemblage génomique en confirmant ou en corrigeant l'agencement des contigs et scaffolds grâce à des marqueurs moléculaires.

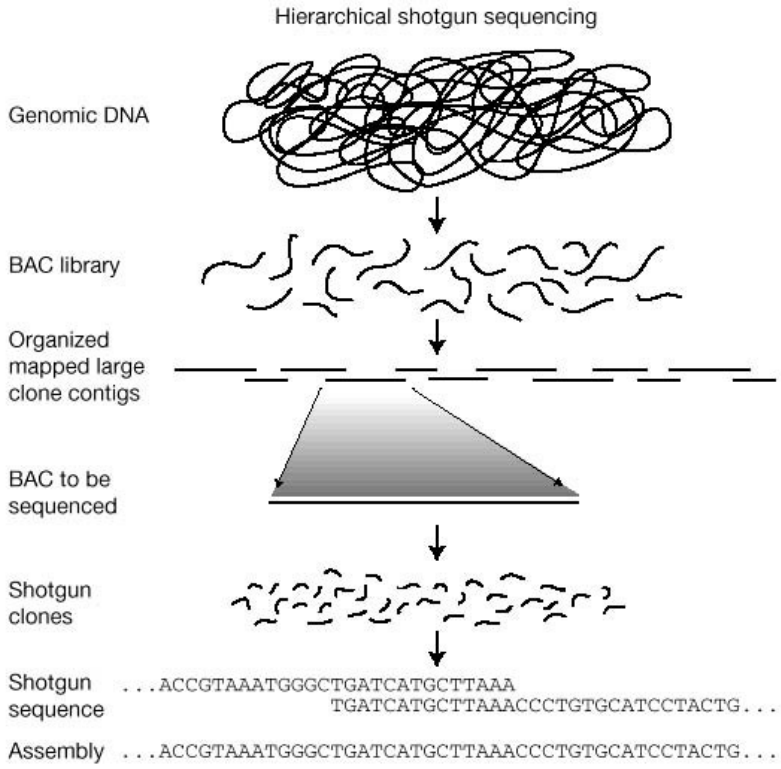



**Figure 12 : Principales étapes d'un séquençage massif aléatoire ou shotgun (d'après Linton *et al.*, 2001).**

### 2.3.2. *Cartographie génétique : principe, construction et intérêts*

Nous n'exposerons pas ici de façon approfondie l'ensemble des concepts et des techniques sur lesquels repose la cartographie génétique ainsi

que la détection de QTL qui lui est liée. Nous synthétiserons simplement la définition, la construction et les intérêts.

Une carte génétique se définit comme une représentation du génome sous la forme d'une succession de balises ou marqueurs, répartis le long de groupes de liaisons établis statistiquement.

Sa construction est basée sur l'étude de la ségrégation de marqueurs dans une descendance (comparaison des génotypes des descendants à ceux des parents). Au cours de la phase réductionnelle de la méiose, des évènements de recombinaison peuvent survenir lors de l'appariement des chromosomes homologues (crossing-over). La probabilité de recombinaison entre deux loci dépend de la distance physique qui les sépare : elle est d'autant plus faible que les deux loci sont physiquement proches. La liaison génétique (linkage) peut être ainsi estimée à partir de la fréquence des recombinaisons méiotiques. Le pourcentage observé de recombinaison entre deux loci permet d'accéder à la distance génétique (exprimée en centiMorgan, cM) entre deux loci. Le taux de recombinaison entre chaque paire de locus est estimé par l'analyse de la distribution des allèles au sein des gamètes ou de descendants. La précision d'une carte dépendra donc du nombre d'évènements méiotiques analysés. L'augmentation du nombre de marqueurs positionnés sur les cartes génétiques a pour conséquence de relier des groupes de liaisons indépendants jusqu'à ce qu'il y ait autant de groupes que de chromosomes. On parle alors de cartes saturées.

Sur le plan pratique, l'établissement d'une carte génétique nécessite de disposer d'abord d'une descendance en ségrégation issue d'un croisement par reproduction sexuée ou comme dans notre cas pour *L. bicolor*, d'une population d'individus haploïdes issu d'un individu diploïde. Cette population haploïde équivaut à une descendance issue de rétrocroisement

(de type backcross en anglais). Dans un deuxième temps, on recherche des marqueurs à hérédité mendélienne qui soient :

- polymorphes, c'est-à-dire montrant au moins un allèle différent entre les parents,
- Nombreux pour augmenter la couverture du génome et
- Co dominants, si possible, afin de distinguer les génotypes hétérozygotes des homozygotes.

Ainsi, les marqueurs génétiques peuvent correspondre à des caractéristiques phénotypiques associées à un locus responsable d'un caractère à transmission mendélienne. Néanmoins, les marqueurs phénotypiques peuvent dépendre du milieu et ne pas être utilisables pour l'établissement d'une carte génétique. De nos jours, on dispose de marqueurs morphologiques, de marqueurs biochimiques (isoenzymes, protéines) et de marqueurs moléculaires (ADN) avec un large éventail de choix : RFLPs, microsatellites, AFLPs, RAPDs, SNPs, etc… Il n'existe pas de hiérarchie absolue entre les différents types de marqueurs. Leur choix dépend des objectifs et de l'espèce étudiée.

Une fois l'ensemble des parents et des descendants typés (dans notre cas pour *L. bicolor*, population haploïde et souche dicaryotique) à l'aide des marqueurs génétiques, la construction d'une carte génétique comporte plusieurs étapes nécessitant l'utilisation d'outils d'analyse statistique :

- pour vérifier la ségrégation,
- pour réunir les marqueurs en groupes de liaison,
- pour ordonner les marqueurs au sein des groupes de liaison et
- pour estimer les distances génétiques entre marqueurs.

Toutes ces étapes sont réalisées à l'aide de logiciels de cartographie tels que Mapmaker (Lander *et al.*, 1987) ou Joinmap v3 (VanOoijen et Voorrips, 2001), ce qui permet d'aboutir à une représentation graphique de la carte génétique.

Les cartes génétiques constituent des outils très puissants de comparaison des génomes. La cartographie comparée permet d'étudier en effet l'organisation et l'évolution des génomes en déterminant et en délimitant les régions chromosomiques homologues conservées chez plusieurs espèces d'un même genre, voire dans des genres phylogénétiquement proches, ex: synténie chez les Graminées (Devos et *al*., 1997).

Les cartes génétiques peuvent aussi être utilisées pour aider à l'assemblage de génomes complets lors de séquençage massif aléatoire, comme dans notre cas, ce qui justifie dans un premier temps l'établissement de la carte génétique de *L. bicolor*. Mais une des principales applications de la cartographie concerne la détection et la localisation de « gènes majeurs » qui peuvent être déterminées entièrement par un seul locus ou par plusieurs loci à effet partiel sur la variation d'un caractère que l'on appelle QTL (Cervera *et al.,* 2000) et justifiant dans un second temps la carte génétique de *L. bicolor*.

### 2.3.3. *Détection de QTL*

La détection de QTL est fondée sur la recherche de déséquilibres de liaison entre des loci caractérisés par des marqueurs moléculaires et des loci contrôlant les caractères quantitatifs étudiés. Pour cela, il faut les critères vu précédemment et nécessaires à la cartographie génétique auxquels s'ajoute la mesure du caractère quantitatif étudié pour chacun des individus de la descendance utilisée. Comme dans notre situation, la localisation précise de QTL les uns par rapport aux autres sur le génome de la plante-hôte nécessite donc de disposer de suffisamment de marqueurs moléculaires afin de saturer la carte génétique. Ce qui n'est pas le cas des cartes obtenues pour les deux parents de la famille 54, totalisant 33 groupes de liaison pour celle de *P. deltoides* et 35 pour celle de *P. trichocarpa* ; le nombre de groupe de liaison

attendu pour chacune de ces cartes à saturation est égal à 19, nombre de paires de chromosomes présentes chez le peuplier.

On distingue deux grands types de méthode de détection de QTL : des méthodes qui considèrent des marqueurs individuellement et des méthodes multi-marqueurs. La première approche consiste à rechercher une relation entre le génotype de chaque locus marqué et la valeur du caractère quantitatif considéré simplement par analyse de variance. L'avantage de cette approche est sa simplicité ; Ces analyses sont robustes aux écarts à la normalité et une carte saturée n'est pas nécessaire. Les méthodes multi-marqueurs présentent l'avantage de positionner les QTL de façon plus précise puisqu'elles permettent de les positionner dans les intervalles entre les marqueurs et d'estimer précisément leurs effets à cette position. Tout d'abord, les méthodes de cartographie d'intervalle (interval mapping) partent de l'hypothèse qu'il y a un QTL au plus dans l'intervalle entre deux marqueurs liés avec un taux de recombinaison r. Cette méthode repose sur le calcul de LOD score (logarithm of the odds ratio) dérivé des méthodes de maximum de vraisemblance. Un LOD se 2 signifie que la présence d'un QTL en un point donné est 100 fois plus probable que son absence, un LOD 3 qu'elle est 1000 fois plus probable, etc. Une courbe de LOD peut donc être tracée en fonction de la position sur un groupe de liaison donné. La position la plus probable d'un QTL étant indiquée lorsqu'elle dépasse un certain seuil.

Il est possible que plus d'un QTL soit présent ou que le QTL détecté résulte de la présence de deux QTL situés hors de l'intervalle étudié et agissant dans le même sens (on parle de QTL fantôme). D'autre part, la présence d'autre QTL hors de l'intervalle considéré peut résulter en un masquage de la présence de QTL à effet faible par la présence d'un QTL à

effet fort. La méthode CIM (composite interval mapping) permet de lever ce problème. Cette méthode combine cartographie d'intervalle et régression sur les marqueurs en utilisant les marqueurs liés à d'autre QTL présents dans le génome comme cofacteur dans le modèle de régression. Les données phénotypiques que nous avons cavons cherché à lier aux données de marquage moléculaire étaient constituées des moyennes clonales de 300 individus de la famille 54 pour la formation d'ectomycorhize chez le peuplier par *L. bicolor*. Nous avons réalisé une détection de QTL en combinant une approche d'interval mapping et de multiple interval mapping.

### 2.3.4. *Les puces à ADN*

Une puce à ADN (DNA microarray en anglais) est constituée de fragments d'ADN immobilisés sur un support solide selon une disposition ordonnée. Son fonctionnement repose sur le même principe que le Southern blot ou le Northern blot couramment utilisés pour détecter et quantifier la présence d'une séquence nucléique spécifique au sein d'un échantillon biologique complexe, par hybridation à une sonde de séquence complémentaire portant un marquage radioactif (Southern, 1975). Les puces à ADN ont permis d'étendre ce principe à la détection simultanée de milliers de séquences en parallèle, une puce comportant quelques centaines à plusieurs dizaines de milliers d'unités d'hybridation (spots en anglais) constituée d'un dépôt de fragments d'ADN ou d'oligonucléotides correspondant à des sondes de séquences données. L'hybridation de la puce avec un échantillon biologique, marqué par un radioélément ou par une molécule fluorescente, permet de détecter et de quantifier l'ensemble des cibles qu'il contient. On distingue plusieurs types de puces selon la densité des spots, le mode de fabrication, la nature des fragments fixés à la surface et les méthodes d'hybridation. Les deux technologies dominantes sont les

puces dites « spottées » (dépôt robotisé de produits de PCR ou de longs fragments oligonucléiques : spotted microarrays) et les puces à oligonucléotides synthétisés *in situ*. Les puces « génome entier » NimbleGen System de peuplier que nous avons utilisées dans ce travail de thèse sont du deuxième type. Ces puces permettent donc de mesurer le niveau d'expression relatif de chaque gène dans un échantillon biologique comparé à une référence. Plus le niveau d'expression est élevé, plus la quantité d'ARNm du gène est importante dans l'échantillon considéré. Classiquement, on considère que le niveau d'expression du gène est significativement plus élevé dans l'échantillon par rapport à la référence si le taux d'expression est supérieur à 2, et inversement qu'il est significativement réduit s'il est inférieur à 0,5. Ce seuil a été originellement établi en supposant que les taux d'expression ont une distribution aléatoire suivant la loi normale, auquel cas l'intervalle de confiance à 95% de la valeur du taux est de ± 1,96 (soit environ ± 2).

Nous avons utilisé cette approche de génomique fonctionnelle en comparant des mycorhizes des deux parents afin d'identifier et/ou de confirmer la régulation de gènes par la mycorhization dans les QTL détectés sur le génome du peuplier.

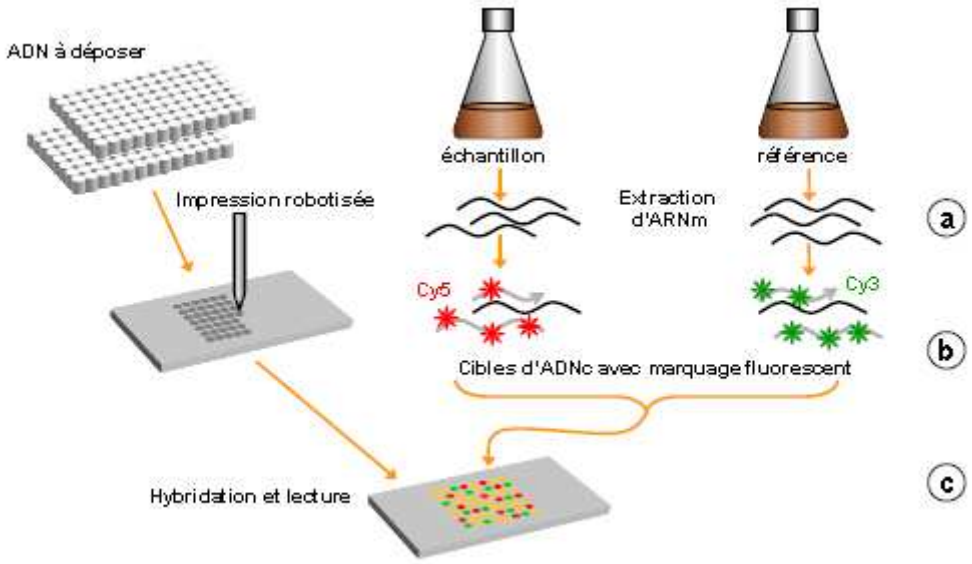



**Figure 13 : Principe général de l'analyse de l'expression transcriptionnelle sur une puce à ADN (d'après la thèse de doctorat de Waka Lin, 2004).** (a) L'ensemble des ARN messagers sont extraits des échantillons à analyser. (b) Une transcription reverse est réalisée en présence de nucléotides modifiés permettant de coupler un marqueur fluorescent. Des cibles d'ADN complémentaires (ADNc) représentatives de l'ensemble des gènes exprimés pour chaque échantillon sont ainsi obtenues. (c) Les deux échantillons sont marqués par des fluorochromes à spectres d'émission distincts (les plus couramment utilisées sont les carbocyanines Cy3 et Cy5) et hybridés simultanément sur une même puce. Au contact de la puce, les brins d'ADNc marqués s'apparient avec les sondes de séquence complémentaire sur la lame. La mesure de l'intensité du signal fluorescent émis sur chaque « spot » permet ainsi d'estimer le taux d'expression différentiel du gène correspondant.

### 2.4. *Protocoles expérimentaux*

#### 2.4.1. *Détection de QTL impliqués dans la mycorhization chez le peuplier*

Expérience 1 – Année 2007

Afin de quantifier l'établissement de la symbiose ectomycorhizienne chez le peuplier, nous avons inoculé plusieurs hybrides interaméricains de la famille 54B de l'INRA d'Orléans par la souche dicaryotique S238N de *L. bicolor*.

Nous avons premièrement réalisé un essai opératoire en utilisant 40 hybrides (liste en Annexe) issus du croisement *P. deltoides* x *P. trichocarpa* ainsi que les deux parents. Les micro-boutures des descendants et des deux

parents (un seul inter-noeud et un bourgeon) ont été mis à enraciner en juin 2007 dans des mini-conteneurs de 80 ml contenant un mélange de tourbe-vermiculite non désinfectée (1 :1 vol. vol., pH 5.5) en serre froide (15-25°C), pendant un mois. Un dispositif de type aspersion a permis un arrosage à l'eau à une fréquence de 10 fois une minute par 24h. En juillet 2007, nous avons ensuite transféré 8 micro-boutures enracinées par clone de peuplier dans des conteneurs de 1 litre contenant un substrat inerte de type attapulgite (Oil Dri US Special, Damolin, Denmark ; http://www.damolin.dk) non désinfecté et 10% d'inoculum tourbe vermiculite S238N (souche dicaryotique). Le dispositif expérimental a été réparti en 8 blocs en serre froide (15-25°C) que l'on déplaçait une fois par semaine dans un souci d'homogénéité. Après deux mois d'arrosage par aspersion (fréquence de 5 fois 30 secondes par 24h), un arrosage hebdomadaire avec une solution nutritive a permis de maintenir la vigueur des plants (Frey-Klett, 1997). Le dispositif expérimental a été démonté en novembre 2007, afin d'observée la variabilité phénotypique des différents génotypes de peuplier étudiés. Pour ce faire, le système racinaire de chaque plant a été coupé en morceau de 1 cm, à partir desquels, 100 fragments de racines ont été observés aléatoirement à la loupe binoculaire. Nous avons distingués les apex racinaires mycorhizés des non mycorhizés, afin de déduire le pourcentage de mycorhization du plant.

Le succès de cet essai nous a permis de confirmer la reproductibilité des travaux de Tagu *et al.* (2005) et de mettre en place la même expérimentation sur un plus grand nombre de génotypes de peupliers.

**Figure 14 : Récapitulatif des différentes étapes de la quantification de la mycorhization (photographies J. Labbé).** Les trois premières photos illustrent l'inoculation et le montage du dispositif expérimental, les deux suivantes le conditionnement et la croissance des plantules de peupliers (jusqu'à trois mois et demi). Enfin les six dernières photos illustrent le dépouiller de l'expérience.

Expérience 2 – Année 2008

Nous avons donc mis en place le même dispositif expérimental en mars 2008, en utilisant cette fois, 300 hybrides *P. deltoides* x *P. trichocarpa* (liste en Annexe), ainsi que les deux parents. Cependant, nous n'avons pas réalisé l'enracinement préalable des 8 micro-boutures par génotypes de peuplier ; l'enracinement s'est directement fait dans le mélange substrat inoculum décrit précédemment. Le protocole opératoire et les conditions utilisées sont identiques à celles décrites dans l'expérience 1 - 2007. Ce dispositif a été démonté en juillet et en août 2008. Le processus d'observation et de comptage des mycorhizes des 2416 plants au total s'est réalisé grâce au travail de 8 observateurs.

### 2.4.3. *Mesure des activités enzymatiques de mycorhizes et de racines non mycorhizées*

Parmi les 300 clones utilisés pour la détection de QTL, un échantillon de 40 clones (8 répétitions par clone), dont les deux parents, a été choisit de façon à couvrir toute la gamme de mycorhization par *L. bicolor*. Sur chacun des 320 peupliers mycorhizés, nous avons prélevé le système racinaire et mesuré, outre les paramètres liés à la croissance des plants issus de boutures, les activités enzymatiques suivantes sur les mycorhizes et les racines non mycorhizées (à partir d'un apex racinaire ; Courty *et al*., 2005) : laccase, phosphatase acide, $\beta$-glucosidase, glucuronidase xylosidase, cellobiohydrolase et N-acétyl-glucosaminidase.

# Chapitre 3. Résultats

## Première partie: Caractérisation et décryptage du génome de *L. bicolor*

## Article 1. The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis

**F. Martin, A. Aerts, D. Ahrén, A. Brun, E. G. J. Danchin, F. Duchaussoy, J. Gibon, A. Kohler, E. Lindquist, V. Pereda, A. Salamov, H. J. Shapiro, J. Wuyts, D. Blaudez, M. Buée, P. Brokstein, B. Canbäck, D. Cohen, P. E. Courty, P. M. Coutinho, C. Delaruelle, J. C. Detter, A. Deveau, S. DiFazio, S. Duplessis, L. Fraissinet-Tachet, E. Lucic, P. Frey-Klett, C. Fourrey, I. Feussner, G. Gay, J. Grimwood, P. J. Hoegger, P. Jain, S. Kilaru, J. Labbé, Y. C. Lin, V. Legué, F. Le Tacon, R. Marmeisse, D. Melayah, B. Montanini, M. Muratet, U. Nehls, H. Niculita-Hirzel, M. P. Oudot-Le Secq, M. Peter, H. Quesneville, B. Rajashekar, M. Reich, N. Rouhier, J. Schmutz, T. Yin, M. Chalot, B. Henrissat U. Kües, S. Lucas, Y. Van de Peer, G. K. Podila, A. Polle, P. J. Pukkila, P. M. Richardson, P.Rouzé, I. R. Sanders, J. E Stajich, A. Tunlid, G. Tuskan & I. V. Grigoriev**

**(publié dans *Nature*)**

**Cet article correspond à l'analyse du génome de *L. bicolor*, suite au séquençage complet de la souche haploïde S238N-H82. Ma contribution à ce travail est principalement illustrée par les figures intégrées dans la partie « supplementary information » et ici insérées à la fin de l'article. J'ai donc participé à l'analyse du génome, notamment à celles des répétitions en tandem et des éléments transposables, puis contribué à l'amélioration de son assemblage en initiant la construction de la carte génétique.**

LETTERS

# The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis

F. Martin[1], A. Aerts[2], D. Ahrén[3], A. Brun[1], E. G. J. Danchin[4], F. Duchaussoy[1], J. Gibon[1], A. Kohler[1], E. Lindquist[2], V. Pereda[1], A. Salamov[2], H. J. Shapiro[2], J. Wuyts[1,5], D. Blaudez[1], M. Buée[1], P. Brokstein[2], B. Canbäck[3], D. Cohen[1], P. E. Courty[1], P. M. Coutinho[4], C. Delaruelle[1], J. C. Detter[2], A. Deveau[1], S. DiFazio[6], S. Duplessis[1], L. Fraissinet-Tachet[8], E. Lucic[1], P. Frey-Klett[1], C. Fourrey[1], I. Feussner[7], G. Gay[8], J. Grimwood[9], P. J. Hoegger[10], P. Jain[11], S. Kilaru[10], J. Labbé[1], Y. C. Lin[5], V. Legué[1], F. Le Tacon[1], R. Marmeisse[8], D. Melayah[8], B. Montanini[1], M. Muratet[11], U. Nehls[12], H. Niculita-Hirzel[13], M. P. Oudot-Le Secq[1], M. Peter[1,14], H. Quesneville[15], B. Rajashekar[3], M. Reich[1,10], N. Rouhier[1], J. Schmutz[9], T. Yin[16], M. Chalot[1], B. Henrissat[4], U. Kües[10], S. Lucas[2], Y. Van de Peer[5], G. K. Podila[11], A. Polle[10], P. J. Pukkila[17], P. M. Richardson[2], P. Rouzé[5,18], I. R. Sanders[13], J. E. Stajich[19], A. Tunlid[3], G. Tuskan[16] & I. V. Grigoriev[2]

**Mycorrhizal symbioses—the union of roots and soil fungi—are universal in terrestrial ecosystems and may have been fundamental to land colonization by plants[1,2]. Boreal, temperate and montane forests all depend on ectomycorrhizae[1]. Identification of the primary factors that regulate symbiotic development and metabolic activity will therefore open the door to understanding the role of ectomycorrhizae in plant development and physiology, allowing the full ecological significance of this symbiosis to be explored. Here we report the genome sequence of the ectomycorrhizal basidiomycete *Laccaria bicolor* (Fig. 1) and highlight gene sets involved in rhizosphere colonization and symbiosis. This 65-megabase genome assembly contains ~20,000 predicted protein-encoding genes and a very large number of transposons and repeated sequences. We detected unexpected genomic features, most notably a battery of effector-type small secreted proteins (SSPs) with unknown function, several of which are only expressed in symbiotic tissues. The most highly expressed SSP accumulates in the proliferating hyphae colonizing the host root. The ectomycorrhizae-specific SSPs probably have a decisive role in the establishment of the symbiosis. The unexpected observation that the genome of *L. bicolor* lacks carbohydrate-active enzymes involved in degradation of plant cell walls, but maintains the ability to degrade non-plant cell wall polysaccharides, reveals the dual saprotrophic and biotrophic lifestyle of the mycorrhizal fungus that enables it to grow within both soil and living plant roots. The predicted gene inventory of the *L. bicolor* genome, therefore, points to previously unknown mechanisms of symbiosis operating in biotrophic mycorrhizal fungi. The availability of this genome provides an unparalleled opportunity to develop a deeper understanding of the processes by which symbionts interact with plants within their ecosystem to perform vital functions in the carbon and nitrogen cycles that are fundamental to sustainable plant productivity.**

The 65-megabase genome of *Laccaria bicolor* (Maire) P. D. Orton is the largest sequenced fungal genome published so far[3–7] (Table 1). Although no evidence for large-scale duplications was observed within the *L. bicolor* genome, tandem duplication occurred within multigene families (Supplementary Fig. 4). Transposable elements comprised a higher proportion (21%) than that identified in the other sequenced fungal genomes and may therefore account for the relatively large genome of *L. bicolor* (Supplementary Table 3). Approximately 20,000 protein-coding genes were identified by combined gene predictions (Supplementary Information Section 2). Expression of nearly 80% (~16,000) of the predicted genes was detected in free-living mycelium, ectomycorrhizal root tips or fruiting bodies (Supplementary Table 4) using NimbleGen custom-oligoarrays (Supplementary Information Section 9). Most genes are activated in almost all tissues, whereas other more specialized genes are only activated in some specific developmental stages, such as the free-living mycelium, ectomycorrhizae or the fruiting body (Supplementary Table 5).

Only 14,464 *L. bicolor* proteins (70%) showed sequence similarity (BLASTX, cut-off *e*-value >0.001) to documented proteins. Most homologues were found in the sequenced basidiomycetes *Phanerochaete chrysosporium*[4], *Cryptococcus neoformans*[5], *Ustilago maydis*[6] and *Coprinopsis cinerea*[7] (Supplementary Table 6). The percentage of proteins found in multigene families was related to genome size and was the largest in *L. bicolor* (Fig. 2). This was mainly owing to the expansion of protein family size, but was also because of the larger number of protein families in *L. bicolor* compared with the other basidiomycetes (Supplementary Table 7). Expansion of protein family sizes in *L. bicolor* was prominent in the lineage-specific multigene families. Marked gene family expansions occurred in those genes predicted to have roles in protein–protein interactions (for example, WD40-domain-containing proteins) and in signal transduction

[1]UMR 1136, INRA-Nancy Université, Interactions Arbres/Microorganismes, INRA-Nancy, 54280 Champenoux, France. [2]US DOE Joint Genome Institute, Walnut Creek, California 94598, USA. [3]Microbial Ecology, Lund University, SE-223 62 Lund, Sweden. [4]Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS-Universités Aix-Marseille I & II, 13288 Marseille Cedex 9, France. [5]Department of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology (VIB), Ghent University, B-9052 Ghent, Belgium. [6]Department of Biology, West Virginia University, Morgantown, West Virginia 26506, USA. [7]Department for Plant Biochemistry, Georg-August-Universität Göttingen, 37077 Göttingen, Germany. [8]Université Lyon 1, UMR CNRS - USC INRA d'Ecologie Microbienne, 69622 Villeurbanne, France. [9]Stanford Human Genome Center, Department of Genetics, Stanford University School of Medicine, 975 California Avenue, Palo Alto, California 94304, USA. [10]Institute of Forest Botany, Georg-August-Universität, 37077 Göttingen, Germany. [11]Department of Biological Sciences, University of Alabama, Huntsville, Alabama 35899, USA. [12]Eberhard-Karls-Universität, Physiologische Oekologie der Pflanzen, 72076 Tübingen, Germany. [13]Department of Ecology & Evolution, University of Lausanne, 1015 Lausanne, Switzerland. [14]Swiss Federal Research Institute WSL, 8903 Birmensdorf, Switzerland. [15]Unité de Recherches en Génomique-Info, INRA-Evry, 91034 Évry Cedex, France. [16]Environmental Science Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA. [17]Department of Biology, The University of North Carolina, Chapel Hill, North Carolina 27599-3280, USA. [18]Laboratoire Associé de l'INRA, Ghent University, B-9052 Gent, Belgium. [19]Department of Plant and Microbial Biology, University of California, Berkeley, California 94720-3102, USA.

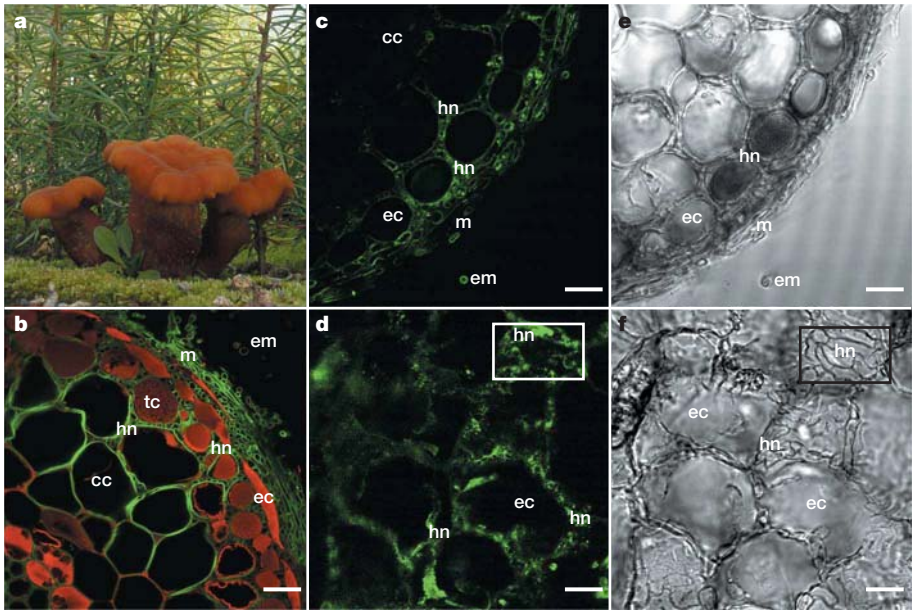



**Figure 1 | Ectomycorrhizal symbiosis and the localization of the SSP MISSP7. a**, Fruiting bodies of *L. bicolor* colonizing seedlings of Douglas fir (photograph courtesy of D. Vairelles). **b**, Laser-scanning confocal microscopy image of a transverse section of *Pseudotsuga menziesii–L. bicolor* ectomycorrhizae showing extramatrical mycelium (em), the mantle (m) and Hartig net hyphae (hn) between epidermal (ec), tannin (tc) and cortical (cc) root cells. Scale, 10 μm. **c–f**, Immunofluorescent localization of MISSP7. Transverse (**c**, **e**) and longitudinal (**d**, **f**) sections of *P. trichocarpa–L. bicolor* ectomycorrhizae. MISSP7 was detected in the hyphae of the mantle (m) and the Hartig net (hn) ensheathing epidermal cells (ec). Rectangles in **d** and **f** show the finger-like, labyrinthine hyphal system accumulating a large amount of MISSP7. **e**, **f**, Phase contrast images. Scale, 10 μm.

mechanisms (Supplementary Table 7). Two new classes of GTPase α genes were found and may be candidates for the complex communication that must occur between the mycobiont and its host plant during mycorrhizae establishment (Supplementary Table 8). Several transcripts coding for expanded and lineage-specific gene families were upregulated in symbiotic and fruiting body tissues, suggesting a role in tissue differentiation (Supplementary Tables 5 and 9).

In our analysis of annotated genes, in particular that of paralogous gene families, we highlighted processes that may be related to the biotrophic and saprotrophic lifestyles of *L. bicolor*. Twelve predicted proteins showed a similarity to known haustoria-expressed secreted proteins of the basidiomycetous rusts *Uromyces fabae*[8] and *Melampsora lini*[9], which are involved in pathogenesis (Supplementary Table 10). Out of the 2,931 proteins predicted to be secreted by *L. bicolor*, most (67%) cannot be ascribed a function, and 82% of these predicted proteins are specific to *L. bicolor*. Within this set, we found a large number of genes that encode cysteine-rich products that have a predicted size of <300 amino acids. Of these 278 SSPs, 69% belong to multigene families, but only nine groups comprising a total of 33 SSPs co-localized in the genome (Supplementary Fig. 5). The structure of two of these clusters is shown in Supplementary Fig. 6. Other SSPs are scattered all over the genome, and we found no correlation between SSP and transposable element genome localization (Supplementary Fig. 5). Transcript profiling revealed that the expression of several SSP genes is specifically induced in the symbiotic interaction (Table 2 and Supplementary Fig. 10). Five of the 20 most highly upregulated fungal transcripts in ectomycorrhizal root tips code for SSPs (Supplementary Table 5). These mycorrhiza-induced cysteine-rich SSPs (MISSPs) belong to *L. bicolor*-specific orphan gene families. Within the MISSPs, we found a family of secreted proteins with a CFEM domain (INTERPRO IPR014005) (Supplementary Figs 7 and 8), as previously identified in the plant pathogenic fungi *M. lini*[9] and *Magnaporthea grisea*[10] (Supplementary Table 10), and proteins with a gonadotropin (IPR0001545) or snake-toxin-like (SSF57302) domains related to the cysteine-knot domain. Expression of several SSPs was downregulated in ectomycorrhizal root tips (cluster E in Supplementary Fig. 10), suggesting a complex interplay between these secreted proteins in the symbiosis interaction.

The rich assortment of MISSPs may therefore act as effector proteins to manipulate host cell signalling or to suppress defence pathways during infection, as suggested for pathogenic rusts[8,9], smuts[6] (*U. maydis*) and *Phytophthora*[11] species. To have a role in symbiosis development, MISSPs should be expressed in *L. bicolor* hyphae colonizing the root tips. To test this assertion, we determined the tissue distribution of the mycorrhiza-induced cysteine-rich SSP of 7 kDa (MISSP7) (JGI identification number 298595) showing the highest induction in ectomycorrhizal tips (Table 2 and Supplementary Table 5). Two peptides, one of which is located in the amino-terminal and the other in the carboxy-terminal part of the mature protein, were selected as antigens for the production of anti-MISSP7 antibodies. The selected peptides were not found in the deduced protein sequences of other *L. bicolor* gene models, nor in the *Populus trichocarpa* genome[12]. MISSP7

**Table 1 | Genome characteristics of *L. bicolor* and other basidiomycetes**

| Genome characteristics | *L. bicolor* | *C. cinerea*[7] | *P. chrysosporium*[4] | *C. neoformans*[5] | *U. maydis*[6] |
|---|---|---|---|---|---|
| Strain | S238N-H82 | Okayama7#130 | RP78 | H99 | 521 |
| Sequencing institution | JGI | Broad | JGI | Broad | Broad |
| Genome assembly (Mb) | 64.9 | 37.5 | 35.1 | 19.5 | 19.7 |
| GC content (%) | 46.6 | 51.6 | 53.2 | 48.2 | 54 |
| Number of protein-coding genes | 20,614 | 13,544 | 10,048 | 7,302 | 6,522 |
| Coding sequence <300 bp | 2,191 | 838 | 163 | 313 | 58 |
| Average gene length (bp) | 1,533 | 1,679 | 1,667 | 1,828 | 1,935 |
| Average coding sequence length (bp) | 1,134 | 1,352 | 1,366 | 1,502 | 1,840 |
| Average exon length (nt) | 210.1 | 251 | 232 | 253 | 1,051 |
| Average intron length (nt) | 92.7 | 75 | 117 | 66 | 127 |

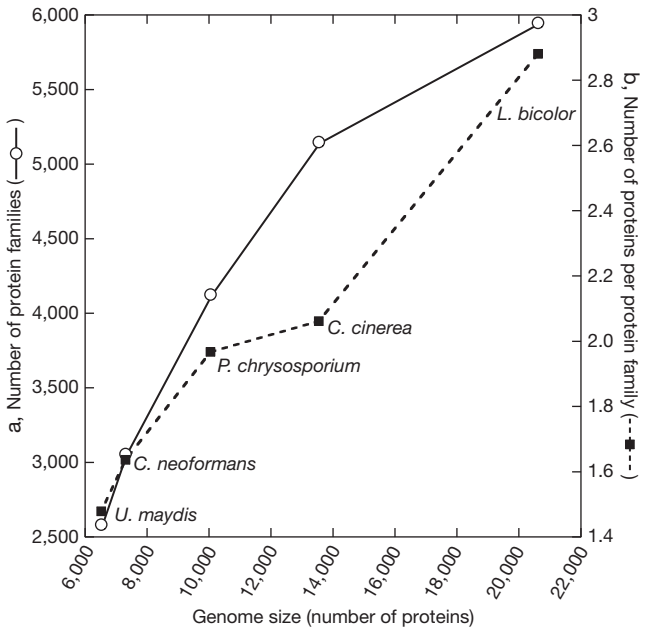



**Figure 2 | Expansion of protein families in *L. bicolor*.** **a**, Relationship between genome size and the number of protein families. **b**, Relationship between genome size and protein family sizes in five sequenced basidiomycetes. Protein sequences predicted from the genome sequences of *L. bicolor*, *C. cinerea*, *P. chrysosporium*, *C. neoformans* and *U. maydis* were clustered into families using the TRIBE-MCL algorithm (see Supplementary Information Section 5 for details).

localization in *L. bicolor–P. trichocarpa* ectomycorrhizal root tips by indirect immunofluorescence is illustrated in Fig. 1 and Supplementary Fig. 11. Control images in which the ectomycorrhizae sections were obtained by replacing primary anti-MISSP7 antibodies with pre-immune immunoglobulin (Ig)G are shown in Supplementary Fig. 12. For cases in which ectomycorrhizae were treated with anti-MISSP7 antibody followed by fluorescent-labelled secondary antibody, fluorescence was localized in the hyphae colonizing short roots (Fig. 1 and Supplementary Fig. 11) and was not detected in the free-living mycelium (Supplementary Fig. 12). Although MISSP7 was detected in the hyphal mantle layers ensheating the root tips, the protein mainly accumulated in the finger-like, labyrinthine branch hyphal system (Hartig net), which provides a very large area of contact between cells of the two symbionts. It accumulated in the cytosol and cell wall of the fungal cells. The MISSP7 protein could therefore interact with the plant components after secretion. MISSP7 shares no sequence similarity or protein motif with other SSPs.

Comparison of the MISSP sequences did not reveal a specific conserved motif that could potentially contribute to their function or to targeting to the host cell, such as the RXLR motif[11] of phytopathogenic *Phytophthora* or the malaria parasite. SSPs with upregulated expression in fruiting bodies (Supplementary Table 5 and Supplementary Fig. 10) may have a role in the differentiation of the sexual tissues and/or the aggregation of sporophore tissues. Interestingly, there is a large set of SSP genes showing significant changes in gene expression in both ectomycorrhizal root tips and fruiting bodies (cluster A in Supplementary Fig. 10), suggesting that both developmental processes recruit similar gene networks (for example, those involved in hyphal aggregation).

Host trees are able to harness the formidable web of mycorrhizal hyphae (which permeates the soil and decaying leaf litter) for their nutritional benefit. A process that is pivotal to the success of ectomycorrhizal interactions is therefore the equitable exchange of nutrients between the symbiont and its host plant[1,2,13]. A comparison with other basidiomycetes (Supplementary Table 12) revealed that the total number of predicted transporters is larger in *L. bicolor* compared with *C. cinerea* and *P. chrysosporium*. Interestingly, *L. bicolor* has multiple ammonia transporters, although it encodes a single nitrate permease. Ammonia is arguably the most important inorganic nitrogen source for ectomycorrhizal fungi[14]. One of the ammonia transporters (*AMT2.2*), for instance, is greatly upregulated in ectomycorrhizae (Supplementary Table 5). Therefore, *L. bicolor* shows an increased genetic potential in terms of nitrogen uptake compared with other basidiomycetes. These capabilities are consistent with *L. bicolor* being exposed to a range of nitrogen sources from the decay of organic matter[15].

Although the *L. bicolor* genome contains numerous genes coding for key hydrolytic enzymes, such as proteases and lipases, we observed an extreme reduction in the number of enzymes involved in the degradation of plant cell wall (PCW) oligosaccharides and polysaccharides. Glycoside hydrolases, glycosyltransferases, polysaccharide lyases, carbohydrate esterases and their ancillary carbohydrate-binding modules were identified using the carbohydrate-active enzyme (CAZyme) classification (http://www.cazy.org/). A comparison of the *L. bicolor* candidate CAZymes with fungal phytopathogens

**Table 2 | Changes in expression of transcripts coding for MISSPs**

| Protein identification (JGI *Laccaria* database) | Family size | Length (amino acids) | Transcript concentration (FLM) | *P. menziesii* ECM/FLM ratio (fold) | *P. trichocarpa* ECM/FLM ratio (fold) | Features |
|---|---|---|---|---|---|---|
| 298595 | sc | 68 | ND | 21,877 | 12,913 | MISSP7 |
| 333839 | 5 | 129 | ND | 7,844 | 1,931 | Glycosyl phosphatidylinositol (GPI)-anchored |
| 298667 | 2 | 70 | ND | 1,906 | 1,407 | |
| 332226 | 8 | 181 | 43 | 847 | 780 | CFEM domain (INTERPRO IPR014005) |
| 311468 | 2 | 59 | ND | 191 | ND | |
| 295737 | 8 | 288 | 131 | 171 | 252 | |
| 334759 | sc | 101 | ND | 109 | 18 | |
| 395403 | 4 | 121 | 24 | 103 | 93 | |
| 333423 | 9 | 120 | 6 | 102 | 72 | Gonadotropin domain (IPR0001545) |
| 312262 | 4 | 106 | 85 | 69 | 53 | |
| 295625 | 4 | 199 | 325 | 66 | 48 | |
| 325402 | 8 | 238 | 310 | 49 | 74 | Snake toxin-like (SSF57302) |
| 316998 | sc | 56 | 137 | 29 | 57 | |
| 333197 | 3 | 148 | 266 | 17 | 8 | |
| 327918 | 2 | 154 | 763 | 13 | 4 | Homologue in *C. cinerea* |
| 307956 | sc | 74 | 336 | 13 | 90 | Whey acidic domain (IPR008197) |
| 327246 | sc | 194 | 1,025 | 10 | 18 | Homologue in *C. cinerea* |
| 303550 | 5 | 98 | 1,365 | 10 | 14 | |
| 300377 | 2 | 291 | 5,499 | 10 | 8 | |
| 293250 | sc | 224 | 127 | 9 | 10 | Homologue in *C. cinerea* |
| 298648 | sc | 64 | 1,108 | 8 | 12 | |
| 298646 | 2 | 73 | 1,028 | 7 | 14 | |
| 293729 | 3 | 210 | 3,000 | 7 | 7 | |

Transcript profiling was performed on free-living mycelium (FLM) and ectomycorrhizal root tips (ECM) of poplar (*P. trichocarpa*) and Douglas fir (*P. menziesii*). See Supplementary Information Section 9 for details. Abbreviations: ND, not detected; sc, single copy.

confirms the adaptation of its enzyme repertoire to symbiosis and reveals the strategy used for the interaction with the host (Supplementary Tables 13 and 14). The reduction in PCW CAZymes affects almost all glycoside hydrolase families, culminating in the complete absence of several key families. For instance, there is only one candidate cellulase (glycoside hydrolase 5, GH5) appended to the sole fungal cellulose-binding module (CBM1) found in the genome, and no cellulases from families GH6 and GH7 (Supplementary Table 14). Similar reductions or loss of hemicellulose- and pectin-degrading enzymes were also noted. These observations suggest that the inventory of *L. bicolor* PCW-degrading enzymes underwent massive gene loss as a result of its adaptation to a symbiotic lifestyle, and that this species is now unable to use many PCW polysaccharides as a carbon source, including those found in soil and leaf litter. The remaining small set of secreted CAZymes with potential action on plant polysaccharides (for example, GH28 polygalacturonases) is probably required for cell wall remodelling during fungal tissue differentiation because their expression was upregulated in both fruiting bodies and ectomycorrhizae (Supplementary Table 15 and Supplementary Fig. 13). By contrast, transcripts coding for proteins with an expansin domain were only induced in ectomycorrhizae, suggesting they may be used by *L. bicolor* for penetrating into the root apoplastic space.

To survive before its mycorrhizal association with its host, *L. bicolor* seems to have developed a capacity to degrade non-plant (for example, animal and bacterial) oligosaccharides and polysaccharides; this is suggested by retention of CAZymes from families GH79, polysaccharide lyase 8 (PL8), PL14 and GH88 (Supplementary Table 14). Interestingly, there is no invertase gene in the *L. bicolor* genome, implying that this fungus is unable to use sucrose directly from the plant. This is consistent with earlier observations[16] that *L. bicolor* depends on its host plant to provide glucose in exchange for nitrogen. We also noticed an expansion of CAZymes involved in the fungal cell wall biosynthesis and rearrangement, almost entirely owing to an increased number of putative chitin synthases and enzymes acting on β-glucans (Supplementary Table 14). Several of the corresponding genes are up- or down-regulated in developmental processes requiring cell wall alterations such as formation of fruiting bodies or mycorrhizae (Supplementary Table 15 and Supplementary Fig. 13).

Ectomycorrhizal fungi have an important role in mobilizing nitrogen from well-decomposed organic matter[2,15]. The hyphal network permeating the soil might therefore be expected to express a wide diversity of proteolytic enzymes. The total number of secreted proteases (116 members) identified (Supplementary Fig. 14) is relatively large compared with that in other sequenced saprotrophic basidiomycetes, such as *C. cinerea* and *P. chrysosporium*. Secreted aspartyl-, metallo- and serine-proteases may have a role in degradation of decomposing litter[15], confirming that *L. bicolor* has also the ability to use nitrogen of animal origin, as suggested previously[17]. They may also have a role in developmental processes because the expression of several secreted proteases is up- or down-regulated in fruiting bodies and ectomycorrhizal root tips (Supplementary Table 16). Mycelial mats formed by *L. bicolor* hyphae colonizing organic matter therefore possess the ability to degrade proteins from decomposing leaf litter.

Our analysis of the gene space reveals a multi-faceted mutualistic biotroph equipped to take advantage of transient occurrences of high-nutrient niches (living host roots and decaying soil organic matter) within a heterogeneous, low-nutrient environment. The availability of genomes from mutualistic, saprotrophic[4] and pathogenic[6] fungi, as well as from the mycorrhizal tree *P. trichocarpa*[12], now provides an unparalleled opportunity to develop a deeper understanding of the processes by which fungi colonize wood and soil litter, and also interact with living plants within their ecosystem, to perform vital functions in the carbon and nitrogen cycles[2] that are fundamental to sustainable plant productivity.

## METHODS SUMMARY

**Genomic sequence.** Scaffolds and assemblies for all genomic sequences generated by this project are also available from the Joint Genome Institute (JGI) portal (http://genome.jgi-psf.org/Lacbi1/Lacbi1.download.ftp.html). A genome browser is available from JGI (http://www.jgi.doe.gov/laccaria). BLAST search of the genome is available at JGI (http://www.jgi.doe.gov/laccaria) and INRA LaccariaDB (http://mycor.nancy.inra.fr/IMGC/LaccariaGenome/).

**Predicted gene models.** Consensus gene predictions, produced by combining several different gene predictors, are available from JGI (http://www.jgi.doe.gov/laccaria) as General Feature Format (GFF) files. These gene models can also be accessed from the Genome Browser in the JGI *L. bicolor* portal (http://www.jgi.doe.gov/laccaria).

**Gene annotations.** Tables compiling KEGG, PFAM, KOG and best BLAST hits for predicted gene models, transposable element and CAZyme data, as well as Tribe-MCL gene families, are available from INRA LaccariaDB (http://mycor.nancy.inra.fr/IMGC/LaccariaGenome/).

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

**Received 10 August; accepted 20 December 2007.**

1. Smith, S. E. & Read, D. J. *Mycorrhizal Symbiosis* 2nd edn (Academic, London, 1996).
2. Read, D. J. & Perez-Moreno, J. Mycorrhizas and nutrient cycling in ecosystems — a journey towards relevance? *New Phytol.* **157,** 475–492 (2003).
3. Galagan, J. E., Henn, M. R., Ma, L. J., Cuomo, C. A. & Birren, B. Genomics of the fungal kingdom: insights into eukaryotic biology. *Genome Res.* **15,** 1620–1631 (2005).
4. Martinez, D. *et al.* Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nature Biotechnol.* **22,** 695–700 (2004).
5. Loftus, B. J. *et al.* The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans. Science* **307,** 1321–1324 (2005).
6. Kämper, J. *et al.* Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis. Nature* **444,** 97–101 (2006).
7. *Coprinus cinereus* database ⟨http://www.broad.mit.edu/annotation/genome/coprinus_cinereus/Home.html⟩.
8. Wirsel, S. G. R., Voegele, R. T. & Mendgen, K. W. Differential regulation of gene expression in the obligate biotrophic interaction of *Uromyces fabae* with its host *Vicia faba. Mol. Plant Microb. Int.* **14,** 1319–1326 (2001).
9. Catanzariti, A. M., Dodds, P. N., Lawrence, G. J., Ayliffe, M. A. & Ellis, J. G. Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *Plant Cell* **18,** 243–256 (2006).
10. Kulkarni, R. D., Kelkar, H. S. & Dean, R. A. An eight-cysteine-containing CFEM domain unique to a group of fungal membrane proteins. *Trends Bioch. Sci.* **28,** 118–121 (2003).
11. Kamoun, S. A. Catalogue of the effector secretome of plant pathogenic oomycetes. *Annu. Rev. Phytopathol.* **44,** 41–60 (2006).
12. Tuskan, G. A. *et al.* The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313,** 1596–1604 (2006).
13. Martin, F., Kohler, A. & Duplessis, S. Living in harmony in the wood underground: ectomycorrhizal genomics. *Curr. Opin. Plant Biol.* **10,** 204–210 (2007).
14. Chalot, M., Blaudez, D. & Brun, A. Ammonia: a candidate for nitrogen transfer at the mycorrhizal interface. *Trends Plant Sci.* **11,** 263–266 (2006).
15. Lindahl, B. D. *et al.* Spatial separation of litter decomposition and mycorrhizal nitrogen uptake in a boreal forest. *New Phytol.* **173,** 611–620 (2007).
16. Nehls, U., Grunze, N., Willmann, M., Reich, M. & Küster, H. Sugar for my honey: carbohydrate partitioning in ectomycorrhizal symbiosis. *Phytochemistry* **68,** 82–91 (2007).
17. Klironomos, J. N. & Hart, M. M. Animal nitrogen swap for plant carbon. *Nature* **410,** 651 (2001).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** The genome sequencing of *L. bicolor* was funded by the US Department of Energy's Office of Science, Biological and Environmental Research Program and the by University of California, the Lawrence Berkeley National Laboratory, the Lawrence Livermore National Laboratory, and the Los Alamos National Laboratory. Annotation and transcriptome analyses were supported by INRA, the US Department of Energy, the US National Science Foundation, the European Commission, the Région Lorraine and the Swedish Research Council. We thank S. Rombauts, L. Sterck, K. Vandepoele, G. Werner and his colleagues, S. Pitluck and K. Zhou, B. Hilselberger and J. Gérard for their assistance. F.M. thanks N. Talbot for critical reading of an early draft of the manuscript.

**Author Contributions** A.A., D.A., A.B., E.G.J.D., F.D., J.G., A. K., E.L., V.P., A.S., H.J.S. and J.W. contributed equally to this work as second authors. M.C., B.H., U.K., S.L., Y.V.d.P., G.K.P., A.P., P.J.P., P.M.R., P.R., I.R.S., J.E.S., A.T., G.T. and I.V.G. contributed equally to this work as senior authors. Individual contributions were as follows. For the sequencing project, F.M. and G.T. initiated the sequencing project in the wake of the *Populus* genome project. F.M., I.V.G. and P.M.R. developed and coordinated the sequencing, annotation and transcriptome projects. S.L. and sequencing staff at

JGI performed the shotgun sequencing. H.J.S. and his staff at JGI performed the JAZZ assembly of the nuclear and mitochondrial genome. J.G. and J.S. performed the Arachne assembly, closed up repetitive gaps and fixed missasembled regions. A.A., A.S., J.W., M.M., P.R., Y.V.d.P. and I.V.G. did the *ab initio* annotation of protein-coding gene models. A.K., E.L., P.B., C.D., A.D., J.C.D., M.P., G.K.P., A.T. and F.M. provided expressed sequence tag/cDNA information for the *ab initio* and manual annotation. Genome statistics was performed by D.A., F.D., J.W., P.R., I.V.G. and F.M. A.A. and I.V.G., and F.D. and M.P.O.-L.S. were responsible for database design and maintenance at JGI and INRA, respectively. For genome analysis, D.C., M.P. and G.K.P. were responsible for DNA extraction and purification; D.A., F.D., Y.C.L., B.R., Y.V.P., P.R., A.S., J.E.S., A.T., I.V.G. and F.M. for comparative genome analysis; B.C., D.A. and A.T. for genome synteny; J.L., T.Y., G.T., F.M. and F.L.T. for construction of the genetic map; S.D.F. for single nucleotide polymorphisms; A.K., F.D. and F.M. for DNA arrays; A.D., B.C., P.F.K. and F.M. for high-GC sequences; J.W., P.R. and F.M. for tRNA, snRNA and rDNA; E.G.J.D., P.M.C., B.H. for carbohydrate active enzymes; A.D. and P.F.-K. for carbohydrate metabolism; J.G., P.H., U.K. and F.M. for cell wall proteins and secretome; L.F.-T., G.G., D.M. and R.M. for cytoskeleton and motor proteins; M.R., I.F. and A.P. for lipid metabolism; and H.N.-H., U.K. and I.R.S. for mating type genes. F.M. was responsible for genome analysis of the mitochondrion; P.E.C., P.H., M.B., S.K. and U.K. for the multi-copper oxidases; M.B., P.E.C. and F.M. for the proteases; N.R. for the redox genes; S.D., P.J. and G.K.P. for the signal transduction pathway; A.B., D.B., C.F., E.L., B.M., U.N., V.P. and M.C. for transporters; J.L., J.W., P.R., F.M. and H.Q. for transposable elements; V.P., J.G., A.B. and V.L. for immunolocalization of SSP; and J.E.S. and P.J.P. for the *Coprinopsis* genome. F.M.organized co-ordination between different groups. F.M. wrote and edited the paper with input from I.V.G., A.T., D.A., P.R., M.C. and B.H.

**Author Information** The whole-genome shotgun project has been deposited at GenBank/EMBL/DDBJ under project accession number ABFE00000000. The version described in this paper including assembly and annotation is the first version, ABFE01000000. The complete expression data set is available as a series under accession number GSE9784 at the Gene Expression Omnibus at NCBI (http://www.ncbi.nlm.nih.gov/geo/). Reprints and permissions information is available at www.nature.com/reprints.  Correspondence and requests for materials should be addressed to F.M. (fmartin@nancy.inra.fr).

## METHODS

**Genome sequencing.** The haploid genome of the strain S238N-H82 from *L. bicolor* (Maire) P. D. Orton was sequenced with the use of a whole-genome shotgun strategy. All data were generated by paired-end sequencing of cloned inserts using Sanger technology on ABI3730xl sequencers. Supplementary Table 1 gives the number of reads obtained per library.

**Genome assembly.** The data were assembled using release 1.0.1b of JAZZ, a JGI whole-genome shotgun assembler. On the basis of the number of alignments per read, the main genome scaffolds were at a depth of 9.88. The amount of sequence in the unplaced reads was 6.5 Mb, which is sufficient to cover the main-genome gaps to a mean depth of 9.9. A total of 64.9 Mb are captured in the scaffold assembly (Supplementary Table 2).

**Genome annotation.** Gene models were predicted using FgenesH[18], homology-based FgenesH$^{+}$ (ref. 18) and Genewise[19], as well as EuGène[20] and TwinScan[21], and alignments of several complementary DNA resources (Supplementary Information Section 3). The JGI pipeline selected a best representative gene model for each locus on the basis of expressed sequence tag support and similarity to known proteins from other organisms, and predicted 20,614 protein-coding gene models. All predicted genes were annotated using Gene Ontology[22], eukaryotic clusters of orthologous groups[23] and KEGG pathways[24]. Protein domains were predicted using InterProScan[25]. Signal peptides were predicted in 2,931 *L. bicolor* proteins by both the hidden Markov and the neural network algorithms of SignalP[26]. After eliminating predicted transmembrane proteins and removal of transposable element fragments, we selected 278 cysteine-rich secreted proteins with a size of $<$300 amino acids. Gene families were built from proteins in *L. bicolor*, *C. cinerea*, *P. chrysosporium*, *C. neoformans* and *U. maydis* using Tribe-MCL tools[27] with default settings.

**Indirect immunofluorescent localization of MISSP7.** The peptides LRALGQASQGGDLHR and GPIPNAVFRRVPEPNF located in the N-terminal and C-terminal parts of the MISSP7 sequence (without the signal peptide) were synthesized and used as antigens for the generation of antibodies in rabbits according to the manufacturer's procedures (Eurogentec). The anti-MISSP7IgG fraction was purified using the MAbTrap kit (GE Healthcare) according to the manufacturer's recommendations. Subsequently, the IgG-containing fraction was desalted using a HiTrap desalting column (GE Healthcare). The concentration of purified IgG from pre-immune serum was determined by Bradford assay using a Bio-Rad protein assay. The final concentration of anti-MISSP7 IgG was 0.16 mg ml$^{-1}$. Immunolocalization was performed essentially as described in refs 28 and 29, with slight modifications (Supplementary Section 10).

**Gene expression.** Average expression levels of genes in different tissues and conditions were analysed using CyberT statistical framework (http://www.igb.uci.edu/servers/cybert/) and hierarchical clustering with EPCLUST (http://ep.ebi.ac.uk/EP/EPCLUST/) (Supplementary Information Section 8).


18. Salamov, A. & Solovyev, V. *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res.* **10,** 516–522 (2000).
19. Birney, E., Clamp, M. & Durbin, R. GeneWise and genomewise. *Genome Res.* **14,** 988–995 (2004).
20. Schiex, T., Moisan, A. & Rouzé, P. in: *Computational Biology* vol. 2066 (eds, Gascuel, O. & Sagot, M. F.) Lecture Notes in Computer Science, publisher: Springer, Berlin/Heidleberg, Germany 111–125.
21. Tenney, A. E. *et al.* Gene prediction and verification in a compact genome with numerous small introns. *Genome Res.* **14,** 2330–2335 (2004).
22. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nature Genet.* **25,** 25–29 (2000).
23. Koonin, E. V. *et al.* A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* **5,** R7.1–R7.28 (2004).
24. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32,** D277–D280 (2004).
25. Zdobnov, E. M. & Apweiler, R. InterProScan — an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17,** 847–848 (2001).
26. Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP, and related tools. *Nature Protocols* **2,** 953–971 (2007).
27. Enright, A. J., van Dongen, S. & Duzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucl. Acids Res.* **30,** 1575–1584 (2002).
28. Blancaflor, E. B., Zhao, L. & Harrison, M. J. Microtubule organization in root cells of *Medicago truncatula* during development of an arbuscular mycorrhizal symbiosis with *Glomus versiforme*. *Protoplasma* **217,** 154–165 (2001).
29. Harrison, M. J., Dewbre, G. R. & Liu, J. Y. A phosphate transporter from *Medicago truncatula* involved in the acquisition of phosphate released by arbuscular mycorrhizal fungi. *Plant Cell* **14,** 2413–2429 (2002).

The *Laccaria* genome is the largest fungal genome published so far (Supplementary Table 3)[11,13,14]; see the Broad Institute's Fungal Genome Initiative (FGI), http://www.broad.mit.edu/annotation/fungi/fgi/ and JGI (http://www.jgi.doe.gov/) web sites]. No evidence for large scale duplications was observed within the *Laccaria* genome, though many examples of short regions of tandem duplication were observed within multigene families (Supplementary Fig. 4). Comparisons of assembled sequences and searches for segments of conserved gene order did not reveal evidence for shared large-scale genome duplications among the released fungal genomes.

After the initial machine and manual annotations were carried out, the WGS data were also assembled using the ARACHNE assembler[15] (http://www.broad.mit.edu/wga/) assuming that the later algorithm was less sensitive to repeated regions, such as transposable elements. The ARACHNE assembly generated larger scaffolds, so-called supercontigs, with less gaps (total scaffold size, 59,935 Mbp and total basepairs present in assembly, 59,275 Mbp). These ARACHNE supercontigs and the JAZZ scaffolds used for the machine and manual annotations were aligned using JGI BlastN and the genome pair rapid dotter, Gepard (http://mips.gsf.de/services/analysis/gepard) (Supplementary Fig. 2). For purposes of integrating the JAZZ and ARACHNE assemblies with the genetic map, the genotypes of 50 polymorphic simple sequence repeat (SSR) length polymorphisms and SCAR markers were scored in a reference mapping population, i.e. monokaryotic progeny of the parental strain *L. bicolor* S238N. As shown in Supplementary Fig. 2, this initial set of marker sequences were mapped on the genome assemblies to validate order and orientation of sequenced JAZZ scaffolds and ARACHNE supercontigs. Additional SSR and AFLP markers continue to be added to the genetic map, furthering the integration of genetic and physical resources. These results will be published elsewhere.

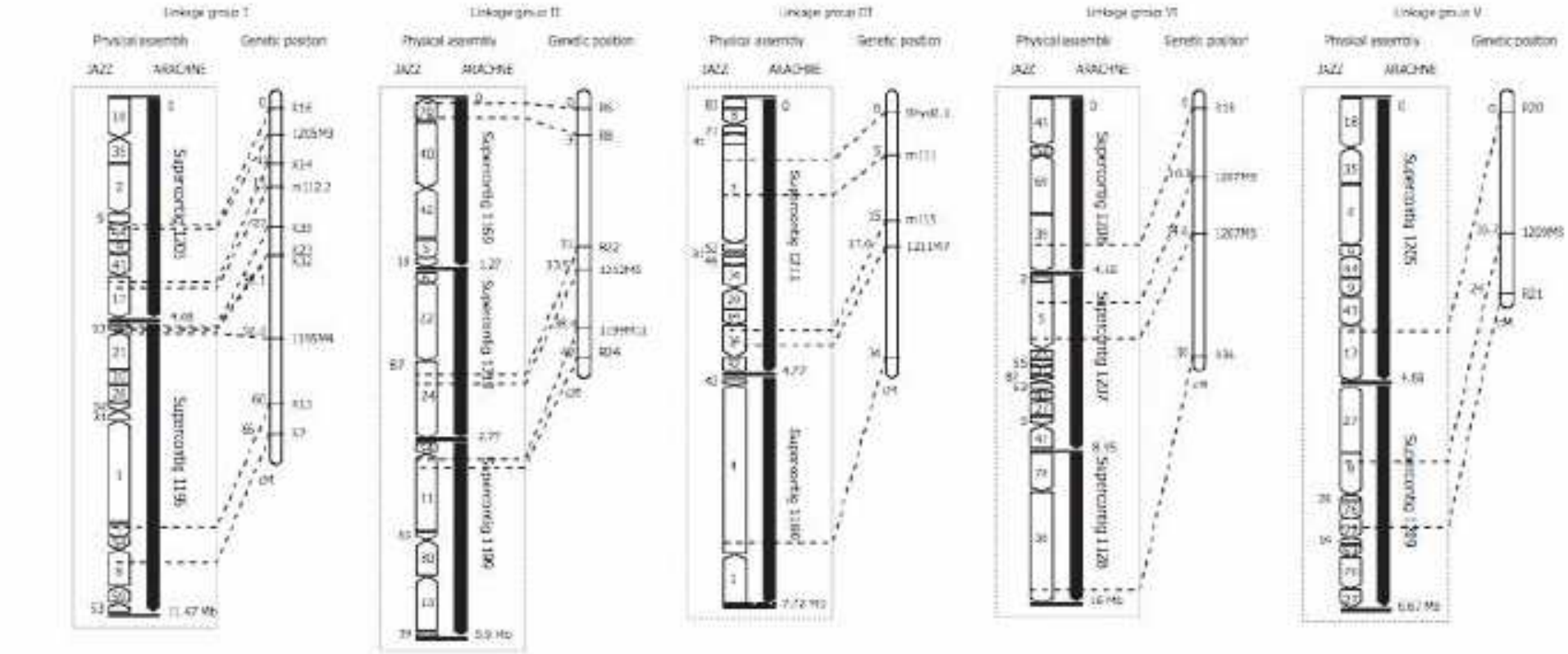


**Figure 2. Maps of the ten linkage groups and major genome assembly scaffolds and supercontigs of *Laccaria bicolor*.** For each putative chromosome (Chr 1-10), the Jazz assembly scaffolds are shown on the left, the Arachne assembly supercontigs on the middle and the known linkage groups (LG) on the right. SSR identifiers are shown to the right of the genetic map bars. The position of SSRs along LGs were calculated on a ratio scale and then converted to base-pair positions. The maps are scaled to genetic distances in centimorgans (cM) and the genome assemblies are depicted in relative physical lengths.

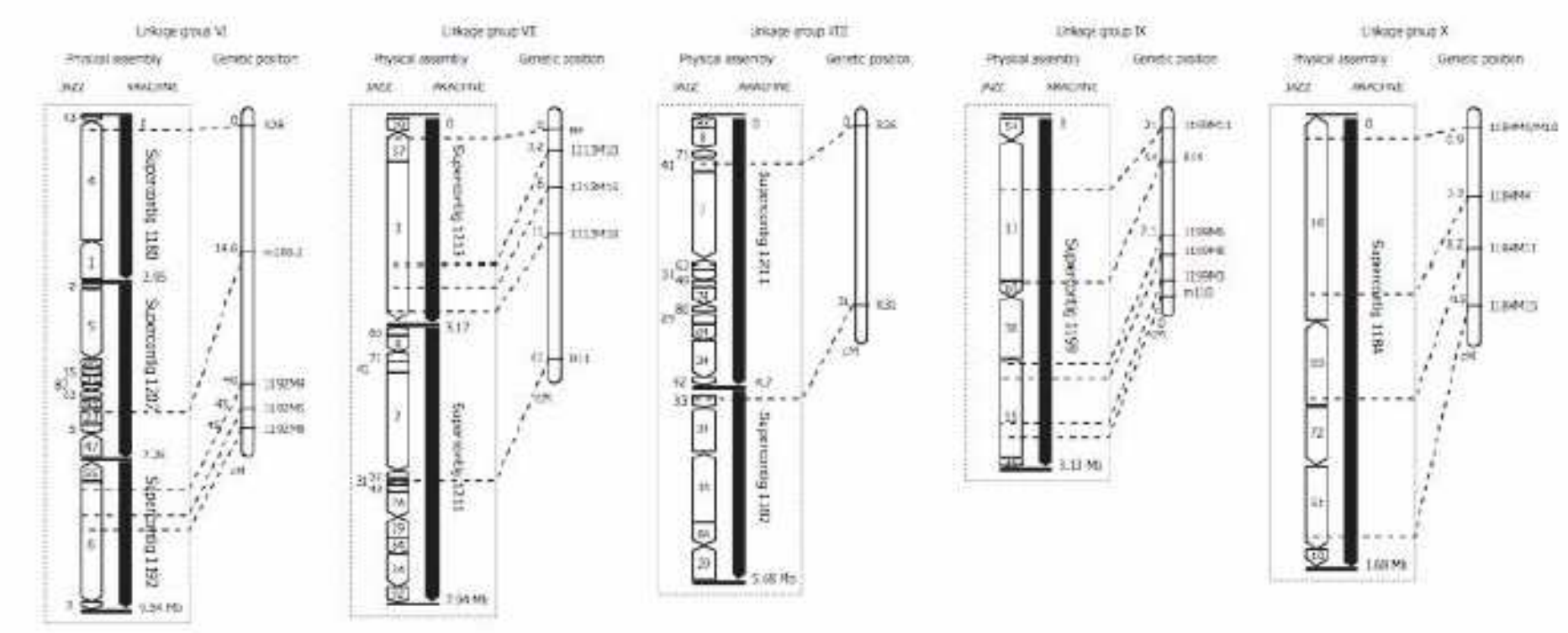


**Figure 2. Maps of the ten linkage groups and major genome assembly scaffolds and supercontigs of *Laccaria bicolor*.** For each putative chromosome (Chr 1-10), the Jazz assembly scaffolds are shown on the left, the Arachne assembly supercontigs on the middle and the known linkage groups (LG) on the right. SSR identifiers are shown to the right of the genetic map bars. The position of SSRs along LGs were calculated on a ratio scale and then converted to base-pair positions. The maps are scaled to genetic distances in centimorgans (cM) and the genome assemblies are depicted in relative physical lengths.

**6. Transposable elements**

Transposable elements (TE) have been predicted anonymously using four different methods: (i) an all-by-all genome comparison with BLASTER using BLASTN followed by chaining with MATCHER and grouping with GROUPER (BLRa), (ii) RECON, using default parameters, (iii) BLASTER using TBLASTX with the entire Repbase Update as the database, followed by chaining with MATCHER (BLRtx), and (iv) a hidden Markov model that detects TE sequences based on nucleotide composition (TE-HMM) (see [47] for details). Twenty-one percent of the *L. bicolor* genome match consensus sequences for the TE (Supplementary Table 3 and Fig. 15). It is apparent that this is an unusual fungal genome. Although previously identified major TE superfamilies found in other fungi[48] are present in *L. bicolor* (Supplementary Fig. 15), the majority of the 215 TE are specific to *Laccaria*. These elements may account for relatively large genome of *Laccaria* compared to other sequenced fungal genomes. The most abundant TE are Class 1 elements which are collectively represented by 962 complete sequences (844 *Copia/Ty1*-like, 118 *Gypsy/Ty3*-like) and over 17,468 remnant degraded copies. Class 2 elements account for a total of 5,738 sequences including 355 complete sequences including MITE, *Pogo, Ant1/Tc1,* and helitrons. This abundance of MITE elements and helitrons has not yet been reported in other fungal genomes. These TE have been and probably still are very active in the *Laccaria* genome as the corresponding transcripts have been detected by oligoarrays (data not shown). The sequence similarity analysis performed within each TE families showed that most of them are ancient, although there are at least 40 different families of TE in which nucleotide substitutions have accumulated to a level of <5%, reflecting a very young age and possible current activity. High genomic plasticity induced by mobile elements might reflect the need for continuous adaptation to changing environments like soil or to host defense mechanisms.

**Table 3. The most abundant transposable element families in *Laccaria bicolor* genome.**

| Transposable element family | Complete members | Sequence length (nt) | Occurrences in the assembled sequence |
|---|---|---|---|
| ReconFam1284 | 2 | 8,383 | 624 |
| ReconFam317 | 18 | 1,620 | 579 |
| ReconFam726 (TIR) | 3 | 2,178 | 490 |
| ReconFam1181 | 1 | 2,701 | 476 |
| T_scf2_3 (TCN3-I, LTR) | 1 | 6,251 | 475 |
| HMMReconFam1457 | 0 | 12,147 | 446 |
| HMMReconFam348 (Helitron-1) | 0 | 10,265 | 444 |
| HMMReconFam691 | 0 | 8,106 | 438 |
| T_scf71_1 (PC retro3_I, LTR) | 1 | 6,617 | 412 |
| ReconFam1186 | 1 | 1,778 | 362 |
| ReconFam419 (TIR) | 2 | 6,643 | 360 |
| ReconFam2212 | 1 | 5,158 | 354 |
| ReconFam144 | 9 | 1,824 | 345 |
| HMMReconFam561 | 1 | 4,018 | 324 |
| T_scf20_2 (PC retro3_I, LTR) | 1 | 6,213 | 320 |
| HMMReconFam1452 (LINE) | 0 | 4,302 | 319 |
| ReconFam350 (LINE) | 0 | 4,848 | 309 |
| MITE_X3 | 182 | 584 | 306 |

Transposable elements have been predicted using four different methods: (i) an all-by-all genome comparison with BLASTER using BLASTN followed by chaining with MATCHER and grouping with GROUPER (BLRa), (ii) RECON, using default parameters, (iii) BLASTER using TBLASTX with the entire Repbase Update as the database, followed by chaining with MATCHER (BLRtx), and (iv) a hidden Markov model that detects TE sequences based on nucleotide composition (TE-HMM)[47].

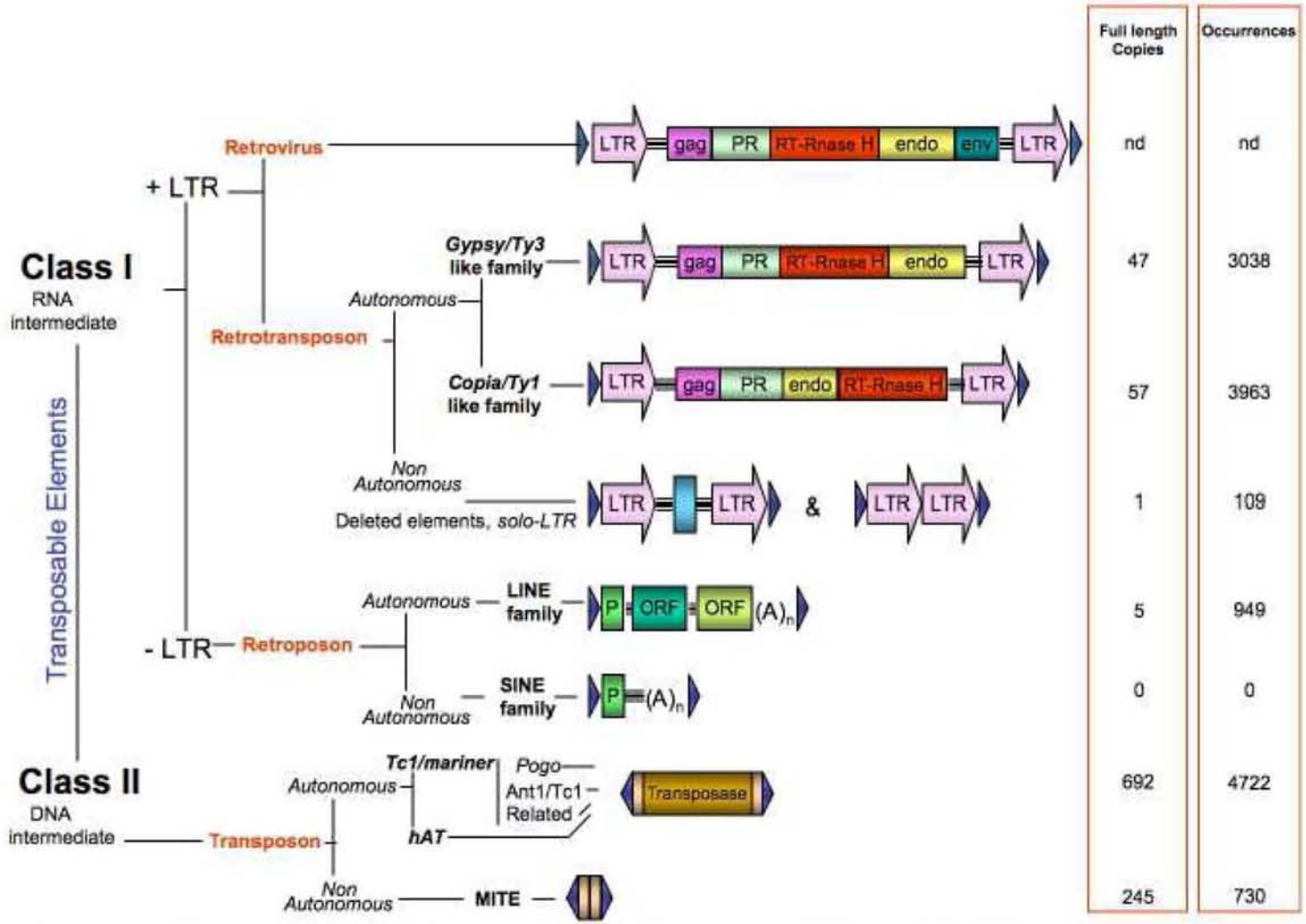



**Figure 15. The diversity and distribution of class I and class II transposable elements in *Laccaria bicolor*.** Number of full length copies and occurrences of fragments in the genome are shown on the right panel. nd, not detected.

Correlation in the co-localization of SSP and TE was assessed using the Kendall method implemented in the R software. Occurrences of SSP genes and TE sequences for the 36 largest scaffolds having SSP showed no correlation, i.e. the highest value was 0.44.

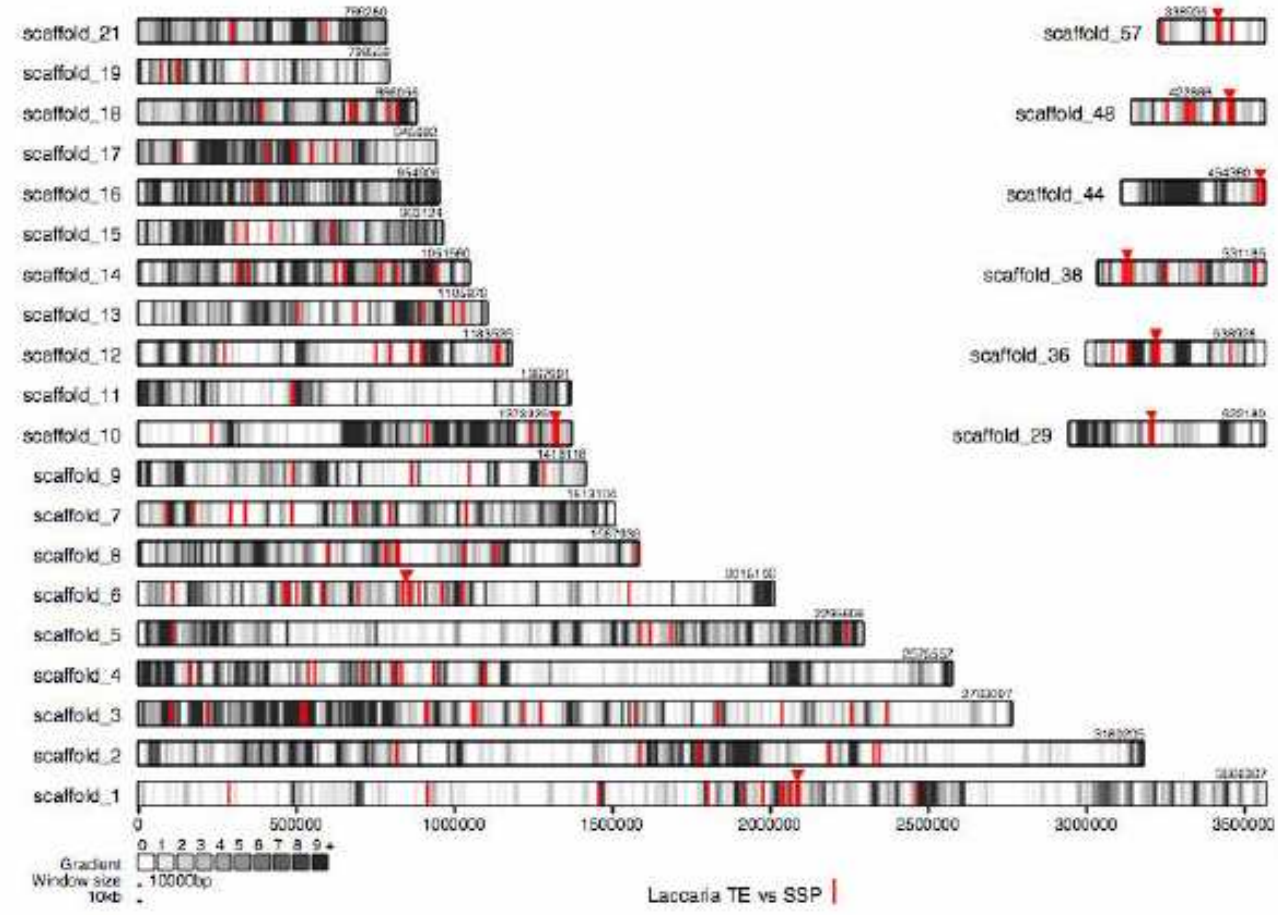



**Figure 5. Genome scale organization of genes coding for small secreted proteins (SSP) and transposable elements (TE) on the largest JAZZ scaffolds.** SSP loci are indicated in by red lines and SSP clusters by red triangles, whereas TE loci are represented by white to black bars according to their density. Distribution of TE is shown with relative density shading for 10-kb window width.

## Article 2. Pattern of simple sequence repeats distribution in the ectomycorrhizal fungus *Laccaria bicolor* genome

**Jessy Labbé, Claude Murat, Emmanuelle Morin, Francis Martin and François Le Tacon**

**(soumis à *BMC Genomics*)**

# Pattern of simple sequence repeats distribution in the ectomycorrhizal fungus *Laccaria bicolor* genome

**Jessy Labbé[§], Claude Murat, Emmanuelle Morin, Francis Martin and François Le Tacon**

UMR 1136, INRA-Nancy Université, Interactions Arbres/Microorganismes, INRA-Nancy, 54280 Champenoux, France

[§]Corresponding author

Email addresses:

JL: labbe@nancy.inra.fr

CM: claude.murat@nancy.inra.fr

EM: Emmanuelle.Morin@nancy.inra.fr

FM: fmartin@nancy.inra.fr

FLT: le_tacon@nancy.inra.fr

# Abstract

### Background

*Laccaria bicolor*, the first sequenced genome of a symbiotic fungus, displays one of the largest fungal genome. High content of repeated sequences (transposable elements, microsatellites) partly explains this large genome size. Although it is becoming clear that simple sequence repeats (SSRs), also called microsatellites are important in genomic organization and have become a large source of genetic markers, little is known on their distribution. We used an *in silico* approach to survey SSRs organization through the genome of the symbiotic ectomycorrhizal basidiomycete *Laccaria bicolor*. Finally, we have developed new SSR markers for genetic mapping and population genetic studies.

### Results

By using the microsatellite identification tool MISA, we detected 277,062 SSRs (≥ 5 repeats) in the *L. bicolor* genome. Mononucleotides outnumbered all other classes (4,152 SSRs/Mb). The SSR density decreased according to the number of nucleotides (31 SSRs/Mb for the dinucleotides, 22.3 SSRs/Mb for the trinucleotides, 4.1 SSRs/Mb for the tetra, penta and hexanucleotides). The majority (73.3%) of SSRs were located in intergenic regions. The highest SSR density (repeats ≥ 5) was observed in transposable elements (TEs) (6,706 SSRs/Mb). In the genic regions, the highest density was in introns. In a *L. bicolor* progeny, SSRs appeared to be the most polymorphic in the intergenic regions outside the TEs. The seventy-eight polymorphic microsatellite markers developed in this study have facilitated the construction of the genetic map of *L. bicolor* and could become powerful tools for ecological studies.

### Conclusions

Among the basidiomycetes analyzed in this study, *L. bicolor* exhibited the highest SSR density corresponding to 8% of the genomic sequence. SSRs were not distributed at random throughout the different genomic regions, showing the highest density in TEs. Nevertheless, the SSR density differed among the various genic regions exhibiting the highest density in introns. Most of genes containing SSR motifs code for unknown protein probably, having a role in the symbiotic interactions remaining to be determined. Our work showed that the 78 SSR markers developed in *L. bicolor* progeny were most polymorphic in the intergenic regions outside the TEs.

## Background

Ectomycorrhizal fungi are symbiotic organisms involved in mutualistic associations with the fine roots of most forest tree species. They play an essential role in the host-tree mineral nutrition and enhance its growth [1, 2]. This nutrient contribution is reciprocated by the allocation of carbohydrates by the trees to the fungus through the root interface, making the relationship a mutualistic association. *Laccaria bicolor,* a worldwide ectomycorrhizal basidiomycete, has been an experimental model for decades and is now used in large-scale mycorrhizal inoculation programs [3, 4]. Moreover, its genome has been recently sequenced [5], facilitating the study of the processes involved in the symbiotic interactions [6]. To date, with a size of 65 megabases (Mb), the *L. bicolor* genome is the largest fungal genome published, comprising a high content of repeated sequences of which 24 % of TEs and 8% of SSRs [5]. This proportion of repeated sequences may account for the relatively large genome of *L. bicolor* and may lead to genome rearrangements.

Microsatellites are formed by 1-6 base pair nucleotide motifs, tandemly repeated at least five times [7]. They are frequently and randomly widespread throughout prokaryotic and eukaryotic genomes although their frequency varies significantly among different organisms [8, 9]. They can be found in any DNA sequence (protein coding region and non-coding regions) and they exhibit a high frequency of length polymorphism due to the high mutation rates affecting the number of repeat units [8, 9]. SSRs are considered as the best genetic markers in human identity testing, diagnosis at an early stage of disease, studies of mating systems, population genetics and genome mapping [7, 10]. Moreover, a recent work has shown that SSRs have many important functions in terms of gene regulation, organism development and evolution [11]. The locations of SSRs determine the potential functional role of genes and can lead to changes in the phenotypes of an organism [12, 13]. Different molecular techniques were used to study *Laccaria* populations in natural or artificial forest ecosystems: the polymorphism of the intergenic spacer (IGS) of the rDNA [6], random amplified polymorphic DNA (RAPD) markers [14], specific characterized amplified region (SCAR) markers [15], and microsatellites [16]. Although SSRs are widely used for

population genetics [7, 16, 17] few studies have been carried out on their genomic organization/distribution in fungi [8, 18].

The availability of the whole-genome sequence (WGS) of *L. bicolor* [5, 6] allows the evaluation of the abundance and the distribution of SSRs inside this genome.

Microsatellites could play an important role in the plasticity and evolution of the genome of *L. bicolor* [5]. The aims of this work were to determine the organization of SSRs in the *L. bicolor* genome, to determine their possible involvement in the polymorphism of this genome, to compare it with those of completely sequenced saprotrophic fungal species, and to generate new microsatellite markers for genetic mapping, genetic population and molecular ecology studies.

# Results

There are several articles dealing with the repartition of the microsatellites according to the regions of the genomes [7, 8, 9, 11, 12, 19], but until now there are very few analyses concerning the fungal genomes [8, 17, 18, 20]. The definition of microsatellites varies according to the authors. We have adopted the most common definition: mono, di, tri, penta and hexa tandems repeated at least 5 times. Nevertheless, to compare our results with those of some other authors, we also have considered other definitions: mono, di, tri, penta and hexa tandems repeated at least 10 times [11] or 12 times [19] (Table 1).

### Comparison of SSRs in the *L. bicolor* genome and in other fungal genomes

We have compared the number and types of SSRs in the *L. bicolor* genome with those of eight other completely sequenced haploid fungal genomes (Table 2). They belong to two different phylogenetic classes, the Basidiomycota (*Phanerochaete chrysosporium, Coprinopsis cinerea, Ustilago maydis, Cryptococcus neoformans*) and the Ascomycota (*Aspergillus nidulans, Magnaporthe grisea, Neurospora crassa, Saccharomyces cerevisiae*). The size of these genomes, ranged from 11.9 Mb (*S. cerevisiae*) to 42.3 Mb (*M. grisea*) (Table 2). The SSR density did not differ between ascomycetes and basidiomycetes, although the highest density was observed in *S. cerevisiae* (Figure 1 and Table 2). Among the basidiomycetes, *L. bicolor* exhibited the highest SSR density. The SSR density appeared to be unrelated with genome size, two

of the smallest genomes (*S. cerevisiae* and *C. neoformans*) [8], exhibiting a high SSR density (6517 and 3620 SSRs/Mb respectively). Mononucleotide motifs outnumbered

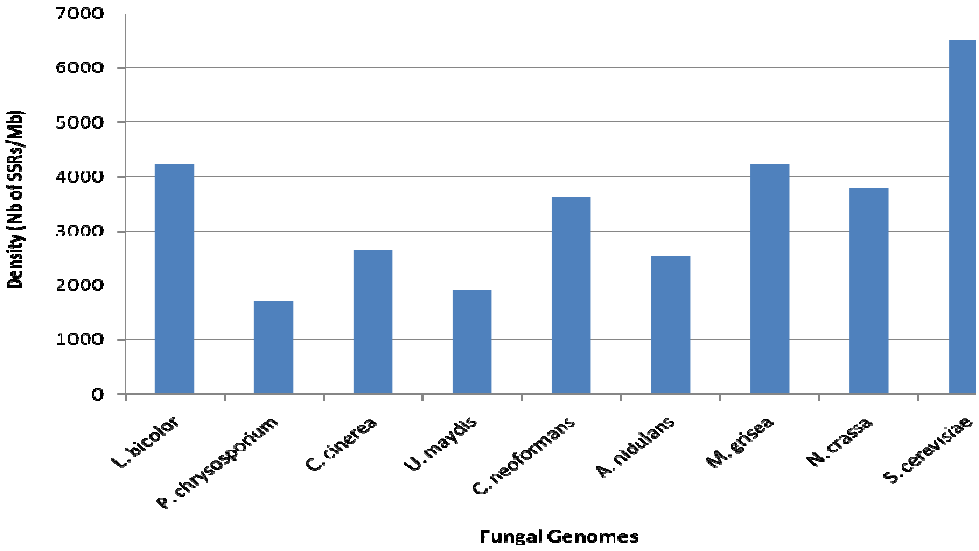



**Figure 1 - Density of SSRs (number of SSRs per Mb) in nine fungal genomes**

all other SSR classes in all fungal genomes, ranging from 1,644 SSRs/Mb in *P. chrysosporium* to 6,425 SSRs/Mb in *S. cerevisiae*. *L. bicolor* displayed one of the greatest densities of mononucleotide (4152 SSRs/Mb) after *S. cerevisiae* (6425 SSRs/Mb). In all genomes, there was a strong overrepresentation of $A_n$ repeats in the mononucleotide class. Their proportion ranged from 93% *(S. cerivisiae*) to 52% (*C. cinerea*), *L. bicolor* showing a percentage of 70%. In the nine analyzed genomes, after the $A_n$ repeats, the $C_n$ motifs were the most frequent mononucleotide microsatellites. Their percentage ranged from 5% in *S. cerevisiae* to 46% in *C. cinerea*, *L. bicolor* exhibiting a percentage of 28%. In most of the nine analyzed fungal genomes, after the mononucleotides, the most represented SSR motifs, as in *L. bicolor* genome, were the dinucleotides and then the trinucleotides. Nevertheless, there was an exception for *M. grisea*, which had a similar density of di- (33 SSRs/Mb) and trinucleotides (34 SSRs/Mb) SSRs and *N. crassa*, which, had a greater density of tri- (95 SSRs/Mb) than that of dinucleotide (78.3 SSRs/Mb) motifs (Table 2). The nucleotide composition of the dinucleotide repeats varied among species. GA/AG repeats were predominant in *L. bicolor*, *C. cinerea, U. maydis* and *C. neoformans*, while AT/TA repeats were

predominant in *A. nidulans*, *N. crassa* and *S. cerevisiae*. GC/GC repeats were only found in *P. chrysosporium* and CA/GA repeats in *M. grisea*.

In the nine analyzed fungal genomes, the tetra-, penta- and hexanucleotide density was very low, compared to the mono-, di-, and trinucleotides. Most of the times, the density of the tetra-, penta and hexanucleotides was equal or inferior to one SSR/Mb. Nevertheless, there were some exceptions. *N. crassa* exhibited a density of 18.6 SSRs/Mb for the tetranucleotides, 4.6 SSRs/Mb for the pentanucleotides and 3.3 SSRs/Mb for the hexanucleotides. *U. maydis* exhibited a density of 3.3 SSRs/Mb, 3.7 SSRs/Mb and 9 SSRs/Mb respectively. The most frequent tri- to hexanucleotide repeats appeared to be extremely variable among the nine fungal genomes and specific to each genome.

**Microsatellite density and most common classes in the *L. bicolor* genome (Table 3)**

The assembly (v.1.0) of 65 Mb haploid genome of *L. bicolor* contained a total of 277,062 SSRs repeated at least five times. Mononucleotide SSRs outnumbered all other classes (4,152 SSRs/Mb). The SSR density decreased according to the number of nucleotides: 31 SSRs/Mb for the dinucleotides, 22.3 SSRs/Mb for the trinucleotides, 4.1 SSRs/Mb for the tetra, penta and hexanucleotides. A chi-square test ($\chi^2$= 71,017; ddl= 3) showed that SSRs were not distributed at random in the genome but that their density was significantly different according to coding and non-coding regions (Figure 2 A, 2B). Twenty-five percent of SSRs (69,943 SSRs) were located in the genic regions, while by difference, 74.8% (207,119 SSRs) were located in the intergenic regions, which represented, 68,7 % of the genome. In this study, the TEs, which may produce transcripts, were considered as part of intergenic regions. These two regions also differed by the density of the mononucleotides: 4,433 SSRs/Mb in the intergenic regions against 3,350 in the genic regions. They also differed by the density of the dinucleotides: 38 SSRs/Mb in the intergenic regions, and 14.7 SSRs/Mb in the genic regions. In the genic regions, the trinucleotides (28.9 SSRs/Mb) were more frequent than the dinucleotides (14.7 SSRs/Mb).

In the genic regions, the SSR density differed among UTR regions, exons and introns (Figure 3). There were no SSRs in the 3'UTR regions, while the 5'UTR regions exhibited a density of 3,802 SSRs per Mb (1,825 SSRs in total). The highest density

was observed in the introns, 5,396 SSRs/Mb against 3,395 SSRs/Mb in the exons. Nevertheless, due to the size of the exons, the total number of SSRs in the exons (69,943) exceeded that found in the intronic regions (37,070). The exons also differed from the other genic regions by their number of trinucleotides (595 in total against 88 in the introns and 8 in the 5'UTR region). As expected, there were few differences in

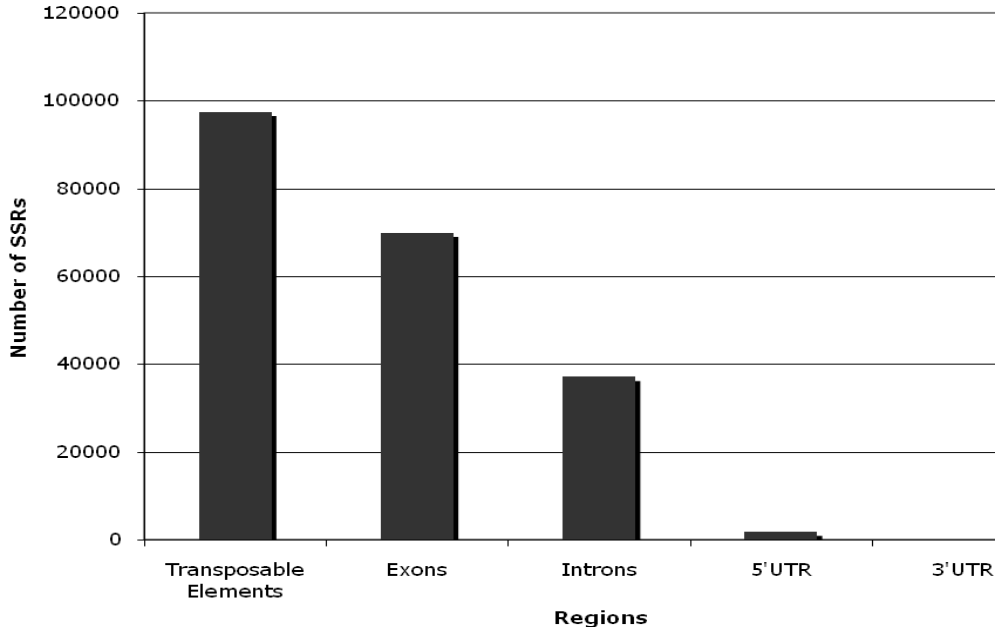



**Figure 3 - Number of SSRs in different regions of the *L. bicolor* genome.**

di- and trinucleotide SSR density between exons and ESTs. The di-, tri- and hexanucleotide density was 14.7 SSRs/Mb, 28.9 SSRs/Mb and 0.45 SSRs/Mb in exons, whereas 22.3 SSRs/Mb, 21 SSRs/Mb and 0.21 SSRs/Mb in ESTs respectively. In order to reveal a positional specificity of SSRs in genes, the density in various area was determined. We divided the genic regions consisting of exons and introns, into the first exon/intron, the second exon/intron, the third exon, the second last exon/intron and the last exon/intron (Figure 4). The SSR density was constant in the various exonic positions while it was higher in the second last intron (4,648 SSRs/Mb) and the last intron (3,802 SSRs/Mb).

SSR motifs were located in 85.8% of the protein-coding genes (17,152). Although 52% of these genes (10,392) exhibited SSRs in both exons and introns, 5% (990) and 28.9% (5,770) of coding-genes only contained SSRs in intronic or exonic

regions, respectively. Eigthy-one percent of the genes (16,162) exhibited SSRs in their exons. Three percent of the genes (580) contained tri- and hexanucleotide motifs (498 only in exons, 73 only in introns, and 9 in both exons and introns). Most of these genes code for unknown proteins. Nevertheless, few genes code for homologues of glycosyltransferase and with proteins containing WD40, leucine-zipper and zinc-finger domains.

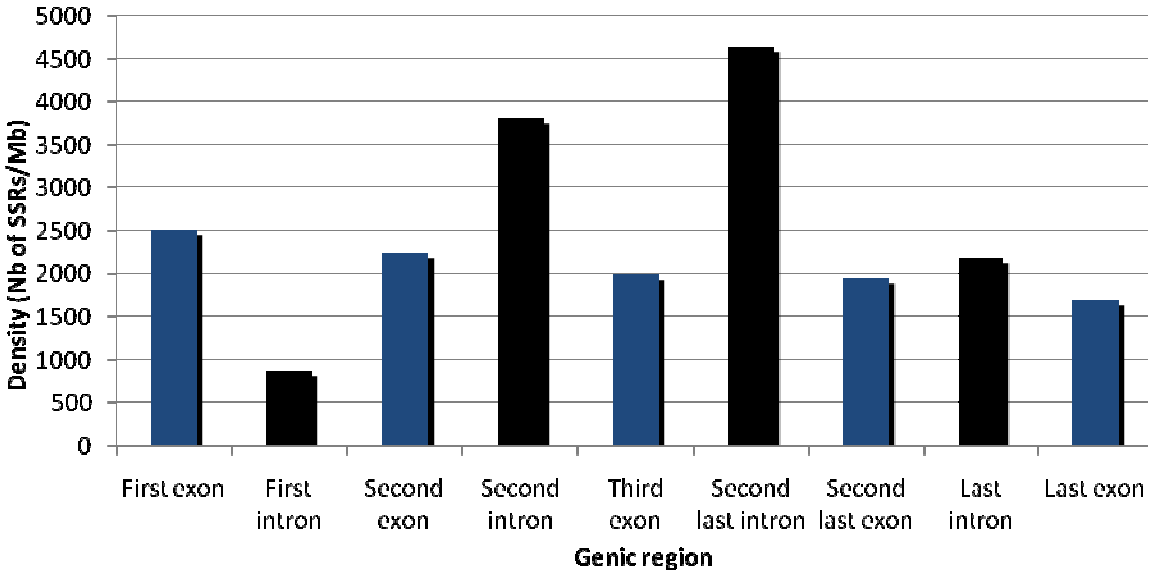



**Figure 4 - SSR density in genic regions (exons and introns).** Exons and introns are classified into the following categories: the first exon/intron, the second exon/intron, the third exon, the second last exon/intron and the last exon/intron. All genic regions throughout the *L. bicolor* genome sequence are considered in measurements. The SSR density is defined as the number of SSRs per Mb.

The main amino-acid repeats corresponding to the tri- or hexanucleotide motifs were proline, serine, threonine, leucine and alanine.

TEs, which represented 24% of the assembly [5], exhibited the highest SSR density (6706 SSRs/Mb) among the different regions of the genome (Figure 3, Table 3) and contained more that one third of the total SSRs (97,240 for a total of 277,062 SSRs). The density of the mononucleotides (6,583 SSRs/Mb) was much higher in the TEs than in all the other regions of the genome. The density of the di- and trinucleotides also was the highest in the TEs, but in a lower proportion. Among all SSRs found within TEs, 28.4% were found in elements of class I (24.5% in long terminal repeat (LTR) retrotransposons, 3.8% in long interspersed nuclear elements (LINE), 0.1% in large retrotransposons derivative (LARD), 26% in elements of class II (25.4% in

terminal inverted repeats (TIRs), 0.5% in Helitrons, 0.1% in miniature inverted-repeat transposable elements (MITE), and 45.6% in TEs of unknown classes specific to *L. bicolor* [5]. SSRs integrated within TEs of class I, were located in both the 5' and 3'ends of LTR, and in the polyA tail of LINE. However, the distribution of some of SSRs was different according to the type of LTR retroelements. In most of the *copia-like* elements, the main regions containing SSRs were the 3' end of the 5' LTR, which starts by a SSR. Similarly, in *gyspy-like* elements, the second LTR always starts in 5' by SSRs of approximately 40 bp. Within class II TEs, SSRs were apparently located into non-coding sequences. Nevertheless, incomplete characterization of these class II elements makes difficult to assign a location to all SSRs.

### Development of microsatellite markers

A pool of 216 SSRs (di- to hexanucleotide motifs with a minimum repeats ≥ 8), distributed throughout the largest scaffolds of the *L. bicolor* genome [21] were selected for primer design. Among these 216 SSRs, 152 were amplified on the dikaryotic strain *L. bicolor* S238N and on the sequenced haploid *L. bicolor* strain S238N-H82. Only 78 of these 152 microsatellite loci appeared to be polymorphic in the *L. bicolor* S238N progeny (Table 4). Thirteen were located in exons, five in introns, 14 in TEs and 46 in remaining regions (Table 5). Amongst the 14 coding sequences containing a SSR, most are coding for proteins with unknown function. However, a histidine-rich protein (JGI ID: 328508) and a SWIFT protein (JGI ID: 328751) contained repeated motifs and the alleles showed a mendelian segragation (3:1) in the progeny.

## Discussion

We have compared the SSR density (≥ 5 repeats) of *L. bicolor* genome with eight other fungal genomes belonging to two phylogenetic classes, the Ascomycota and the Basidiomycota (Table 2). These two classes did not differ by their SSR density, *S. cerevisiae* in the Ascomycota and *L. bicolor* in the Basidiomycota exhibiting the highest SSR density. The differences in SSR density among the different fungal species could be explained by different genomic organization among the species. This differential distribution among the fungal species could be the result of differences in mutability or bias in efficiency of the mismatch repair system [22]. It appeared that, in fungal genomes, SSR abundance was neither inversely nor directly proportional to the genome

size, as reported in other studies [8].

Mononucleotide motifs outnumbered all other SSR classes in all fungal genomes. The proportion of A was always greater than that of C with no differences between the two classes. However, the proportion of A and C mononucleotides was almost equal (52% versus 46%) in the basidiomycete *C. cinerea*. In most of the nine fungal genomes, after the mononucleotides, the most represented SSR motifs, were the dinucleotides, the GA motif being found in 3 basidiomycetes and 2 ascomycetes. From the tri- to the hexanucleotides, there were no common motifs. These repeat types appeared to be characteristic of the species examined. The diversity of microsatellite motifs seems to give to each species a unique signature.

In *L. bicolor* genome, the highest SSR density (repeats ≥ 5) was observed in TEs (6,706 SSRs/Mb) (Figure 3, Table 3). Several studies have shown that SSRs are often associated with retrotransposons [24, 25, 26]. Some are component of active TEs spreading throughout the genome. Others could act as a 'landing pad' for transposable element insertion. A third category, could arise after the integration of an extended and polyadenylated retro-transcript into the genome [24]. Among the LTR retrotransposon sequences analyzed, we have identified SSR extended to both 5' and 3' ends. Thus, it would seem that these SSRs have arisen by mutation followed by expansion of the same proto-SSR [27], then spread throughout the *L. bicolor* genome as a component of an active retrotransposable element. SSRs and TEs were not distributed at random in *L. bicolor* genome. They often showed links, revealing highly targeted regions for SSR insertion in TEs. The positioning of these SSRs in TEs allowed us to classify them into four subtypes: SSRs located in 3' at the beginning of a TE, SSRs located in 5' at the beginning of a TE, SSRs positioned on each side and SSRs located inside the TE.

The major part of the SSRs into the *L. bicolor* genome (≥ 5 repeats) was represented by mono-, di- and trinucleotides accounting for up to 80% of all SSRs found (Table 1). Tóth *et al.* [19] have examined SSR abundance in several taxonomic groups. Tri and hexanucleotides ≥ 12 repeats generally prevailed in exons of all taxa including yeast and other fungi [19, 28]. According to our results, even if we also take into account the ≥ 12 repeats (Table 1), tri- and hexanucleotides were never predominant in *L. bicolor* exons. The tri- and hexanucleotide density in *L. bicolor* exons were confirmed by the SSR densities found in *L. bicolor* ESTs. The tri- and hexanucleotide (≥ 5 repeats) densities respectively were 28.9 and 0.45 SSRs/Mb in exons against 21 and 0.21 SSRs/Mb in

ESTs. The weak discrepancy observed between exon and EST SSR density is easily explain by the fact that ESTs only represented 45.6% of the total exons.

The present survey revealed a higher proportion of SSRs (repeats ≥ 5) in the intergenic regions of the *L. bicolor* genome than in the genic regions (exome). The density of SSRs decreased slightly from the first exon to the last exon (Figure 4), being almost constant feature of mammalian showed in the study of Fujimori *et al.,* [29]. Assuming repeats ≥ 10 per motif, the SSR density decreased to 58, 41 and 151 SSRs/Mb in the 5'UTR, exons and introns respectively. This density is much lower than in Arabidopsis (2364, 334 and 814 SSRs/Mb respectively) or Rice (3971, 658 and 634 SSRs/Mb) [11]. A feature of the genome of *L. bicolor* is a lack of SSRs in 3'UTR. The average amount of ≥ 5 repeats per gene in *L. bicolor* genome was ~5.3, while it was ~0.0125 if we only considered the ≥ 10 repeats, which was very low compared to Arabidopsis or Rice. It is known that SSRs may play a role in gene transcription, regulation, mRNA splicing, and gene silencing [12]. They are involved in several human diseases, for example by activation of "toxic" proteins. These possible deleterious effects likely explain why SSR density is relatively low in protein-coding regions. On the other hand, the presence of polymorphic SSRs in exons, introns or UTRs could alter the protein structure and/or gene expression and then, play a role in the mechanisms of adaptation, survival and evolution of the organisms. Most of the *L. bicolor* genes exhibiting tri- and hexanucleotide insertion in their exons code for unknown proteins. However, some of these SSR-containing genes code for homologues of a glycosyltransferase and proteins containing WD40, leucine-zipper and zinc-finger domains. Transcript profiling showed that the genes corresponding to glycosyltransferase was expressed during the symbiosis as well as identified genes coding for proteins containing leucine-zipper domains which were also up-regulated [5]. In contrast, the identified genes coding for proteins containing WD40 and zinc-finger domains were not expressed and not regulated (Kohler *et al.,* unpublished data) in the same conditions. It remains to be determined whether those proteins have a role in the symbiotic interactions.

SSRs are often used as genetic markers helping the study of genetic population [7, 9, 16, 17, 20] the construction of genetic maps [21] and the identification of species or individuals [9, 16, 17]. Our work showed that SSRs in a *L. bicolor* progeny were the most polymorphic in the intergenic regions outside the TEs. If it is generalizable, this

result could facilitate the search of polymorphic SSRs, for the construction of genetic maps in other species.

## Conclusions

The assembly (v.1.0) of 65 Mb haploid genome of *L. bicolor* contained a total of 277,062 SSRs corresponding to 8% of the genomic sequence. This is the basiodiomycetes genome analyzed here having the highest density of SSRs. The highest SSR density was observed in TEs. This could support the hypothesis that the *L. bicolor* genome is characterized by a large plasticity, allowing it to adapt both to saprotrophic and symbiotic ways of life. The average amount of ≥ 5 repeats per gene in *L. bicolor* genome was ~5.3. According to Tóth *et al.* [19], tri- and hexanucleotides ≥ 12 repeats generally prevailed in exons of all eukaryote taxa. Althougth, tri- and hexanucleotide were not predominant in exons of *L. bicolor*, according to our results they could play a role in the code of repeated structure of some protein domains involved in transfer processes. Moreover, the presence of SSRs in the exons of 81% of the coding genes of *L. bicolor* could mean that these SSRs have played a key role in the evolution of this symbiotic fungus. Finally, our work showed that the 78 SSR markers developed in *L. bicolor* progeny were most polymorphic in the intergenic regions outside the TEs.

## Methods

### Sequence data

The *L. bicolor* genome (JAZZ assembly v.1.0) was downloaded from publicly available database on the Joint Genome Institute (JGI) web site [23], while the ARACHNE assembly (v. 1.0) of *L. bicolor* genome was downloaded from the INRA *LaccariaDB* [29]. All other fungal genome sequences used in this analysis were downloaded from publicly available databases listed in Table 6.

### ESTs

The *L. bicolor* EST containing 14,312 tentative consensi was downloaded from the INRA *LaccariaDB* [30].

### Transposable elements

TEs were identified by similarity with other fungal genomes and compiled in the INRA *LaccariaDB* [5, 30].

### SSR scanning and analysis of sequences containing SSRs

Using the *L. bicolor* JAZZ and ARACHNE assemblies, the EST and transposon files, the SSRs were located in translated and untranslated regions, in various genic regions (3'UTR, 5'UTR, exons and introns) and in transposable elements. This identification was carried out with specific written PYTHON scripts. On the other hand, all SSR genome scanning were carried out using the microsatellite identification tool (MISA) available from the Plant Genome Resources Center (PGRC) [31]. MISA was used to identify systematically all possible mono- to hexanucleotide motifs of 5, 10 or 12 repeats in the databases selected. Because of the presence of polyA tails, we could not determined, the mononucleotides from the EST file. The MISA tool recorded repeat number and SSR types in ouput files. The data were processed with Microsoft Excel 2004. We have obtained the total number of all perfect repeat types and calculated the number of SSRs per Mb to allow comparison among genome of different sizes.

The coding genes of *L. bicolor* exhibiting tri and hexa SSR motifs in their exons, were blasted (BLAST X) against the publicly available database on the Joint Genome Institute (JGI) web site [23].

### Development of microsatellite markers

### Primer design

The SSRs used as molecular markers were detected on the *L. bicolor* JAZZ assembly (v.1.0) [31] and confirmed on the *L. bicolor* ARACHNE assembly (v.1.0) [30]. With the exception of mononucleotide and compound motifs $\geq$ 8 repeat SSRs, were sample on the 20 larger JAZZ scaffolds and on the 15 larger ARACHNE supercontigs. The online Primer3 tool [32] was used to design primers on sequences flanking microsatellite motifs. Primers were synthesized by Invintrogen (Cergy Pontoise, France).

### Fungal strain and culture conditions

Ninety-one individual homokaryotic (haploid) mycelia, including the S238-H82 strain from which the genome was sequenced [5], were used according to conditions described in Labbé *et al.*, [21].

### DNA isolation

We used the protocol of DNA isolation as described in Labbé *et al.*, [21, 33].

### SSR analysis

We have performed amplification tests, using unlabeled primers (Invitrogen) designed for 216 SSRs firstly chosen, by following the same amplification protocol described below for the polymorphism test. PCR products were analyzed on 2% agarose gels. Then, the 152 successfully amplified SSRs were analyzed for the detection of polymorphism on the progeny of *L. bicolor* S238N, using the following high resolution procedure: SSRs were amplified with one fluorescent dye-labeled primer (D2-PA, D3-PA, D4-PA, WellRED dyes, Proligo, Paris). PCR reactions were carried out in a total volume of 10 µl containing 10 ng of template DNA, 10 ng forward and reverse oligonucleotide primers (Invitrogen), 200 µM of each dNTP, 0.5 U *Taq* DNA polymerase (Qbiogene), 1 µl Qbiogene buffer 10x containing 25 mM $MgCl_2$. PCR was conducted in a Perkin-Elmer Cetus thermocycler 9700 (Applied Biosystems) with 30 cycles of 94˚C for 30 s, 60˚C for 1 min, and 72˚C for 2 min combined with a final extension at 72˚C for 10 min. PCR products were detected, on a CEQ 8000 XL sequencer using the CEQ 8000 V9.0 genotyping module. Before analysis, each amplification product was diluted in 1:10 with deionized water. One µl of this dilution was mixed with 0.5 µl DNA size standard (DNA size standard kit 600 bp, Beckman Coulter) and with 30 µl SLS buffer (Beckman Coulter). The seventy-eigth polymorphic SSRs identified by this procedure were listed in Table 4, with the corresponding primer sequences and their positions, on the *L. bicolor* JAZZ assembly.

## Authors' contributions

JL and CM contributed equally in the *in silico* survey and analysis of SSRs. JL carried out the molecular analyses, performed data acquisition, data analysis and interpretation. EM run the computer analysis and wrote PYTHON scripts for parsing the data. FM and FLT, supervised and contributed to the data analysis. JL, FLT, FM wrote the manuscript.

## Acknowledgements

This work was supported by the Lorraine Region, and INRA through a PhD Scholarship to JL. Funds were also provided by INRA, the European Commission Network of Excellence EVOLTREE. We would like to thank Benoît Barrès, Axelle Andrieux, Christine Delaruelle and Pascal Frey, for their assistance and helpful discussions.

## References

1. Smith SE, Read DJ: **Mycorrhizal Symbiosis**. Second edn. *Academic Press*, London 1997.
2. Martin F, Kohler A, Duplessis S: **Living in harmony in the wood underground: ectomycorrhizal genomics.** *Current Opinion in Plant Biology* 2007, **10**:204-210
3. Martin F, Tuskan Ga, Difazio SP, Lammers P, Newcombe G, Podila GK: **Symbiotic sequencing for the *Populus* mesocosm.** *New Phytologist* 2004, **161**:330-335
4. Le Tacon F, Alvarez IF, Bouchard D, Henrion B, Jackson RM, Luff S, Parlade JI, Pera J, Stenstrom E, Villeneuve N, Walker C: **Variation in field response of forest trees to nursery ectomycorrhizal inoculation in Europe**. In: Read DJ, Lewis DH, Fitter AH, Alexander DJ (eds). *Mycorrhizas in Ecosystems* International, Wallingford 1997, 131-154
5. Martin F, Aerts A, Ahrén D, Brun A Danchin EGJ, Duchaussoy F *et al.*: **Symbiosis insights from the genome of the mycorrhizal basidiomycete *Laccaria bicolor*.** *Nature* 2008, **452**:88-92

6. Henrion B, Di Battista C, Bouchard D, Vairelles V, Thompson BD, Le Tacon F, Martin F: **Monitoring the persistence of *Laccaria bicolor* as an ectomycorrhizal symbiont of nursery-grown Douglas fir by PCR of the rDNA intergenic spacer.** *Molecular Ecology* 1994, **3**:571-580
7. Ashley MV, Dow BD: **The use of microsatellite analysis in population biology: Background, methods and potential applications.** In: Schierwater B, Streit B, Wagner GP, Desalle R (Eds). *Molecular Ecology and Evolution*, Approaches and Applications. 1994, 185-201
8. Karaoglu H, Lee CMY, Meyer W: **Survey of simple sequence repeats in completed fungal genomes.** *Molecular Biology and Evolution* 2005, **22:**639-649
9. Selkoe Ka, Toonen RJ: **Microsatellites for ecologists: pratical guide to using and evaluating microsatellite markers.** *Ecology Letter* 2006, **9:**615-629
10. Estoup A, Angers B: **Microsatellites and minisatellites for molecular ecology: theoretical and empirical considerations.** *Advances in Molecular Ecology*. IOS Press 1998, 55-86
11. Lawson MJ, Zhang L: **Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes.** *Genome Biology* 2006, **7:**R14
12. Li Yc, Korol AB, Fahima T, Nevo E: **Microsatellites within genes: Structure, function, and evolution.** *Mol Biol Evol* 2004, **21:**991-1007
13. Hancock JM: **Microsatellites and other simple sequences: genomic context and mutational mechanisms.** Goldstein DB and Schlotterer C, eds. Microsatellites: evolution and applications. Oxford University Press 1999, 1-9
14. Selosse MA, Jacquot D, Bouchard D, Martin F, Le Tacon F: **Temporal persistence and spatial distribution of strain of the ectomycorrhizal basidiomycete *Laccaria bicolor* in a French forest plantation.** *Molecular Ecology* 1998, **7**:561-573
15. Weber J, Díez J, Selosse MA, Tagu D, Le Tacon F: **SCAR markers to detect mycorrhizas of an American *Laccaria bicolor* strain inoculated in European Douglas-fir plantations.** *Mycorrhiza* 2000, **12**:19-
16. Jany JL, Bousquet J, Gagné A, Khasa DP: **Simple sequence repeat (SSR) markers in the ectomycorrhizal fungus *Laccaria bicolor* for environmental**

**monitoring of introduced strains and molecular ecology applications.** *Mycological Research* 2006, **110**:51-59

17. Prospero S, Black JA, Winton ML: **Isolation and characterization of microsatellite markers in *Phytophtora ramorum*, the causal agent of sudden oak death.** *Molecular Ecology* 2004, **4**: 672-674
18. Lim S, Notley-McRobb L, Lim M, Carter DA: **A comparison of the nature and abundance of microsatellites in 14 fungal genomes**. *Fungal Genet Biol* 2005, **41**: 1025-1036
19. Tóth G, Gáspári Z, Jurka J: **Microsatellites in different eukaryotic genomes: survey and analysis.** *Genome Res* 2000, **10:**967-981
20. Garnica DP, Pinzon AM, quesada-Ocampo LM, Bernal AJ, Barreto E, Grunwald NJ, Restrepo S: **Survey and analysis of microsatellites from transcript sequences in *Phytosphtora* species: frequency, distribution, and potential as markers for the genus.** *BMC Genomics* 2006, **7**: 245-262
21. Labbé J, Zhang X, Yin T, Schmutz J, Grimwood J, Martin F, Tuskan GA, Le Tacon F: **A genetic linkage map for the ectomycorrhizal fungus *Laccaria bicolor* and its alignement to the whole-genome sequence assemblies.** *New Phytologist* 2008, **180:**316-328
22. Harr B, Todorova J and Schlotter J: **Mismatch repair-driven mutational bias in *D. melanogaster*.** *Mol. Cel.* 2002, **10:**199-205
23. **Joint Genome Institute (JGI) *Laccaria bicolor Home*** [http://genome.jgi-psf.org/Lacbi1/Lacbi1.home.html]
24. Ramsay Luke, Macaulay M, Cardle L, Morgante M, Ivanissevich S, Maestri E, Powell W, Waugh R: **Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley.** *Plant Journal* 1999, **17:**415-425
25. Venturi S, Dondini L, Donini P and Sansavini S: **Retrotransposon characterisation and fingerprinting of apple clones by S-SAP markers.** *Theorical and Applied Genetics* 2006, **112:**440-444
26. Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, and McCouch S: **Computational and Experimental Analysis of Microsatellites in Rice (Oryza sativa L.): Frequency, Length Variation, Transposon**

**Associations, and Genetic Marker Potential.** *Gen. Research* 2001, **11:**1441-1452

27. Weber JL: **Informativness of human (dC-dA)n.(dG-dT)n polymorphysm**. *Genomics* 1990, **7**: 524-530
28. Metzgar D, Bytof J, Wills C: **Selection against frameshift mutations limits microsatellite expansion in coding DNA.** *Genome Res* 2000, **10:**72-80
29. Fujimori S, Washio T, Higo K, Ohtomo Y, Murakami K, Matsubara K, Kawai J, Carninci P, Hayashizaki Y, Kikushi S, Tomita M: **A novel feature of microsatellites in plants: a distribution gradient along the direction of transcription.** *FEBS Letters* 2003, **554**: 17-22
30. **INRA *LaccariaDB*** [http://mycor.nancy.inra.fr/IMGC/LaccariaGenome/]
31. **Plant Genome Resources Center (PGRC)** [http://pgrc.ipk-gatersleben.de]
32. Koressaar T, Remm M: **Enhancements and modifications of primer design program Primer3.** *Bioinformatics* 2007, **23**: 1289-1291
33. Di Battista C, Selosse MA, Bouchard D, Stenström E, Le Tacon F: **Variations in symbiotic efficiency, phenotypic characters and ploidy level among different isolates of the ectomycorrhizal basidiomycete *Laccaria bicolor* strain S238**. *Mycol Res* 1996, **100**: 1315-1324

**Figure 2 - Percentage and density of genes, transposable elements and SSRs found on the linkage group 1 (LG1) of *L. bicolor*.**
**(A)** Percentage of genes (blue), TEs (red) and SSRs (yellow). A1 Relative percentage of genes, TEs and SSRs. A2 Relative percentage of genes and SSRs. A3 Relative percentage of TEs and SSRs. **(B)** The density is graphically represented as the number of genes, TEs and SSRs found in a window of 10 kb.

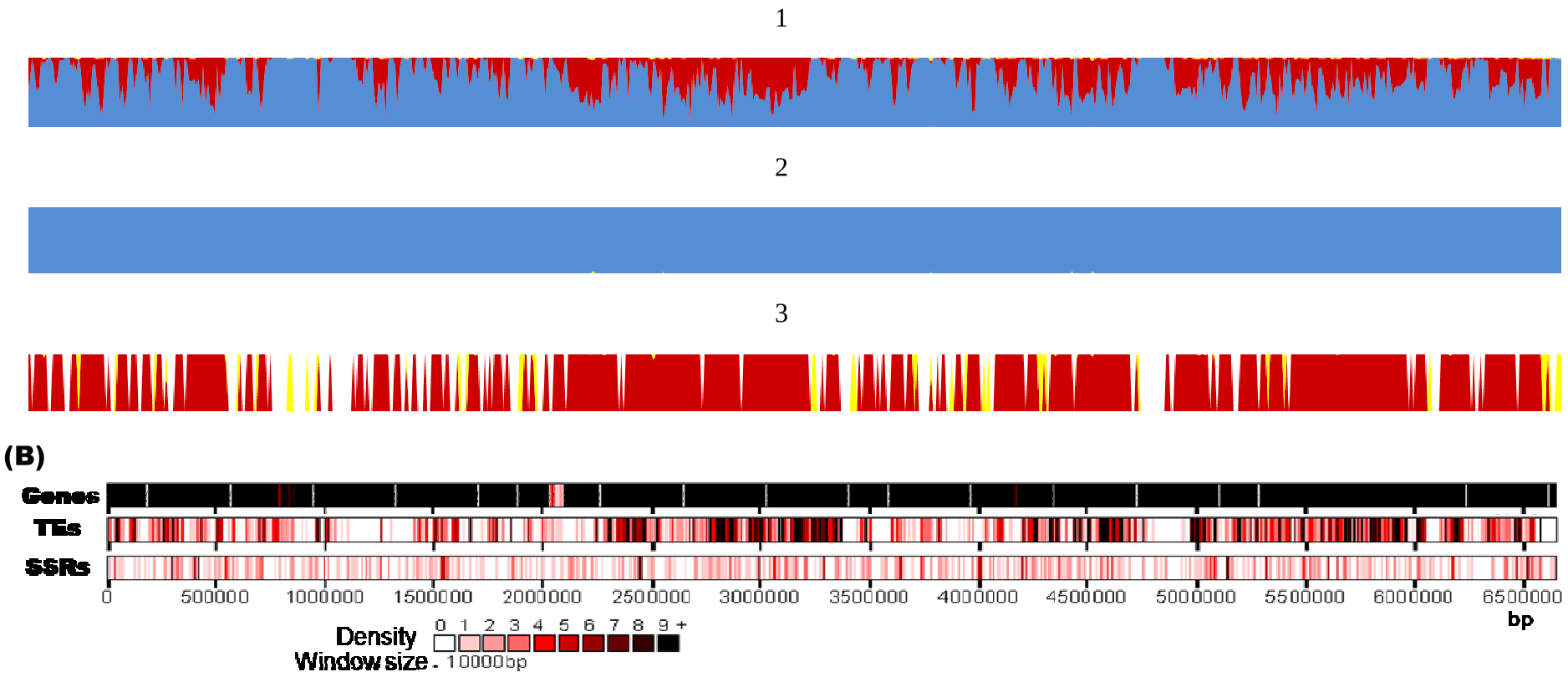


**Table 1 - Number of SSRs in 5' UTR, exons and introns of the *L. bicolor* genome according to the number of repeats.**

| Regions | Mono | | | Di | | | Tri | | | Tetra | | | Penta | | | Hexa | | | All SSRs | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ≥ 5 repeats | ≥ 10 repeats | ≥ 12 repeats | ≥ 5 repeats | ≥ 10 repeats | ≥ 12 repeats | ≥ 5 repeats | ≥ 10 repeats | ≥ 12 repeats | ≥ 5 repeats | ≥ 10 repeats | ≥ 12 repeats | ≥ 5 repeats | ≥ 10 repeats | ≥ 12 repeats | ≥ 5 repeats | ≥ 10 repeats | ≥ 12 repeats | ≥ 5 repeats | ≥ 10 repeats | ≥ 12 repeats |
| 5' UTR | 3,565 | 39.5 | 14.6 | 20.8 | 2 | 0 | 16.7 | 0.26 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3,802 | 58 | 14.6 |
| Exons | 3,350 | 6.60 | 2.23 | 14.7 | 3.64 | 0.05 | 28.9 | 0.87 | 0.87 | 0.24 | 0 | 0 | 1.30 | 0.05 | 0 | 0.45 | 0.20 | 0.15 | 3,395 | 41 | 3.3 |
| Introns | 5,335 | 58.6 | 41 | 45 | 16.3 | 1.75 | 12.8 | 0.35 | 0 | 1.86 | 0.15 | 0 | 1.31 | 0.3 | 0 | 0.29 | 0.48 | 0.15 | 5,396 | 151 | 42.9 |

**Table 3 - Number, density per Mb and percentage of SSRs in the different categories and in the different regions of the *L. bicolor* genome.** ND: not determined

| Regions | Genome size Mb | Percentage of genome | Mono | | | Di | | | Tri | | | Tetra | | | Penta | | | Hexa | | | All SSRs | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Number | N/Mb | % | Number | N/Mb | % | Number | N/Mb | % | Number | N/Mb | % | Number | N/Mb | % | Number | N/Mb | % | Number | N/Mb | % |
| 5' UTR | 0.48 | 0.74 | 1,807 | 3,565 | 0.66 | 10 | 20.8 | 0.49 | 8 | 16.7 | 0.54 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1,825 | 3,802 | 0.66 |
| Exons | 20.6 | 31.3 | 69,004 | 3,350 | 25.3 | 303 | 14.7 | 14.8 | 595 | 28.9 | 41 | 5 | 0.24 | 9 | 26 | 1.30 | 13 | 10 | 0.45 | 19 | 69,943 | 3,395 | 25.2 |
| Introns | 6.87 | 10.5 | 36,651 | 5,335 | 13.4 | 309 | 45 | 15 | 88 | 12.8 | 6 | 11 | 1.86 | 20 | 9 | 1.31 | 4.5 | 2 | 0.29 | 3.8 | 37,070 | 5,396 | 13.4 |
| 3' UTR | 0.32 | 0.49 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Transposable elements | 14.5 | 22 | 95,459 | 6583 | 35 | 759 | 52.3 | 37 | 794 | 55 | 54 | 29 | 2 | 53 | 187 | 12.9 | 27 | 12 | 0.83 | 23 | 97,240 | 6,706 | 35 |
| Intergenic regions | 45.2 | 68.7 | 200,354 | 4,433 | 73.3 | 1,716 | 38 | 83.8 | 850 | 18.8 | 58 | 50 | 1.11 | 91 | 175 | 3.9 | 88 | 43 | 0.95 | 81 | 207,118 | 4,582 | 74.8 |
| ESTs | 9.4 | 14.3 | ND | ND | ND | 210 | 22.3 | 10.3 | 196 | 21 | 13.3 | 3 | 0.32 | 5.5 | 9 | 0.95 | 4.5 | 2 | 0.21 | 3.8 | ND | ND | ND |
| Total | 65.8 | 100 | 273,238 | 4,152 | 98.62 | 2,047 | 31 | 0.74 | 1,469 | 22.3 | 0.53 | 55 | 0.08 | 0.02 | 200 | 3 | 0.07 | 53 | 0.8 | 0.02 | 277,062 | 4211 | 100 |

**Table 2 - Density and percentage of the most frequent SSR motifs in nine fungal genomes.** *[a]Percentage inside the given SSR category.*

| Organisms | Analyzed genome size (Mb) | Density of SSRs (Nb/Mb) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Mono[a] | Di | Tri | Tetra | Penta | Hexa | Total |
| *L. bicolor* | 65.8 | $A_n$(70%) | GA(4.5%) | CCA | AGGG | ACTGG | CTAACC | |
| | | $C_n$(28%) | AT (4%) | GAC | CCTC | CAGTC | AACCAC | |
| | | **4,152** | **31** | **22.3** | **0.8** | **3** | **0.8** | **4,211** |
| *P. chrysosporium* | 35.8 | $A_n$(62%) | CG (9%) | CGC | CCGC | AGTCA | CGAGGA | |
| | | $C_n$(35%) | GC (8%) | GCC | AACG | CCTCC | AAAAGG | |
| | | **1,644** | **38.4** | **15.6** | **0.4** | **00.8** | **0.4** | **1,699** |
| *C. cinerea* | 36.8 | $A_n$(52%) | GA (8%) | CGA | TGAA | AAGGT | CCCTAA | |
| | | $C_n$(46%) | AG (5%) | GAC | ACCA | AAGTA | AAAGGA | |
| | | **2,594** | **30.1** | **21.4** | **1** | **0.4** | **0.7** | **2,647** |
| *U. maydis* | 19.9 | $A_n$(69%) | AG (6%) | GCA | AGTC | CAGCA | AATCAC | |
| | | $C_n$(25%) | GA (5.5) | AGC | GCAA | ACAGC | CAGCAA | |
| | | **1,795** | **50** | **37.6** | **3.3** | **3.7** | **9** | **1,898** |
| *C. neoformans* | 19.5 | $A_n$(66%) | GA(8.2%) | GAA | CATC | AAGGA | AAGGAA | |
| | | $C_n$(32%) | AG(6.4%) | AAG | ATGG | - | AGAAAG | |
| | | **3,564** | **33** | **19** | **2.4** | **0.05** | **1.2** | **3,620** |
| *A. nidulans* | 30.4 | $A_n$(61%) | AT (4.5%) | AAG | AAAG | AAACC | CCCTAA | |
| | | $C_n$(37%) | GA (4.4%) | GGA | AAAC | AAAAG | AAGAAA | |
| | | **2,506** | **23.3** | **8.9** | **0.6** | **0.7** | **0.6** | **2,540** |
| *M. grisea* | 42.3 | $A_n$(68%) | CA (3.1%) | GCC | GGTA | AAAAC | ACTTCG | |
| | | $C_n$(30%) | GA (3.1%) | CAG | CCTA | GTACA | GCCAGA | |
| | | **4,148** | **33** | **34** | **4.8** | **0.7** | **0.9** | **4,221** |
| *N. crassa* | 39.8 | $A_n$(57%) | TA (14%) | CAA | GGTA | GTGAA | AGGTAC | |
| | | $C_n$(38%) | AT (8%) | ACC | CTAC | CTACA | GGACAA | |
| | | **3,567** | **78.3** | **95** | **18.6** | **4.6** | **3.3** | **3,767** |
| *S. cerevisiae* | 11.9 | $A_n$(93%) | TA (9%) | CAA | AAAT | AACAA | AAAACA | |
| | | $C_n$ (5%) | AT (9.5%) | ACC | AATA | AATAC | AAGAAC | |
| | | **6,425** | **60** | **29** | **1.3** | **0.5** | **1.2** | **6,517** |

**Table 4 - Set of primers designed to amplify polymorphic SSR regions in *L. bicolor* genome.**

| SSR Name | SSR Type | Primer sequence | Position on *L. bicolor* genome v1 |
|---|---|---|---|
| *INRAjl03* | $(GAGCAG)_{25}$ | (F)CTGAGCTAGAGCAGCTGGTA<br>(R)GACAATCGCACATTGGATAG | Scaffold 34<br>334200-334406 |
| *1184M15* | $(GTCCTT)_5$ | (F)GTTTGACTTTCCAGCGCTTC<br>(R)CGCCTCTGATGAGGATGAGT | Scaffold 35<br>222848-223068 |
| *1184M16* | $(CTGGTG)_5$ | (F)TATGCCGTTGCGTTTTCATA<br>(R)CCCTTTGCAAGTTCTCTTCG | Scaffold 16<br>384272-611895 |
| *1184M4* | $(AGG)_9$ | (F)GCTTCGCAAACAAAGAAAGC<br>(R)AACAGGGGCATTGCTCTACA | Scaffold 16<br>611337-611584 |
| *1184M11* | $(CCATA)_5$ | (F)GGAAATAGAATCAAGTCTTGTTGTTG<br>(R)GGGTTTTCTGACTTCCTCTGC | Scaffold85<br>1410-1728 |
| *1205M21* | $(CTGGA)_5$ | (F)CGTCGTTGAGTGAAGTGTGG<br>(R)TGTGTTGGGAGAGTCCAGTG | Scaffold 18<br>217884-218081 |
| *1205M17* | $(ACTGG)_7$ | (F)TGGACTCCGACTGGACTCAT<br>(R)CATTCCACCACTGTTCATGC | Scaffold 18<br>219444-219463 |
| *1205M14* | $(TCC)_{11}$ | (F)TTGATGCTCTCCGTCCTCTT<br>(R)GCCTGGTTTGCTTCCTGTAG | Scaffold 18<br>322388-322591 |
| *1205M16* | $(ACTGG)_5$ | (F)ACCTGGGGGAAGTATGGATT<br>(R)TGTGTTGGGAGAGTCCAGTG | Scaffold 18<br>829525-829643 |
| *INRAjl107* | $(TCA)_{14}$ | (F)TCCAACAATTCCAGCTCCTC<br>(R)GATTCATCTTGGGGACCAAC | Scaffold 10<br>679442-679630 |
| *INRAjl112* | $(CAC)_{12}$ | (F)CCCCTTGCTATGGGATGTG<br>(R)AAATTTGAGGGCCTTGCAG | Scaffold 17<br>313582-313770 |
| *1205M12* | $(CAC)_{12}$ | (F)CTGCCTGAGGTGCCTTTTT<br>(R)GGGTCTGCAATTTGGAACAG | Scaffold 17<br>313506-313726 |
| *1205M18* | $(ACTGG)_7$ | (F)CTGGGACTGGACTCAGGAAA<br>(R)AGCAGGAGGTGGTGGTAGTG | Scaffold 17<br>283452-283658 |
| *1199M3* | $(GT)_{12}$ | (F)ATACCAAAGGCCCTCGAACT<br>(R)CAGTATCTGCAACCCAAGCA | Scaffold 15<br>944503-944700 |
| *1199M9* | $(AGG)_9$ | (F)AGAACTTCGCGAGTTTTGGA<br>(R)GGGCGTCTTGTTATTCGGTA | Scaffold 15<br>481624-481488 |
| *1199M6_1* | $(CTT)_9$ | (F)AGGCTGAAATGCAAGCCAT<br>(R)CATTGAAGGGATCGGAGAAA | Scaffold 15<br>243706-243928 |
| *1199M11* | $(TCC)_8$ | (F)GGATCCCGTTTCTTCTTCCT<br>(R)GATGAAGGGCATCCTGACAC | Scaffold 11<br>870603-870828 |
| *1209M8* | $(TGG)_8$ | (F)CTTTGAGATACGGCGGACTT<br>(R)ATTACGGGGTGTCATCATCC | Scaffold 8<br>630699-630878 |
| *1209M4* | $(CAGTC)_6$ | (F)TGTGTTGGGAGAGTCCAGTG<br>(R)TATCCCTGAGGGCCAATACA | Scaffold 64<br>66123-66364 |
| *1209M3* | $(CATAC)_6$ | (F)AATCCCATGTTGCATTCTCA<br>(R)TTTTCTGACTTCCTCTGCTTGA | Scaffold 84<br>1413-1662 |
| *1213M6* | $(GT)_{17}$ | (F)GCCAAATTACCTGAGCAAGG<br>(R)CGTCAACGTATTTGGGGACT | Scaffold 37<br>469159-468977 |
| *1206M3* | $(AG)_{11}$ | (F)GAATCCCATCCCCTCAAGTT<br>(R)CATCTCGGGCTAGGATTGAA | Scaffold 75<br>165228-166050 |
| *1206M5* | $(AG)_{11}$ | (F)TGAGGAGGGTGGAGTACCAG<br>(R)ATGACCGACATTTTCGAGGT | Scaffold 19<br>120592-120850 |
| *1206M2* | $(GGA)_9$ | (F)ATACCAAAGGCCCTCGAACT<br>(R)CATAGGTGCAGGTCAGAGCA | Scaffold 19<br>136336-136595 |
| *1207M10* | $(GAC)_{10}$ | (F)CGCGTCACGTCAACACAC<br>(R)TTGACATGTCGCCGTCGT | Scaffold 47<br>231939-232190 |
| *1207M4* | $(CA)_{12}$ | (F)AGGCTGAGAAAGTGGGAACA<br>(R)CCCTCAAACTTGTGCACTGA | Scaffold 23<br>434731-434912 |
| *1214M3* | $(GT)_{17}$ | (F)ATACCAAAGGCCCTCGAACT<br>(R)CTCTACCCGTACCCCATGAA | Scaffold 4<br>849407-849626 |
| *1206M1* | $(GT)_{12}$ | (F)CCCTTGAACTTGTGCACTGA<br>(R)AACAAATCCTGGAGGGCATT | Scaffold 19<br>155532-155673 |

Table 4 continued

| SSR Name | SSR Type | Primer sequence | Position on *L. bicolor* genome v1 |
|---|---|---|---|
| *INRAjl07* | $(CATAC)_{15}$ | (F)CTGTTCTCGCGAAGAAGGTC<br>(R)GGGAAATGTGATGAACCGGC | Scaffold 47<br>330033-330468 |
| *1214M6* | $(GA)_{11}$ | (F)TCTTTCACGATGACCCAGTG<br>(R)GCTTTGTTTGGCATCCATTC | Scaffold 12<br>955770-956008 |
| *INRAjl06* | $(AAAAAG)_{24}$ | (F)CGAGTATCCAGTATCTCGCTTTCCTAC<br>(R)AGCATCGTGAGATGAGCAGCTGCTT | Scaffold 9<br>611086-611543 |
| *1168M2* | $(CCT)_{9}$ | (F)TTCTCGACATGGTTGGGATT<br>(R)GGTGCACAAGAGGAAGAAGG | Scaffold 9<br>531727-531953 |
| *1168M4* | $(GAG)_{10}$ | (F)CGAGCTTTGGGAGAAGTTTG<br>(R)AGGATTTATGCCACCACAGC | Scaffold 9<br>479751-479919 |
| *1195M6* | $(GTG)_{9}$ | (F)GACAACGTCAAGGCCAAGAT<br>(R)TTGTAGTCGGAGAGGGTTCG | Scaffold 10<br>924213-924395 |
| *1192M5* | $(GGT)_{10}$ | (F)ACCTTTTTGTTGCTCGTTCG<br>(R)AGTCAGGACAACCGTGAAGG | Scaffold 6<br>694909-695111 |
| *1207M9* | $(CTC)_{8}$ | (F)TCTAGACGCTTCGGATCCAT<br>(R)AAGGCCAAGAAGGGAAAGAG | Scaffold 5<br>1771432-1771670 |
| *1207M5* | $(AGA)_{9}$ | (F)GGAGCCAAACGTAAATTGGA<br>(R)GCGGACTTGATCCTCACTCT | Scaffold 34<br>399735-399904 |
| *1192M4* | $(GGT)_{8}$ | (F)TGACGAAGATGGCTCAGATG<br>(R)GCATGCTTGTAGCCTTGTCA | Scaffold 6<br>1318168-1318340 |
| *1192M5* | $(GGT)_{10}$ | (F)AGTCAGGACAACCGTGAAGG<br>(R)ACCTTTTTGTTGCTCGTTCG | Scaffold 6<br>694890-695073 |
| *1192M8* | $(ATGGG)_{6}$ | (F)TGGGAGTGTTGGTATGGTATC<br>(R)AATCCCTTGTTGCACTCTCA | Scaffold 3<br>390536-392448 |
| *1213M10* | $(CA)_{13}$ | (F)TGAAGTGGTGTCCGGTGATA<br>(R)ATACCAAAGGCCCTCGAACT | Scaffold 3<br>168919-169097 |
| *1213M15* | $(TCG)_{8}$ | (F)GTCACTGGTGGTGGTCGTC<br>(R)ACCAAACAAATGCAACACCA | Scaffold 3<br>270319-270516 |
| *1213M18* | $(ATGGT)_{5}$ | (F)AGAACTTCGCGAGTTTTGGA<br>(R)CCCTCATACCAATTCCATCG | Scaffold 3<br>425400-756328 |
| *1199M8* | $(GGT)_{10}$ | (F)CGTCTGGGTCTTCTCTGAGG<br>(R)ATGGCCCATCACTGTTCTTC | Scaffold 15<br>618334-618167 |
| *INRAjl115* | $(TTG)_{12}$ | (F)TCCTCAGATGGGGATCATTC<br>(R)ACAACCACCATGCAACAGAC | Scaffold 34<br>347048-347235 |
| *INRAjl114* | $(CAA)_{13}$ | (F)TTAGTGGGCTTTGCTTGTCC<br>(R)GGGTAGTATTTGGGGAGCAG | Scaffold 7<br>410499-410679 |
| *INRAjl106* | $(GTA)_{13}$ | (F)ACAATGGGTGGTGTCGTTG<br>(R)GCCTACATAGAGGCCAAAGG | Scaffold 23<br>492416-492435 |
| *INRAjl108* | $(AC)_{15}$ | (F)ACATGTTGCATGGGGATGAC<br>(R)TACCAAAAGCCCTCGAACTC | Scaffold 3<br>685905-686035 |
| *INRAjl110* | $(GAG)_{14}$ | (F)ATGAGGATGAGGATGGCAAG<br>(R)AGTGGCAGAGATCGTTTTCG | Scaffold 65<br>42219-42446 |
| *INRAjl111* | $(TGG)_{12}$ | (F)TGAGGTTGATTGACGGTGTG<br>(R)TCAACCACCTCATCTCCATTC | Scaffold 7<br>1221384-1221549 |
| *INRAjl113* | $(CAA)_{14}$ | (F)ACCACGACCACCACACAAC<br>(R)GAAAGTTGCCATCGACGTG | Scaffold 7<br>1397966-1398100 |
| *INRAjl120* | $(TC)_{11}$ | (F)CCCCGTCAGACATACCTCTC<br>(R)TCGTTGAATTGGGGTAAAGC | Scaffold 75<br>165169-165352 |
| *INRAjl121* | $(AC)_{12}$ | (F)CACGACGATACGCTGATACG<br>(R)AATACCAAAGGCCCTCGAAC | Scaffold 14<br>916425-916487 |
| *INRAjl123* | $(ACA)_{32}$ | (F)CACAAACGGCAACAGCAAC<br>(R)TCGTCCTCACATGTCGTGTC | Scaffold 21<br>572180-572360 |
| *INRAjl124* | $(ACC)_{8}$ | (F)CGAGTTCTGAGTGGGGTTTC<br>(R)GGGAGATGACCAAGATGCAG | Scaffold 1<br>1319636-1319810 |
| *INRAjl125* | $(GTT)_{19}$ | (F)CAACGGCTCCTTTATGCTTG<br>(R)GGCCCAATATTCCTCACTTG | Scaffold 8<br>126215-126391 |

Table 4 continued

| SSR Name | SSR Type | Primer sequence | Position on *L. bicolor* genome v1 |
|---|---|---|---|
| *1168M9* | (TCA)6 | (F)TCATCAAAGCCTTACCCATTG<br>(R)TGCAAAATCGTGAGCATTTC | Scaffold 9<br>433917-434154 |
| *1195M21* | (ACA)10 | (F)GATGACGACGTGCAAGGAT<br>(R)GTTCGGCACCTCAAGCGCCTGTGCA | Scaffold 59<br>73921-74716 |
| *1195M22* | (GAAG)9 | (F)AAGGGCGGAAAGAAGAGAAG<br>(R)TTGTGATGACGCATCAGAAGGAG | Scaffold 53<br>112199-112422 |
| *1213M23* | (CCA)7 | (F)TCGTCGAAACGTCCAACATA<br>(R)GTAAGAAGGCGCCAAGGAG | Scaffold 3<br>1098015-1098211 |
| *1213M24* | (GT)7 | (F)TGTCGTGCTATCGCTGTTTC<br>(R)ACCAATGGGTGGTGAAGAAG | Scaffold 29<br>315768-315988 |
| *1213M25* | (CAC)7 | (F)CCTACCGGGACAAACGAATA<br>(R)CGACGACTCGGGAGTTAGAG | Scaffold 3<br>2619127-2619345 |
| *1213M26* | (ACCCTA)5 | (F)CTTGGAATCAAATGGCAGTTA<br>(R)GCAGCGAAGAGGTTTGCTAT | Scaffold 86<br>97001-97783 |
| *1214M15* | (CCTAAC)5 | (F)CCCAGACCTAACCCTAACCC<br>(R)CATGTCAGACTTCACTTGCAC | Scaffold 10<br>933884-934117 |
| *1207M20* | (CAG)6 | (F)AGGACAGCCAGGACAACAAC<br>(R)TATGGTCTGTGGCTGCTGAG | Scaffold 5<br>512541-512692 |
| *1207M21* | (GGT)6 | (F)ATGGGCAGCCTAACACTCAC<br>(R)GGTGCCTAAACAGTCCATCC | Scaffold 5<br>780119-780330 |
| *tel_1_1192* | (G)5 (A)6 | (F)TTCGTCAAAATGTCGGAGTG<br>(R)TCCGAAATCGACGAAATTCT | Scaffold 6<br>1105773-1105933 |
| *tel_1_1181* | (CCTAAC)7 | (F)GGCTTAAGGCCGTAAAGGAT<br>(R)CCGACTGATTTGGTGGGTTA | Scaffold 257<br>5-126 |
| *tel_1_1168* | (C)6 (A)6 | (F)CAGGAGCGAAATCACCATCT<br>(R)GAGATCCCTGAGCATGAGGA | Scaffold 9<br>559208-559365 |
| *tel_1_1195* | (A)6 C (A)9 | (F)GGCATGCACCTTCCTTTTT<br>(R)GAAGTGCTCCCATTGTGGAT | Scaffold 77<br>67254-67416 |
| *tel_1_1211* | (A)5 (C)5 | (F)CGCATGATTTGACCTAACGA<br>(R)CGCCAACTTGTTTTCTGACA | Scaffold 83<br>77181-77392 |
| *tel_1_1206* | (A)5 C (A)5 | (F)CATGGTCGGTCTGTCAAATG<br>(R)TCGAGATCCTCCAGATGTCC | Scaffold 10<br>806352-806555 |
| *tel_2_1206* | (G)5 T (G)5 | (F)TCATTGGATCAACGCAGAGA<br>(R)ACGGGGAGCGAAATTATTCT | Scaffold 75<br>90145-90366 |
| *tel_1_1184* | (G)5 (A)5 | (F)TGATACTGCGTCTGGAGGTG<br>(R)ACGCAGAACTCCTGCACTTT | Scaffold 16<br>802661-802863 |
| *tel_2_1184* | (G)5 (A)5 | (F)GACTGCCTGGAAATTTCGTC<br>(R)CACCAGGTCAGTCGGAAAAT | Scaffold 16<br>320387-320525 |
| *tel_3_1184* | (C)7 (A)6 | (F)TCGACGGCACATAGACACTC<br>(R)CGTCGCACGTAGTAACTCCA | Scaffold 16<br>26872-27042 |
| *tel_1_1208* | (T)5 (C)5 | (F)GACGAAACCACCAGGTCAGT<br>(R)GAAGGTGTGACTGCCTGAAA | Scaffold 75<br>157287-157422 |
| *tel_1_1169* | (T)5 (C)5 | (F)GGAATTGCACTTACCCACCT<br>(R)TCCACCTAGGCACATGAACA | Scaffold 5<br>58244-58432 |

**Table 5 - Number of successfully amplified SSRs and number of polymorphic SSRs according to the genome regions among the progeny of *L. bicolor*.**

| Region | Amplified SSRs | Polymorphic SSRs |
|---|---|---|
| 5'UTR | 0 | 0 |
| Exon | 30 | 13 |
| Intron | 7 | 5 |
| 3'UTR | 0 | 0 |
| Transposable Elements | 55 | 14 |
| Others | 60 | 46 |

**Table 6 - List of analyzed fungal genomes: their web site and genome size.**

| Organism | Public Web Site | Genome Size (Mb) |
|---|---|---|
| *L. bicolor* | http://genome.jgipsf.org/Lacbi1/Lacbi1.home.html | 65 |
| *P. chrysosporium* | ftp://ftp.jgi-psf.org/pub/JGI_data/WhiteRot/whiterot.020216.fasta | 30 |
| *C. cinerea* | http://www.broad.mit.edu/cgi-bin/annotation/fungi/coprinus_cinereus | 37.5 |
| *U. maydis* | http://www.broad.mit.edu/cgi-bin/annotation/fungi/ustilago_maydis | 20 |
| *C. neoformans* | http://www.broad.mit.edu/cgi-bin/annotation/fungi/cryptococcus_neoformans | 19.5 |
| *A. nidulans* | http://www.broad.mit.edu/cgi-bin/annotation/fungi/aspergillus | 31 |
| *M. grisea* | http://www.broad.mit.edu/cgi-bin/annotation/fungi/magnaporthe | 40 |
| *N. crassa* | http://www.broad.mit.edu/cgi-bin/annotation/fungi/neurospora | 43 |
| *S. cerevisiae* | ftp://genome-ftp.stanford.edu/pub/yeast/data_download | 12 |

## Article 3. Characterization of Transposable Elements in the *Laccaria bicolor* Genome

**Jessy Labbé, Claude Murat, Benoit Hiselberger, Emmanuelle Morin, Marie-Pierre Oudot-Le Secq, Jan Wuyst, Igor Grigoriev, Pierre Rouzé, Hadi Quesneville, François Le Tacon and Francis Martin**

**(à soumettre à *BMC Genomics*)**

# Characterization of Transposable Elements in the *Laccaria bicolor* Genome

Jessy Labbé[1]*, Claude Murat[1], Benoit Hiselberger[1], Emmanuelle Morin[1], Marie-Pierre Oudot-Le Secq[1], Jan Wuyst[2], Igor Grigoriev[3], Pierre Rouzé[2], Hadi Quesneville[4], François Le Tacon[1] and Francis Martin[1]

[1] UMR 1136, INRA-Nancy Université, Interactions Arbres/Microorganismes, INRA-Nancy, 54280 Champenoux, France.

[2] Department of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology (VIB), Ghent University, B-9052 Ghent, Belgium.

[3] US DOE Joint Genome Institute, Walnut Creek, CA 94598.

[4] Unité de Recherches en Génomique-Info, Tour Évry 2, 91034 Évry Cedex.

*corresponding author

Email: labbe@nancy.inra.fr

## Summary

- The recent *Laccaria bicolor* genome sequencing has provided a considerable genomic resource allowing the identification of TEs in this symbiotic etcomycorrhizal fungus. Using a TE annotation pipeline we have characterized and analyzed all TEs in the *L. bicolor* S238N-H82 genome.
- TEs represent 24 % of the 60 Mb *L. bicolor* genome, totalizing 25,787 full-length and partial copies distributed within 172 families. The most abundant elements were the *Copia*-like. TEs are not randomly distributed across the genome, but are tightly nested or clustered. The majority of TEs are ancient except some terminal inverted repeats elements (TIRS), long terminal repeats elements (LTRs) and a large retrotransposon derivative element (*LARD*). There were three main periods of acquirement of TEs in *L. bicolor*: the first one from 57 to 10 Ma, the second one from 5 to 1 Ma and the last one from 500,000 years ago until now. LTR retrotransposons are closely related between them and to retrotransposons found in another basidiomycete, *Coprinopsis cinerea.*
- This analysis represents an initial characterization of TEs in the *L. bicolor* genome. It contributed to facilitate its assembly and to better understand the role played by TEs in genome evolution.



**Key words:** *Laccaria bicolor*, transposable elements and genome organization.

## Introduction

The basidiomycete *Laccaria bicolor* (Maire) P.D. Orton is an ectomycorrhizal fungus living in symbiosis with various woody host plants. Symbiotic plant-fungus association plays a fundamental role in biology and ecology of forest trees, affecting growth, water and nutrient absorption, and providing protection from root diseases (Harley & Smith, 1983). *L. bicolor* is characterized by its ubiquity, resulting in lack of strict host specificity. This feature requires a large plasticity of the genome in order to adapt to different hosts.

Transposable elements (TEs), which are foreign DNA segments inserted in eukaryote genomes, are characterized by their ability to move within these genomes. They had a tremendous influence on the evolution of eukaryote genomes (Kidwell & Lisch, 2000). These elements often constitute a large proportion of the genome, for instance, ~ 45 % of the human genome (Lander, 2001), 50 to 80 % of some grass genomes (Meyers, 2001) and only 9.7 % in the *Magnaporthe grisea* genome (Deane *et al*, 2005). TEs were first identified in fungi in the yeast *Saccharomyces cerevisiae* (Boeke, 1989), with the first evidence for their presence in filamentous fungi coming from conventional genetic studies with *Ascobolus immersus* mutants unstable for spore-staining (Decaris *et al*., 1978). In 2003, *Ty1-Copia* retroelements were reported in three ectomycorrhizal basidiomycetes, *L. bicolor*, *Pisolithus macrocarpus* and *Pisolithus* sp., by Diez *et al*. (2003). Recombination between elements at different sites can lead to large-scale chromosomal rearrangements. Some TE insertions may even become domesticated to play roles in the normal functions of the host genome (Kidwell, 2002). Indeed, TE insertions next to genes can modify gene expression patterns while insertions within genes can directly change gene structure. Thus, proliferation of TEs induces potentially deleterious effects. To avoid that eukaryotes have developed mechanisms limiting TE activity. However, TE existence depends on the survival of eukaryote genome. They have themselves developed mechanisms directing their integration to specific parts of the genome, and minimizing damages (Sandmeyer, 1998). Thus, there is no doubt that eukaryote genomic DNA has evolved in close association with TEs.

Eukaryotic TEs are divided into two classes: DNA transposable elements and retroelements or retrotransposons (Feschotte *et al.*, 2002). There are numerous families of TEs within each class, based on features (length and target site), and whether they are

autonomous or non autonomous. The retrotransposons, most commonly isolated from fungi, transpose via an RNA intermediate (class I elements; Boeke & Stoye, 1997). There are two broad classes: the major divisions are the LTR retrotransposons characterized by long terminal repeats (LTRs) at each end and those that lack these repeats, the Non-LTR retrotransposons (also known as LINE and SINE elements; Deninger, 1989). The transposons (class II elements) transpose via DNA intermediates using either a "cut-and-paste" mechanism for those which have terminal inverted repeats (TIRs) or a rolling circle replicon mechanism, for *Helitrons* elements similar to some known prokaryotic transposition mechanism.

Fungi have small genomes, usually with limited amounts of repetitive DNA. The Ascomycota and Basidiomycota have a strong tendency towards streamlined genomes. The majority of them contain no more than 10-15% of repetitive DNA (Kempken and Kück 1998; Wöstemeyer and Kreibich 2002). However, the recent sequencing of the haploïd genome of *L. bicolor* highlighted the largest sequenced fungal genome published (Martin *et al.*, 2008). The 60 megabases (Mb) *L. bicolor* genome comprises a higher proportion of repetitive sequences and elements than that identified in the other fungal genomes.

As TEs are mobile, high-copy, variable and dispersed throughout the genomes, they can induce serious difficulties in sequence assembly. To assemble *L. bicolor* genome, it was necessary to determine the abundance and distribution of TEs.

Here, we report the results of a computer-assisted analysis which provides information on the abundance (copy number), diversity and distribution for all TE types and temporal feature for the major TE types: the LTR retrotransposons. Moreover, we have compared LTR distribution and insertion age of full length LTRs in *L. bicolor* and *Coprinopsis cinerea*, a saprophyte basidiomycete.

## Materials and Methods

**Sequences analyses** were performed using the WGS of the strain S238N-H82 from *L. bicolor* (Maire) P.D. Orton downloaded from the Joint Genome Institute website (http://genome.jgi-psf.org/Lacbi1/Lacbi1.home.html) or from the *L. bicolor* genome resources (http://mycor.nancy.inra.fr/fr/index.html). The WGS of *Coprinopsis cinerea* was downloaded from the Broad Institute website (http://www.broadinstitute.org/). The TE reference set used was constructed from RepeatMasker repeat library (http://www.repeatmasker.org), from Repbase on the

Genetic Information Research Institute website (GIRI) (http://www.girinst.org/repbase/index.html) and sequences from NCBI (http://www.ncbi.nlm.nih.gov/).

**Sequence annotation:** TEs were annotated in *L. bicolor* genome using the RMBLR procedure from the TE annotation pipeline described by Quesneville *et al.*, 2005. Then, consecutive fragments on both the genome and the same reference TE were automatically joined if they were separated by a sequence of which more than 80% consisted of other TE insertions. Simple repeats were found using the Tandem Repeat Finder program (Benson G, 1999) and used to filter out spurious hits. All TE annotations that were less than 20bp after removing any regions that over-lopped simple repeat regions were eliminated. Finally, 215 consensus sequences (complete and incomplete) which could belong to any class or type of TEs were obtained. LTR retrotransposons sequences were confirmed by a second identification using the program LTR_STRUC (McCarthy & McDonald, 2003). To facilitate manual curation, we automatically promoted the 215 consensus sequences of RMBLR to be the candidate annotation (defined as a set of one or more joined fragments and by using sequence alignments with CLUSTALW, http://www.ebi.ac.uk/clustalw), which could then be validated or modified by the curator in Artemis v11 freeware. A database of TE of L. bicolor is available on the L. bicolor genome resources (http://mycor.nancy.inra.fr/fr/index.html).

TE copy number estimation and distribution: the copy number for each TE type in the 60 Mb of *L. bicolor* genome sequence was calculated on the basis of the elements obtained from the TE annotation pipeline, TBLASTN searches and the RepeatMasker analysis used in the previous pipeline. The TE distribution in *L. bicolor* genome was established on the RepeatMasker analysis with the TE database of *L. bicolor* made after annotation. Specific python scripts were used for the mapping of TE density on the updated linkage groups from the genetic map of *L. bicolor* (Labbé *et al.*, 2008). Multiple sequence alignments by CLUSTALW (http://www.ebi.ac.uk/clustalw) allowing estimate the age of TEs according to the sequence identity.

**Sequence alignments and comparison:** to study the phylogenetic relationship between the major elements form in *L. bicolor* and *C. cinerea*, the reverse transcriptase (RT) amino-acidic sequences were aligned for LTR elements. Multiple sequence alignments were constructed with CLUSTAL_X (Thompson *et al.*, 1997). Phylogenetic trees were calculated on the basis of the neighbor-joining method (Saitou, 1997) using

PHYLIP with default parameters (Felsenstein, 1989). Bootstrap values were calculated for each tree from 1000 replicates. The phylogram graphics were obtained using the freeware FigTree v1.2.3 (http://tree.bio.ed.ac.uk/software/figtree/).

**The insertion age** of full length LTR retrotransposons was determined from the evolutionary distance between 5' and 3' solo LTR derived from a CLUSTALW alignment of the two solo LTR by the Kimura two-parameter method (emboss distmat, http://emboss.sourcesforge.net/). The insertion age was determined for each family of a category of retroelements (i.e., LTR *Gypsy* family 1 and LTR *Gypsy* family 2 found in *L.* bicolor). For the conversion of distance to insertion age, a substitution rate of 1.3E-8 mutations per site per year was used (Ma & Bennetzen, 2004).

## RESULTS

### TE abundance, diversity and distribution

Consensus sequences obtained by TE annotation pipeline (Quesneville *et al.* 2005) used as queries in BLASTN searches of *L. bicolor* genomic sequence did not detect any novels repeats, indicating that the vast majority of TEs had been identified. The results of these analyses are summarized in Table 1. On the 60 Mb of *L. bicolor* genomic sequence, ~15 Mb (24 %) derived from TEs. Among these 15 Mb, TEs of class I represent ~ 5 Mb (8 % of the genome) and TEs of class II represent ~ 3 Mb (5 % of the genome). The 7 remaining Mb (11 % of the genome), which did not exhibit matches, corresponded to truncated and unclassified TEs.

From the 215 consensus sequences obtained by the RMBLR procedure, without counting TE fragments, we have identified 112 families of TEs in *L. bicolor* showing 912 full length copies and 24875 incomplete copies. The major class was the class I (84 families) in which the *Copia* families were the most numerous (46), followed by the *Gypsy* (27). The third major families were represented by the TIRs (23, class II). We have also identified other families: three LINEs, four MITEs, three large retrotransposon derivatives (*LARD*), two *ERVs*, one *DIRS* and one *Helitron*.

The identification of thousands of TEs in *L. bicolor* genome provided the raw material to address questions about TE diversity and temporal aspect of TE amplification. We have compared all TE types and families sequences previously identified by multiple alignments using CLUSTALW. According to sequence diversity, most of these elements could be considered as ancient except some of them, namely four TIRs, two MITEs, three LTRs and one *LARD* (Supplementary Table 1). These

elements (5 % of total; threshold: 95% of identity) contained nearly identical members, suggesting recent and ongoing transposition.

**Table 1** Transposable elements characterized in *L. bicolor* genome

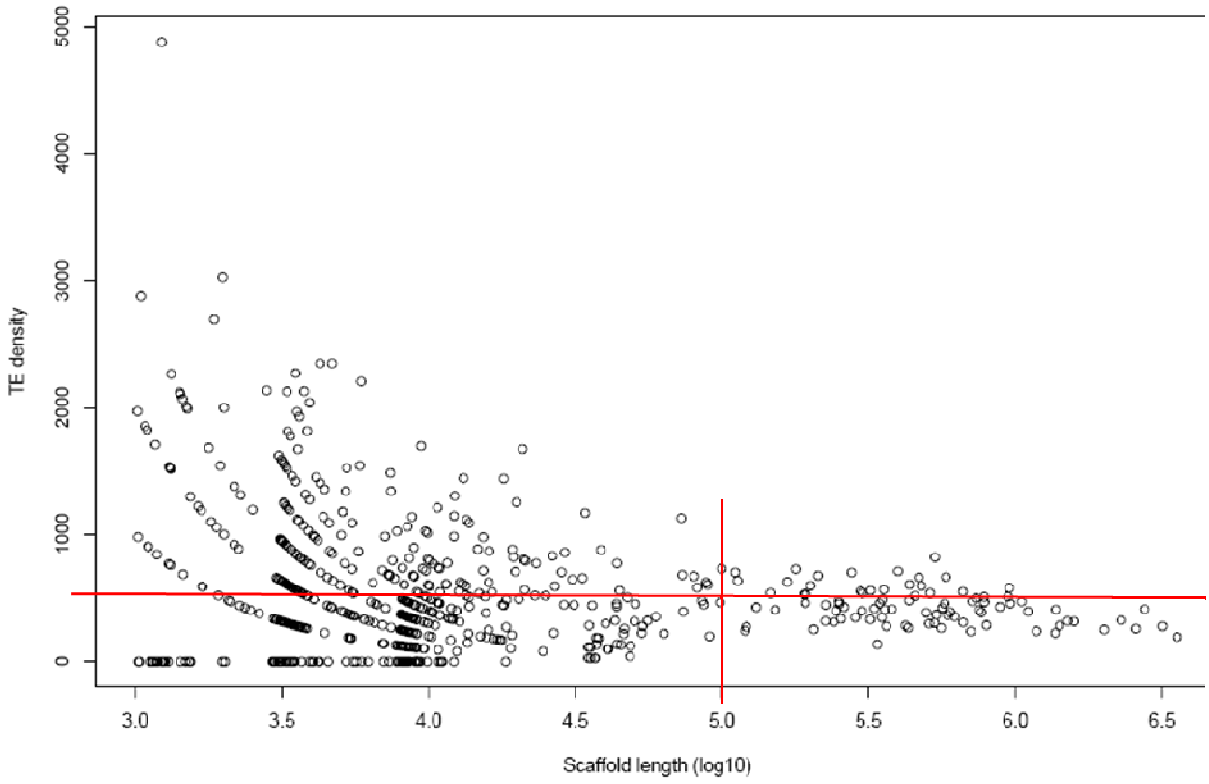
| TE type | Family | Copy number | Full length / incomplete | DNA amount (Mb) | % TE composition in genome |
|---|---|---|---|---|---|
| ***Class I*** | | | | | |
| LTR | | 6585 | 105 / 6480 | 3.99 | 6.65 |
| *ERV* | 2 | 7 | 4/3 | 0.03 | - |
| *Copia*-like | 46 | 3382 | 57 / 3325 | 2.23 | 3.72 |
| *Gypsy*-like | 27 | 3085 | 47 / 3038 | 1.71 | 2.85 |
| *LARD* | 3 | 118 | 1 / 117 | 0.05 | - |
| *DIRS* | 1 | 19 | 1/ 18 | 0.01 | - |
| LINEs | 5 | 1249 | 5 / 1244 | 0.58 | 0.96 |
| Total Class I | 84 | 7860 | 115 / 7745 | 4.57 | 7.62 |
| Class II | | | | | |
| TIRs | 23 | 6369 | 692 / 5677 | 2.85 | 4.75 |
| *Helitrons* | 1 | 215 | 2 / 213 | 0.08 | 0.13 |
| MITEs | 4 | 249 | 103 / 146 | 0.06 | 0.10 |
| Total Class II | 28 | 6833 | 797 / 6036 | 2.99 | 4.98 |
| TE fragments[a] | 60 | 11094 | — | 6.96 | 11.6 |
| Total TE | 172 | 25787 | 912 / 24875 | 14.56 | 24.2 |

[a] Others elements truncated, unclassified and unfound matches (named no cat).

The TE density was higher in the small scaffolds, whereas the average density throughout the whole genomic sequence was 430 copies per Mb (Figure 1). To determine whether the TEs might show differential accumulation in the genome, we have analyzed their distribution within 10 kb intervals along the updated and anchored genetic-physical linkage groups of *L. bicolor* 47.7 Mb (Figure 2). Due to their small length, the unmapped scaffolds were not included in this analysis. As shown in Figure 2, TEs were not uniformly distributed within each linkage group. Similar results were obtained within 100 kb intervals (data not shown). Most of linkage groups contain 20 % of TEs and two, LG 8 and LG 9 respectively contained 60 and 62 % of TEs. On each linkage group, we observed regions exhibiting high TE density, mainly *Copia, Gypsy* and TIRs. In majority, these regions corresponded to proximal ends of the linkage groups and were correlated with telomere regions for LG 2, LG 4, LG 6 and LG 9. As it

is commonly found in plant genomes, TEs inserted within another transposable element are named nests. Groups of TEs located within 10 kb of each other are defined as clusters (Tikhonov *et al.* 1999; Fu *et al*, 2001). We found 382 nests or clusters of TEs containing 138 families (Figure 3). This

**Figure 1** Graphic representation of transposable elements (TE) density function to the scaffold length of the genomic sequence of *L. bicolor*




means that about 80 % of TEs of *L. bicolor* were either inserted into another element or positioned adjacently to another element. Among the 382 nests or clusters, 153 were within the proximal 1 Mb region of the linkage groups. Most of TEs belonging to nests or clusters were incomplete. LTRs *Gypsy* and *Copia* (34 %) were nested or clustered most often than either TIRs (25 %) or LINE (16 %).

**Phylogenetic analysis of the LTR retrotransposons**

LTR retrotransposons being the most frequent TEs in the *L. bicolor* genome (*Copia* > *Gypsy*), we have analyzed the relationships between these elements and those of *C. cinerea* where retrotransposons also are over-represented (*Gypsy* > *Copia*). Whave constructed phylogenetic trees based on multiple alignments of various RT domains including all identified families of LTR retrotransposons on both genome sequences by

LTR_STRUC (McCarthy & McDonald, 2003) (Figures 4a, 4b and 4c). *L. bicolor* and *C. cinerea* TEs exhibited the same phylogenetic structure with three independent monophyletic groups. In *L. bicolor*, the group 1 and 2 mainly comprised *Copia*, while the group 3 mainly comprised *Gypsy*. In *C. cinerea*, the three groups mainly contained *Gypsy* (Figure 4a, Figure 4b). Most of *Gypsy* and *Copia* families appeared to be closely related to each other in both genomes (Figure 4c). Indeed, the group 1 and 3 and the 3 subgroups of group 2 contained both *L. bicolor* and *C. cinerea Copia* or *Gypsis*. Nevertheless, the subgroup 3 almost exclusively contained *C. cinerea* Gypsy.

**The insertion age of LTR retrotransposons**

As shown by Figure 5a and 5b, the first LTR retrotranspsosons were respectively inserted in *L. bicolor* and *C. cinerea* genome 57 and 59 million years ago (Ma). In *L. bicolor*, there were three main periods of acquirement of TE families: the first one from 57 to 10 Ma, the second one from 5 to 1 Ma and the last one from 500,000 years ago until now. In *C. cinerea*, there were only two main periods of high TE acquirement: one from 59 to 10 Ma and a second one from 500,000 years ago until now. In *L. bicolor*, during the three main periods of TE *Copia* were more abundant than *Gypsy*. During the period of low acquirement (0.5 to 1 Ma), *Gypsy* were more abundant than *Copia*. In contrast, in *C. cinerea* during the two main periods of TE acquirement, *Gypsy* were more abundant than *Copia*.

**Figure 5a** Number of LTR retrotransposons families function to their insertion age expressed in million years in the *L. bicolor* genome

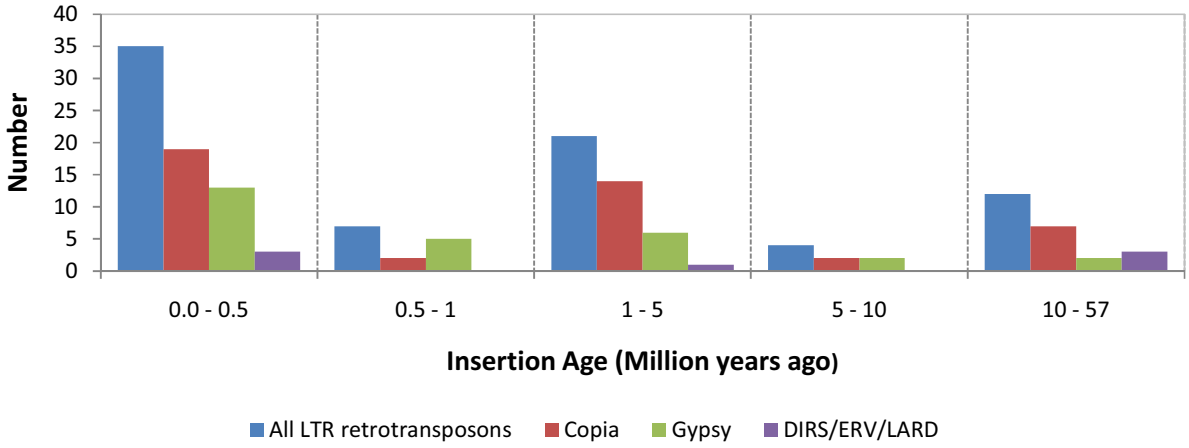
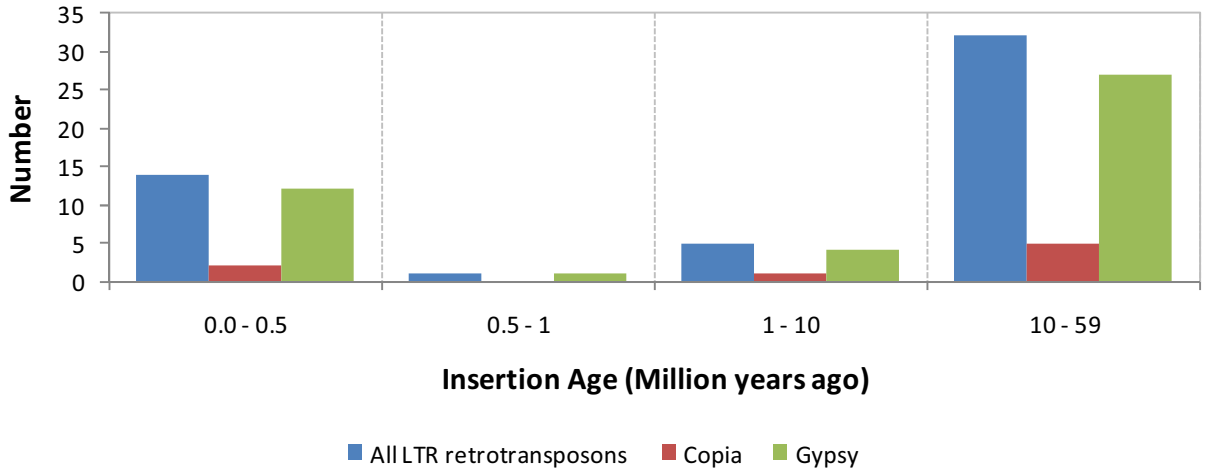


**Figure 5b** Number of LTR retrotransposons families function to their insertion age expressed in million years in the *C. cinerea* genome




## Discussion

We have described and characterized 172 TE families in the symbiotic fungus *L. bicolor* genome. The localization of these repetitive elements has first greatly improved the assembly of the *L. bicolor* genome sequence. In a second step, our results have contributed to a better knowledge of the role of transposable elements the structure and the evolution of this genome.

The *L. bicolor* genome is rich in TEs (24 % of the genome), which explains its large size compared to other fungi. The major class was the class I with a high proportion of LTR retrotransposons (*Copia* > *Gypsy*). Nevertheless, some other types were found such as LINE, *ERV* and *LARD*. A large proportion of class II elements also were found, mainly TIRs, *Helitron* and MITE. Among the 172 recorded families, 40 exhibited intact structure, which suggests that these elements are still active. However, according to their sequences, most of the families are ancient and only 5 % of them could be recent and still could have the capacity to transpose.

Phylogenetic analysis of the LTR retrotransposons revealed that these elements are closely related between them and to retrotransposons found in another basidiomycete, *C. cinerea*. The estimation of the insertion age of LTR retotransposons suggests that *L. bicolor* had three main periods of acquirement, including a recent one. *C. cinerea* had only two main periods of acquirement, which could explain the higher frequency of LTRs observed in *L. bicolor*. The first main period of TE insertion in *L. bicolor* dates from 57 Ma. This is congruent with the age of the oldest known fossil of

ectomycorrhiza, which was found in the middle of the Eocene and dates back to about 50 Ma (Lepage *et al.*, 1997). But it is probable that ectomycorhizal basidiomycetes are more ancient (Selosse & Le Tacon, 1998; Ducousso *et al.* 2004).

TEs are not randomly distributed across the genome, but are tightly nested or clustered. The analysis of the TE localization on the linkage groups revealed that the nests or clusters are positioned at the proximal ends in the telomeric regions. Most of TEs belonging to nests or clusters are partial and are mainly LTRs (*Copia* and *Gypsy*), and with a least degree, TIRs and LINE. We propose several sources of selective pressure to explain this type of distribution. The ectopic exchange model states that repetitive DNA promotes deleterious ectopic recombination events. Since negative selection would be reduced in regions of low recombination rate, repetitive DNA would accumulate in regions with low recombination rate. Under the insertion model, TE insertions generally have a deleterious effect on fitness, and are thus likely to be lost from the population. In chromosomal segments that have low recombination rates, deleterious mutations are more likely to be genetically linked to neutral genes, decreasing their negative fitness effect. Otherwise, the insertion of TE may trigger a runaway process, since it will provide a target into which other elements may insert without deleterious consequences (Walbot *et al.*, 2001).

Finally, as all plant and animal and fungal genomes characterized to date, the *L. bicolor* genome has a distinctive TE composition and evolution. The high proportion, the large diversity and the nested and clustered distribution of TE of *L. bicolor* may suggest rearrangements promoted by their presence. This nested and clustered distribution could result from potential insertion sites formed by the SSRs flanking the transposable elements (Metzgar *et al.*, 2000). This nested structure is predominant in weakly recombinant telomeric regions, where TE insertion can produce rearrangements without deleterious effects. However, TEs are distributed throughout the *L. bicolor* genome and could be major contributors to the genesis of new genes or to the adaptation of existing genes. The high density of TEs in *L. bicolor* genome could explain its plasticity and its evolution from a saprophytic to a symbiotic way of life.

## Acknowledgements

This work was supported by INRA and Région Lorraine through a PhD Scholarship to JL. We would like to thank Emilie Tisserant and Yao Cheng for their assistance and helpful discussions.

## References


Benson G. 1999. Tandem repeat finder: a program to analyse DNA sequences. *Nucleic Acids Res* 27(2): 573-580.

Boeke JD. 1989. Transposable elements in *Saccharomyces cerevisiae*. In: Berg DE and M Howe (eds) Mobile DNA. ASM, Washington DC, pp 335-374.

Boeke JD, Stoye JP. 1997. Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In retroviruses, coffins JM, Hughes SH, Varmus HE (eds). *Cold Spring Harbor Laboratory Press*: New York: 343-435.

Deane RA , Talbot NJ, Ebbole DJ, Farman ML, Mitchel TK, Orbach MJ, Thon M, Kulkarni R, Xu JR, Pan H, Read ND, Lee YH, Carbone I, Brown D, Oh YY, Donofrio N, Jeong JS, Soanes DM, Djonovic S, Kolomiets E, Rehmeyer C, Li W, Harding M, Kim S, Lebrun MH, Bohnert H, Coughlan H, Butler J, Calvo S, Ma LJ, Nicol R, Purcell S, Nusbaum C, Galagan JE & Birren BW. 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 434: 980-986.

Decaris B, Francou F, Lefort C and Rizet G. 1978. Unstable ascospore color mutants of *Ascobolus immersus*. *Mol Gen Genet* 162: 69-81.

Deninger PL. 1989. SINEs: Short interspersed repeated DNA elements in higher eukaryotes. *American Society of Microbiology* 19: 619-637.

Diez J, Beguristain Th, Le Tacon F, Casacuberta JM and Tagu D. 2003. Identification of Ty1-*copia* retrotransposons in three ectomycorrhizal basidiomycetes: evolutionary relationships and use as molecular markers. *Current Genetics* 43, 34-44.

Ducousso M, Béna G, Bourgeois C, Buyck B, Eyssartier G, Vincelette M, Rabevohitra R, Randrihasipara L, Dreyfus B & Prin Y. 2004. The last common ancestor of Sarcolaenaceae and Asian dipterocarp trees was ectomycorrhizal before the India–Madagascar separation, about 88 million years ago. *Molecular Ecology*, 13, 231-236.

Feschotte C, Jiang N, Wessler S. 2002. Plant transposable elements: where genetics meets genomics. *Nature Reviews Genetics* 3:329–341.

Felsentein J. 1989. PHYLIP – phylogeny inference package (version3.2). *Cladistics* 5: 164-166.

Fu H, Park W, Yan X, Zheng Z, Shen B, Dooner HK: The highly recombinogenic *bz* locus lies in an unusually gene-rich region of the maize genome.
*Proc Natl Acad Sci USA* 2001, 98:8903-8908.

Harley JL, Smith SE. 1983. *Mycorrhizal Symbiosis*. Academic Press. London

Kidwell MG, Lisch DR. 2000. Transposable elements and host genome evolution. *Trends Ecol Evol* 15: 95–99.

Kempken F, Jacobsen S and Kück U. 1998. Distribution of the Fungal Transposon *Restless:* Full-Length and Truncated Copies in Closely Related Strains. *Fungal Genetics and Biology* 25(2): 110-118.

Kidwell MG. 2002. Transposable elements and the evolution of genome size in eukaryotes. *Genetica* 115: 49–63.

Labbé J, Zhang X, Yin T, Schmutz J, Grimwood J, Martin F, Tuskan GA, Le Tacon F: A genetic linkage map for the ectomycorrhizal fungus *Laccaria bicolor* and its alignement to the whole-genome sequence assemblies. *New Phytologist* 2008, 180:316-328.

Lander EM, Linton LM, Birren B, Nusbaum C, Zody MC. 2001. Initial sequencing and analysis of human genome. *Nature* 409: 860–921.

LePage BA, Currah RS, Stockey RA and Rothwell GW. 1997. Fossil ectomycorrhizae from the Middle Eocene. *American Journal of Botany* 84: 410-410.

Ma J & Bennetzen JL. 2004. Rapid recent growth and divergence of rice nuclear genomes. *PNAS* 101(34): 12404-12410.

Martin F, Aerts A, Ahrén D, Brun A, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V, Salamov A, Shapiro HJ, Wuyts J, Blaudez D, Buée M, Brokstein P, Canbäck B, Cohen D, Courty PE, Coutinho PM, Danchin EGJ, Delaruelle C, Detter JC, Deveau A, DiFazio S, Duplessis S, Fraissinet-Tachet L, Lucic E, Frey-Klett P, Fourrey C, Feussner I, Gay G, Grimwood J, Hoegger PJ, Jain P, Kilaru S, Labbé J, Lin YC, Legué V, Le Tacon F, Marmeisse R, Melayah D, Montanini B, Muratet M, Nehls U, Niculita-Hirzel H, Oudot-Le Secq MP, Peter M, Quesneville H, Rajashekar B, Reich M, Rouhier N, Schmutz J, Yin T, Chalot M, Henrissat B, Kües U, Lucas S, Van de Peer Y, Podila G, Polle A, Pukkila PJ, Richardson PM, Rouzé P, Sanders IR, Stajich JE, Tunlid A, Tuskan G, Grigoriev IV. 2008. Symbiosis insights from the genome of the mycorrhizal basidiomycete *Laccaria bicolor*. *Nature* 452:88-92.

McCarthy EM and McDonald JF. 2003. LTR_STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics* 19(3): 362-367.

Meyers BC, Tingey SV, Morgante M. 2001. Abundance, distribution and transcriptional activity of repetitive elements in the maize genome. *Genome Res.* 11: 1660-1676.

Metzgar D, Bytof J, Wills C. 2000. Selection Against Frameshift Mutations Limits Microsatellite Expansion in Coding DNA. *Genome Res.* 10: 72-80.

Quesneville H, Bergam CM, Andrieu O, Autard D, Nouaud D, Ashburner M, Anxolabehere D. 2005. Combined evidence annotation of transposable elements in genome sequences. *PLos Comp Biol* 1(2): e22.

Saitou N, and Nei M.1987.The neighbor-joining method: a new method for reconstructin phylogenetic trees. Mol. Biol. Evol. 4: 406–425.

Sandmeyer S. 1998. Targeting transposition: at home in the genome. *Genome Res* 8: 416-418.

Selosse MA and Le Tacon F. 1998. The land flora: a phototroph - fungus partnership. *Trends in Ecology and Evolution*, 13, 15-20.

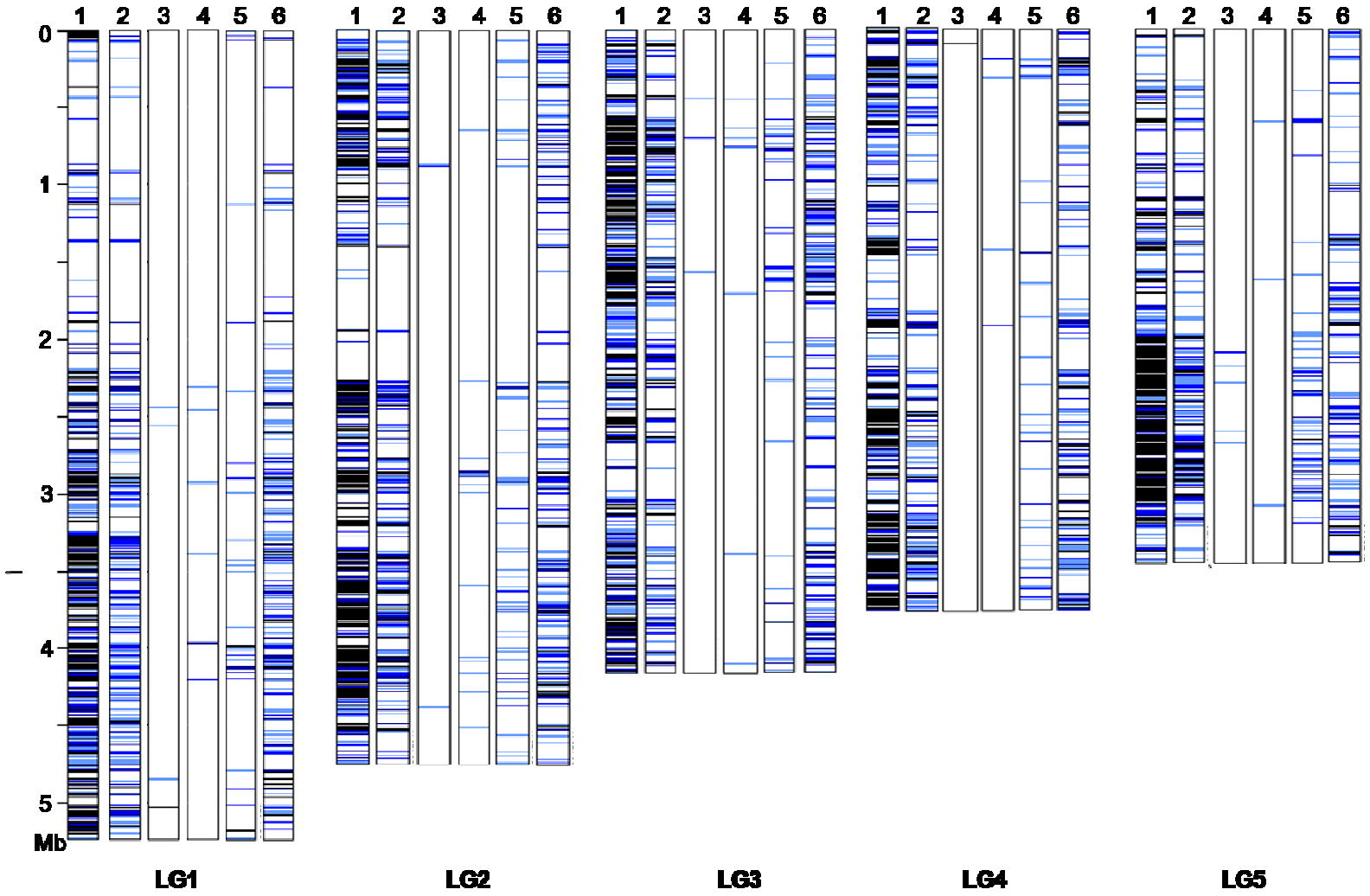
Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*25: 4876-4882.

Tikhonov AP, SanMiguel PJ, Nakajima Y, Gorenstein NM, Bennetzen JL, Avramova Z: Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. *Proc Natl Acad Sci USA* 1999, 96:7409-7414.

Walbot V, Petrov DA: Gene galaxies in the maize genome. *Proc Natl Acad Sci USA* 2001, 98:8163-8164

Wöstemeyer J & Kreibich A. 2002. Repetitive DNA elements in fungi (Mycota): impact on genomic architecture and evolution. *Curr Genet* 41: 189–198.

**Figure 2** Density of transposable elements (TE) along the linkage groups of *L. bicolor*. The density is graphically represented as the number of TE copy found in a window of 10 kb. Lane 1 represents the density of all TE; Lane 2 represents the density of *Gypsy* and *Copia;* Lane 3 represents the density of *Helitron*; Lane 4 represents the density of *LARD*; Lane 5 represents the density of LINE; Lane 6 represents the density of TIRs.



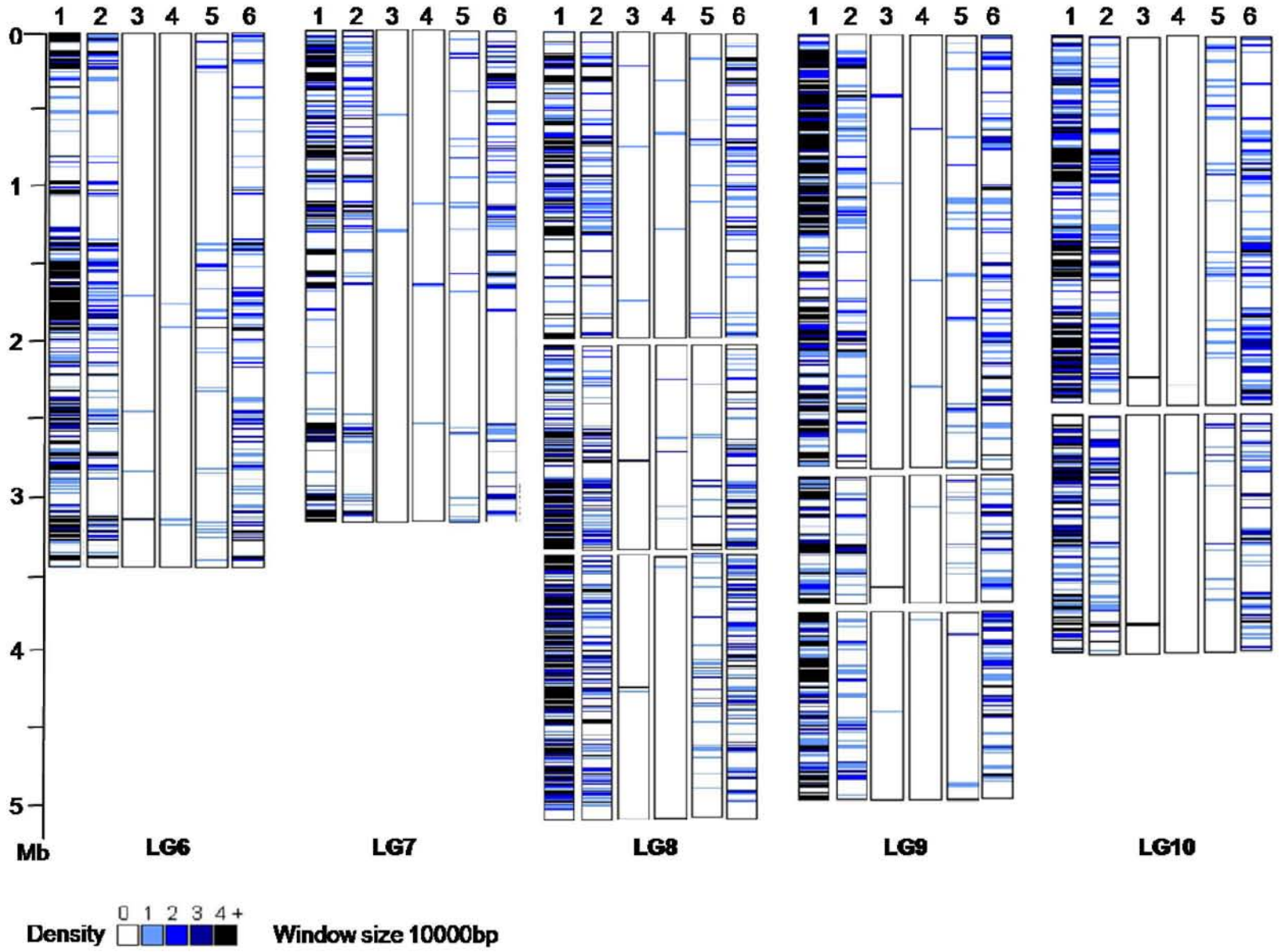
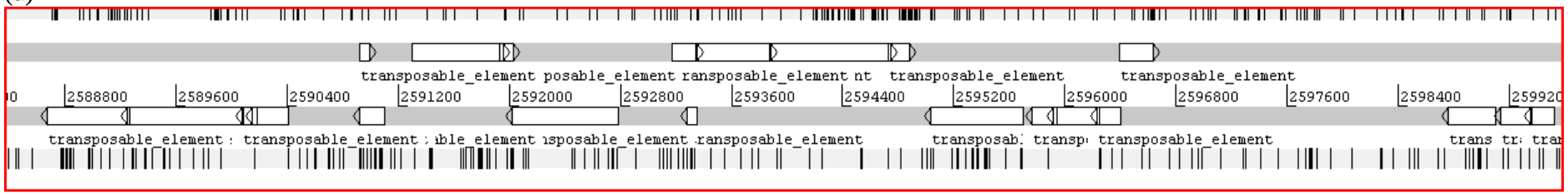

1 2 3 4 5 6
0
1
2
3
4
5
Mb
LG6
LG7
LG8
LG9
LG10
Density
0 1 2 3 4+
Window size 10000bp

**Figure 3** Screen capture from Artemis v11 reprensenting  two samples of the nested and clustered distribution of TE (a) on the LG 5 and (b)on the LG 1

(a)




(b)

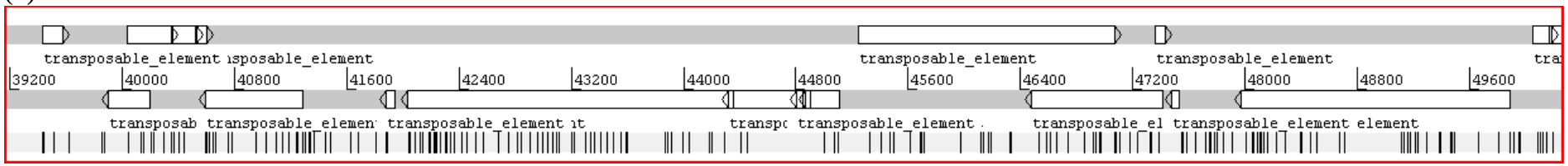


**Figure 4 a** Relationships of LTR retrotransposons of *L. bicolor*. This phylogenetic tree is based on reverse transcriptase amino-acidic sequences of all the LTR retrotransposons identified by LTR_STRUC. In green: group 1, in orange: group 2, in blue: group 3.

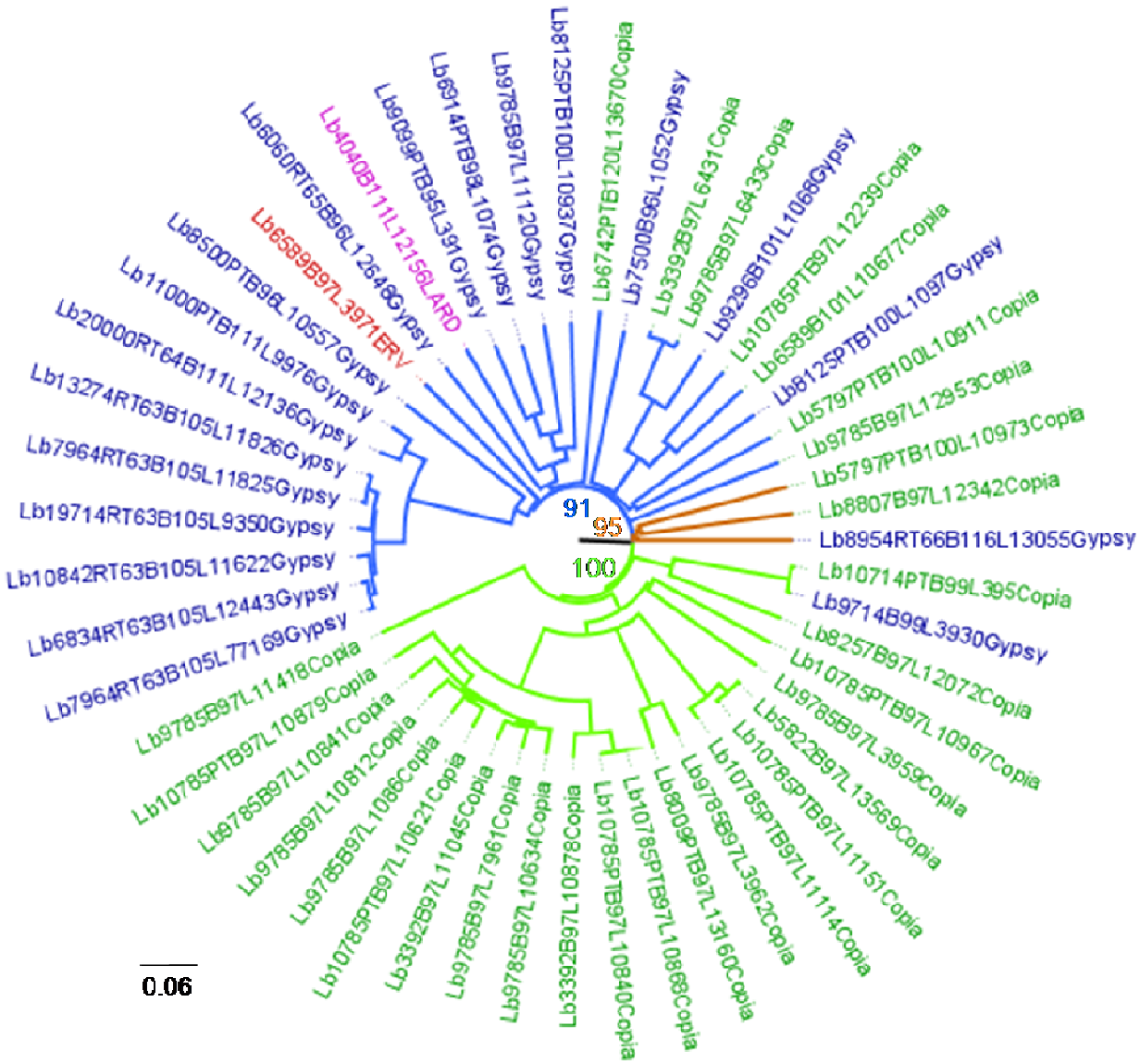


**Figure 4b** Relationships of LTR retrotransposons of *C. cinerea*. This phylogenetic tree is based on reverse transcriptase amino-acidic sequences of all the LTR retrotransposons identified by LTR_STRUC. In green: group 1, in orange: group 2, in blue: group 3.

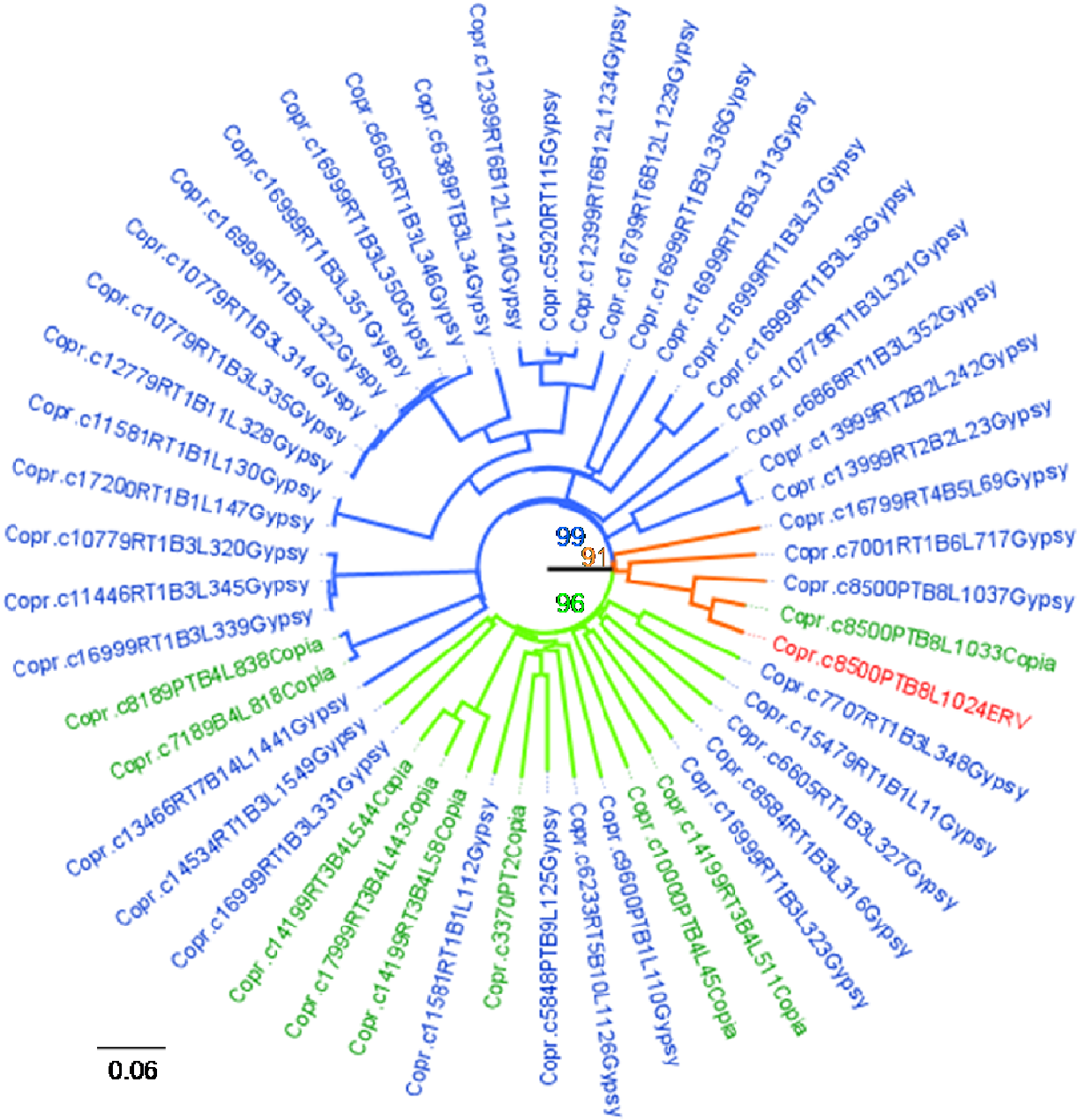
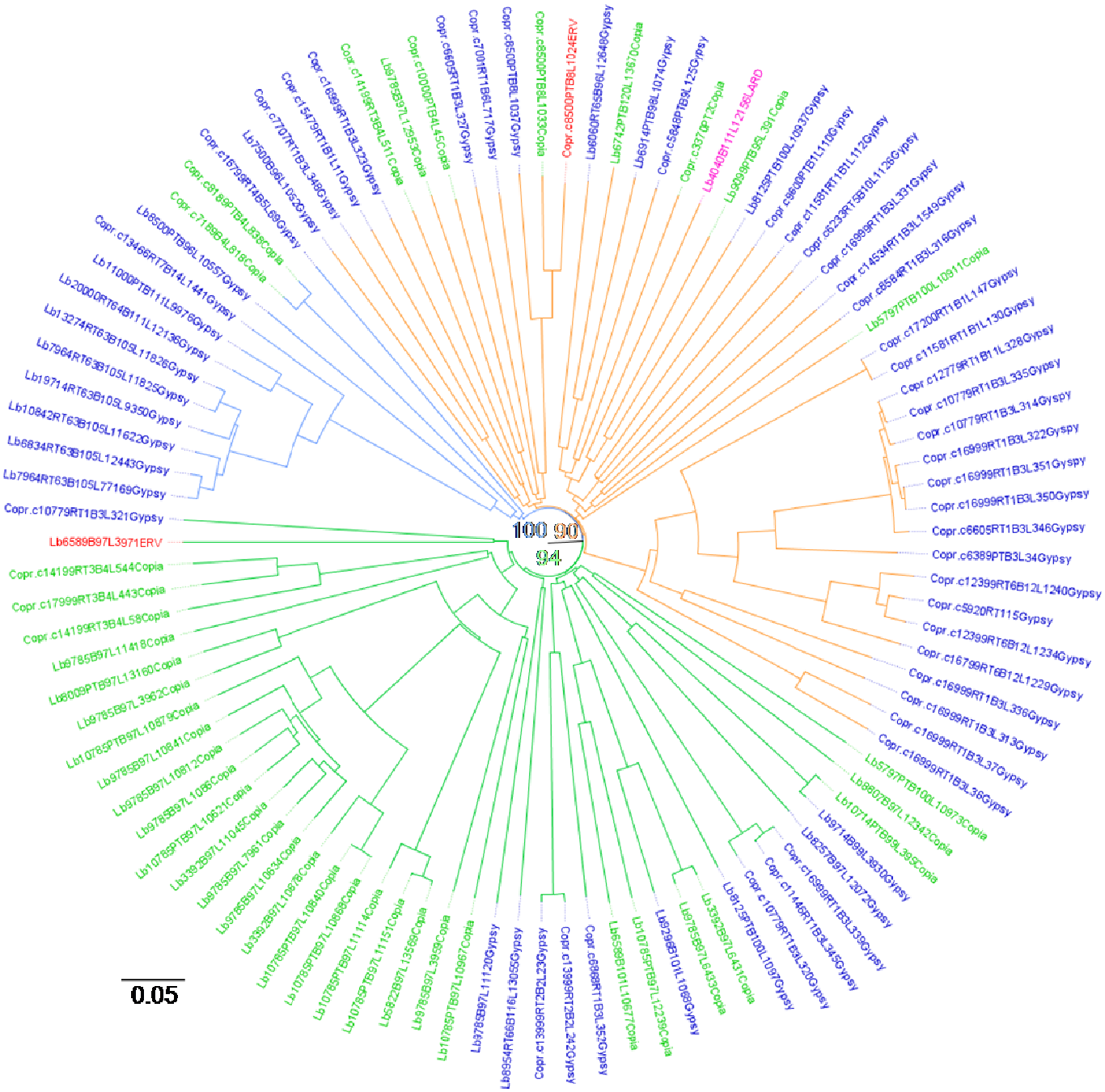


**Figure 4c** Relationships of LTR retrotransposons of *L. bicolor* and *C. cinerea*. This phylogenetic tree is based on reverse transcriptase amino-acidic sequences of all the LTR retrotransposons identified by LTR_STRUC. In green: group 1, in orange: group 2, in blue: group 3.



**Supplementary Table 1** Table showing the percentage of identity of multiple alignment of TE copies corresponding to 215 consensus TE families identified by the RMBLR procedure.

| TE family ID | Class | Type | Pairwise multiple alignment identity (%) |
|---|---|---|---|
| ReconFam1562 | II | TIR | 98 |
| T_scf9_3 | I | LTR | 97 |
| ReconFam30 | II | TIR | 97 |
| ReconFam33 | II | TIR | 96 |
| ReconFam40 | II | TIR | 96 |
| ReconFam654 | I | LTR | 96 |
| T_scf60_1 | I | LTR | 96 |
| MITE_X2 | II | MITE | 95 |
| MITE_X3 | II | MITE | 95 |
| ReconFam383 | I | LARD | 95 |
| ReconFam1037 | II | TIR | 95 |
| T_scf20_2 | I | LTR | 94 |
| ReconFam43 | II | TIR | 94 |
| ReconFam938 | II | TIR | 94 |
| T_scf49_1 | I | LTR | 93 |
| ReconFam437 | II | TIR | 93 |
| T_scf6_3 | I | LTR | 93 |
| ReconFam243 | II | TIR | 92 |
| HMMReconFam82 | no cat | no cat | 92 |
| ReconFam641 | II | TIR | 92 |
| ReconFam399 | no cat | no cat | 91 |
| HMMReconFam1523 | no cat | no cat | 91 |
| HMMReconFam986 | no cat | no cat | 91 |
| ReconFam374 | II | TIR | 90 |
| ReconFam1184 | II | TIR | 90 |
| HMMReconFam219 | no cat | no cat | 90 |
| ReconFam1670 | II | TIR | 90 |
| T_scf4_5 | I | LTR | 90 |
| HMMReconFam622 | no cat | no cat | 90 |
| ReconFam877 | II | TIR | 89 |

| TE ID | Class | Type | Pairwise multiple alignment identity (%) |
|---|---|---|---|
| ReconFam2358 | II | TIR | 88 |
| T_scf21_1 | I | LTR | 88 |
| T_scf7_1 | I | LTR | 88 |
| ReconFam2355 | II | TIR | 88 |
| ReconFam416 | I | LTR | 88 |
| HMMReconFam572 | no cat | no cat | 87 |
| T_scf52_1 | I | LTR | 87 |
| HMMReconFam3152 | no cat | no cat | 87 |
| ReconFam116 | no cat | no cat | 87 |
| HMMReconFam1417 | no cat | no cat | 86 |
| ReconFam2610 | II | TIR | 86 |
| HMMReconFam228 | no cat | no cat | 86 |
| HMMReconFam611 | no cat | no cat | 86 |
| ReconFam1841 | II | TIR | 85 |
| HMMReconFam691 | no cat | no cat | 85 |
| HMMReconFam315 | no cat | no cat | 85 |
| HMMReconFam395 | no cat | no cat | 85 |
| ReconFam346 | II | TIR | 85 |
| ReconFam924 | II | TIR | 85 |
| ReconFam17 | II | TIR | 85 |
| ReconFam125 | II | TIR | 85 |
| T_scf3_2 | I | LTR | 85 |
| ReconFam3527 | I | LTR | 85 |
| ReconFam559 | no cat | no cat | 85 |
| T_scf91_1 | I | LTR | 84 |
| ReconFam2694 | II | TIR | 84 |
| ReconFam807 | II | TIR | 84 |
| T_scf1_6 | I | LTR | 84 |
| HMMReconFam311 | no cat | no cat | 84 |
| ReconFam18213 | II | MITE | 84 |
| T_scf6_2 | I | LTR | 83 |
| T_scf42_1 | I | LTR | 83 |
| ReconFam316 | I | LTR | 83 |
| HMMReconFam973 | no cat | no cat | 83 |
| HMMReconFam782 | no cat | no cat | 83 |
| HMMReconFam343 | no cat | no cat | 83 |
| T_scf78_1 | I | LTR | 83 |
| ReconFam285 | II | TIR | 83 |
| ReconFam2359 | II | TIR | 83 |

| TE ID | Class | Type | Pairwise multiple alignment identity (%) |
|---|---|---|---|
| ReconFam629 | II | TIR | 83 |
| HMMReconFam1041 | no cat | no cat | 83 |
| ReconFam547 | no cat | no cat | 83 |
| T_scf6_1 | I | LTR | 83 |
| HMMReconFam2191 | I | LARD | 82 |
| T_scf43_1 | I | LTR | 82 |
| HMMReconFam1286 | no cat | no cat | 82 |
| ReconFam838 | II | TIR | 82 |
| T_scf144_1 | I | LTR | 82 |
| ReconFam1049 | II | TIR | 82 |
| MITE_X1 | II | TIR | 82 |
| ReconFam223 | no cat | no cat | 82 |
| HMMReconFam354 | no cat | no cat | 82 |
| HMMReconFam296 | no cat | no cat | 82 |
| T_scf90_1 | I | LTR | 81 |
| T_scf80_2 | I | LTR | 81 |
| HMMReconFam355 | no cat | no cat | 81 |
| HMMReconFam24 | no cat | no cat | 81 |
| HMMReconFam610 | no cat | no cat | 81 |
| ReconFam1019 | I | LTR | 81 |
| T_scf4_4 | I | LTR | 81 |
| T_scf8_1 | I | LTR | 81 |
| HMMReconFam353 | no cat | no cat | 81 |
| ReconFam1142 | no cat | no cat | 81 |
| ReconFam24640 | II | MITE | 81 |
| ReconFam28 | II | TIR | 81 |
| T_scf80_1 | I | LTR | 81 |
| T_scf1_5 | I | LTR | 81 |
| HMMReconFam1053 | no cat | no cat | 81 |
| T_scf8_2 | I | LTR | 80 |
| HMMReconFam229 | no cat | no cat | 80 |
| T_scf23_1 | I | LTR | 80 |
| ReconFam190 | II | TIR | 80 |
| ReconFam2178 | no cat | no cat | 80 |
| MITE_LaTEX1 | II | MITE | 80 |
| ReconFam737 | no cat | no cat | 80 |
| Restless-like | II | TIR | 80 |
| T_scf2_3 | I | LTR | 80 |
| TIR-ET1 | II | TIR | 80 |
| ReconFam158 | II | TIR | 80 |
| ReconFam138 | II | TIR | 80 |

| TE ID | Class | Type | Pairwise multiple alignment identity (%) |
|---|---|---|---|
| HMMReconFam1555 | no cat | no cat | 80 |
| T_scf10_2 | I | LTR | 80 |
| T_scf41_1 | I | LTR | 80 |
| restless? | II | TIR | 80 |
| HMMReconFam1465 | no cat | no cat | 80 |
| T_scf41_2 | I | LTR | 80 |
| ReconFam16 | II | TIR | 79 |
| T_scf14_1 | I | LTR | 79 |
| T_scf1_3 | I | LTR | 79 |
| MITE_X4 | II | MITE | 79 |
| ReconFam730 | II | TIR | 79 |
| T_scf21_3 | I | LTR | 79 |
| ReconFam447 | no cat | no cat | 79 |
| HMMReconFam445 | no cat | no cat | 79 |
| ReconFam301 | II | TIR | 79 |
| HMMReconFam1527 | no cat | no cat | 79 |
| T_scf3_1 | I | LTR | 79 |
| T_scf4_1 | I | LTR | 79 |
| ReconFam244 | no cat | no cat | 79 |
| ReconFam10701 | II | MITE | 79 |
| ReconFam2208 | I | LTR | 79 |
| T_scf21_2 | no cat | no cat | 79 |
| ReconFam217 | I | LTR | 79 |
| ReconFam1180 | II | TIR | 78 |
| ReconFam281 | II | TIR | 78 |
| ReconFam1195 | I | LTR | 78 |
| T_scf18_1 | I | LTR | 78 |
| ReconFam520 | II | TIR | 78 |
| ReconFam391 | II | TIR | 78 |
| HMMReconFam2190 | no cat | no cat | 78 |
| T_scf71_1 | I | LTR | 78 |
| ReconFam419 | II | TIR | 78 |
| ReconFam280 | I | LTR | 78 |
| T_scf74_1 | I | LTR | 78 |
| HMMReconFam233 | no cat | no cat | 78 |
| T_scf38_1 | I | LTR | 78 |
| T_scf40_1 | I | LTR | 78 |
| HMMReconFam693 | I | LINE | 78 |
| ReconFam1269 | I | LTR | 78 |
| T_scf1_1 | I | LTR | 78 |
| T_scf1_4 | I | LTR | 78 |

| TE ID | Class | Type | Pairwise multiple alignment identity (%) |
|---|---|---|---|
| ReconFam1725 | II | TIR | 78 |
| ReconFam702 | I | LINE | 77 |
| HMMReconFam386 | I | LINE | 77 |
| ReconFam1305 | II | Helitron | 77 |
| Fot1-like | II | TIR | 77 |
| HMMReconFam4263 | no cat | no cat | 77 |
| T_scf63_1 | I | LTR | 77 |
| T_scf11_2 | I | LTR | 77 |
| ReconFam510 | no cat | no cat | 76 |
| ReconFam928 | I | LINE | 76 |
| HMMReconFam2229 | no cat | no cat | 76 |
| T_scf24_1 | I | LTR | 75 |
| HMMReconFam1457 | no cat | no cat | 75 |
| T_scf1_2 | I | LTR | 75 |
| HMMReconFam1504 | no cat | no cat | 75 |
| MITE_LaTEX2_master | II | MITE | 75 |
| T_scf47_1 | I | LTR | 75 |
| ReconFam350 | I | LINE | 75 |
| T_scf24_2 | no cat | no cat | 75 |
| HMMReconFam1079 | no cat | no cat | 75 |
| T_scf17_3 | I | LTR | 75 |
| HMMReconFam1300 | no cat | no cat | 74 |
| ReconFam2212 | no cat | no cat | 74 |
| ReconFam726 | II | TIR | 74 |
| T_scf10_1 | I | LTR | 74 |
| TIR-ET3 | II | TIR | 73 |
| T_scf43_2 | I | LTR | 73 |
| tc1-like | II | TIR | 73 |
| HMMReconFam472 | no cat | no cat | 73 |
| HMMReconFam487 | I | LTR | 73 |
| ReconFam1284 | no cat | no cat | 73 |
| HMMReconFam786 | no cat | no cat | 73 |
| ReconFam567 | no cat | no cat | 72 |
| T_scf4_3 | I | LTR | 72 |
| T_scf41_3 | I | LTR | 72 |
| T_scf11_1 | I | LTR | 72 |
| ReconFam1186 | no cat | no cat | 72 |
| HMMReconFam1452 | I | LINE | 71 |
| Nht1-like | II | TIR | 71 |
| T_scf9_1 | I | LTR | 71 |

| TE ID | Class | Type | Pairwise multiple alignment identity (%) |
|---|---|---|---|
| T_scf27_1 | I | LTR | 70 |
| ReconFam1181 | no cat | no cat | 69 |
| HMMReconFam3186 | I | LTR | 69 |
| ReconFam349 | no cat | no cat | 69 |
| T_scf9_2 | I | LTR | 68 |
| ReconFam317 | no cat | no cat | 68 |
| ReconFam1227 | no cat | no cat | 68 |
| T_scf37_1 | I | LTR | 31 |
| T_scf15_2 | I | LTR | 29 |
| T_scf2_2 | I | LTR | 28 |
| T_scf61_1 | I | LTR | 28 |
| T_scf20_1 | I | LTR | 26 |

## Article 4. Gene organization of the mating type regions in the ectomycorrhizal fungus *Laccaria bicolor* reveals distinct evolution between the two mating type loci

**Hélène Niculita-Hirzel, Jessy Labbé, Annegret Kohler, François le Tacon, Francis Martin, Ian R. Sanders and Ursula Kües**

**(publié dans *New Phytologist*)**

Cet article présente des travaux menés en collaboration avec une équipe suisse et une équipe allemande sur l'organisation des régions du génome responsables de la compatibilité sexuelle chez *L. bicolor*. J'ai contribué à ces analyses en fournissant les souches fongiques à nos collaborateurs et en participant au séquençage de contrôle des différents gènes utilisés dans cette étude. Pour finir, j'ai contribué par une analyse approfondie des éléments transposables et des fréquences de recombinaison des régions cibles.

New Phytologist

Research

# Gene organization of the mating type regions in the ectomycorrhizal fungus *Laccaria bicolor* reveals distinct evolution between the two mating type loci

Hélène Niculita-Hirzel[1], Jessy Labbé[2], Annegret Kohler[2], François le Tacon[2], Francis Martin[2], Ian R. Sanders[1] and Ursula Kües[3]

[1]Department of Ecology and Evolution, University of Lausanne, Biophore Building, CH–1015 Lausanne, Switzerland; [2]UMR 1136, Interactions Arbres/Microorganismes, INRA-Nancy, F–54280 Champenoux, France; [3]Molecular Wood Biotechnology and Technical Mycology, Büsgen-Institute, Georg-August-University Göttingen, D–37077 Göttingen, Germany

Author for correspondence
*Hélène Niculita-Hirzel*
*Tel: +41 21 692 4247*
*Fax: +41 21 692 4265*
*Email: helene.hirzel@unil.ch*

Received: *3 April 2008*
Accepted: *24 April 2008*

## Summary

• In natural conditions, basidiomycete ectomycorrhizal fungi such as *Laccaria bicolor* are typically in the dikaryotic state when forming symbioses with trees, meaning that two genetically different individuals have to fuse or 'mate'. Nevertheless, nothing is known about the molecular mechanisms of mating in these ecologically important fungi.

• Here, advantage was taken of the first sequenced genome of the ectomycorrhizal fungus, *Laccaria bicolor*, to determine the genes that govern the establishment of cell-type identity and orchestrate mating.

• The *L. bicolor* mating type loci were identified through genomic screening. The evolutionary history of the genomic regions that contained them was determined by genome-wide comparison of *L. bicolor* sequences with those of known tetrapolar and bipolar basidiomycete species, and by phylogenetic reconstruction of gene family history.

• It is shown that the genes of the two mating type loci, *A* and *B*, are conserved across the Agaricales, but they are contained in regions of the genome with different evolutionary histories. The *A* locus is in a region where the gene order is under strong selection across the Agaricales. By contrast, the *B* locus is in a region where the gene order is likely under a low selection pressure but where gene duplication, translocation and transposon insertion are frequent.

**Key words:** gene order conservation, mating type loci, mushroom-forming fungi, recombination, sex chromosome evolution.

*New Phytologist* (2008) **doi**: 10.1111/j.1469-8137.2008.02525.x



## Introduction

The growth of boreal, temperate and montane forest trees relies on root colonization by ectomycorrhizal fungi, Basidiomycetes and to a lesser extent Ascomycetes, that provide the host plant with nutrients and water in exchange for photosynthetically derived carbohydrates. Basidiomycetes display a life cycle that is predominantly dikaryotic with two separate haploid nuclei in the hyphal cells, the mating process between two monokaryons taking place, in natural conditions, before the establishment of the symbiotic association (Harley & Smith, 1983) or before the infection of the plant by pathogenic Basidiomycetes such as *Ustilago maydis* (for review see Banuett, 2007). In the laboratory, some basidiomycete monokaryotic mycelia have the potential to form ectomycorrhizas (Kropp & Fortin, 1988; Gay *et al.*, 1994) but their symbiotic potential was shown to be lower than that of dikaryons (Kropp & Fortin, 1988; Gay *et al.*, 1994). Consequently, understanding

how the specialized regions of the genome that control mating in fungi (the mating type loci) regulate the formation of the dikaryons, and if they regulate the establishment of the symbiotic association, is important. Although the mating type loci have been studied in a large number of basidiomycete species (for review see Casselton & Olesnicky, 1998; Fraser & Heitman, 2004; Fraser *et al.*, 2007), very little is known about the genetics of mating in ectomycorrhizal fungi. Recently the first draft genome of a basidiomycete ectomycorrhizal fungus, *Laccaria bicolor*, has been released (Martin *et al.*, 2008) and can be compared with other closely related basidiomycete genomes in which the mechanisms that give rise to mating compatibility are understood.

In all Ascomycetes, and some Basidiomycetes, mating compatibility is determined by different alleles of a single mating type (*MAT*) locus (mating type system called bipolar). In a large number of basidiomycete species (55–65% of the Agaricomycotina) the mating compatibility is controlled by different alleles of two unlinked regions of the genome (Whitehouse, 1949; Raper & Flexer, 1971), commonly known as the *A* and *B* loci (called tetrapolar mating system). The number of potentially compatible mating type alleles per mating type locus gives rise to thousands of potentially compatible mating types in some tetrapolar fungi (e.g. *c.* 12 000 in *Coprinopsis cinerea* and *c.* 20 000 in *Schizophyllum commune*; Raper, 1966). The population genetic consequences of the tetrapolar versus the bipolar system is a reduction in the amount of interbreeding among the haploid progeny of one individual (*A1B1* can mate with *A2B2*, but not with *A1B1*, *A1B2* or *A2B1*).

The genes encoded by the *A* and *B* loci have been characterized in a number of Basidiomycetes. In all studied higher Basidiomycetes, the *A* mating type locus contains genes for two types of homeodomain transcription factors (HD1 and HD2) and the genes at the *B* mating type locus encode lipopeptide pheromones and G-protein-coupled pheromone receptors (STE3-like pheromone receptors with seven transmembrane domains) (for review see Casselton & Olesnicky, 1998; Fraser *et al.*, 2007). More than one copy of these genes is frequently observed in agaricomycete species. The model species *C. cinerea* contains three gene copies for each type of homeodomain transcription factor (gene pair *a*, *b* and *d*), at least six copies of mating type pheromones and three copies of STE3-like pheromone receptors (subfamily 1, 2 and 3). By contrast, the model species *S. commune* possesses two gene copies for STE3-like pheromone receptors and six pheromone genes and at least two genes copies for each type of transcription factor (for review see Casselton & Olesnicky, 1998; Casselton & Kües, 2007). Each unique combination of alleles at the different subloci specifies a unique mating type, making the subloci redundant in function. Receptors and pheromones diverged rapidly in sequence within and between species. Only the seven transmembrane domains share enough similarity in sequence between species to help the assignment of new STE3-like receptors. By contrast, the mating type pheromones share only a CaaX motif at the *C*-terminus and this makes them difficult to detect. Similarly, the homeodomain transcription factors diverged in sequence within and between species while certain domains for essential functions remain conserved.

The Agaricomycetes use an elegant strategy to ensure that mating-dependent developmental pathways are activated only after fusion between compatible mates. Two functional domains of their mating type-specific transcription factors are separated into two proteins encoded at the *A* locus by divergently transcribed pairs of the *HD1* and *HD2* genes, respectively. The products of the *HD2* gene family carry the crucial DNA binding domain, the homeodomain motif HD2. By contrast, the products of the *HD1* gene family carry an atypical homeodomain motif HD1, with low or no affinity for DNA. In *C. cinerea*, HD1 proteins were shown to also carry the nuclear localization signal (NLS) and the activation domain (AD), thought to be required for transcriptional activation of target genes (Asante-Owusu *et al.*, 1996). Moreover, dimerization motifs were characterized in the *N*-terminal domains of *HD1* and *HD2* gene products, respectively. Indeed, heterodimerization of *HD1* and *HD2* gene products from the different mates plays an important role in bringing together the various domains required for the formation of a functional transcription factor complex (Banham *et al.*, 1995; Spit *et al.*, 1998). If the full function of HD1/HD2 heterodimers is conserved across Agaricales, the functional motifs described in *C. cinerea* are also expected to be detected in *L. bicolor* homologues.

The *L. bicolor* mating system is considered to be tetrapolar (Fries & Mueller, 1984; Kropp & Fortin, 1988; Doudrick & Anderson, 1989) with at least 45 *A* and 24 *B* mating type alleles (Raffle *et al.*, 1995). In order to understand the mechanisms involved in compatibility and incompatibility during mating in *L. bicolor*, it is essential to know which genes are located at each of the mating type loci. Furthermore, understanding how allelic diversity can be generated at these loci requires the analysis of the evolution of these two loci and, in particular, whether recombination plays a major role in the diversity and arrangement of genes at these loci. Indeed, an increase in recombination rate in the vicinity of mating type loci could reduce the association and potential selective conflicts between mating type genes and linked genes (Hsueh *et al.*, 2006). Alternatively, recombination activity within the gene could mix existing variations in a combinatorial fashion to contribute to the evolution of novel alleles (Badrane & May, 1999). Such a mechanism has been suggested to explain the evolution of shared clusters of genes among major histocompatibility complex alleles (Hughes *et al.*, 1993; Yuhki & O'Brien, 1994; Zangenberg *et al.*, 1995) and alleles of the gametophytic self-incompatibility locus (Awadalla & Charlesworth, 1999; Vieira *et al.*, 2003). Recombination within a gene may also have an advantage if it increases the effectiveness with which deleterious mutations are eliminated (Kondrashov, 1984; Rice, 2002). Such an effect may be most relevant among

interacting genes (Rice, 1998) or among sex-determining alleles (Beye *et al.*, 1999).

The aim of the present study was to identify the *A* and *B* loci in whole-genome sequences of *L. bicolor* and compare the genes and their arrangement to loci known in other Basidiomycetes such as *C. cinerea*, *Phanerochaete chrysosporium* or *Pleurotus djamor*. We then analysed whether the two loci have evolved with recombination in or around the loci and within genes at these two loci.

## Materials and Methods

### Strains and culture conditions

A fruiting body of *L. bicolor* (Maire) P.D. Orton (common name: bicoloured deceiver) was collected in 1976 under *Tsuga mertensiana* in the Crater Lake National Park (OR, USA) by J. Trappe and R. Molina and deposited at the Forest Service (Corvallis, OR, USA). A subculture of this strain, so-called S238-O, was transferred to INRA-Nancy in 1980 and renamed S238N (Di Battista *et al.*, 1996). For this study, spores of *L. bicolor* were obtained from sporophores collected under *Pseudotsuga menziesii*, that had been inoculated with *L. bicolor* strain S238N in a glasshouse or in a nursery (Di Battista *et al.*, 1996) and germinated according to Fries (1983). Forty-two different sib-monokaryotic mycelia were used in this study, including the H82 line whose genome was sequenced (Martin *et al.*, 2008). All material was stored and subcultured at INRA-Nancy (Champenoux, France). All monokaryotic strains were grown in Petri dishes containing a modified Pachlewski agar medium (Di Battista *et al.*, 1996) and incubated at 25°C.

### *MAT* loci annotation

The *L. bicolor* genome sequencing, assembly and annotation were described in Martin *et al.* (2008). All *L. bicolor* sequences from the monokaryon H82 are available at the Joint Genome Institute website: http://genome.jgi-psf.org/Lacbi1/Lacbi1.home.html. Complete *C. cinerea* (= *Coprinus cinereus*) DNA sequences and protein sequences were obtained from the *C. cinerea* strain Okayama 7 genome database (*Coprinus cinereus* Sequencing Project, Broad Institute of MIT and Harvard, http://www.broad.mit.edu; Stajich *et al.*, 2006), while sequence data for *P. chrysosporium* (Martinez *et al.*, 2004) were obtained from the Joint Genome Institute website (http://genome.jgi-psf.org/Phchr1/Phchr1.home.html) and the sequence data for *Pleurotus djamor* from National Center for Biotechnology Information (NCBI) GenBank (accession numbers AY462111 and AY462110 for *A* and *B* locus, respectively). The evolutionary relationships between species used in the present study are shown in Supplementary material Fig. S1. We carried out BLAST translated DNA and amino acid sequence similarity searches of the predicted proteins from *C. cinerea* and *P. chrysosporium* mating type regions against the *L. bicolor* translated genome and vice versa using BLOSUM62 substitution matrices and the DUST and SEG filters for low-complexity regions (Altschul *et al.*, 1997). All reciprocal hits were tested for functional homology by PSI-BLAST searches against all NCBI database (http://www.ncbi.nlm.nih.gov/BLAST/) and SMART (http://smart.embl-heidelberg.de/index2.cgi) searches for putative conserved domains. Genes were annotated in *L. bicolor* based on homology to the existing annotation in other fungi. Correction on some *L. bicolor*, *C. cinerea* and *P. chrysosporium* gene annotation as the annotation of new genes for *P. djamor* and new pheromone-like genes in *L. bicolor* and *C. cinerea* are detailed in the Supplementary Material, Text S1.

Syntenic relationships were illustrated with CHROMOMAPPER (http://www2.unil.ch/biomapper/chromomapper/, Niculita-Hirzel & Hirzel, 2008). COILS program (http://www.ch.embnet.org/software/COILS_form.html; Lupas *et al.*, 1991) was used to locate dimerization motifs at a window size of 14 sites. HMMTOP 2.0 (Tusnády & Simon, 2001), TMHMM (Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark) and PHOBIUS (Käll *et al.*, 2007) programs were used to predict the transmembrane helices in STE3-like pheromone receptor proteins.

### Phylogenetic reconstruction

Protein-coding DNA sequences were aligned using the ClustalW algorithm and edited by hand in BIOEDIT version 7.0.5.3 (Hall, 1999). To reconstruct the evolutionary relationships of *L. bicolor* STE3-like pheromone receptor sequences with that of other fungal species the full-length coding sequences of the STE3-like pheromone receptor genes deduced from the *L. bicolor* genome (http://genome.jgi-psf.org/Lacbi1/Lacbi1.home.html; Martin *et al.*, 2008), the *C. cinerea* strain Okayama 7 genome (http://www.broad.mit.edu/annotation/genome/coprinus_cinereus/Home.html), the *P. chrysosporium* genome (http://genome.jgi-psf.org/Phchr1/Phchr1.home.html), as well as from the *Coprinellus disseminatus* and *P. djamor B* locus sequences available at the NCBI GenBank were used. The *Cryptococcus neoformans STE3* sequence was added as an outgroup. The accession numbers of all these sequences are given in the Supplementary Material section. To estimate phylogenetic relationships, the alignment was analysed using sequential and parallel maximum likelihood-based inference as implemented in the RAXML software, version 4.0.0 (Stamatakis, 2006), with a general time-reversible model of nucleotide substitution and additionally assuming a percentage of invariant sites and gamma-distributed substitution rates at the remaining sites (GTR + I + G). Branch support was inferred from 1000 replicates of nonparametric bootstrapping (Felsenstein, 1985). Additionally, a Bayesian Markov chain Monte Carlo (MCMC) analysis was performed using MRBAYES 3.1 (Ronquist & Huelsenbeck, 2003). We ran two independent

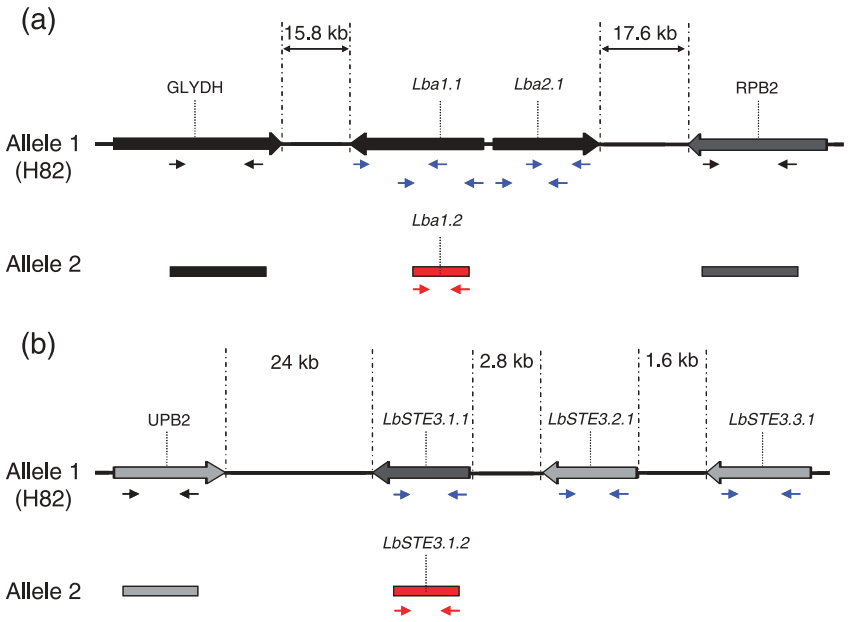



**Fig. 1** Relative position in the mating type genes of the primers used in the allele-specific polymerase chain reaction (PCR) test. Primer pairs designed for *STE3.1.1*, *STE3.2.1* *STE3.3.1*, *Lba1.1* and *Lba2.1* genes amplify specifically the allele 1 of the dikaryon S238N. Two primer pairs were designed for *Lba1.1 and Lba2.1*, one covering the 5′ region of the gene and the other the 3′ region of the gene. Primer pairs designed for *STE3.1.2* and *Lba2.2* genes amplify specifically the allele 2 of the dikaryon S238N. Primer pairs designed for *GLYDH*, *RPB2* and *UPB2* amplify the two alleles present in the dikaryon S238N. Primers are indicated as thin arrows. Genes are indicated as bold arrows, with orientation indicating the direction of transcription.

MCMC analyses, each involving six incrementally heated chains over one million generations, using the GTR + I + G model of nucleotide substitution and starting from random trees. Model parameters were not fixed but sampled during MCMC. Trees were sampled every 100 generations, resulting in an overall sampling of 10 000 trees per run, from which the first 5000 trees of each run were discarded. The remaining 5000 trees sampled from each run were pooled and used to compute a majority rule consensus tree to get estimates for the posterior probabilities. Stationarity of the process was assessed using the TRACER software (Rambaut & Drummond, 2003).

## Gene expression

Expression of predicted mating type genes was detected in free-living mycelium of *L. bicolor* S238N, ectomycorrhizal root tips (Douglas fir ECM and Poplar ECM) and fruiting bodies of *L. bicolor* S238N using the *L. bicolor* whole-genome expression array data described in Martin *et al.* (2008).

## DNA extraction

Mycelium for DNA extraction was prepared by growing isolates at 28°C in 20 ml of Pachlewski liquid medium according to Henrion *et al.* (1992), until stationary phase. The mycelium was removed from the growth medium, rinsed in $H_2O$, and lyophilized. Approximately 80 mg were used for DNA extraction, using the DNeasy Plant Mini Kit (Qiagen, Courtaboeuf, France) according to the manufacturer's instructions, and DNA was recovered in 50 µl of distilled water. Quality of DNA extracts was tested by polymerase chain reaction (PCR) amplification of the ribosomal DNA internal transcribed spacer (ITS) using the following primers: ITS1 (5′-TCCTCCGCTTATTGATATGC-3′) and ITS4 (5′-TCCGTAGGTGAACCTGCGG-3′) (Henrion *et al.*, 1992).

## Recombination analysis

In order to determine the recombination rate between and within *L. bicolor* mating type loci, the inheritance of the allele 1 or allele 2 of each mating type gene present in the dikaryon S238N was determined in 42 sib-monokaryons, ($F_1$ progeny of the dikaryon S238N) by a allele specific PCR test. The allele 1 corresponds to the monokaryon H82 sequence whose genome was sequenced (Martin *et al.*, 2008). A partial sequence of the second allele of *Lba1* (*Lba1.2*) and *LbSTE3.1* (*LbSTE3.1.2*) were obtained from the dikaryon *L. bicolor* S238N cDNA library constructed and sequenced as described in the Supplementary Material section. Contigs with alleles *Lba1.2* and with *LbSTE3.1.2* were generated from the accession numbers EYMTT9U01EFVE6, EYMTT9U01DQOK6 and EYMTT9U01BO6Z1, EYMTT9U01CEA01, EYMTT9U01-DUGTU, EYMTT9U01BX66W, EYMTT9U01BX9M2, EYMTT9U01BXYKF, EYMTT9U01D5QSG and EYMTT-9U01DNINT, respectively.

Allele-specific primer pairs were designed for amplifying mating types genes *Lba1.1*, *Lba2.1*, *STE3.1.1*, *STE3.2.1* or *STE3.3.1* alleles or for genes *Lba1.2* or *STE3.1.2* from the opposite mating type, respectively. Primer pairs were also designed for detection of either of the respective alleles of two genes surrounding the A locus: *GLYDH*, *RPB2*, and of one gene localized in the proximity of the *B* locus: *UPB2* (Fig. 1, see the Supplementary Material, Table S1).

Two independent PCRs were conducted for each gene in each monokaryon. The DNA from the monokaryon H82 and the dikaryon S238N was used as a positive control. A

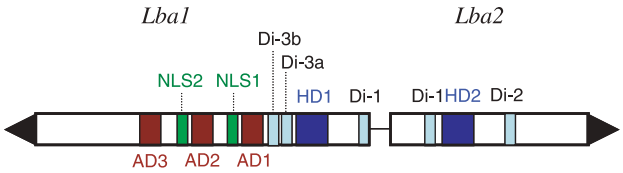



**Fig. 2** Schematic depiction of *Lba1* and *Lba2* with their regions encoding conserved and divergent protein domains: homeodomain (HD1 and HD2), nuclear localization signals (NLS1 and NLS2), activation domains (AD1 to AD3) and dimerization motifs (Di-1 to Di-4). Genes are indicated as arrows, with orientation indicating the direction of transcription.

PCR was carried out on 15 ng of genomic DNA in 25 μl, with final concentrations of 250 μM of each dNTP, 0.2 μM of each of the primers (Microsynth, Balgach, Switzerland), Qiagen PCR buffer containing 25 mM $MgCl_2$ 1×, 1.5 units of *Taq* DNA Polymerase (Qiagen). Reactions were performed in a T1 Thermocycler (Biometra, Châtel-St-Denis, Switzerland) under the following conditions: initial denaturation at 94°C for 30 s, annealing at specific temperature (Supplementary Material, Table S1) for 30 s and extension at 72°C for 90 s. After 40 cycles, a terminal extension of 7 min at 72°C was carried out. Finally the PCR products were run in 1% agarose gel. In order to confirm if the sequence is belonging to the allele 1 or allele 2 of a particular gene, amplicons from five monokaryons were purified using a MinElute PCR purification kit (Qiagen, Hombrechtikon, Switzerland) and directly sequenced on both strands using the forward and reverse primers. Sequencing reactions utilized the BigDye sequencing kit (Applied Biosystems, Rotkreuz, Switzerland) and were analysed on an ABI3700 DNA sequencer. The two alleles of *GLYDH*, *RPB2* and *UPB2* in the progeny dikaryon S238N were always found to be identical in sequence.

## Results

### Characterization of the *L. bicolor* A locus

We screened the *L. bicolor* genome of the H82 monokaryon for *HD1* and *HD2* sequences and found one gene encoding the HD1 domain (*Lba1*, ID: 301103) and one gene encoding the HD2 domain (*Lba2*, ID: 379291). The two genes are linked and divergently transcribed (Fig. 2) as seen in *A* mating type loci of other basidiomycete species (Casselton & Olesnicky, 1998; James *et al.*, 2004, 2006). As expected, HD1 and HD2 domains were both highly conserved between the Basidiomycetes, including *L. bicolor*. However, few other motifs were found to be conserved among products of the *HD1* gene family (see the Supplementary Material, Fig. S2). The absence of motifs conserved among products of the *HD2* gene family (see the Supplementary Material, Fig. S3) is in accordance with the finding in *C. cinerea* that almost all sequence *C*-terminal to the HD2 homeodomain can be deleted without a loss of function (Kües *et al.*, 1994).

The conserved motifs between the HD1 proteins correspond to nuclear localization signals, transactivation domains and dimerization domains already characterized in *C. cinerea*. Indeed, the two nuclear localization signals (NLS1 and NLS2) present in the *C*-terminal region of the *C. cinerea* HD1 proteins (Tymon *et al.*, 1992, Asante-Owusu *et al.*, 1996, Spit *et al.*, 1998) were both detected in the *L. bicolor* Lba1 protein (Fig. 2, Fig. S2). These observations suggest a functional conservation of both nuclear localization motifs in Agaricales.

Four transactivation domains, AD1–AD4, were predicted in *C. cinerea* HD1 proteins (Badrane & May, 1999), but only three of them (AD1, AD2 and AD4) appear to be of functional importance (Kües *et al.*, 1994, Asante-Owusu *et al.*, 1996). The similarity in the serine, threonine and proline composition (Mermod *et al.*, 1989; Tymon *et al.*, 1992) and the sequence conservation of five to seven amino acids between AD1 or AD2 transactivation domains of *L. bicolor*, *C. cinerea*, *C. disseminatus* (a1 protein only) and *Coprinopsis scobicola* HD1 proteins, suggest a functional conservation of these two domains through all these species. By contrast, the low similarity in amino acids composition and the absence of sequence conservation observed between AD4 of *L. bicolor* and the other fungal species suggest that this transactivation domain might not be functional in *L. bicolor* (Fig. S2).

Because a heterodimeric interaction between HD1 and HD2 proteins is required for sexual compatibility (Banham *et al.*, 1995; Spit *et al.*, 1998), we looked for coiled/coil motifs using predictive algorithms (Lupas *et al.*, 1991). The robustness of these predictions has already been tested by Badrane & May (1999) on *C. cinerea* proteins. Two coiled/coil motifs, Di-1 and Di-2, were predicted in the Lba2 protein with a high probability ($P = 0.90$ and $P = 0.98$, respectively; Fig. S3), one on each side of the HD2 domain (Fig. 2). Interestingly, only the Di-1 motif is at a similar position in all Agaricales proteins and was shown to be functional in *C. cinerea* proteins suggesting that only this motif has an important role in Lba1/Lba2 dimerization (Banham *et al.*, 1995).

For the Lba1 protein, three coiled/coil motifs were predicted with distinct probabilities: one on *N*-terminal and two on *C*-terminal of HD1 domain (Fig. 2). The *N*-terminal motif, Di-1, was predicted with a low probability ($P = 0.30$) as in *C. cinerea* alleles ($P = 0.088$–$0.35$). However, Di-1 was proved to have a functional importance for heterodimerization and protein discrimination in *C. cinerea* (Banham *et al.*, 1995). Consequently, it seems that the Di-1 motif could also participate in Lba1/Lba2 protein dimerization. The *C*-terminal coiled/coil motifs, Di-3a and Di-3b, of Lba1 were predicted with a strong probability of 0.90 and 0.98, respectively (Fig. S2) and their positions are conserved between the Lba1 and two proteins of *C. disseminatus* or Lba1 and most *C. cinerea* HD1 proteins. Nevertheless, *in vitro* protein dimerization studies with *C. cinerea* proteins suggest that Di-3b does not participate in heterodimerization

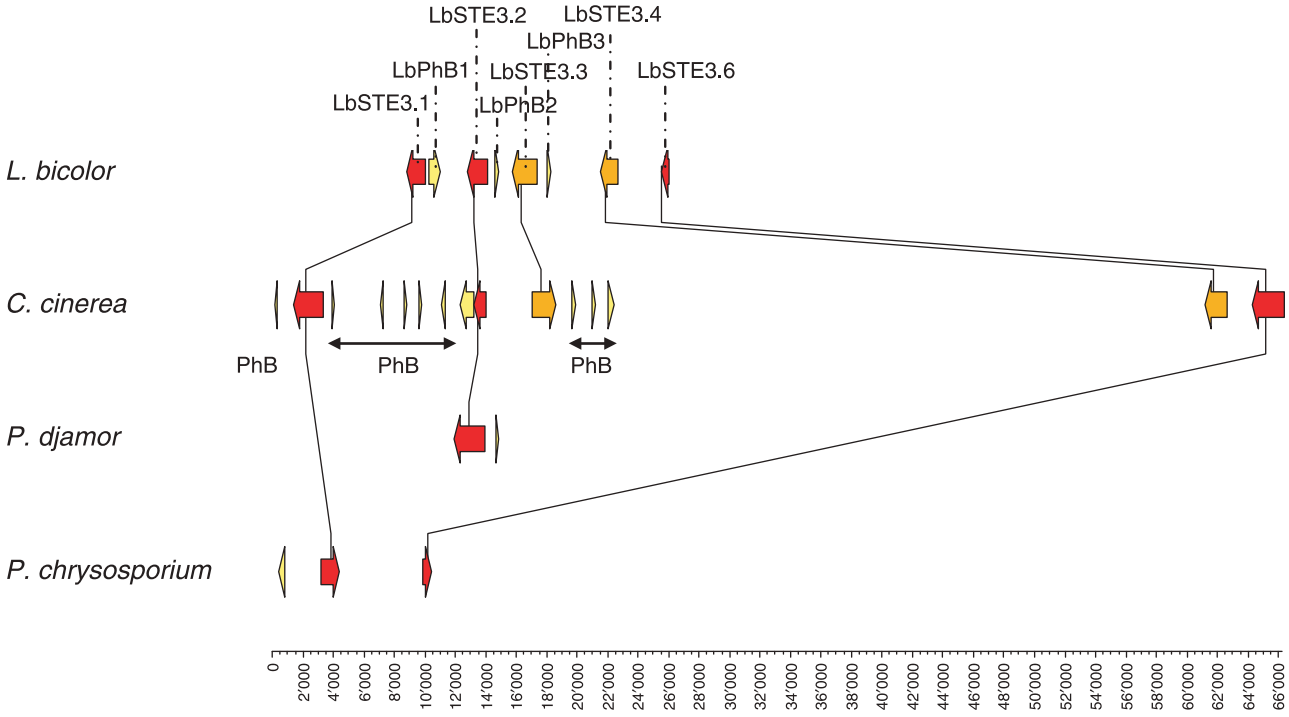



**Fig. 3** Schematic comparison of the *B* locus from *Laccaria bicolor*, *Coprinopsis cinerea*, *Pleurotus djamor* and *Phanerochaete chrysosporium* drawn with CHROMOMAPPER (Niculita-Hirzel & Hirzel, 2008). Only the pheromones genes and STE3-like receptors were drawn. The species names are labelled on the left. Arrows represent individual genes. The colour gradient indicates in how many species the gene is present, from yellow (one species) to red (all species). Connectors join orthologous genes, as determined by the phylogenetic analysis performed in the present study. The DNA scale (bp) is indicated on the bottom. The order of *the C. cinerea* genes from left to right is as follows: *CcPhB1*, *CcSTE3.1* (*rcb1*[43]), *CcPhB2* (*phb1*[43]), *CcPhB3*, *CcPhB4* (*phb2.3*[43]), *CcPhB5* (*phb2.2*[43]), *CcPhB6* (*phb2.1*[43]), *CcSTE3.2a* (*rcb2*[43]), *CcSTE3.2b*, *CcSTE3.3* (*rcb3*[43]), *CcPhB7* (*phb3.1*[43]), *CcPhB8* (*phb3.2*[43]), *CcPhB9* (*phb3.3*[43]), *CcSTE3-2151* and *CcSTE3-2153* (names in brackets refer to the nomenclature used previously for genes as identified by Riquelme *et al*., 2005). The order of the *P. djamor* genes from left to right is *Pdphb1* and *PdSte3.3* (James *et al*., 2004) and the order of the *P. chrysosporium* genes *PcSTE3-432*, *PcSTE3-276*, *PcSTE3-430* (this study). Note that there are no pheromone genes linked to the pheromone receptor genes in the bipolar basidiomycete *P. chrysosporium*.

(Banham *et al*., 1995). These contradictory data prevented us from concluding on an eventual function of the Di3-a motif in Lba1/Lba2 dimerization.

## Characterization of the *L. bicolor B* locus

In a tetrapolar species, such as *L. bicolor*, the *B* locus is expected to contain at least one mating type pheromone gene and one STE3-like pheromone receptor gene. The *L. bicolor* genome was screened with all known Agaricales STE3-like sequences and thirteen STE3-like pheromone receptor genes were identified. Their distribution in the *L. bicolor* genome is given in the Supplementary Material (Table S2). Only eight of these genes encode all the seven transmembrane helices characteristic of STE3-like receptors (as predicted by HMMTOP, TMHMM and PHOBIOUS programs). The other five genes correspond to partial duplication of pheromone receptor genes and they are not functional (Table S2). Mating type specific pheromone genes (*PhB1*, *PhB2* and *PhB3*; Fig. 3), encoding the characteristic CaaX motifs at the *C*-terminus of their proteins, were detected only in the close proximity of *LbSTE3.1*, *LbSTE3.2* and *LbSTE3.3*. These three STE3-like pheromone receptors/mating type pheromone gene pairs are clustered in one locus of scaffold 56 (Fig. 3).

Interestingly, the molecular phylogeny of STE3-like pheromone receptor sequences from *L. bicolor*, *C. cinerea*, *C. disseminatus*, *P. djamor* and *P. chrysosporium* shows that LbSTE3.1, LbSTE3.2 and LbSTE3.3 belong to the three independent subfamilies of STE3-like receptors known to be involved in mating in *C. cinerea* (Riquelme *et al*., 2005, Fig. 4). In *C. cinerea*, the genes for these three STE3-like receptors (*CcSte3.1*, *CcSte3.2b* and *CcSte3.3*) are grouped into a single locus encompassing approx. 9.5 kb, with two or more mating-type-specific pheromone genes being associated with each of them and an extra gene coding for another CcSte3.2 receptor (*CcSte3.2a*) in between that appeared to be obtained through a gene duplication event (Figs 3, 4). Surprisingly, the order of the genes coding for the three independent subfamilies of STE3-like receptor is conserved between *C. cinerea* strain Okayama 7 and *L. bicolor* H82 monokaryon despite the lack of homology in the surrounding regions between the two species.

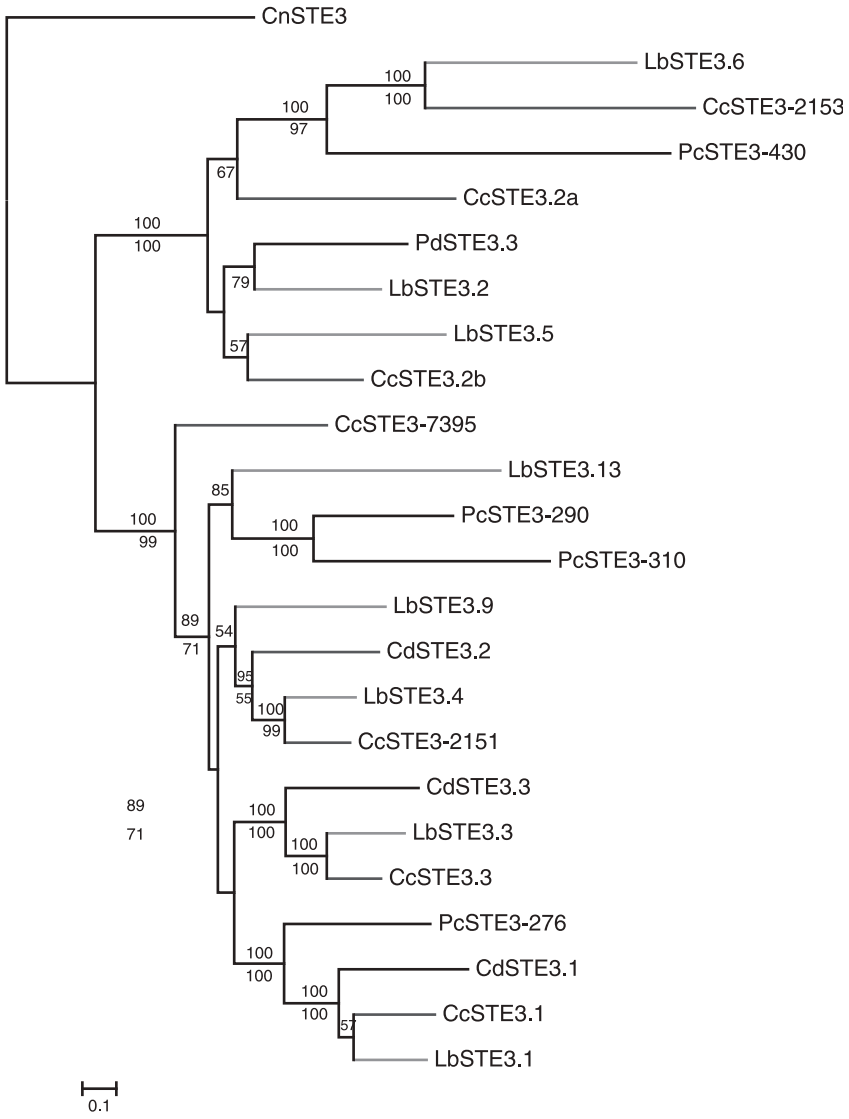



**Fig. 4** Phylogenetic relationships between the STE3-like pheromone receptors. An alignment of the seven transmembrane domains of the coding STE3-like sequences was used. The phylogram is derived by heuristic maximum-likelihood (ML) analysis using a time-reversible model of nucleotide substitution, additionally assuming a portion of invariable sites and gamma-distributed substitution rates at the remaining sites (GTR + I + G). Branch support values were calculated from Bayesian Markov chain Monte Carlo analysis (numbers above) and from 1000 replicates of nonparametric ML bootstrap analysis (numbers below). Branch lengths are scaled in terms of expected numbers of nucleotide substitutions per site. STE3.1, STE3.2 and STE3.3 indicate, respectively, the three subfamilies of STE3-like receptors involved in mating. Cc, *Coprinopsis cinerea*; Cd, *Coprinellus disseminatus*; Cn, *Cryptococcus neoformans*; Lb, *Laccaria bicolour*; Pc, *Phanerochaete chrysosporium*; Pd, *Pleurotus djamor*.

Four of the five other full-length *L. bicolor* STE3-like receptors genes (*LbSTE3.4*, *LbSTE3.6* and *LbSTE3.9*, *LbSTE3.13*) appear to be homologues to three other subfamilies of STE3 receptors that are not known to play a role in mating (Fig. 4). Moreover, no pheromone genes were found in the proximity of these receptors suggesting that they do not assume the *B* locus function. Nevertheless, 45 more pheromone-like genes encoding peptides with a CaaX motifs at their *C*-terminus, were detected spread throughout the *L. bicolor* genome, of which 36 look functional (Table S3). The unprocessed precursors of all these pheromone-like peptides have a shorter *N*-terminal domain than the mating type-specific pheromone precursors and, surprisingly, they have multiple homologues in the *C. cinerea* genome (Fig. S4). The sequence conservation of these pheromone-like peptides, as well as of nonmating type STE3-like pheromone receptors, suggests that the function of these shared components of signalling pathways is conserved between *C. cinerea* and *L. bicolor*.

To examine whether all the full-length STE3-like receptors of *L. bicolor* are functional, transcript profiling was carried out using the custom NimbleGen oligoarray. The eight full-length STE3-like receptor genes were all expressed in free-living mycelium, ectomycorrhizal root tips and fruiting bodies (Table S4). But only *LbSTE3.1*, *LbSTE3.2* and *LbSTE3.6* genes were strongly repressed, as were the three mating type pheromone genes and the transcription factor genes *Lba1* and *Lba2* in the fruiting body of the S238N strain (Table S4). This finding is interesting because the fruiting body is the tissue in which changes from a dikaryotic stage to a diploid stage occurs.

## Recombination within and around *A* and *B* mating type loci

Analysis by allele-specific PCR test of *A* and *B* allele segregation in S238N progeny showed that the two mating

type loci are unlinked. Moreover, when the two loci are mapped on linkage group J in the *L. bicolor* genetic map of Doudrick *et al*. (1995), their position corresponds to that expected for the mating type loci (J. Labbé *et al*., unpublished). These observations confirmed that the newly characterized *A* and *B* loci correspond to the functional *A* and *B* loci. Interestingly, the *A* and *B* mating type loci are located on chromosomes of very different size – the *A* locus on the largest chromosome and the *B* locus on one of the smallest chromosomes (Martin *et al*., 2008; J. Labbé *et al*., unpublished) – as is the case in *C. cinerea* (O'Shea *et al*., 1998). In *C. cinerea*, the linkage map is similar in length between the two chromosomes containing the mating type loci despite the huge difference between their physical maps (Muraguchi *et al*., 2003). In consequence, the organization around the *A* locus is expected to be maintained by a low recombination frequency, whereas that around the *B* locus is predicted to be much less conserved because of a high recombination frequency (Casselton & Kües, 2007).

To examine whether the *L. bicolor A*- and *B*-containing regions are in genomic regions with different recombination rates, we explored different genomic characteristics that are known to illustrate the recombination frequency such as the length of intergene distance or the presence of G + C rich sequences. Indeed, although recombination events would be counterselected within clusters of essential genes with short intergene distances (Pal & Hurst, 2003; Poyatos & Hurst, 2007), they will be frequent in clusters of genes with large intergene distances (Cromie *et al*., 2007). Moreover, in the basidiomycete *C. neoformans* frequent recombination events, detected in the border of the mating type locus, are associated with regions with a high G + C content (Hsueh *et al*., 2006).

To address this issue, we first determined the median intergenic distance. It was found that the intergene spacers had different sizes between the two regions of the genome. The distance between genes was small in the 450 kb surrounding the *A* locus (median = 229 bp; Supplementary material Fig. S5a). By contrast, it was longer in the 350 kb surrounding the *B* locus (median = 1.4 kb; Supplementary Material Fig. S5c). Notably, the intergenic distances within the *B* locus *(LbSTE3.1*, *PhB1*, *LbSTE3.2*, *PhB2*, *LbSTE3.3*, *PhB3*) were relatively short, those within *B* subloci being shorter than those between *B* subloci (464 bp and 819 bp, respectively; Fig. S5c). This difference in the intergene length between the two regions containing the *MAT* loci suggests differences in recombination rates between these two regions, the *A* locus being in region subject to lower rates of recombination.

Second, we monitored the presence of potential recombination hotspots by determining the G + C content in the two genomic regions with a 4-kb sliding window. No G + C-rich sequences were detected in the proximity of the *A* locus, the closest one being situated 250 kb downstream of the *A* locus (Fig. S5b). However, no G + C peak was detected surrounding the *B* locus either (Fig. S5d). These features would suggest an absence of recombination and, in particular, of hotspot recombination in the proximity of both loci.

Recombination is known to be suppressed within the mating type loci of different basidiomycetes (Fraser *et al*., 2007) or within the regions encoding the protein domains mediating self/nonself recognition (Badrane & May, 1999). When recombination occurs a local enrichment in G and C can often be observed (Duret *et al*., 2006). Therefore, we calculated the G + C content in the two *MAT* loci with a small sliding window of 100 bp. Short G + C rich sequences (over 60%) were detected within both loci: in the *B* locus upstream of the *LbSTE3.1, LbSTE3.2* and *LbSTE3.3* genes, and in the *A* locus in the 3′end of gene *Lba2* (Supplementary Material, Fig. S6). The authenticity of these recombination sites was tested by an allele-specific PCR test for mating type genes. This test monitored the recombination events occurring at the *MAT* loci of the 42 S238N sib-monokaryotic progeny between *LbSTE3.1*, *LbSTE3.2* and *LbSTE3.3* genes as well as between and within genes *Lba1* and *Lba2* (Fig. 1, Table 1). Surprisingly, in most of the recombining individuals only one of the two genotypes, expected after a reciprocal exchange, is observed, suggesting that these products are an outcome of a nonreciprocal exchange (gene conversion). The excess of one parental genotype to the other infers the same hypothesis. A high number of recombination events were detected in the 3′ end of the *Lba2* gene and upstream of the *LbSTE3.2* and *LbSTE3.3* genes, as predicted by the high G + C content (Table 1). Recombination events occurring upstream of the *LbSTE3.1* gene could not be detected by the PCR test. No crossing-over events were detected in the *A* locus, but one was detected in the *B* locus between the first and second *B* subloci.

## Gene order conservation

Because gene pairs with short intergene spacers are less likely to have been rearranged during evolution (Poyatos & Hurst, 2007) we expected a higher conservation of the gene order in the *A* mating type containing region than in the *B* mating type containing region, with a conservation of the gene order restricted to the *B* locus itself. This prediction was confirmed by genomic comparison between *L. bicolor* and *C. cinerea*, the closest species to *L. bicolor* with a fully sequenced genome. The syntenic region containing the *A* locus spans more than 350 kb (331 kb in *L. bicolor* and 363 kb in *C. cinerea*) with 119 genes being homologues between the two species (Fig. 5; 36–80% of sequence identity depending on the genes with the exceptions of *HD1* and *HD2* genes that present only 20% and 15% of sequence identity, respectively). Just two insertions of one and three genes and one deletion of one gene (*FMNOR*) are observed in this part of the *L. bicolor* genome compared with the *C. cinerea* genome. The shared gene order is also relatively conserved at a longer phylogenetic distance, such as with *P. djamor* and *P. chrysosporium* (Fig. 5). The inversion events that disturbed the synteny in this region

**Table 1** Recombination events detected in the mating type loci of $F_1$ progeny of S238N

| Locus | Genotype | Number observed | Observed frequency |
|---|---|---|---|
| B | STE3.1.1 STE3.2.1 STE3.3.1 | 13 | 31.0% |
| B | **STE3.1.2 STE3.2.2 STE3.3.2** | 7 | 16.7% |
| B | **STE3.1.2** STE3.2.1 **STE3.3.2** | 8 | 19.0% |
| B | STE3.1.1 **STE3.2.2** STE3.3.1 | 0 | 0.0% |
| B | STE3.1.1 STE3.2.1 **STE3.3.2** | 12 | 28.6% |
| B | **STE3.1.2 STE3.2.2** STE3.3.1 | 0 | 0.0% |
| B | **STE3.1.2** STE3.2.1 STE3.3.1 | 1 | 2.4% |
| B | STE3.1.1 **STE3.2.2 STE3.3.2** | 1 | 2.4% |
| A | A2C.1 A2N.1 A1N.1 A1C.1 | 8 | 20.5% |
| A | **A2C.2 A2N.2 A1N.2 A1C.2** | 12 | 30.8% |
| A | A2C.1 A2N.1 A1N.1 **A1C.2** | 0 | 0.0% |
| A | **A2C.2 A2N.2 A1N.2** A1C.1 | 2 | 5.1% |
| A | A2C.1 **A2N.2 A1N.2 A1C.2** | 13 | 33.3% |
| A | **A2C.2** A2N.1 A1N.1 A1C.1 | 0 | 0.0% |
| A | A2C.1 **A2N.2 A1N.2** A1C.1 | 1 | 2.6% |
| A | **A2C.2** A2N.1 A1N.1 **A1C.2** | 0 | 0.0% |
| A | A2C.1 **A2N.2** A1N.1 A1C.1 | 3 | 7.7% |
| A | **A2C.2** A2N.1 **A1N.2 A1C.2** | 0 | 0.0% |

The parental alleles are noted 1 (H82 monokaryon) and 2 (in bold type) for each locus. N refers to the 5′ region of the gene and 'C' to the 3′ region of the gene.

during evolution are clearly visible. It is interesting to note that all the genes in this conserved region are single copies in the fungal genomes and that several of them, including *pab1* and *ade8*, are critical for fungal fitness in nature (Fig. 5; Kües *et al*., 2001; James *et al*., 2004).

In contrast to the *A* locus-containing region, little gene order conservation was observed around the *B* locus between *L. bicolor* and the other Agaricomycotina (Supplementary Material, Fig. S7).

Two STE3-like pheromone receptor genes, *LbSTE3.4* and *LbSTE3.6*, are located at the proximity of *LbSTE3.1*, *LbSTE3.2* and *LbSTE3.3* in the *L. bicolor* genome. Similarly, their homologues in *C. cinerea*, *CcSTE3-2151* and *CcSTE3-2153* are located 38 kb downstream of the *C. cinerea B* locus. One gene, UPB2, coding for a conserved protein of unknown function, is found in *C. cinerea* and *P. djamor* in close vicinity to the *B* mating type genes and is still linked to the *L. bicolor B* locus but it is highly duplicated upstream to the pheromone receptor genes of the *L. bicolor B* locus (Fig. S7). Two genes linked to the *B* locus in *P. djamor* and *C. cinerea* (*RPS19* and *CLA4*) were not found in *L. bicolor* in the region surrounding of the *B* locus, but in another scaffold (scaffold 5) which also contains a gene coding for a nonmating-specific STE3-like pheromone receptor (*LbSTE3.9*) at a distance of 424 kb. By contrast, two genes located 5′ (*2083* and *2084*) and two genes located 3′ (*2089* and *2114*) to genes *RPS19* and *CLA4* in *C. cinerea* were also found 240 kb upstream of the *L. bicolor B* locus (Fig. S7). Interestingly, in addition to gene duplication and translocation events, the surrounding of the *L. bicolor B* locus showed an accumulation of transposable elements (Fig. S7). These observations suggest high rearrangement rates in the genomic regions surrounding the *B* locus of *L. bicolor*.

## Discussion

We have investigated the genetics underlying of the tetrapolar mating system of the ectomycorrhizal fungus *L. bicolor*. Using genome comparison, molecular phylogeny, allele-specific PCR tests and gene expression, we demonstrated that the mating specificity in this species is encoded by the two mating type loci known in Agaricomycotina, the *A* mating type locus (homeodomain transcription factor genes) and the *B* mating type locus (pheromones/pheromone receptor genes). The two loci are located in two unlinked genomic regions and have a very different evolutionary history.

### Conservation of mating type gene functions

Comparisons of gene structures between distinct lineages have been instrumental in understanding the conservation of structural motifs necessary for protein function and, further, in developing predictive algorithms that relate secondary structure and protein function (Lupas *et al*., 1991). We used the sequence variation at the *A* and *B* mating type loci to test the correlation between known function and sequence conservation. We showed that, despite more than 100 million years of divergence (Berbee & Taylor, 2001), the *L. bicolor* and *C. cinerea* mating type proteins share sequence conservation for all *C. cinerea* functional domains. The Lba2 protein carries the classical homeodomain that binds target DNA sequences. The Lba1 protein carries the atypical homeodomain, two nuclear localization domains, as well as at least two of the functional transactivation (AD1 and AD2) domains described in *C. cinerea* HD1 proteins AD1 and AD2 (Kües *et al*., 1994; Asante-Owusu *et al*., 1996; Badrane & May, 1999). The

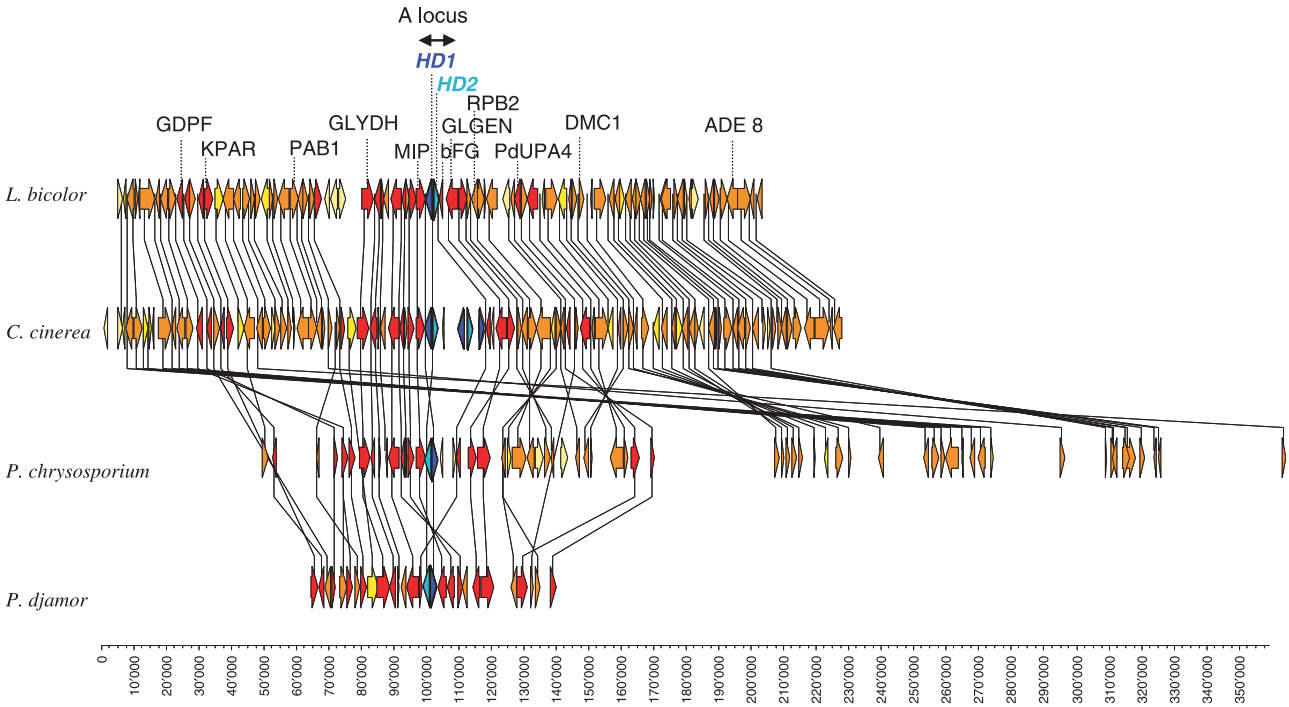



**Fig. 5** Schematic comparison of genomic regions containing the *A* locus from *Laccaria bicolor*, *Coprinopsis cinerea*, *Pleurotus djamor* and *Phanerochaete chrysosporium* drawn with CHROMOMAPPER (Niculita-Hirzel & Hirzel, 2008). Some genes conserved across species are of known function such as mitochondrial intermediate peptidase (*MIP*), the essential subunit of Sec61 complex that forms a channel for SRP-dependent protein import and retrograde transport of misfolded proteins out of the ER (*Sec61*), glycine dehydrogenase (*GLYDH*), ketopantoate reductase-like protein (*KPAR*), GDP-fucose transporter-like protein (*GDPF*), *para*-amino acid synthase (*pab1*) and aminoimidazole ribonucleotide synthetase (*ade8*). The species names are labelled on the left. Arrows represent individual genes. The colour gradient indicates in how many species the gene is present, from yellow (one species) to red (all species). Dark and light blue arrows refer to *HD1* and *HD2* genes, respectively. Connectors join orthologous genes. The DNA scale (bp) is indicated on the bottom. Data stem from *P. djamor* (positions 1–75 000 from GenBank accession no. AY462111; James *et al*., 2004), *C. cinerea* (positions 382 826–488 764 from contig 26, http://www.broad.mit.edu/annotation/genome/coprinus_cinereus/Home.html), *L. bicolor* (positions 1 143 076–1 245 371 from scaffold 1, JAZZ annotation v.1.0, http://genome.jgi-psf.org/Lacbi1/Lacbi1.home.html), *P. chrysosporium* (positions 1 012 000–1 106 000 from scaffold_13, http://genome.jgi-psf.org/Phchr1/Phchr1.home.html).

position of dimerization domains is also conserved between *L. bicolor* and *C. cinerea* in HD1 and HD2 proteins. Conservation is, however, surprising for the AD3 transactivation domain as this motif is less conserved between *C. cinerea* and its closer relatives, *C. scobicola* and *C. disseminatus*, than between *C. cinerea* and *L. bicolor* (Fig. S2). Moreover, this domain was not found to be functionally essential in *C. cinerea* HD1 proteins (Kües *et al*., 1994; Asante-Owusu *et al*., 1996). Site-directed mutagenesis combined with biochemical assays will be necessary to test if this domain provides a specific function to *Lba1* alleles.

In the *B* locus, STE3-like pheromone receptor genes encode for a seven transmembrane protein which couple to heterotrimeric guanine nucleotide-binding proteins (α, β, and γ GTPase proteins) to effect intracellular signalling (Casselton & Olesnicki, 1998). Despite the dissimilarity in primary sequence, the *L. bicolor* STE3-like pheromone receptor genes have a common tertiary structure with those of the other Agaricales, namely, a short *N*-terminal extracellular domain, seven hydrophobic α-helices (the transmembrane domains), three extracellular and three intracellular loops, a *C*-terminal intracellular tail (Supplementary Material, Fig. S8). The particularity of *L. bicolor* receptors lies in the relatively short *C*-terminal tail compared with those of *C. cinerea* or of those of the ascomycete *Saccharomyces cerevisiae*. Such a short terminal tail means that an interaction between the terminal tail and the GTPase α subunit is unlikely. In the *S. cerevisiae* STE3 receptor, the second and third intracellular loops of the protein were predicted to interact with the GTPase α subunit (reviewed by Bourne, 1997) and the first and third extracellular loops with the mating type specific pheromone (Sen & Marsh, 1994). If these functions are conserved in Agaricales, a higher conservation of the domain interacting with the GTPase protein (the second and third intracellular loop of STE3-like receptor) compared with those interacting with pheromones (first and third extracellular loops of STE3-like receptor) is expected, each pheromone and its respective receptor being engaged in long-term coevolution. Surprisingly, the first and third extracellular loops are more conserved than the second intracellular loop between all mating type specific and

nonmating type specific STE3-like paralogues. No conservation was observed for the third intracellular loop. These features make it difficult to predict which domains interact with the mating type pheromones. However, they suggest a functional innovation for the STE3-like receptor in Agaricales: the capacity to interact by their extracellular domains with proteins other than the mating type pheromones, for example, the conserved nonmating pheromones whose genes were detected in this study in the *L. bicolor* genome. The sequence variability at the internal domain could reflect receptor-specific interactions with different classes of GTPase α proteins. Indeed, eight classes of GTPase α genes were found in the *L. bicolor* genome, four being specific to Agaricales (Martin *et al.*, 2008; S. Duplessis *et al.*, unpublished).

## Evolution of mating type loci

The molecular characterization of the *L. bicolor A* locus highlights that this locus encodes only one pair of homeodomain transcription factors as in the bipolar *P. chrysosporium* and the tetrapolar species *P. djamor*. This distinguishes *L. bicolor* from *C. cinerea,* which encodes three functional pairs of homeodomain transcription factors. Interestingly, the number of *A* alleles observed in *C. cinerea* populations (124–164 mating-types; Raper, 1966; May & Matzke, 1995) is three times more than in *L. bicolor* (*c.* 45 mating-types; Raffle *et al.*, 1995) or *P. djamor* (*c.* 58 mating-types; James *et al.*, 2004) populations. This suggests similar rates of evolution of each *A* sublocus within species under the assumption of a low recombination rate between the *A* subloci in *C. cinerea*. The low recombination rate and the small number of rearrangement events observed in the region containing the *A* locus would explain the homogeneity in the evolutionary rate of the *A* locus in Agaricales. Situated in a low recombining region with a very low transposable element density, the divergence in the self-recognition domains between the products of the *A* locus alleles is favoured, the multiple alleles of this region being maintained in the population by balancing selection (May *et al.*, 1999). By contrast, the DNA-binding domains of the HD2 mating type proteins are essential and should not be able to tolerate the accumulation of deleterious mutations. Biased gene conversion has to purge such mutations, as shown in *C. cinerea* (Badrane & May, 1999) and as suggested by the present study.

The organization of the *B* locus is more complex in *L. bicolor* than the organization at the *A* locus. Three tightly linked subloci, born from tandem duplication of pheromone/receptor units, were identified. These two duplication events have been traced by phylogeny from the common ancestor to *C. cinerea* and *L. bicolor*. The *B* mating type genes of *L. bicolor* encode proteins that contain structural domains (seven transmembrane domains for the STE3-receptor) and self-recognition domains. Balancing selection is expected to maintain linked allelic diversity between a given class of receptors and a given class of pheromones. Indeed, short intergene spacers counterselect recombination events within subloci. Unexpectedly, the three *B* subloci must have remained linked for a long time since the general gene order is conserved between *L. bicolor* and *C. cinerea*. This suggests that these three genes are functionally important or that recombination was suppressed between *B* subloci. Nevertheless, in *C. cinerea* each one of these subloci is sufficient for mating (Riquelme *et al.*, 2005). The conservation of two other STE3-like receptor genes, *LbSTE3.4* and *LbSTE3.6*, in the proximity of the *B* locus and the similarity of the *LbSTE3.6* expression profile to that of the mating type genes in *L. bicolor* open the discussion on the existence of mating type STE3-like receptors other than those already characterized in *C. cinerea*.

Unlike the *A* locus, the *B* locus is surrounded by genes with relatively long intergene distance. This feature allows sequence translocation, gene duplication and transposon insertion (Fig. S7) that has caused divergences among related species. The rapid evolution of the region surrounding the *B* locus would favour a species-specific rate of *B* locus diversification (*c.* 24 mating-types in *L. bicolor*, Raffle *et al.*, 1995; *c.* 79 mating-types in *C. cinerea*, Riquelme *et al.*, 2005; *c.* 231 mating-types in *P. djamor*, James *et al.*, 2004).

It is interesting to note the striking difference between evolution at the *A* and *B* loci. Despite typical features characteristic of sex-determining genes (fast-evolving genes, suppression of recombination, punctual gene conversion to protect the functional domains against erosion), the genomic regions containing the two mating type loci seem to have different evolutionary histories. The *A* locus is in a region where the gene order is under strong selection across the Agaricales. Consequently, the genes linked to the *A* locus evolve at a constant rate in each species. By contrast, the *B* locus is in a region where the gene order is likely to be under low selection pressure but where gene duplication, translocation and transposon insertion are frequent. Consequently, the genes surrounding the *B* locus evolve at a species-specific rate, depending on the recombination rate variation.

Understanding the differences in mating type locus evolution is important for understanding the theoretical basis for sex chromosome formation. Indeed, a widely accepted scheme for the evolution of Y chromosomes in the XX/XY system consists of three major events: acquisition of the sex-determining loci, suppression of recombination, and genetic degeneration driven by evolutionary processes such as Muller's ratchet (Charlesworth, 1978; Charlesworth & Charlesworth, 2000). However, the evolution of sex chromosomes of a haploid system was predicted to be different from that in a diploid system (Bull, 1983). In a haploid organism, degeneration should not occur, because it would impair essential genes. In *L. bicolor*, further investigations consisting of a full comparative genomic analysis of the mating type containing regions from compatible monokaryons are needed to determine whether gene loss and degeneration occur in the mating type containing regions of

mating compatible partners. Nevertheless, the frequent gene duplication and inversion events as well as the accumulation of transposable elements combined with gene degenerations observed in the *B* mating type containing region in the present study show that the *B*-containing region evolves faster than the *A*-containing region. These events resemble those that shaped the *Cryptococcus* mating type region (Fraser *et al.*, 2004) and those that shaped the sex chromosomes in animals and plants (Fraser & Heitman, 2004).

## Acknowledgements

We would like to thank Timothy James and the three anonymous referees for critical comments on an earlier version of this paper. We acknowledge the Joint Genome Institute and the *Laccaria* Genome Consortium for access to the *L. bicolor* genome sequence before publication and the Broad Institute for releasing the *C. cinerea* genomic data to the public. This work was supported by the Swiss National Science Foundation (SNSF Grant No. 3100A0-105790/1), which is gratefully acknowledged. The transcript profiling was funded by INRA and the Network of Excellence EVOLTREE. UK was funded by the Deutsche Bundesstiftung Umwelt (DBU).

## References

**Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997.** Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389–3402.

**Asante-Owusu RN, Banham AH, Böhnert HU, Mellor EJC, Casselton LA. 1996.** Heterodimerization between two classes of homeodomain proteins in the mushroom *Coprinus cinereus* brings together potential DNA-binding and activation domains. *Gene* **172**: 25–31.

**Awadalla P, Charlesworth D. 1999.** Recombination and selection at *Brassica* self-incompatibility loci. *Genetics* **152**: 413–425.

**Badrane H, May G. 1999.** The divergence-homogenization duality in the evolution of the *b1* mating type gene of *Coprinus cinereus*. *Molecular Biology and Evolution* **16**: 975–986.

**Banham AH, Asante-Owusu RN, Göttgens B, Thompson S, Kingsnorth CS, Mellor E, Casselton LA. 1995.** An *N*-terminal dimerization domain permits homeodomain proteins to choose compatible partners and initiate sexual development in the mushroom *Coprinus cinereus*. *Plant Cell* **7**:773–783.

**Banuett F. 2007.** History of the mating types in *Ustilago maydis*. In: Heitman J, Kronstand JW, Taylor JW, Casselton LA, eds. *Sex in fungi*. Washington DC, USA: ASM Press, 351–375

**Berbee ML, Taylor JW. 2001.** Fungal molecular evolution: gene trees and geologic time. In: McLaughlin DJ, McLaughlin EG, Lemke PA, eds. *The Mycota: a comprehensive treatise on fungi as experimental systems for basic and applied research. Systematics and Evolution, Part B. Volume VII.* Berlin Heidelberg, Germany: Springer-Verlag, 229–245.

**Beye M, Hunt GJ, Page RE, Fondrk MK, Grohmann L, Moritz RF. 1999.** Unusually high recombination rate detected in the sex locus region of the honey bee (*Apis mellifera*). *Genetics* **153**: 1701–1708.

**Bourne HR. 1997.** How receptors talk to heterotrimeric G proteins. *Current Opinion in Genetics and Development* **9**: 134–142.

**Bull JJ. 1983.** *Evolution of sex determining mechanisms*. Menlo Park, CA, USA: Benjamin-Cummings.

**Casselton LA, Kües U. 2007.** The origin of multiple mating types in the model mushrooms *Coprinopsis cinerea* and *Schizophyllum commune*. In: Heitman J, Kronstand JW, Taylor JW, Casselton LA, eds. *Sex in fungi*. Washington DC, USA: ASM Press, 283–300.

**Casselton LA, Olesnicky NS. 1998.** Molecular genetics of mating recognition in basidiomycete fungi. *Microbiology and Molecular Biology Reviews* **62**: 55–70.

**Charlesworth B. 1978.** Model for evolution of Y chromosomes and dosage compensation. *Proceedings of the National Academy of Sciences, USA* **75**: 5618–5622.

**Charlesworth B, Charlesworth D. 2000.** The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London, Series B* **355**: 1563–1572.

**Cromie GA, Hyppa RW, Cam HP, Farah JA, Grewal SI, Smith GR. 2007.** A discrete class of intergenic DNA dictates meiotic DNA break hotspots in fission yeast. *PLoS Genetics* **3**: e141.

**Di Battista C, Selosse MA, Bouchard D, Stenström E, Le Tacon F. 1996.** Variations in symbiotic efficiency, phenotypic characters and ploidy level among different isolates of the ectomycorrhizal basidiomycete *Laccaria bicolor* strain S 238. *Mycological Research* **100**: 1315–1324.

**Doudrick RL, Anderson NA. 1989.** Incompatibility factors and mating competence of two *Laccaria spp*. (Agaricales) associated with black spruce in northern Minnesota. *Phytopathology* **79**: 694–700.

**Doudrick RL, Raffle VL, Nelson CD, Furnier GR. 1995.** Genetic analysis of homokaryons from a basidiome of *Laccaria bicolor* using random amplified polymorphic DNA (RAPD) markers. *Mycological Research* **99**: 1361–1366.

**Duret L, Eyre-Walker A, Galtier N. 2006.** A new perspective on isochore evolution. *Gene* **385**: 71–74.

**Felsenstein J. 1985.** Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.

**Fraser JA, Diezmann S, Subaran RL, Allen A, Lengeler KB, Dietrich FS, Heitman J. 2004.** Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. *PLoS Biology* **2**: e384.

**Fraser JA, Heitman J. 2004.** Evolution of fungal sex chromosomes. *Molecular Microbiology* **51**: 299–306.

**Fraser JA, Hsueh YP, Findley KM, Heitman J. 2007.** Evolution of the mating type locus: the Basidiomycetes. In: Heitman J, Kronstand JW, Taylor JW, Casselton LA, eds. *Sex in fungi*. Washington DC, USA: ASM Press, 19–34.

**Fries N. 1983.** Spore germination, homing reaction and intersterility groups in *Laccaria laccata* (Agaricales). *Mycologia* **75**: 221–227.

**Fries N, Mueller GM. 1984.** Incompatibility systems, cultural features and species circumscriptions in the ectomycorrhizal genus *Laccaria* (Agaricales). *Mycologia* **76**: 633–642.

**Gay G, Normand L, Marmeisse R, Sotta B., Debaud JC. 1994.** Auxin overproducer mutants of *Hebeloma cylindrosporum* Romagnesi have increased mycorrhizal activity. *New Phytologist* **128**: 645–657.

**Hall TA. 1999.** BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95–98.

**Harley JL, Smith SE. 1983.** *Mycorrhizal symbiosis*. London: Academic Press.

**Henrion B, Le Tacon F, Martin F. 1992.** Rapid identification of genetic variation of ectomycorrhizal fungi by amplification of ribosomal RNA genes. *New Phytologist* **122**: 289–298.

**Hsueh YP, Idnurm A, Heitman J. 2006.** Recombination hotspots flank the *Cryptococcus* mating type locus: implications for the evolution of a fungal sex chromosome. *PLoS Genetics* **2**: e184.

**Hughes AL, Hughes MK, Watkins DI. 1993.** Contrasting roles of interallelic recombination at the *HLA-A* and *HLA-B* loci. *Genetics* **133**: 669–680.

**James TY, Kües U, Rehner SA, Vilgalys R. 2004.** Evolution of the gene encoding mitochondrial intermediate peptidase and its cosegregation with the *A* mating type locus of mushroom fungi. *Fungal Genetics and Biology* **41**: 381–390.

**James TY, Srivilai P, Kües U, Vilgalys R. 2006.** Evolution of the bipolar mating system of the mushroom *Coprinellus disseminatus* from its tetrapolar ancestors involves loss of mating-type-specific pheromone receptor function. *Genetics* **172**: 1877–1891.

**Käll L, Krogh A, Sonnhammer ELL. 2007.** Advantages of combined transmembrane topology and signal peptide prediction – the PHOBIUS web server *Nucleic Acids Research* **35**: W429–432.

**Kondrashov AS. 1984.** Deleterious mutations as an evolutionary factor. 1. The advantage of recombination. *Genetical Research* **44**: 199–217.

**Kropp BR, Fortin JA. 1988.** The incompatibility system and relative ectomycorrhizal performance of monokaryons and reconstituted dikaryons of *Laccaria bicolor*. *Canadian Journal of Botany* **68**: 289–294.

**Kües U, Asante-Owusu RN, Mutasa ES, Tymon AM, Pardo EH, O'Shea SF, Göttgens B, Casselton LA. 1994.** Two classes of homeodomain proteins specify the multiple *A* mating types of the mushroom *Coprinus cinereus*. *Plant Cell* **6**: 1467–1475.

**Kües U, James TY, Vilgalys R, Challen MP. 2001.** The chromosomal region containing *pab-1*, *mip*, and the *A* mating type locus of the secondarily homothallic homobasidiomycete *Coprinus bilanatus*. *Current Genetics* **39**: 16–24.

**Lupas A, van Dyke M, Stock J. 1991.** Predicting coiled coils from protein sequences. *Science* **252**: 1162–1164.

**Martin F, Aerts A, Ahrén D, Brun A, Danchin EGJ, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V *et al.* 2008.** Symbiosis insights from the genome of the mycorrhizal basidiomycete *Laccaria bicolor*. *Nature* **452**: 88–92.

**Martinez D, Larrondo LF, Putnam N, Gelpke MD, Huang K, Chapman J, Helfenbein KG, Ramaiya P, Detter JC, Larimer F *et al.* 2004.** Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nature Biotechnology* **22**: 695–700.

**May G, Matzke E. 1995.** Recombination and variation at the *A* mating-type of *Coprinus cinereus*. *Molecular Biology and Evolution* **12**: 794–802.

**May G, Shaw F, Badrane H, Vekemans X. 1999.** The signature of balancing selection: fungal mating compatibility gene evolution. *Proceedings of the National Academy of Sciences, USA* **96**: 9172–9177.

**Mermod N, O'Neill EA, Kelly TJ, Tjian R. 1989.** The proline-rich transcriptional activator of CTF/NF-I is distinct from the replication and DNA binding domain. *Cell* **58**: 741–753.

**Muraguchi H, Ito Y, Kamada T, Yanagi SO. 2003.** A linkage map of the basidiomycete *Coprinus cinereus* based on random amplified polymorphic DNAs and restriction fragment length polymorphisms. *Fungal Genetics and Biology* **40**: 93–102.

**Niculita-Hirzel H, Hirzel AH. 2008.** Visualizing the gene order conservation among genomes with ChromoMapper. *International Journal of Computational Intelligence in Bioinformatics and Systems Biology*, in press.

**O'Shea SF, Chaure PT, Halsall JH, Olesnicky NS, Leibbrandt A, Connerton IF, Casselton LA. 1998.** A large pheromone and receptor gene complex determines multiple *B* mating type specificities in *Coprinus cinereus*. *Genetics* **148**: 1081–1090.

**Pal C, Hurst LD. 2003.** Evidence for co-evolution of gene order and recombination rate. *Nature Genetics* **33**: 392–395.

**Poyatos JF, Hurst LD. 2007.** The determinants of gene order conservation in yeasts. *Genome Biology* **8**: R233.

**Raffle VL, Anderson NA, Furnier GR, Doudrick RL. 1995.** Variation in mating competence and random amplified polymorphic DNA in *Laccaria bicolor* (Agaricales) associated with three tree host species. *Canadian Journal of Botany* **73**: 884–890.

**Rambaut A, Drummond A. 2003.** Tracer MCMC trace analysis tool, version 1.4. http://tree.bio.ed.ac.ukdownload.html?name=tracer&version=v1.4&id=54&num=2

**Raper JR. 1966.** *Genetics of sexuality in higher fungi*. New York, NY, USA: Ronald Press.

**Raper JR, Flexer AS. 1971.** Mating systems and evolution of the Basidiomycetes. In: Petersen RH, ed. *Evolution in the higher Basidiomycetes*. Knoxville, TN, USA: University of Tennessee Press, 149–167.

**Rice WR. 1998.** Requisite mutational load, pathway epistasis and deterministic mutation accumulation in sexual versus asexual populations. *Genetica* **102/103**: 71–81.

**Rice WR. 2002.** Experimental tests of the adaptive significance of sexual recombination. *Nature Reviews Genetics* **3**: 241–251.

**Riquelme M, Challen MP, Casselton LA, Brown AJ. 2005.** The origin of multiple *B* mating specificities in *Coprinus cinereus*. *Genetics* **170**: 1105–1119.

**Ronquist F, Huelsenbeck JP. 2003.** MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

**Sen M, Marsh L. 1994.** Noncontiguous domains of the α-factor receptor of yeasts confer ligand specificity. *Journal of Biological Chemistry* **269**: 968–973.

**Spit A, Hyland RH, Mellor EJ, Casselton LA. 1998.** A role for heterodimerization in nuclear localization of a homeodomain protein. *Proceedings of the National Academy of Sciences, USA* **95**: 6228–6233.

**Stajich JE, Birren B, Burns C, Casselton LA, Dietrich F, Fargo DC, Gathman AC, James TY, Kamada T, Lilly WW *et al.* 2006.** Genomic analysis of *Coprinus cinereus. Proceedings of the International Symposium on Mushroom Science*. Akita, Japan: Akita Prefectural University, 59–74.

**Stamatakis A. 2006.** RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.

**Tusnády GE, Simon I. 2001.** The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**: 849–850.

**Tymon AM, Kües U, Richardson WV, Casselton LA. 1992.** A fungal mating type protein that regulates sexual and asexual development contains a POU-related domain. *EMBO Journal* **11**: 1805–1813.

**Vieira CP, Charlesworth D, Vieira J. 2003.** Evidence for rare recombination at the gametophytic self-incompatibility locus. *Heredity* **91**: 262–267.

**Whitehouse HLK. 1949.** Multiple allelomorph heterothallism in the fungi. *New Phytologist* **48**: 212–244.

**Yuhki N, O'Brien SJ. 1994.** Exchanges of short polymorphic DNA segments predating speciation in feline major histocompatibility complex class I genes *Journal of Molecular Evolution* **39**: 22–33.

**Zangenberg G, Huang MM, Arnheim N, Erlich H. 1995.** New HLA-DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm. *Nature Genetics* **10**: 407–414.

## Supplementary Material

The following supplementary material is available for this article online:

**Text S1** (Complements the Materials and Methods section): Study species; Complement to *MAT* loci annotation section; Accession numbers of the sequences used in the phylogenetic reconstruction; Construction and sequencing of the dikaryon S238N cDNA library.

**Fig. S1** Phylogeny of basidiomycete species compared with *Laccaria bicolor* in the present study.

**Fig. S2** Amino acid alignment of HD1 mating type proteins.

**Fig. S3** Amino acid alignment of HD2 mating type proteins.

**Fig. S4** Amino acid alignment of pheromone-like and mating type specific pheromone sequences encoding CaaX motifs at their *C*-terminus in *Laccaria bicolor* and *Coprinopsis cinerea*.

**Fig. S5** The intergene distances have different size between the genomic regions containing the *A* and *B* loci of *Laccaria bicolor*.

**Fig. S6** The G + C content variation in the *A* and *B* mating type loci of *Laccaria bicolor.*

**Fig. S7** Schematic comparison of the region surrounding the *B* locus from *Laccaria bicolor*, *Coprinopsis cinerea*, *Pleurotus djamor* and *Phanerochaete chrysosporium.*

**Fig. S8** Amino acid alignment of STE3-like pheromone receptor sequences.

**Table S1** Polymerase chain reaction (PCR) primers used to survey recombination events in $F_1$ progeny of S238N

**Table S2** Distribution of STE3-like pheromone receptor genes in *Laccaria bicolor* genome

**Table S3** Distribution of pheromone-like genes in *L. bicolor* genome

**Table S4** Changes in the expression of transcripts coding for predicted mating-type genes in free-living mycelium and Poplar and Douglas-fir ectomycorrhizal root compared to fruiting bodies

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1469-8137.2008.02525.x
(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the journal at *New Phytologist* Central Office.

## Article 5. A genetic linkage map for the ectomycorrhizal fungus *Laccaria bicolor* and its alignment to the whole-genome sequence assemblies

J. Labbé, X. Zhang, T. Yin, J. Schmutz, J. Grimwood, F. Martin, G. A. Tuskan and F. Le Tacon

(publié dans *New Phytologist*)

New Phytologist

Research

# A genetic linkage map for the ectomycorrhizal fungus *Laccaria bicolor* and its alignment to the whole-genome sequence assemblies

J. Labbé[1]*, X. Zhang[2,3]*, T. Yin[2,3]*, J. Schmutz[4], J. Grimwood[4], F. Martin[1], G. A. Tuskan[2,3] and F. Le Tacon[1]

[1]UMR 1136, INRA-Nancy Université, Interactions Arbres/Microorganismes, INRA-Nancy, 54280 Champenoux, France; [2]Environmental Sciences Division, Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, TN 37831-6422, USA; [3]Joint Genome Institute, 2500 Mitchell St, Walnut Creek, CA 94250, USA; [4]Stanford Human Genome Center, Department of Genetics, Stanford University School of Medicine, 975 California Avenue, Palo Alto, CA 94304, USA

Author for correspondence:
*J. Labbé*
*Tel: +33 3 83 39 40 80*
*Fax: +33 3 83 39 40 69*
*Email: labbe@nancy.inra.fr*

Received: *07 March 2008*
Accepted: *08 August 2008*

## Summary

• A genetic linkage map for the ectomycorrhizal basidiomycete *Laccaria bicolor* was constructed from 45 sib-homokaryotic haploid mycelial lines derived from the parental S238N strain progeny. For map construction, 294 simple sequence repeats (SSRs), single-nucleotide polymorphisms (SNPs), amplified fragment length polymorphisms (AFLPs) and random amplified polymorphic DNA (RAPD) markers were employed to identify and assay loci that segregated in backcross configuration.

• Using SNP, RAPD and SSR sequences, the *L. bicolor* whole-genome sequence (WGS) assemblies were aligned onto the linkage groups. A total of 37.36 Mbp of the assembled sequences was aligned to 13 linkage groups. Most mapped genetic markers used in alignment were colinear with the sequence assemblies, indicating that both the genetic map and sequence assemblies achieved high fidelity.

• The resulting matrix of recombination rates between all pairs of loci was used to construct an integrated linkage map using JoinMap. The final map consisted of 13 linkage groups spanning 812 centiMorgans (cM) at an average distance of 2.76 cM between markers (range 1.9–17 cM).

• The WGS and the present linkage map represent an initial step towards the identification and cloning of quantitative trait loci associated with development and functioning of the ectomycorrhizal symbiosis.

**Key words:** genetic map, gene positioning, *Laccaria bicolor*, sequence assemblies.

*New Phytologist* (2008) **doi**: 10.1111/j.1469-8137.2008.02614.x



## Introduction

Underlying a tree's ability to generate large amounts of biomass or store carbon is its interactions with soil microbes known as ectomycorrhizal fungi, a symbiotic organism that excels at procuring necessary, but scarce, nutrients such as phosphate and nitrogen. The fungus within the root is protected from competition with other soil microbes and gains preferential access to carbohydrates within the plant, while concurrently transferring most of its obtained nutrients to the tree, and thus a mutualistic relationship is established (Martin *et al.*, 2007). The ectomycorrhizal fungus *Laccaria bicolor* (Maire) P.D. Orton (Basidiomycota, Agaricales, Hydnangiaceae) (Matheny *et al.*, 2007) forms symbiotic associations with a wide variety of tree species in the northern hemisphere (Mueller, 1982, 1991). In Europe, *L. bicolor* is mainly associated with Pinaceae, but it is sometimes found under deciduous trees such as *Quercus* or *Fagus*. In North America, *L. bicolor* appears to be primarily associated with members of Pinaceae. *L. bicolor* has been a major experimental model for decades (Martin *et al.*, 2004), and has been used in large-scale mycorrhizal inoculation programs (Le Tacon *et al.*, 1992). The elucidation

*These authors contributed equally to this work as first authors.

New Phytologist

of its genome, which has been sequenced under the initiative of the US Department of Energy (Martin *et al.*, 2008), will be of interest to communities studying everything from physiology and ecology in forest ecosystems to fundamental questions in evolution and development of host and symbionts.

The whole-genome sequence (WGS) reads of *L. bicolor* S238N-H82 were initially assembled using the JAZZ assembler (J. Chapman *et al.*, unpublished). After excluding redundant and short segments of assembled sequences or scaffolds, there remained 65.4 Mb of sequence (665 scaffolds with 248 > 10 kb in length), of which 6.6 Mb (10.1%) existed as unassembled sequences or captured gaps (Martin *et al.*, 2008). After initial computational and manual annotations, the WGS data were re-assembled using the ARACHNE assembler (Batzoglou *et al.*, 2002). The ARACHNE algorithm was less sensitive to repeat regions, such as abundant transposable elements. The ARACHNE assembly generated larger scaffolds, so-called supercontigs, with fewer gaps (total supercontig size, 59.9 Mb with 0.6 Mb gaps; 82 supercontigs > 10 kb), suggesting inconsistencies between genome assemblies. As a result, it was difficult to assess the degree of mis-assembly, as no independent comparator, such as a genetic linkage map, was detailed enough to verify the released WGS assemblies. Most eukaryotic genome sequencing projects are often preceded by the construction of physical and genetic (meiotic) maps. For the *L. bicolor* genome project, there was no physical BAC map, and although a *L. bicolor* linkage map was constructed by Doudrick *et al.* (1995), the polymorphic marker sequences were not available, thus precluding genetic map-assembled sequence integration. Moreover, there were no reports on the chromosome number for *L. bicolor*; however, in *L. montana*, a close species, the haploid number of chromosomes appeared to be $n = 9$ (Mueller *et al.*, 1993).

In addition to WGS, another important component of ecological genomics and association genetics programs is the development of a linkage and quantitative trait loci (QTL) map. Linkage maps in ascomycetous fungi have been shown to be useful for identifying QTLs, candidate gene mapping, and comparative mapping between species (Hulbert *et al.*, 1988; Tzeng *et al.*, 1992; Debets *et al.*, 1993; Arnau *et al.*, 1994; Nitta *et al.*, 1997; Gale *et al.*, 2005). However, among basidiomycetes genetic linkage maps are scarce. Thus far, there are genetic maps for *Phanerochaete chrysosporium* (Raeder *et al.*, 1989), *Agaricus bisporus* (Callac *et al.*, 1997), *Coprinopsis cinerea* (Muraguchi *et al.*, 2003), *Cryptococcus neoformans* (Marra *et al.*, 2004) and *Pleurotus ostreatus* (Park *et al.*, 2006). One feature of basidiomycetes that facilitates genetic mapping is the presence of homokaryotic haploid meiotic spores produced in the fruiting body. Assay of mycelium issued from these spores directly reveals the products of meiosis (essentially, behaving like a cross to a homozygous testor strain), allowing an efficient mapping of genes.

In the present study, we generated a genetic linkage map for *L. bicolor* based on a mapping panel comprising 45 individual haploid siblings from a progeny of the parental S238N dikaryon strain. For map construction, 294 simple sequence repeats (SSRs), single-nucleotide polymorphisms (SNPs), amplified fragment length polymorphisms (AFLPs) and random amplified polymorphic DNA (RAPD) markers were employed to identify loci that segregated in backcross configuration. The JAZZ and ARACHNE WGS assemblies were anchored to this genetic linkage map using 49 sequence-tagged markers. Finally, key structural genes coding for mating-type loci, nuclear rDNA, laccase *lcc6* and hydrophobin *LbH2* were mapped on the linkage map. The release of the WGS and development of a linkage map for *L. bicolor* would enable us to map and clone QTLs associated with symbiosis development and functioning, and acquisition of scarce soil nutrients.

## Materials and Methods

### Fungal strain and culture conditions

Spores were sampled from caps of *Laccaria bicolor* (Maire) P.D. Orton fruiting bodies growing beneath *Pseudotsuga menziesii* (Mirb.) Franco seedlings inoculated with *L. bicolor* stain S238N in a glasshouse or in a nursery (Di Battista *et al.*, 1996) and germinated according to Fries (1983). Ninety-one individual homokaryotic (haploid) mycelia were used in this study, including the S238-H82 strain from which the genome was sequenced (Martin *et al.*, 2008). All homokaryotic mycelia were subcultured in Petri dishes containing a Pachlewski agar medium (Henrion *et al.*, 1992) and stored at 4°C with yearly subculturing. To provide material for DNA isolation, mycelium was grown in glass tubes containing 20 ml of Pachlewski liquid medium for 3 wk at 25°C. All strains are stored and subcultured at INRA-Nancy.

### DNA isolation

Homokaryotic mycelium was removed from the growth medium, rinsed in $H_2O$ and frozen in liquid nitrogen. Approximately 80 mg (FW) were used for DNA isolation using the DNeasy Plant Mini Kit (Qiagen, Courtaboeuf, France) according to the manufacturer's instructions. DNA was recovered in 50 µl of deionized water. Taxonomic identity of strains and quality of DNA were ascertained by PCR amplification and sequencing of the internal transcribed spacer (ITS) of the nuclear ribosomal DNA, using the following primers: ITS1 (5′-TCCTCCGCTTATTGATATGC) and ITS4 (5′-TCCGTAGGTGAACCTGCGG). The PCR was performed in a Perkin-Elmer Cetus thermocycler 9700 (Applied Biosystems, Foster City, CA, USA) according to Di Battista *et al.* (2002).

### Marker development

**Detection, scoring and sequencing of RAPD fragments** The sequence of 23 10-*mer* primers (Table 1) used by Doudrick

**Table 1** Sequences of the 23 random amplified polymorphic DNA (RAPD) primers used for the genetic linkage map construction

| Primer | Sequence | Annealing temperature (°C) |
|---|---|---|
| RAPD 157 | 5′-CGTGGGCAGG-3′ | 36 |
| RAPD 213 | 5′-CAGCGAACTA-3′ | 30 |
| RAPD 499 | 5′-GGCCGATGAT-3′ | 32 |
| RAPD 300 | 5′-GGCTAGGGCG-3′ | 36 |
| RAPD 159 | 5′-GAGCCCGTAG-3′ | 34 |
| RAPD 271 | 5′-GCCATCAAGA-3′ | 30 |
| RAPD 551 | 5′-GGAAGTCCAC-3′ | 32 |
| RAPD 222 | 5′-AAGCCTCCCC-3′ | 34 |
| RAPD 308 | 5′-AGCGGCTAGG-3′ | 34 |
| RAPD 429 | 5′-AAACCTGGAC-3′ | 30 |
| RAPD 190 | 5′-AGAATCCGCC-3′ | 32 |
| RAPD 258 | 5′-CAGGATACCA-3′ | 30 |
| RAPD 156 | 5′-GCCTGGTTGC-3′ | 34 |
| RAPD 288 | 5′-CCTCCTTGAC-3′ | 32 |
| RAPD 556 | 5′-ATGGATGACG-3′ | 30 |
| RAPD 375 | 5′-CCGGACACGA-3′ | 34 |
| RAPD 532 | 5′-TTGAGACAGG-3′ | 30 |
| RAPD 608 | 5′-GAGCCCCAAA-3′ | 32 |
| RAPD 567 | 5′-AGACACCTGA-3′ | 30 |
| RAPD 184 | 5′-CAAACGGCAC-3′ | 32 |
| RAPD 503 | 5′-ATCGTCCAAC-3′ | 30 |
| RAPD 248 | 5′-GAGTAAGCGG-3′ | 32 |
| RAPD 256 | 5′-TGCAGTCGAA-3′ | 30 |

*et al.* (1995) was retrieved from the UBC primer set (http://www.michaelsmith.ubc.ca/services/NAPS/Primer_Sets/) available at the Nucleic Acid Protein Service Unit (University of British Columbia, Canada). RAPD analysis was carried out on the 91 sibling homokaryons and the parental S238N dikaryon. PCR amplification of the DNA template was as per the protocol of Doudrick *et al.* (1995), except that PCR was performed using a Perkin-Elmer Cetus thermocycler 9700 using 0.5 U of Taq DNA polymerase (Qbiogene, Strasbourg, France) and the Qbiogene 10× buffer containing 25 mM $MgCl_2$. After amplification, RAPD fragments were resolved by electrophoresis (5 V $cm^{-1}$) for 2 h in 2% agarose gels (one-third agarose (Qbiogene), two-thirds wide-range agarose (Sigma-Aldrich, Saint-Quentin Fallavier, France)) in TBE buffer (45 mM Tris base, 45 mM boric acid, and 1 mM EDTA, pH 8). RAPD fragments were detected by staining with ethidium bromide (0.5 mg $ml^{-1}$) for 20 min. RAPD fragments were scored as dominant segregating markers (presence/absence) when amplicons of identical size were detected in three replicates for each polymorphic fragment. RAPD markers were identified by the letter R followed by a number (i.e. R1). The markers were imported in JoinMap v 3.0 (Van Ooijen and Voorrips, 2001) and checked for deviation from 1 : 1 ratios. Polymorphic RAPD fragments were excised from agarose gel using QIAquick Gel extraction kit (Qiagen), cloned using TOPO TA cloning Kit pCR 2.1TOPO vector (Invitrogen, Cergy-Pontoise, France) and sequenced on both strands using M13 forward and M13 reverse primers, the CEQ Dye-labeled Dideoxy-Terminator Cycle Sequencing kit (Beckman Coulter, Fullerton, CA, USA) and the automated CEQ 8000 XL sequencer (Beckman Coulter). Sequences of these RAPD fragments are available at http://mycor.nancy.inra.fr/IMGC/LaccariaGenome/GeneticMap/RAPDMarkerSequences and were located by similarity search on the JAZZ and ARACHNE genome assemblies using the BLAST algorithm at the JGI *L. bicolor* genome website (www.jgi.doe.gov/laccaria) or INRA LaccariaDB website (http://mycor.nancy.inra.fr/IMGC/LaccariaGenome/).

**SSR analysis** The SSR markers were detected on the *L. bicolor* ARACHNE genome assembly (v. 1.0) (http://mycor.nancy.inra.fr/IMGC/LaccariaGenome/Annotation/index.php?select=fast) using MAGELLAN 1.1 software (Lim *et al.*, 2005). Details on the SSR distribution analysis will be presented elsewhere (J. Labbé *et al.*, unpublished). We selected all SSR types except for the compound motifs in an effort to sample the 20 larger ARACHNE supercontigs. SSR primers were designed using the online Primer3 tool (Koressaar & Remm, 2007). The SSR analyses were conducted at Oak Ridge National Laboratory (ORNL) and INRA-Nancy using two different protocols.

At ORNL, SSRs were analyzed with Fluorescein-12-dUTP (Enzo Roche, Penzberg, Germany). PCR reactions were carried out in a total volume of 15 µl containing 25 ng of template DNA, 7.5 ng forward and reverse oligonucleotide primers (Operon Technologies, Alameda, CA, USA), 20 µM of each dNTP, 0.5 U Taq DNA polymerase (New England Biolabs, Beverly, MA, USA), 1.5 µl 10× buffer (containing 100 µM Tris-HCl, pH 8.3, 500 mM KCl, 20 mM $MgCl_2$ and 10.0 g $l^{-1}$ bovine serum albumin). PCR was conducted in a Perkin-Elmer Cetus thermocycler 9700, using 10 touchdown cycles performed with annealing temperature starting at 60°C and ending at 50°C with a 1°C decrease each cycle. After the touchdown cycles, there followed 20 cycles of 94°C for 30 s, 50°C for 30 s, and 72°C for 1 min combined with a final extension at 72°C for 7 min. PCR products were detected on an ABI 3730XL sequencer using the standard microsatellite genotyping module. The amplification products were mixed in appropriate ratios and diluted 1 : 10 with loading buffer (91% deionized formamide, 9% internal standard GeneScan 450ROX (Applied Biosystems)), then denaturated at 95°C for 5 min followed by rapid cooling on ice.

At INRA-Nancy, SSRs were analyzed with one fluorescent dye-labeled primer (D2-PA, D3-PA, D4-PA, WellRED dyes; Proligo, Paris, France). PCR reactions were carried out in a total volume of 10 µl containing 10 ng of template DNA, 10 ng forward and reverse oligonucleotide primers (Invitrogen), 200 µM of each dNTP, 0.5 U Taq DNA polymerase (Qbiogene), 1 µl Qbiogene buffer 10× containing 25 mM $MgCl_2$. PCR was conducted in a Perkin-Elmer Cetus thermocycler 9700 (Applied Biosystems) with 30 cycles of 94°C for 30 s, 60°C for 1 min, and 72°C for 2 min combined with a final

extension at 72°C for 10 min. PCR products were detected on a CEQ 8000 XL sequencer using the CEQ 8000 V9.0 genotyping module (Supporting information, Fig. S3). Before analysis, each amplification product was diluted 1 : 10 with deionized water. One microliter of this dilution was mixed with 0.5 μl DNA size standard (DNA size standard kit 600 bp, Beckman Coulter) and with 30 μl SLS buffer (Beckman Coulter).

**AFLP analysis** The AFLP analyses, carried out at the ORNL, have been applied in map construction for many different pedigrees (Yin *et al.*, 2002, 2004). In this study, we used three replicates of the DNA template extracted from the dikaryon S238N and H82 (the sequenced homokaryon). Two initial screening steps were performed before genotyping the entire progeny set. We initially used S238N and H82 to screen primer combinations to identify discernible amplification profiles. In a second step, we used S238N, H82 and six additional homokaryons to screen primer combinations for reproducible amplification profiles. These eight individuals were then included in the progeny genotyping step. Thus, the reproducibility of the AFLP amplification profiles was confirmed during each of the three steps listed above.

The AFLP procedure was performed as described by Vos *et al.* (1995) with the following modifications. Pre-amplification reactions (15 μl) were performed for 3 μl of the diluted DNA template using 20 pmol each of a pair of AFLP primers (Operon Technologies) with no selective 3′ nucleotides on the 'E' primer and 1 'C'-selective 3′ nucleotide on the 'M' primer. Reaction conditions were identical to those described for SSRs. The pre-amplified products were diluted 1 : 30 as DNA template for selective amplification. Selective amplification was carried out in a volume of 15 μl reaction mixture, containing 3 μl diluted pre-amplification product, 0.5 pmol 'E' primer with two selective nucleotides (Hex-labeled) and 5 pmol 'M' primer with three selective nucleotides (Operon Technologies), 200 μM each dNTP, 0.5 U Taq DNA polymerase (Promega, Madison, WI, USA), 1.5 μl 10× buffer (100 μM Tris-HCl, pH 8.3, 500 mM KCl, 20 mM $MgCl_2$), 10.0 g $l^{-1}$ BSA and 1% (v/v) deionized formamide. Thermocycling conditions for selective amplification were 12 cycles of 94°C for 30 s, 65°C for 30 s decreasing by 0.7°C per cycle, and 72°C for 60 s, followed by 23 cycles of 94°C for 30 s, 56°C for 30 s and 72°C for 60 s. PCR products were detected use the ORNL SSR genotyping protocol discussed earlier on the ABI 3730XL sequencer. The restriction site sequences of AFLP fragments were too short for aligning the AFLP markers on the genome sequence assemblies.

## Mapping of selected nuclear genes, SNPs and mating-type genes

Five structural loci were mapped: hydrophobin (*LbH2*), laccase (*lcc6*), mating-type A and B (*MATa* and *MATb*), and nuclear ribosomal DNA intergenic spacer 1 (*IGS1*). PCR amplification of the genes coding for *LbH2* (JGI ID 399267) and *lcc6* (JGI ID 399748) on 91 monokaryotic genomic DNA was conducted in a Perkin-Elmer Cetus thermocycler 9700 with 30 cycles of 94°C for 30 s, 53°C for 30 s, and 72°C for 2 min combined with a final extension at 72°C for 10 min. Primers (Table 2) were designed using the online Primer3 tool (Koressaar & Remm, 2007). PCR reactions were carried out in a total volume of 25 μl containing 25 ng of template DNA, 20 ng forward and reverse oligonucleotide primers (Invitrogen), 250 μM of each dNTP, 0.5 U Taq DNA polymerase (Qbiogene), and 2.5 μl Qbiogene buffer 10× containing 25 mM $MgCl_2$. PCR products were sequenced on both strands using the respective forward and reverse PCR primers, the CEQ Dye-labeled Dideoxy-Terminator Cycle Sequencing kit and the automated CEQ 8000 XL sequencer according to the manufacturer's instructions. Sequences were then assembled using CLUSTALW (http://bioinfo.hku.hk/services/analyseq/cgi-bin/clustalw_in.pl) and Sequencher 4.2 (Gene Code Corporation, USA) software. SNPs were detected for *lcc6* at positions 1758, 1869, and 1940, and for *LbH2* at position 478.

Analysis of meiotic segregation of the intergenic spacer 1 (*IGS1*) of the nuclear rDNA was carried out as per Selosse *et al.* (1996). The *IGS1* haplotypes, α or β, were amplified from 91 monokaryotic genomic DNA using the primer 5SA (5′-CAGAGTCCTATGGCCGTGGAT) and the fluorescent dye-labeled primer INRAjLIGS1 (5′-CAGTGGAGTAAGT-CAG-D4-PA) (D2-PA, D3-PA, D4-PA, WellRED dyes, Proligo). PCR amplification was conducted in a Perkin–Elmer Cetus thermocycler 9700 with 35 cycles of 94°C for 30 s, 50°C for 30 s, and 72°C for 5 min combined with a final extension at 72°C for 10 min. PCR reactions were carried out in a total volume of 10 μl containing 10 ng of template DNA, 10 ng forward and reverse oligonucleotide primers (Invitrogen), 200 μM of each dNTP, 0.5 U Taq DNA polymerase (Qbiogene), and 1 μl Qbiogene buffer 10× containing 25 mM $MgCl_2$. Amplified *IGS1* homo- and heteroduplexes (Selosse *et al.*, 1996) were detected on the CEQ 8000 XL sequencer using the fragment detection module. Before analysis, each amplification product was diluted in 1 : 10 with deionized water. One microlitre of this dilution was mixed with 0.5 μl DNA size standard (DNA size standard kit 600 bp, Beckman Coulter) and with 30 μl SLS Beckman buffer. The *IGS1* sequence, being part of the rDNA tandem repeat (18S-ITS1-5.8S-ITS2-28S-IGS1-5S-IGS2) (Martin *et al.*, 1999), was assumed to co-map with the whole rDNA repeat region. The mating-type genes *STE3, HD1* and *HD2* were mapped using the segregation analysis amplification products carried out in Niculita-Hirzel *et al.* (2008).

## Linkage analysis and map construction

Owing to the numbers of used markers (i.e. 301), the pre-mapping diagnostics showed that these markers can be resolved

**Table 2** Sequenced-tagged markers used for anchoring whole-genome sequence (WGS) assemblies on the genetic linkage map

| Marker ID | Primer sequence | Positions on ARACHNE supercontigs | Positions on JAZZ scaffolds | Linkage groups (genetic distance, cM) |
|---|---|---|---|---|
| *Shyd2.1* | (F)GTGCTGCTCCGAGCCTTACC | S 1211 | Sc 7 | LG 1 |
| | (R)GGAGAGCATCCGACAACGACC | 2191423 | 1206777 | (45.1) |
| *μ 03.4* | (F)CTGAGCTAGAGCAGCTGGTA | S 1211 | Sc 34 | LG 1 |
| | (R)GACAATCGCACATTGGATAG | 4254317–4254523 | 334200–334406 | (59.3) |
| *1184M15* | (F)GTTTGACTTTCCAGCGCTTC | S 1184 | Sc 51 | LG 1 |
| | (R)CGCCTCTGATGAGGATGAGT | 1503565–1503781 | 222848–223068 | (117.6) |
| *1184M16* | (F)TATGCCGTTGCGTTTTCATA | S 1184 | Sc 16 | LG 1 |
| | (R)CCCTTTGCAAGTTCTCTTCG | 328207–526772 | 384272–611895 | (118.2) |
| *1184M4* | (F)GCTTCGCAAACAAAGAAAGC | S 1184 | Sc 16 | LG 1 |
| | (R)AACAGGGGCATTGCTCTACA | 328518–328784 | 611337–611584 | (120.5) |
| *1184M11* | (F)GGAAATAGAATCAAGTCTTGTTGTTG | S 1184 | Sc 85 | LG 1 |
| | (R)GGGTTTTCTGACTTCCTCTGC | 816179–819095 | 1410–1728 | (124.7) |
| *1205M21* | (F)CGTCGTTGAGTGAAGTGTGG | S 1205 | Sc 18 | LG 2 |
| | (R)TGTGTTGGGAGAGTCCAGTG | 218443–218640 | 217884–218081 | (18.4) |
| *1205M17* | (F)TGGACTCCGACTGGACTCAT | S 1205 | Sc 18 | LG 2 |
| | (R)CATTCCACCACTGTTCATGC | 220003–220062 | 219444–219463 | (19.9) |
| *1205M14* | (F)TTGATGCTCTCCGTCCTCTT | S 1205 | Sc 18 | LG 2 |
| | (R)GCCTGGTTTGCTTCCTGTAG | 294315–294150 | 322388–322591 | (20.8) |
| *1205M16* | (F)ACCTGGGGGAAGTATGGATT | S 1205 | Sc 18 | LG 2 |
| | (R)TGTGTTGGGAGAGTCCAGTG | 792190–792308 | 829525–829643 | (31.9) |
| *μ 107.3* | (F)TCCAACAATTCCAGCTCCTC | S 1198 | Sc 10 | LG 2 |
| | (R)GATTCATCTTGGGGACCAAC | 1911207–1911395 | 679442–679630 | (43) |
| *μ 112.1* | (F)CCCCTTGCTATGGGATGTG | S 1205 | Sc 17 | LG 2 |
| | (R)AAATTTGAGGGCCTTGCAG | 4065293–4065481 | 313582–313770 | (0) |
| *μ 112.2* | (F)CCCCTTGCTATGGGATGTG | S 1205 | Sc 17 | LG 2 |
| | (R)AAATTTGAGGGCCTTGCAG | 4065293–4065481 | 313582–313770 | (15.9) |
| *1205M12* | (F)CTGCCTGAGGTGCCTTTTT | S 1205 | Sc 17 | LG 2 |
| | (R)GGGTCTGCAATTTGGAACAG | 4065217–4065428 | 313506–313726 | (20.8) |
| *1205M18* | (F)CTGGGACTGGACTCAGGAAA | S 1205 | Sc 17 | LG 2 |
| | (R)AGCAGGAGGTGGTGGTAGTG | 4031269–4031475 | 283452–283658 | (25.7) |
| *1199M3* | (F)ATACCAAAGGCCCTCGAACT | S 1199 | Sc 15 | LG 3 |
| | (R)CAGTATCTGCAACCCAAGCA | 168676–168873 | 944503–944700 | (7.7) |
| *1199M9* | (F)AGAACTTCGCGAGTTTTGGA | S 1199 | Sc 15 | LG 3 |
| | (R)GGGCGTCTTGTTATTCGGTA | 2626779–2626965 | 481624–481488 | (11.5) |
| *μ 03.3* | (F)CTGAGCTAGAGCAGCTGGTA | S 1199 | Sc 15 | LG 3 |
| | (R)GACAATCGCACATTGGATAG | 2849726–2850115 | 252454–252843 | (0) |
| *1199M6_1* | (F)AGGCTGAAATGCAAGCCAT | S 1199 | Sc 15 | LG 3 |
| | (R)CATTGAAGGGATCGGAGAAA | 2858807–2859030 | 243706–243928 | (11.4) |
| *1199M11* | (F)GGATCCCGTTTCTTCTTCCT | S 1199 | Sc 11 | LG 3 |
| | (R)GATGAAGGGCATCCTGACAC | 580742–580967 | 870603–870828 | (0) |
| *1209M8* | (F)CTTTGAGATACGGCGGACTT | S 1209 | Sc 8 | LG 3 |
| | (R)ATTACGGGGTGTCATCATCC | 880080–880642 | 630699–630878 | (5) |
| *1209M4* | (F)TGTGTTGGGAGAGTCCAGTG | S 1209 | Sc 64 | LG 3 |
| | (R)TATCCCTGAGGGCCAATACA | 1736375–1736618 | 66123–66364 | (30.1) |
| *1209M3* | (F)AATCCCATGTTGCATTCTCA | S 1209 | Sc 84 | LG 3 |
| | (R)TTTTCTGACTTCCTCTGCTTGA | 1513518–1513863 | 1413–1662 | (54.1) |
| *1213M6* | (F)GCCAAATTACCTGAGCAAGG | S 1213 | Sc 37 | LG 4 |
| | (R)CGTCAACGTATTTGGGGACT | 280255–280475 | 469159–468977 | (0) |
| *1206M3* | (F)GAATCCCATCCCCTCAAGTT | S 1206 | Sc 75 | LG 4 |
| | (R)CATCTCGGGCTAGGATTGAA | 1610490–1610667 | 165228–16605 | (26.8) |
| *1206M5* | (F)TGAGGAGGGTGGAGTACCAG | S 1206 | Sc 19 | LG 4 |
| | (R)ATGACCGACATTTTCGAGGT | 789453–789711 | 120592–120850 | (66.8) |
| *1206M2* | (F)ATACCAAAGGCCCTCGAACT | S 1206 | Sc 19 | LG 4 |
| | (R)CATAGGTGCAGGTCAGAGCA | 803810–804059 | 136336–136595 | (66.8) |
| *1206M1* | (F)CCCTTGAACTTGTGCACTGA | S 1206 | Sc 19 | LG 4 |
| | (R)AACAAATCCTGGAGGGCATT | 822987–822846 | 155532–155673 | (66.8) |
| *1207M10* | (F)CGCGTCACGTCAACACAC | S 1207 | Sc 47 | LG 5 |
| | (R)TTGACATGTCGCCGTCGT | 325759–3257846 | 231939–232190 | (0) |

**Table 2** *continued*

| Marker ID | Primer sequence | Positions on ARACHNE supercontigs | Positions on JAZZ scaffolds | Linkage groups (genetic distance, cM) |
|---|---|---|---|---|
| *μ 07* | (F)CTGTTCTCGCGAAGAAGGTC | S 1207 | Sc 47 | LG 5 |
| | (R)GGGAAATGTGATGAACCGGC | 4006848–4007283 | 330033–330468 | (15) |
| *1207M4* | (F)AGGCTGAGAAAGTGGGAACA | S 1207 | Sc 23 | LG 5 |
| | (R)CCCTCAAACTTGTGCACTGA | 3323790–3323971 | 434731–434912 | (30.6) |
| *1214M3* | (F)ATACCAAAGGCCCTCGAACT | S 1214 | Sc 4 | LG 6 |
| | (R)CTCTACCCGTACCCCATGAA | 3096597–3096792 | 849407–849626 | (26.8) |
| *1214M6* | (F)TCTTTCACGATGACCCAGTG | S 1214 | Sc 12 | LG 6 |
| | (R)GCTTTGTTTGGCATCCATTC | 1298208–1298446 | 955770–956008 | (64.4) |
| *μ 06* | (F)CGAGTATCCAGTATCTCGCTTTCCTAC | S 1168 | Sc 9 | LG 8 |
| | (R)AGCATCGTGAGATGAGCAGCTGCTT | 1297215–1297672 | 611086–611543 | (8.3) |
| *1168M2* | (F)TTCTCGACATGGTTGGGATT | S 1168 | Sc 9 | LG 8 |
| | (R)GGTGCACAAGAGGAAGAAGG | 1223201–1223427 | 531727–531953 | (14.4) |
| *1168M4* | (F)CGAGCTTTGGGAGAAGTTTG | S 1168 | Sc 9 | LG 8 |
| | (R)AGGATTTATGCCACCACAGC | 211567–211735 | 479751–479919 | (19.3) |
| *IGS1.1* | (F)CAGAGTCCTATGGCCGTGGAT | S 1216 | Sc 1067 | LG 7 |
| *IGS1.2* | (R)CAGTGGAGTAAGTCAG-D4 | 785280–785572 | 160-452 | (0) |
| *1195M6* | (F)GACAACGTCAAGGCCAAGAT | S 1195 | Sc 10 | LG 9 |
| | (R)TTGTAGTCGGAGAGGGTTCG | 1233145–1233327 | 924213–924395 | (4.9) |
| *SLaC14.1* | (F)CCCGGGTTTCGCTGCAAACTCTCC | S 1215 | Sc14 | Pair 2 |
| *SLaC14.2* | | | | (0) |
| *SLaC14.3* | (R)AATTATTCCACGCAGCAGCG | 3379947–3382098 | 429733–431884 | (2.1) unlinked |
| *STE3.1* | (R)TATTCCACGCAGCAGCGCGATC | S 1168 | Sc 56 | Pair_4 |
| | (F)CCGGAAAAGGAAGAAGGGACAATGA | 638167–638191 | 285965–385989 | (0) |
| *STE3.2* | (R)CTTTTCGCGGTGTAGAGTGGTTTGG | S 1168 | Sc 56 | Pair_4 |
| | (F)CTTCACTCAGGCCTGCTTCCACTTC | 635410–639634 | 284522–288746 | (17) |
| *STE3.3* | (R)TTTCTAGTCCCGCTTGCGCTCTTTG | S 1168 | Sc 56 | Pair_4 |
| | (F)TGTAACAAAGAGCGCAAGCGGGACT | 630872–630901 | 293255–293284 | |
| *1192M5* | (R)AGTCAGGACAACCGTGAAGG | S 1192 | Sc 6 | Unlinked |
| | (F)ACCTTTTTGTTGCTCGTTCG | 921896–922079 | 694909–695111 | |
| *HD1* | (F)GCTGTGGTGGAAGTGGTTTACGTTGTG | S 1195 | Sc 1 | LG 1 |
| | (R)ATGTCCGGAGTCTGGTTTGGGTGTATG | 3564435–3566855 | 1169650–1172070 | |
| *HD2* | (F)AGATTATCGGCCATTGGGAAGAGGTTG | S 1195 | Sc 1 | LG 1 |
| | (R)GAGGGAAGGCGGCTGAAGTTTGTTATC | 3566677–3568724 | 1169802–1167807 | |

(F) and (R) indicate the forward and reverse primers used to amplify the 49 sequence-tagged markers, respectively. CM values indicate the genetic distance on linkage groups in centiMorgan from the proximal end.

with confidence until a minimum of mapping population size of 39 (JoinMap $X^2$ test, Van Ooijen & Voorrips, 2001). The mapping population consisted of a *L. bicolor* S238N progeny of 45 haploid homokaryons. The haploid nature of the members of this population allowed the application of a backcross model for handling data. Linkage analysis between markers, estimation of recombination frequencies and determination of the linear order of loci, were performed using JoinMap V3.0 software (Van Ooijen and Voorrips, 2001). Recombination rates were converted to genetic distances in centiMorgans (cM) using the Kosambi's mapping function (Kosambi, 1944). This mapping function assumes crossover interference met in multipoint analysis, meaning that the presence of one crossover reduces the probability of another in the area. Thus, when multi-markers were analyzed, the crossover disturbing coefficient $C$ was introduced into the mapping function to adjust the genetic distance on linkage groups. Thus, for marker A-B-C:

$$C = R_r/R_e = R_r/(r_1 r_2 n)$$

where $R_r$ is the observed recombination rate, $R_e$ is the expected recombination rate, and $r_1$ and $r_2$ are the independent recombination rates between marker A-B and marker B-C. The standard deviation of $C$ is:

$$\mathrm{Sc} = \pm\sqrt{(C/n)\cdot[(1 - Cr_1 - Cr_2 - Cr_1r_2 + 2C^2r_1r_2)/(r_1r_2)]}\ .$$

$C$ is a value between 0 and 1. If $C = 1$, there is not a crossover disturbance; if $C = 0$, there is complete crossover disturbance. Then, $C$ is integrated into calculation of the Kosambi's mapping function (Kosambi, 1944). Thus, the genetic

distances between markers were adjusted by this parameter to a multipoint analysis, which is consequently different from a single two-point analysis.

Referring to the genome assembly, we adapted the grouping criteria, reaching final minimal LOD scores of 3.0 for small linkage groups (LGs) and 5.0 for large LGs, and a maximum recombination fraction of 0.4. In some cases, because of the lack of adequate recombination and LOD information, complete maps for some groups could not be created. In these cases, markers within these groups were split based on linkage relationships; two or more maps were generated and linkage between smaller groups and larger groups assessed. When smaller groups linked within larger groups, the smaller groups were removed. A third round of marker addition was performed, but without reordering first- and second-generation maps. LGs with ≥ four loci were retained as major LGs to represent the *L. bicolor* genome. We calculated total map distance that covered all LGs and average distance between markers as total map distance divided by the number of mapped markers. Several markers were linked as pairs or triplets, but not integrated in the map (e.g. linkage pair 4). The integration of the AFLP markers into our preliminary linkage map (Martin *et al.*, 2008), constructed using only RAPD and SSR markers, led to the exclusion of several markers. The positions of these excluded sequence-tagged markers were determined by alignment on the ARACHNE assembly.

### Alignment of JAZZ and ARACHNE assemblies

Masked sequences for whole-genome JAZZ and ARACHNE assemblies (version 1.0) were downloaded from the JGI (www.jgi.doe.gov/laccaria) and the INRA LaccariaDB (http://mycor.nancy.inra.fr/IMGC/LaccariaGenome/Annotation/), respectively. The JAZZ assembly scaffolds were aligned on the ARACHNE assembly supercontigs by using the BLASTN algorithm (99% identity, *E* value 0.0) at the INRA LaccariaDB BLAST server, and the GEnome PAir – Rapid Dotter Gepard (http://mips.gsf.de/services/analysis/gepard) (Krumsiek *et al.*, 2007). The sequenced genetic map markers were located on this assembly alignment with BLASTN, and the genetic map and the genome assemblies aligned manually. Average recombination rates were obtained by dividing the total linkage distance (cM) by the total physical length (Mb) for each LG (pseudochromosome). These estimates were not adjusted for gaps between LGs or supercontigs or for differences in marker density.

## Results

Of the 144 SSR primers screened using the parental S238N (N+N) dikaryon and a pilot set of seven randomly selected haploid homokaryons, 63 yielded PCR amplicons. Among these 63 SSRs, we retained 38 polymorphic SSRs (13 codominant and 25 dominant). For map construction, we selected 45 homokaryons yielding one of the two alternative alleles inherited from the dikaryon, by using the 13 codominant SSRs. Homokaryons displaying none of these two alternative alleles were exluded based on assuming a segregation distortion (possibility of aneuploidy) (Table 3, Fig. S2). Ultimately, 37 SSRs were placed on the genetic map, with one SSR remaining unlinked. The primers and the sequences of these 37 SSRs were then used to anchor the SSR markers on the JAZZ/ARACHNE genome assemblies (see later). In the AFLP analysis, from the 99 AFLP primer combinations, 46 primer pairs revealed 254 polymorphisms (average of five loci per primer combination). These AFLP markers were used to improve map coverage in regions of the genome that lacked mapped SSRs or sequence-tagged RAPD markers. The mapped marker list is presented in Table 2. In the RAPD analysis, 23 RAPD primers, chosen from Doudrick *et al.* (1995) for their profile clarity, generated 39 segregating RAPD markers. Only eight were finally kept for mapping calculation because of ambiguous segregation in the remaining 31 markers.

**Table 3** The codominant microsatellite markers (simple sequence repeats, SSRs) segregating in the *Laccaria bicolor* mapping pedigree identified during SSR primer screening

| SSR name | SSR type |
|---|---|
| 1199M11 | (TCC)8 |
| 1168M2 | (CCT)9 |
| 1168M4 | (GAG)10 |
| 1195M6 | (GTG)9 |
| 1205M12 | (CAC)12 |
| 1205M14 | (TCC)11 |
| 1192M5 | (GGT)10 |
| 1199M9 | (AGG)9 |
| 1206M3 | (AG)11 |
| 1206M5 | (GGA)9 |
| 1209M8 | (TGG)8 |
| 1213M6 | (GT)17 |
| 1184M15 | (GTCCTT)5 |

Among the initial set of 91 homokaryons, 45 were selected by using 13 codominant SSRs. Primer sequences are given in Supporting Information, Table S1.

### Genetic map construction

Thus, a total of 301 markers that segregated in a 1 : 1 ratio in a *L. bicolor* progeny of 45 selected haploid homokaryons were used to calculated average pairwise LOD and recombination data. The initial calculation with JoinMap was able to group 294 markers (four SNPs, eight RAPDs, 37 SSRs, 243 AFLPs, two mating-type genes) at LOD thresholds of three (small groups) and five (large groups) with a maximum recombination fraction of 0.4. These LOD thresholds were chosen as they produced the minimum number of LGs, while maintaining the integrity of the map. Adding AFLP markers significantly

increased the map size from 400 to 812 cM. The final map contained 287 markers positioned on 13 LGs, four marker pairs and one marker triplet (Fig. S1). LGs contained four to 66 markers and their sizes ranged from 10 to 124.7 cM, covering a total genetic length of 812 cM, with an average distance of 2.76 cM between the adjacent markers (range 1.9–17 cM). Thus, the 287 markers in the current map may be assumed to provide a reasonably comprehensive coverage of the *L. bicolor* genome. Owing to unknown chromosome number, the strict correspondence between LG ID and *L. bicolor* chromosomes remained unresolved.

### Integration of genome assemblies and genetic map

To anchor the JAZZ and ARACHNE assembled sequences to the *L. bicolor* genetic map, we aligned the sequences of 49 mapped SSR, RAPD and SNP markers onto the assembled genomic sequences using BLASTN. The number of markers located on the assemblies ranged from one (LG 9) to nine (LG 1) per linkage group. No sequenced markers were available to align assembled sequences to LG 10, 11 and 12. Table 2 summarizes the total number of JAZZ scaffolds and ARACHNE supercontigs anchored on the genetic map.

Schematics of the assembled sequences-genetic map integration are shown in Fig. 1. For example, based on sequence alignments of nine mapped markers, ARACHNE supercontigs 1211 and 1184, and large regions of supercontigs 1199 and 1195 were anchored to LG 1 (124 cM, 6.38 Mb) (Fig. 1). Eight scaffolds (e.g. 83, 8, 71, 41 and 7) from the JAZZ assembly were similarly integrated into LG 1. Based on the alignment of nine markers, LG 2 (97 cM, 4.68 Mb), comprising two subgroups, was anchored to ARACHNE supercontig 1205, partly to ARACHNE supercontig 1198 and to several scaffolds (18, 35, 2, 9, 44, 10 and 17) from the JAZZ assembly. In total, *c.* 38 Mb of the ARACHNE assembled sequences were integrated into pseudochromosome units along 10 LGs. The majority of the mapped markers used in alignment were colinear with the sequence assembly (Table 2, Fig. 1). The 10 LGs aligned with 16 ARACHNE supercontigs accounted for 63% of this genome sequence assembly. Other supercontigs were too short to be genetically oriented. Several unoriented JAZZ scaffolds (e.g. 41, 71, 8 and 83 on LG 1) were placed on pseudochromosomes, but their orientation was determined using BLASTN against the ARACHNE supercontig.

The ARACHNE assembly displayed some discrepancies that can be corrected by the present genetic map. For instance, it is clear that regions of ARACHNE supercontigs 1195 and 1199 anchored on LG 1 (Fig. 1) should be removed from the rest of their assembled sequences that were anchored on LG 3 and LG 9. Similarly, based on the organization of LG 2 (Fig. 1), a part of the ARACHNE supercontig 1198, which is mostly anchored to LG 3, should be moved into ARACHNE supercontig 1205. According to Martin *et al.* (2008), the largest ARACHNE supercontig 1195 (6.8 Mb) likely belongs to LG 9.

### Recombination rate

The ratio of genetic distances to physical lengths provides an estimate of the recombination rate. In *L. bicolor*, the ratio averages 20.92 cM $Mb^{-1}$ for the anchored part of the genome (Fig. 1). In many species there is a large variation in the recombination rate among linkage groups and a general tendency for the smallest linkage groups to recombine more than the large ones (Solignac *et al.*, 2007). In *L. bicolor*, the recombination rate is very similar for all LGs, varying from 15.58 cM $Mb^{-1}$ on LG 5 to 25.14 on LG 4. A region with a high recombination rate was identified on LG 5 from approximately locus *1207M10* to locus *TA-CCC498*, whereas a region of low recombination rate was noted on LG 6 from approximately *TC-CAT72* to *TA-CCA73*. The two regions of the mating-type loci also showed distinct recombination rates.

### Mapping structural genes

Nucleotide sequencing of *LbH2* and *lcc6* (P. E. Courty, unpublished) genes from the *L. bicolor* S238N homokaryons allowed us to detect a single SNP in *LbH2* intron 3 (*Shyd2.1)* and three SNPs in *lcc6* coding sequence (Table 2). By using these SNPs, we located the corresponding loci on the linkage map. *LbH2* (identified by *Shyd2.1*) mapped on LG 1 (supercontig 1211) (Fig. 1), whereas *lcc6* (identified by S*Lac14.1, SLac14.2*) mapped on pair 2 (supercontig 1215) (Fig. 1). The mating-type locus *MATa* (homeodomain transcriptional factors, *HD1 and HD2*) was mapped on LG1 (supercontig 1195) (Figs S1, S4). This result suggests that LG1 and LG9 are tandem ends of the same chromosome. *MATb* (*STE3*-like pheromone receptors) was mapped on nonintegrated linkage pair 4 on supercontig 1168; the latter likely belonging to LG 8 (see previous map version in Martin *et al.* 2008).

The whole nuclear rDNA repeat is not part of the large JAZZ and ARACHNE assembled sequences currently available, because of the repetitive nature of this region. A BLASTN analysis using known nuclear rDNA tandem repeat sequences (Martin *et al.*, 1999) as queries revealed that portions of the rDNA located onto unassembled JAZZ scaffolds (Martin *et al.*, 2008). Sequence reads corresponding to the consensus sequence of the rDNA unit were found at *c.* 500 copies in raw sequence traces (NCBI Trace Archive), whereas single copy genes were present at 10 copies, indicating that the *L. bicolor* genome contains 50 repeats of the 10 kbp-rDNA unit, that is, 500 kbp. The linkage analysis of the segregating rDNA IGS1 heteroduplexes ($\alpha$ or $\beta$) allowed the mapping of this rDNA tandem repeat locus on the distal end of LG 7 (Fig. 1).

## Discussion

The present genetic linkage map is composed of 287 markers on 13 LGs covering a total genetic length of 812 cM, with an

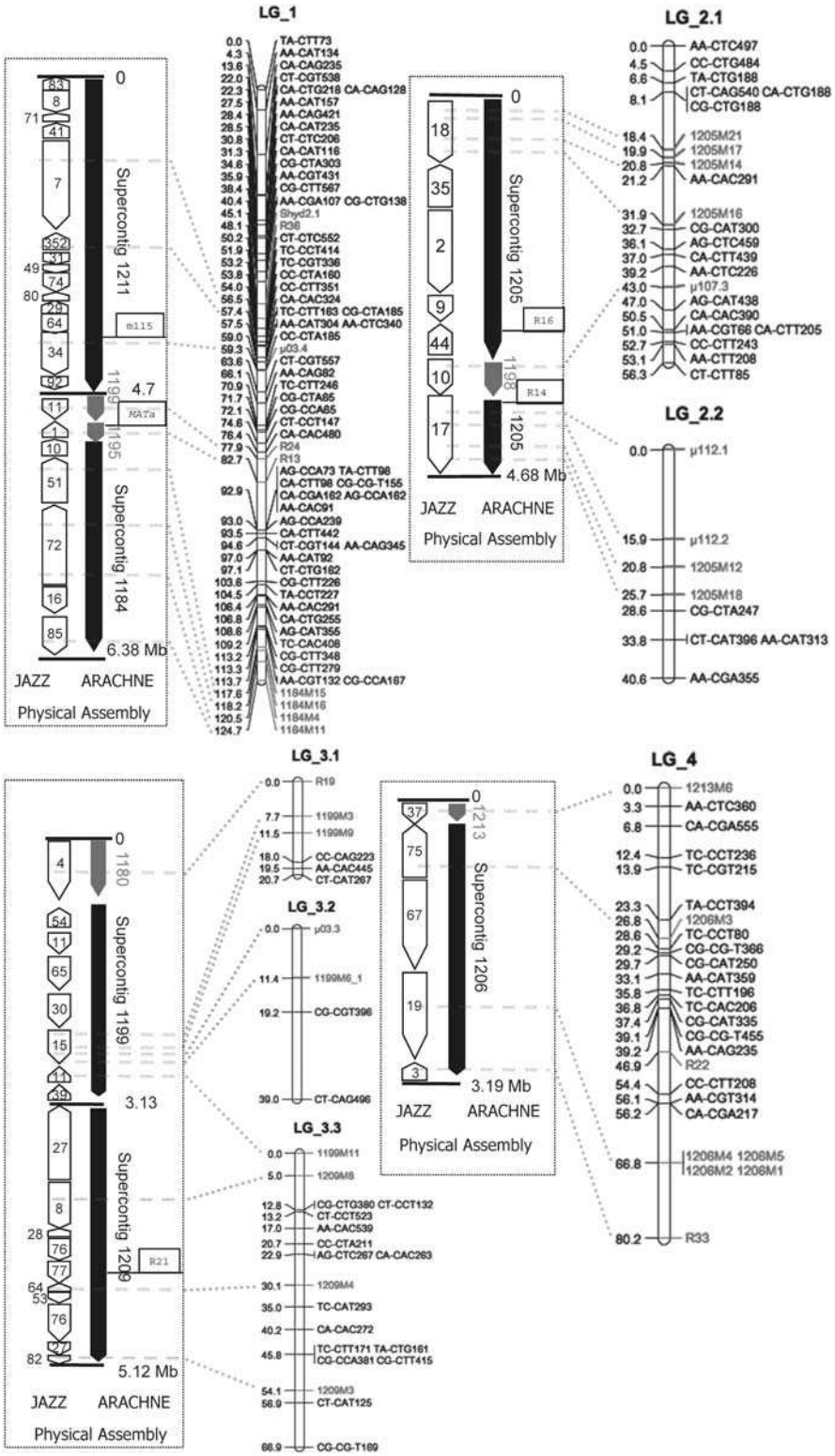



**Fig. 1** Alignment of whole-genome sequence (WGS) assemblies and the genetic map resulting in chromosome maps for *Laccaria bicolor*. JAZZ scaffolds and ARACHNE supercontigs anchored to the genetic map are drawn on the left of linkage groups. Maps are scaled to represent the relative physical length of the pseudochromosomes. The genetic unit in centiMorgan (cM) is indicated on the left of LGs. The physical size in Mb (or kb) is shown on the right of the supercontigs. AFLP markers are shown but they were not anchored on the WGS assembled sequences. Additional markers from a previous version of the linkage map (Martin *et al*., 2008) are also shown by boxes (i.e. *MATa*). *LbH2* gene marker is identified by *Shyd2.1*. *HD1* and *HD2* genes are colocalized in mating-type A locus and are shown by the box *MATa*.

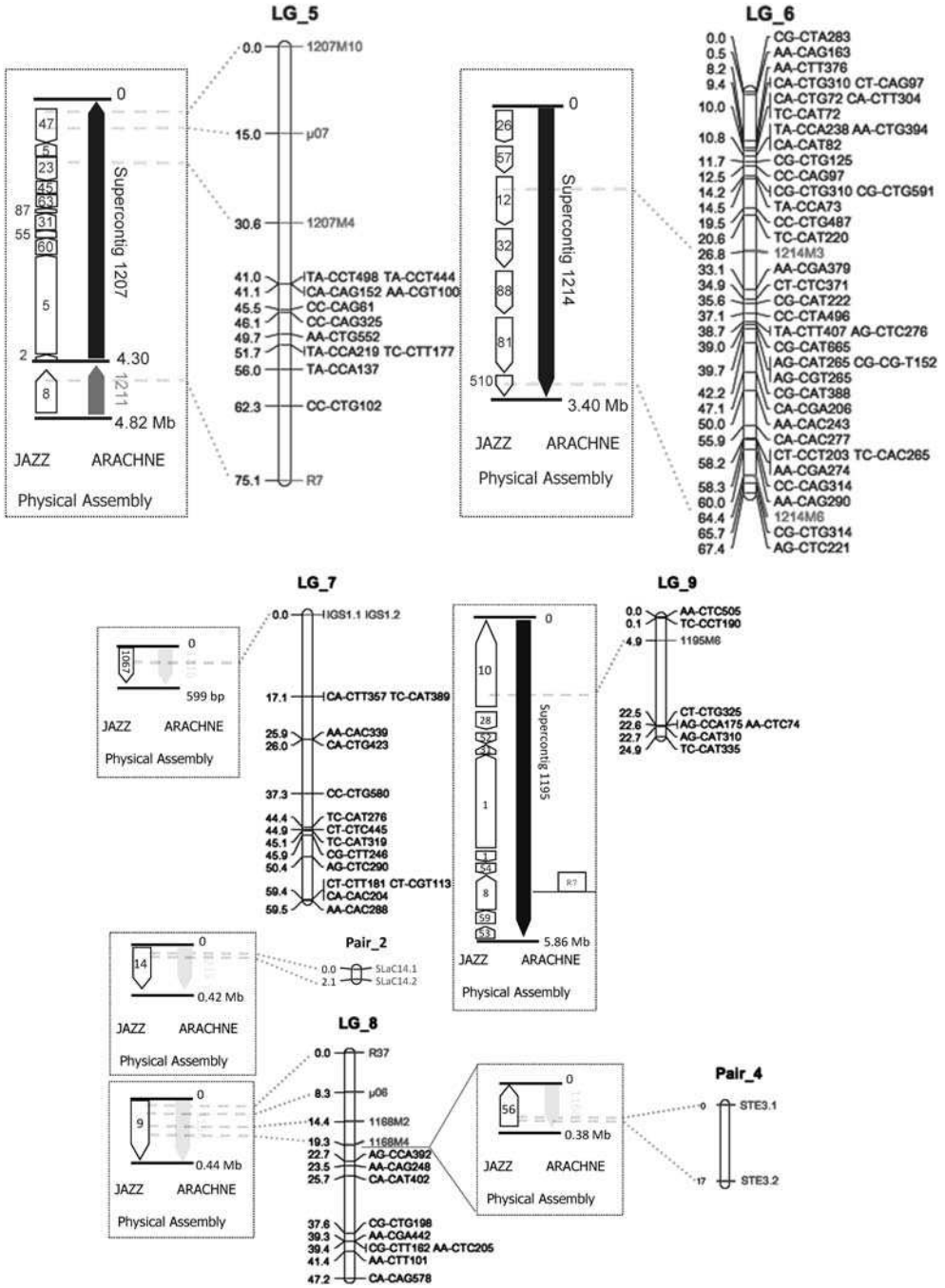



**Fig. 1 continued**

average of 2.76 cM between adjacent markers. The advantage of this linkage map over that generated in Doudrick *et al*. (1995) is greater genome coverage and integration of the assembled sequences of the *L. bicolor* S238N-H82 WGS. Nucleotide sequences of the RAPD markers used by Doudrick *et al*. (1995) are unavailable and therefore the two *L. bicolor* genetic maps could not be reconciled. We were, however, able to create and map sequence tags for 39 RAPD markers derived from 23 Doudrick *et al*. RAPD primers, but only eight of these markers were integrated into the current genetic map.

In total, we aligned 48 sequence-based markers with the JAZZ and ARACHNE genome sequence assemblies; nine for LG 1, LG 2 and LG 3, four for LG 4, LG 5 and LG 8, two for LG 6, pairs 2 and 4, and one for LG 7 and LG 9 (Fig. 1). This validated the current large-scale structure of the draft

WGS assemblies, although additional finishing sequencing will be required to improve the assembly of highly polymorphic transposon-rich regions and the integration of telomeric and centromeric regions. The majority of the mapped markers used in our alignment were colinear with the sequence assembly, indicating that both the genetic map and sequence scaffolds achieved high fidelity. The genetic map anchored *c*. 38 Mb (66%) of the ARACHNE assembly. For example, the ARACHNE supercontig 1184 and corresponding JAZZ scaffolds, were anchored through four markers and the ARACHNE supercontig 1211 was anchored by three markers on LG 1 (Fig. 1). The anchoring of the distal end of LG 1, corresponding by similarity searches to the JAZZ scaffold 83, 8, 71 and 41, was not confirmed by a genetic linkage marker. The number of markers allowing the anchoring of LGs on the genome assemblies is still limited and, as a consequence, a large number of the ARACHNE assembled sequences remained poorly anchored, or even unanchored (e.g. anchoring of supercontig 1214 on LG 6, Fig. 1).

In most cases, the linkage map and physical sequence reciprocally validated marker order and position; it also revealed some problems in each. Improving the genetic map-assembled sequence integration will require an increased number of assignable markers designed to target poorly aligned regions. Additional SNP, SSR and AFLP markers will continue to be added to the genetic map, furthering the integration of genetic and physical resources. However, filling the gaps in the WGS assemblies will be limited in the centromeric and heterochromatic regions (Grewal & Klar, 1997). Similarly, regions having a high density of protein-coding genes (Poyatos & Hurst, 2007) will require additional measures.

Based on the JAZZ and ARACHNE assemblies, the haploid genome size of *L. bicolor* was estimated to be 59.9–64.9 Mb (Martin *et al*., 2008). An average of 46.7 kb $cM^{-1}$ was obtained by dividing the size of ARACHNE assembled genome integrated into the genetic map by the total genetic size. This ratio was close to those reported in other Agaricales, 48.5 kb $cM^{-1}$ in *Agaricus bisporus* (Kerrigan, 1993) and 35.1 kb $cM^{-1}$ in *Pleurotus ostreatus* (Larraya *et al*., 2000). Thus, the estimated size of the *L. bicolor* genome contained within the integrated genetic map and sequence assembly is 37.36 Mb.

Several regions of relatively high or low recombination were discernible when the genetic linkage distances were compared with the physical sequence distances (Fig. 1). Recombination suppression has been reported in the basidiomycetes *Ustilago hordei* (Lee *et al*., 1999) and *Cryptococcus neoformans* (Lengeler *et al*., 2002) for sex chromosomes where mating-type loci were present in large genomic regions that fail to pair properly during meiosis. A similarly low recombination rate for the region surrounding the *L. bicolor* mating-type A (Fig. S4) was proposed by Niculita-Hirzel *et al*. (2008) based on the determination of intergenic distance. Niculita-Hirzel *et al*. (2008) reported that the *STE3*-like pheromone receptor genes (*LbSTE3.1*, *LbSTE3.2*, *LbSTE3.3*, *LbSTE3.4*) were clustered in one locus (i.e. *MATb*) of JAZZ scaffold 56. Our data suggest that *MATb* is located on linkage pair 4, likely belonging to LG 8. The size of this locus is *c*. 9.5 kb with three sub-loci, each containing two genes that encode for a mating-type-specific pheromone and the corresponding receptor (Niculita-Hirzel *et al*., 2008). Analysis by PCR of the *MATa* and *MATb* loci showed that the two mating-type loci were unlinked in contrast to previous findings (Doudrick *et al*., 1995). In our study, the *MATa* locus was localized on ARACHNE supercontig 1195 (JAZZ scaffold 1) anchored onto LG 1 (Fig. 1).

We located the ribosomal DNA at the distal end of LG 7. Based on the number of copies (i.e. 50) per haploid genome and the size of the rDNA tandem repeat (10 kb, Martin *et al*., 1999), the size of the rDNA locus is *c*. 0.5 Mb. The occurrence of numerous telomeric motifs (TTAGGG) in the intergenic spacer (IGS) (Martin *et al*., 1999) is in agreement with the suggested peritelomeric location.

The current genetic map reduced the number of LGs from 15 (Doudrick *et al*., 1995) to 13. Cytological chromosome counts for *Laccaria montana*, a closely related species, suggested a minimal haploid number of nine chromosomes (Mueller *et al*., 1993). Ongoing cytological chromosome counts and electrophoretic karyotyping will confirm the haploid number of chromosomes in *L. bicolor* S238N. Based on the integration of the linkage map into WGS assemblies (i.e. unaligned supercontigs and scaffolds), we postulate that the total haploid number of *L. bicolor* chromosomes is less than 13.

The principal interest of genetic maps is to localize and then identify Mendelian genes or quantitative trait loci (QTLs) in association mapping studies. The current linkage map is therefore a precursor to several genomics applications, including QTL mapping, comparisons of genetic maps among *Laccaria* species and interspecific transfer of genomic information, and candidate-gene association studies. Moreover, this study supplies useful marker resources with known physical and genetic position for population genetic studies that could be used to differentiate local genetic populations (Tuskan *et al*., 1990) or to conduct future genome-wide association studies. The availability of a genetic linkage map integrated to the WGS (Martin *et al*., 2008) in *L. bicolor* will enhance our ability for comprehensive analyses of the regulation of chromosome recombination, the mechanisms underlying fruiting body and ectomycorrhiza formation, ecological symbiosis fitness and evolutionary interactions between the symbiont model *L. bicolor* and its various hosts.

## Acknowledgements

This project was supported by the Lorraine Region and INRA through a PhD. scholarship to JL. Funds were also provided by INRA, the European Commission Network of Excellence EVOLTREE, the Bioenergy Center Program at ORNL and the US Department of Energy. Oak Ridge National Laboratory

(ORNL) is managed by UT-Battelle, LLC for the US Department of Energy under contract no. DE-AC05-00OR22725. We would like to thank Christine Delaruelle, Véronique Jorge, Brett Mommer, Lee Gunter, Hélène Niculita-Hirzel and Joseph Armento for their assistance and helpful discussions. We also thank Pierre-Emmanuel Courty and Julien Gibon for the permission to use *LbH2* and *lcc6* sequences. We thank the Joint Genome Institute and the Stanford Human Genome Sequencing Center for making the *L. bicolor* genome assemblies available before publication.

## References

**Arnau J, Housego AP, Oliver RP. 1994.** The use of RAPD markers in genetic analysis of the plant pathogenic fungus *Cladosporium fulvum*. *Current Genetics* **25**: 438–444.

**Batzoglou S, Jaffe DB, Stanley K, Butler J, Gnerre S, Mauceli E, Berger B, Mesirov JP, Lander ES. 2002.** ARACHNE: a whole-genome shotgun assembler. *Genome Research* **12**: 177–189.

**Callac P, Desmerger C, Kerrigan RW, Imbernn M. 1997.** Conservation of genetic linkage with map expansion in distantly related crosses of *Agaricus bisporus*. *FEMS Microbiology Letters* **146**: 235–240.

**Debets F, Swart K, Hoekstra RF, Bos CJ. 1993.** Genetic maps of eight linkage groups of *Aspergillus niger* based on mitotic mapping. *Current Genetics* **23**: 47–53.

**Di Battista C, Bouchard D, Martin F, Genere B, Amirault J, Le Tacon F. 2002.** Survival after outplanting of the ectomycorrhizal fungus *Laccaria bicolor* S238N inoculated on Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) cuttings. *Annals of Forest Science* **59**: 81–92.

**Di Battista C, Selosse MA, Bouchard D, Stenström E, Le Tacon F. 1996.** Variations in symbiotic efficiency, phenotypic characters and ploidy level among different isolates of the ectomycorrhizal basidiomycete *Laccaria bicolor* strain S238. *Mycological Research* **100**: 1315–1324.

**Doudrick RL, Raffle VL, Nelson CD, Furnier GR. 1995.** Genetic analysis of homokaryons from basidiome of *Laccaria bicolor* using RAPD markers. *Mycological Research* **99**: 1361–1366.

**Fries N. 1983.** Spore germination, homing reaction, and intersterility groups in *Laccaria laccata* (Agaricales). *Mycologia* **75**: 221–227.

**Gale LR, Bryant JD, Calvo S, Giese H, Katan T, O'Donnell K, Suga H, Taga M, Usgaard TR, Ward TJ *et al.* 2005.** Chromosome complement of the fungal plant pathogen *Fusarium graminearum* based on genetic and physical mapping and cytological observations. *Genetics* **171**: 985–1001.

**Grewal SIS, Klar AJS. 1997.** Recombinationally repressed region between *mat2* and *mat3* loci shares homology to centromeric repeats and regulates directionality of mating-type switching in fission yeast. *Genetics* **146**: 1221–1238.

**Henrion B, Le Tacon F, Martin F. 1992.** Rapid identification of genetic variation of ectomycorrhizal fungi by amplification of ribosomal RNA genes. *New Phytologist* **122**: 289–298.

**Hulbert SH, Ilott TW, Legg EJ, Lincoln SE, Lander ES, Michelmore RW. 1988.** Genetic analysis of the fungus, *Bremia lactucae*, using restriction fragment length polymorphisms. *Genetics* **120**: 947–958.

**Kerrigan RW. 1993.** Meiotic behavior and linkage relationships in the secondarily homothallic fungus *Agaricus bisporus*. *Genetics* **133**: 225–236.

**Koressaar T, Remm M. 2007.** Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**: 1289–1291.

**Kosambi DD. 1944.** The estimation of map distances from recombination values. *Annals of Eugenics* **12**: 172–175.

**Krumsiek J, Arnold R, Rattei T. 2007.** Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* **23**: 1026–1028.

**Larraya LM, Perez G, Ritter E, Pisabarro AG, Ramirez L. 2000.** Genetic linkage map of the edible basidiomycete *Pleurotus ostreatus*. *Applied and Environmental Microbiology* **66**: 5290–5300.

**Le Tacon F, Alvarez IF, Bouchard D, Henrion B, Jackson JM, Luff S, Parlade JI, Pera J, Stenström E, Villeneuve N *et al.* 1992.** Variation in field response of forest trees to nursery ectomycorrhizal inoculation in Europe. In: Read DJ, Lewis DH, Fitter AH, Alexander IJ, eds. *Mycorrhizas in ecosystems*. Wallingford, UK: CAB International, 119–134.

**Lee N, Bakkeren G, Wong K, Sherwood JE, Kronstad JW. 1999.** The mating-type and pathogenicity locus of the fungus *Ustilago hordei* spans a 500-kb region. *Proceedings of the National Academy of Sciences, USA* **96**: 15026–15031.

**Lengeler KB, Fox DS, Fraser JA, Allen A, Forrester K, Dietrich FS, Heitman J. 2002.** Mating-type locus of *Cryptococcus neoformans*: a step in the evolution of sex chromosomes. *Eukaryotic Cell* **1**: 704–718.

**Lim S, Notley-McRobb L, Lim M, Carter DA. 2005.** A comparison of the nature and abundance of microsatellites in 14 fungal genomes. *Fungal Genetics and Biology* **41**: 1025–1036.

**Marra R, Huang JC, Fung E, Nielsen K, Heitman J, Vilgalys R, Mitchell TG. 2004.** A genetic linkage map of *Cryptococcus neoformans variety neoformans* serotype D. *Genetics* **167**: 619–631.

**Martin F, Aerts A, Ahrén D, Brun A, Danchin EGJ, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V *et al.* 2008.** Symbiosis insights from the genome of the mycorrhizal basidiomycete *Laccaria bicolor*. *Nature* **452**: 88–92.

**Martin F, Kohler A, Duplessis S. 2007.** Living in harmony in the wood underground: ectomycorrhizal genomics. *Current Opinion in Plant Biology* **10**: 204–210.

**Martin F, Lammers P, Tuskan GA, DiFazio SP, Podila GK. 2004.** Symbiotic sequencing for the *Populus* mesocosm: DOE tackles the genomes of endomycorrhizal *Glomus intraradices* and ectomycorrhizal *Laccaria bicolor*. *New Phytologist* **161**: 330–335.

**Martin F, Selosse MA, Le Tacon F. 1999.** The nuclear rDNA intergenic spacer of the ectomycorrhizal basidiomycete *Laccaria bicolor*: structural analysis and allelic polymorphism. *Microbiology* **145**: 1605–1611.

**Matheny PB, Curtis JM, Hofstetter V, Aime MC, Moncalvo JM, Ge ZW, Slot JC, Ammirati JF, Baroni TJ, Bougher NL *et al.* 2007.** Major clades of Agaricales: a multilocus phylogenetic overview. *Mycologia* **98**: 982–995.

**Mueller GJ, Mueller GM, Shih LH, Ammirati JF. 1993.** Cytological studies in *Laccaria* (Agaricales). *American Journal of Botany* **80**: 316–321.

**Mueller GM. 1982.** *The genus Laccaria in North America excluding Mexico*. PhD thesis, University of Tennessee, Knoxville, TN, USA.

**Mueller GM. 1991.** The Swedish taxa *Laccaria* (Agaricales) with notes on their distribution. *Nordic Journal of Botany* **10**: 665–680.

**Muraguchi H, Ito Y, Kamada T, Yabagi SO. 2003.** A linkage map of the basidiomycete *Coprinus cinereus* based on RAPD and RFLP. *Fungal Genetics and Biology* **40**: 93–102.

**Niculita-Hirzel H, Labbé J, Kohler A, Le Tacon F, Martin F, Sanders I, Kües U. 2008.** Gene organization of the mating type regions in the ectomycorrhizal fungus *Laccaria bicolor* reveals distinct evolution between the two mating type loci *New Phytologist*. doi: 10.1111/j.1469-8137.2008.02525.x

**Nitta N, Farman ML, Leong SA. 1997.** Genome organization of *Magnaporthe grisea*: integration of genetic maps, clustering of transposable elements and identification of genome duplications and rearrangements. *Theoretical and Applied Genetics* **95**: 20–22.

**Park SK, Penas MM, Ramirez L, Pisabarro AG. 2006.** Genetic linkage map and expression analysis of genes expressed in the lamellae of the edible basidiomycete *Pleurotus ostreatus*. *Fungal Genetics and Biology* **43**: 376–387.

**Poyatos JF, Hurst LD. 2007.** The determinants of gene order conservation in yeasts. *Genome Biology* **8**: 233–236.

**Raeder U, Thompson W, Broda P. 1989.** RFLP-based genetic map of *Phanerochate chrysosporium* ME446: lignin peroxidase genes occur in clusters. *Molecular Microbiology* **3**: 911–918.

**Selosse MA, Costa G, Di Battista C, Le Tacon F, Martin F. 1996.** Meiotic segregation and recombination of the intergenic spacer of the ribosomal DNA in the ectomycorrhizal basidiomycete *Laccaria bicolor*. *Current Genetics* **30**: 332–337.

**Solignac M, Mougel F, Vautrin D, Monnerot M, Cornuet JM. 2007.** A third-generation microsatellite-based linkage map of the honey bee, *Apis mellifera*, and its comparison with the sequence-based physical map. *Genome Biology* **8**: R66.

**Tuskan GA, Walla JA, Lundquist JE. 1990.** Genetic-geographic variation in western gall rust in North Dakota. *Phytopathology* **80**: 857–861.

**Tzeng TH, Lyngholm LK, Ford CF, Bronson CR. 1992.** A restriction fragment length polymorphism map and electrophoretic karyotype of the fungal maize pathogen *Cochliobolus heterostrophus*. *Genetics* **130**: 81–96.

**Van Ooijen JW, Voorrips RE. 2001.** *JoinMap 3.0, software for the calculation of genetic linkage maps*. Wageningen, the Netherlands: Plant Research International. [http://www.plant.wageningen-ur.nl].

**Vos P, Hogers R, Bleeker M, Reijans M, Vandelee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M *et al.* 1995.** AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**: 4407–4414.

**Yin T, Difazio S, Gunter LE, Riemenschneider D, Tuskan GA. 2004.** Large-scale heterospecific distortion in *Populus* revealed by a dense genetic map. *Theoretical and Applied Genetics* **109**: 451–463.

**Yin T, Zhang X, Huang M, Wang M, Zhuge Q, Tu S, Zhu L, Wu R. 2002.** Molecular linkage map of the *Populus* genome. *Genome* **45**: 541–555.

## Supporting Information

Additional supporting information may be found in the online version of this article.

**Fig. S1** Genetic linkage map of *Laccaria bicolor* S238N genome.

**Fig. S2** Electrophoretic profile of the co-dominant SSR marker *1205M12*.

**Fig. S3** Electrophoretic profile of rDNA *IGS1* homo- and heteroduplices generated by the co-occurrence of the α and β haplotypes in the PCR mixes.

**Fig. S4** Mapping of the region carrying the mating-type gene loci *MATa* (genes *HD1* and *HD2*) and *MATb* (genes *STE3.1*, *STE3.2*, *STE3.3*).

**Table S1** Sequenced-tagged markers excluded in the current genetic map (as a result of segregation distortion) and used for anchoring whole-genome sequence (WGS) assemblies on the genetic linkage map showed in Martin *et al.* (2008)

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

L'amélioration de la carte génétique de *L . bicolor* est en cours. Une nouvelle version sera disponible dans quelques semaines. Actuellement elle comporte 10 groupes de liaison au lieu de 13 dans l'article publié.

# Chapitre 3. Résultats

## Deuxième partie : recherche des gènes impliqués dans la formation des ectomycorhizes chez les deux partenaires

## Article 6. Genetic variability of ectomycorrhizal development and functionning in an interspecific F1 poplar cross inoculated with *Laccaria bicolor*

**Courty PE, Labbé J, Marçais B, Bastien C, Churin JL, Garbaye J, Martin F, Le Tacon F**

**(à soumettre à *New phytologist*)**

Cet article présente une étude révélant la variabilité génétique au sein d'un croisement entre deux espèces de peuplier et du fonctionnement de l'association symbiotique avec *L. bicolor* par la mesure d'activité enzymatiques. Comme il est indiqué en première page de cet article, j'ai contribué à égalité avec le premier auteur à la rédaction de ce manuscrit et aux expérimentations ; notamment en participant à la conception, à la mise en place de l'expérience et à l'entretien des plants en cultures, mais aussi aux analyses statistiques et aux calculs des héritabilités et des hétérosis.

**Genetic variability of ectomycorrhizal development and functionning in an interspecific F1 poplar cross inoculated with *Laccaria bicolor***

Courty PE[1,2,4*], Labbé J[1*], Marçais B[1], Bastien C[3], Churin[1]JL, Garbaye[1] J, Martin F[1], Le Tacon F[1]

[1] UMR 1136 INRA-Nancy Université, Interactions Arbres / Microorganisms, INRA-Nancy, 54280 Champenoux, France.

[3] INRA Orléans, Unité Amélioration, Génétique et Physiologie forestières, Ardon, BP 20619, 45166 Olivet Cedex, France.

[4] Present address: Botanical Institute, University of Basel, CH-4056 Basel, Switzerland

[*] These authors contributed equally to this work.

[2] Corresponding author: Courty Pierre-Emmanuel

Mail: pierre.courty@unibas.ch

tél: + 33 3 41 61 267 23 17

fax: + 33 3 41 61 267 23 30

**Abstract**

From an interspecific *P. deltoides* X *P. trichocarpa* controlled cross, 38 F1 individuals were inoculated with the ectomycorrhizal fungus *Laccaria bicolor*. The colonization of poplar roots by *L. bicolor* completely modified their capacity to secrete enzymes involved in organic matter breakdown or organic phosphorus mobilization such as N-acetylhexosaminidase, glucuronidase, cellobiohydrolase, glucosidase, xylosidase, laccase and acid phosphatase.

The enzymatic activities expressed in mycorrhizal roots differed significantly between the two parents, while it did not differ in non-mycorrhizal roots. Significant differences were found between poplar genotypes for all enzymatic activities measured on ectomycorrhizas except for laccase activity. On the contrary, no significant differences were found between poplar genotypes for enzymatic activities of non-mycorrhizal root tips except for acid phosphatase activity. The level of enzyme secretion by the associated fungus in mycorrhizal root tips is under the genetic control of the host. Moreover, poplar heterosis is expressed through the enzymatic activities of the fungal partner.

**Introduction**

The fine roots of social tree species in temperate and boreal forests are symbiotically associated with fungi, forming composite organs called ectomycorrhizas (ECM) (Smith & Read, 2008). The establishment and the functioning of ECM lead to complex morphological and physiological changes in both the plant and the fungus (Martin *et al.*, 2007). The ECM symbiosis has been classically described as a mutualistic association where the autotrophic plant supplies photosynthates to the heterotrophic fungus, which in turn supplies water and nutrients to the host (Smith & Read, 2008). Several studies also have shown that ectomycorrhiral fungi (ECMf) are capable to produce extracellular enzymes involved in the direct mobilization of nutrients from organic substrates (Burke & Cairney, 2002; Courty *et al.*, 2005, 2006; Lindahl *et al*., 2005; Koide *et al.*, 2008). In addition, a given species may contribute to significant functional variations of function through physiological plasticity (Buée *et al*., 2007; Courty *et al*., 2009).

The ecological performance and the physiological plasticity of ECMf depend on their genotypes, on environmental factors (Smith and Read, 1997; van der Heijden and Sanders 2002), on host plant genotypes (Barker *et al.*, 2002; Linderman and Davis, 2004), and on the interactions between all these factors (Khasa *et al.*, 2002; Karst *et al*., 2009). Recent studies also suggest that host plant genome may play a role in determining the dominant mycorrhizal type in dually colonized hosts (Walker and McNabb, 1984; Van der Heijden and Kuyper, 2001; Khasa *et al*., 2002). However, no

studies have simultaneously examined the host plant genetics and the functional plasticity of one ECMf species in controlled conditions.

The capability of ECM fungal species to form mycorrhiza with trees is a critical dimension of the biotic interactions. In the ECM symbiosis *Laccaria bicolor*/poplar, Tagu *et al*. (2005) have shown for the first time that the host genotype impacts on root colonization by the fungus. The use of poplar as host tree model was motivated by the availability of large genetic and genomic resources (Andersson *et al*., 2004; Brunner *et al*., 2004). Moreover, the study of heritability and variability of physiological parameters (i.e. water use efficiency, dry weight, Leaf Maximum Area) at family level were intensively studied in poplar (Brendel *et al*., 2002; Marron *et al*., 2005). Heritability of mycorrhizal colonization of poplar was also studied (Tagu *et al*., 2001, 2005). However, the functional plasticity of one ECM fungal genotype colonizing different genotypes of the same host species was never studied.

The purpose of this study was first to determine how the association of Poplar with the ECMf *L. bicolor* modifies the secretion by the root system of enzymes involved in the mobilisation of nutrient from organic matter. The second objective was to determine whether the enzymatic activities expressed in mycorrhizal roots differed significantly between two parents, *P. deltoides* and *P. trichocarpa*, and the different poplar genotypes (*P. deltoides* X *P. trichocarpa*). The third and the last objective was to determine the degree of poplar genetic control on fungal traits by calculating the heritability of enzymatic activities in mycorrhizal and non-mycorrhizal root tips and by calculating for these traits a possible heterosis among the progeny.

## Materials and Methods

*Plant material, strain and culture conditions*

Poplar material consisted of cloned 38 F1 individuals from an interspecific *P. deltoides* (female clone from Illinois, no. 73028-62) and *P. trichocarpa* (male clone from Washington, no. 101-74) controlled cross (family 54B) (Lefèvre *et al.,* 1998; Tagu *et al.*, 2001; Tagu *et al.*, 2005). We have tested the two parents and the 38 progenies to form mycorrhizas by inoculating them with *Laccaria bicolor* S238N (Di Battista *et al.*, 1996; Tagu *et al.,* 2001). The 38 F1 genotypes were chosen at random among the 336 genotypes used for the construction of genetic map (Faivre-Rampant *et al.,* 1999; Cervera *et al.,* 2001; Jorge *et al.,* 2005). The *L. bicolor* S238N fungal strain, coming from the INRA-Nancy collection of ECMf, was maintained on Pachlewski's medium

(Pachlewski and Pachlewska 1974). This model fungal strain was chosen for its ability to form ECMs with poplar and for the avaibility of genomic ressources (Tagu *et al.*, 2001; Martin *et al.*, 2008). The inoculum of *L. bicolor* S238N was prepared by aseptically growing the mycelium in a peat-vermiculite nutrient mix in glass jars for 2 months in the dark at 25 °C, and kept at 4° C before use (Le Tacon and Bouchard, 1986).

*Inoculation*

Cuttings of one internode of each of the 38 poplar progenies and the two parents were rooted and individually inoculated at the same time, in 1-l pots containing a mixture of fungal inoculum (1:9 vol/vol) and calcinated attapulgite (Oil Dri US Special, Damolin, Denmark ; http://www.damolin.dk) during three months and a half, in greenhouse in spring with day-night temperatures of 15 and 28°C respectively. Plantlets were watered during the whole experiment until measurements. From the second month, a nutrient solution was applied weekly (Frey-Klett *et al.*, 1997). In order to control environmental heterogeneity in the greenhouse, 8 replicates were done for each poplar genotype and were randomly distributed in 8 blocks. Each block contained one pot of each 38 progenies and the two parents.

*Root colonization*

Entire root systems (except roots present 1 cm depth from cal) were carefully washed under tap water and cut into approximately 1-cm pieces. For each root system, 100 randomly selected root tips were examined and assessed as mycorrhizal or non-mycorrhizal under a stereomicroscope (magnification, 40).

*Chlorophyll content, leaf morphological measurement and dry weight*

Before harvesting plantlets, chlorophyll content was measured with a Minolta SPAD chlorophyll meter (Minolta Corp., Ramsey, N.J.). The SPAD chlorophyll meter measures the absorbance by plant tissues of wavelengths in the visible spectrum, which gives the relative internal concentration of chlorophylls a and b. Three SPAD measurements were done on three leaves of each plants and then averaged (Monje and Bugbee, 1992). SPAD measures were then converted into chlorophyll content.

Once mycorrhizal infection had been determined, leaves, stems and roots were separated. The leaves were placed in plastic bags and kept at 4°C until leaf morphological measurements were completed. The leaf area ($cm^2$) of all leaves of each

plantlet was measured by using a LI-COR 3100 (Li-Cor Inc., Lincoln, NE, USA). Then, leaves, stems and roots were dried at 70 °C for 1 week (Mettler, Toledo balance). The leaf mass area (LMA) was calculated for each clone using the relationship between the area of each leaf and its corresponding dry weight.

*Enzymatic activity profiling of ectomycorrhizal and non-mycorrhizal root tip*

One mycorrhizal root tip and one non-mycorrhizal root tip were collected from each of the 320 cuttings in order to determine their potential enzymatic activities, using the high-throughput photometric and fluorimetric microplate assays described and detailed in Courty et al. (2005), and applied in previous studies (Buée et al., 2007; Rineau et al., 2009; Courty et al., 2009). Each well of the 96-well micro-titration plate contained either one ectomycorrhizal root tip or one non-ectomycorrhizal root tip. Seven activities were successively measured on root tips: xylosidase (EC 3.2.1.37), glucuronidase (EC 3.2.1.31), cellobiohydrolase (EC 3.2.1.91), N-acetylglucosaminidase (EC 3.2.1.14), ß-glucosidase (EC 3.2.1.3), acid phosphatase (EC 3.1.3.2), and laccase (EC 1.10.3.2) activities. The enzymes activities were expressed as $pmol.min^{-1}.mm^{-2}$ of developed surface area of root tips.

*Statistical analysis*

The percentage of mycorrhizal colonization was transformed by arcsin $\sqrt{X/100}$ function prior to variance analysis (ANOVA). Xylosidase, glucuronidase, cellobiohydrolase, chitinase, ß-glucosidase, acid phosphatase and laccase activities, root, shoot and stem dry weight and LMA were also submitted to ANOVA. The following mixed linear model was applied on an individual basis to detect significant differences among the clones:

$Y_{ijk} = \mu + B_i + G_j + \varepsilon_{ijk}$ where $\mu$ is the overall mean, B is the block effect (fixed), G is the genotype effect (random), and $\varepsilon$ is the random residual error. Restricted maximum likelihood estimates of genetic, block and residual variance components ($\sigma^2_G$, $\sigma^2_B$ and $\sigma^2_\varepsilon$) were computed, and for each trait, individual broad sense heritability ($h^2$) was estimated as follows:

$h^2 = \sigma^2_G / (\sigma^2_G + \sigma^2_\varepsilon / n)$ where n is the average number of replicates per genotype. Standard deviations (SD) of were derived from classic estimation of SD for a ratio x/y where $x = \sigma^2_G$ and $y = \sigma^2_G + \sigma^2_\varepsilon / n$.

All analyses were performed with the statistical programs JMP 5.0 (SAS Institute Inc., Cary, NC, USA) and R version 1.8.0 (R. Development Core Team, 2003, www.R-project.org).

The genetic coefficient of variation ($CV_G$) was used (Cornelius, 1994) to compare the relative amounts of genetic variation of traits with different means:

$CV_G = \sqrt{( \sigma^2_G / \mu)}$

Relationships between the different traits were also analysed by Pearson linear correlations.

**Results**

A total of 320 plants were harvested and studied in this experiment. Only 3 cuttings died and were not used in the analysis. No significant block effect was found for any measured traits.

*Poplar ecophysiological traits*

Significant differences were found between plant genotypes for all measured traits (chlorophyll content, leaf maximum area, stem and root dry weight; Table 1).

*Effect of poplar genotype on root colonization by L. bicolor*

Three months and half after inoculation, progenies and parental clones were exclusively mycorrhizal with L. bicolor. No contaminant ectomycorrhizas were observed on the roots. The two parents greatly differed in their mycorrhizal development, *P. trichocarpa* exhibiting a rate of colonization of 40 % ± 8, and P. deltoides a rate of 16 % ± 4. The percentages of colonization of the different genotypes varied from 12 % ± 8 to 64 % ± 6, with an average of 31 % (Figure 1). The variance analysis revealed a significant genotype effect and no block effect.

*Enzymatic activity patterns of ectomycorrhizal and non-mycorrhizal root tips*

For each plantlet, the seven enzymatic activities were measured successively on one mycorrhizal and on one non-mycorrhizal root tip (40 poplar genotypes x 8 plant replicates x 2 root tips). For activities measured on ECM root tips of the 8 replicates of the same Poplar genotype, the standard deviation (n = 8) was around 15 % for very low activities (e.g. glucuronidase) and below 10 % for medium and high activities (e.g. cellobiohydrolase) (Figure 1). For activities measured on non-mycorrhizal root tips, the standard deviation (n = 8) was between 15 to 50 %. Mycorrhizal root tips never lost

their ability to secrete the seven enzymes in the conditions of the test, even if sometimes at very low level (e.g. for glucuronidase, laccase).

For the 40 clones, six of the enzymatic activities differed significantly between mycorrhizal and non mycorrhizal roots, while no laccase activity could be detected in non mycorrhizal roots. Compared to non mycorrhizal root tips, N-acetylhexosaminidase activity was multipled by more than 100 in mycorrhizas, while glucuronidase, cellobiohydrolase and glucosidase activities were multiplied by a factor ranging between 50 and 100, and xylosidase and acid phospatase between 15 and 50 (Table 2).

*Enzymatic activities of the two parents*

The six enzymatic activities expressed in non mycorrhizal roots did not differ significantly between the two parents, both expressed a similar profile. The seven enzymatic activities measured on ECM root tips differed significantly between the two parents (Figure 2): five activities (cellobiohydrolase, chitinase, ß-glucosidase, acid phosphatase, and laccase) were overexpressed in *P. trichocarpa* and two (xylosidase and glucuronidase) were overexpressed in *P. deltoides*.

*Enzymatic activities of the hybrids*

Enzyme activity patterns of mycorrhizal and non-mycorrhizal root tips of the parents and of their progeny differed significantly (Figure 3, Table 3). Significant differences were found between plant genotypes for all activities measured on ECM root tips except for laccase activity (Table 3). No significant differences were found between plant genotypes for any activities measured on non-mycorrhizal root tips except for acid phosphatase activity (Table 3).

*Heritability*

Heritability values ($h^2$) of plant phenotypic characters ranged from 0.21 to 0.48 (Table 4). The highest values of heritability were found for leaf maximum area (0.48 ± 0.01), chlorophyll content (0.45 ± 0.01) and stem dry weight (0.43 ± 0.01). The lowest value was found for root dry weight (0.21 ± 0.01). A value of 0.45 ± 0.02 was found for the percentage of mycorrhizal colonization (Table 4). Heritability values of enzymatic activities were similar for ectomycorrhizal and non-mycorrhizal root tips, except for laccase activity, which was not detected on non-mycorrhizal root tips (0.29 ± 0.01 in mycorrhizal root tips; 0 in non-mycorrhizal root tips). The highest heritabilities were found for N-acetylhexosaminidase (mycorrhizal root tips, 0.42 ± 0.01, non mycorrhizal

root tips 0.40 ± 0.01), acid phosphatase (mycorrhizal root tips, 0.41 ± 0.01; non mycorrhizal root tips, 0.40 ± 0.01), glucosidase (mycorrhizal root tips 0.36 ± 0.01; (non mycorrhizal root tips, 0.34 ± 0.01) and cellobiohydrolase (mycorrhizal root tips, 0.33 ± 0.02; non mycorrhizal root tips, 0.31 ± 0.02) activity. The lowest value was found for glucuronidase activity (mycorrhizal root tips 0.04 ± 0.01; non mycorrhizal root tips, 0.04 ± 0.01).

*Heterosis*

For each trait, we have calculated the ratio between the average of the hybrids and the best parent (a) or between the average of the hybrids and the average of the two parents (b) (Table 5). The leaf maximum area exhibited a high positive heterosis (a = + 46, b = + 51), while the leaf dry weight exhibited a negative one (a = - 41, b = - 36). The percentage of mycorrhizal colonization also exhibited a positive heterosis (a = + 25, b = + 75). In mycorrhizal and non mycorrhizal root tips, all the enzymatic activities displayed a positive heterosis at least for the b values, with the exception of laccase activity in mycorrhizal root tips (a = - 15, b = - 8) and acid phosphatase activity in non mycorrhizal root tips (a = - 4, b = - 3).

*Correlation between the different traits*

No poplar trait was correlated with enzymatic activities of non-mycorrhizal root tips (Table 6). LMA, stem and root dry weights were not correlated with any activities either from ectomycorrhizal and non-mycorrhizal root tips (Table 6). Chlorophyll content was significantly negatively correlated with xylosidase, glucosidase and cellobiohydrolase activities, three enzymes involved in cellulose and hemi-cellulose catabolism.

Enzymatic activities of ectomycorrhizal root tips were not correlated with those of non mycorrhizal root tips. All the enzymatic activities of mycorrhizal root tips were correlated between them, except for laccase activity. Similarly, except for acid phosphatase, all the enzymatic activities of non-mycorrhizal root tips were correlated between them. Xylosidase activity of mycorrhizal roots was the only activity positively correlated with the percentage of mycorhizal infection (Table 6). Stem and root dry weights also were significantly correlated with the percentage of root colonization.

## Discussion

*Enzymatic activities of root tips*

The colonization of poplar roots by *L. bicolor* completely modified their capacity to secrete enzymes involved in organic matter breakdown or organic phosphorus mobilization. Compared to non-mycorrhizal root tips, N-acetylhexosaminidase activity was multiplied by more than 100 in mycorrhizas, while glucuronidase, cellobiohydrolase and glucosidase activities were multiplied between 50 and 100 and xylosidase and acid phospatase between 15 and 50. Moreover, laccase activity could not be detected on non-mycorrhizal roots. It is well known that ECMf are able to secrete enzymes such as chitinases and phosphatases, which allow the release of mineral nutrients from soil organic matter (Tibbett and Sanders, 2002). But it is the first time that the width of the modifications induced by the symbiotic association in the potential enzymatic secretion by the root system is measured. Although the complete sequences of the *P. trichocarpa* and *L. bicolor* genome are available, few studies are actually available on the characterization of genes coding for the enzymes used in our experiment. Nevertheless, we know that *L. bicolor* has a poor cellulolytic and hemicellulolytic machinery (Martin *et al.*, 2008). Concerning chitin catabolism, it appears that many genes code for secreted exo- and endo-chitinases as the N-acetylglucosaminidase. However, their functions could be multiple such as the mobilization of nitrogen from chitin or defense against other fungi. More information on the genes coding for laccases and acid phosphatases are available. In *L. bicolor*, the MCO (Multi Copper oxidases) family comprised 11 genes divided into two subfamilies corresponding to laccases *sensu stricto* (*lcc*1 to *lcc*9) and two ferroxidases (*lcc*10 and *lcc*11). Lcc3 and lcc8 transcripts are very abundant in *L. bicolor*/poplar ectomycorrhizas (Martin *et al.*, 2008; Courty *et al.*, 2009). Several MCOs, found in the *P. trichocarpa* genome, seem to be involved in the production of melanin and in the polymerization of lignin precursor (McCaig *et al.* (2005). This could explain why we did not find any activities on non-mycorrhizal root tips.

Acid phosphatases are widespread in living organisms. They form a complex family of enzymes and play many different roles, most of them remaining unknown. However, several of them, particularly in microorganisms, are able to free phosphate groups from complex organic compounds. Both ECMf and plants secrete acid phosphatases in the mycorhizosphere. Nevertheless, mycorrhizas secrete more phosphatases than non colonized roots (Colpaert et aL, 1997, Conn and Dighton, 2000). Ectomycorrhizal fungi exhibit high phosphatase release in their environment, particularly under mineral

phosphorus deficiency (Dighton, 1983). The *L. bicolor* genome comprises only one gene coding for acid phosphatase, while *P. trichocarpa* genome contains seven. We can make the assumption that the numerous acid phosphatases potentially existing in poplar roots are stored in the cells, explaining the absence of phosphatase activity on the root surface. On the contrary, the unique *Laccaria* acid phosphatase seems to be intensively secreted outside the fungal cells, giving to mycorrhizas access to phosphorus bound in organic compounds.

*Host genetic control of ECM enzyme secretion*

The enzymatic activities expressed in mycorrhizal roots differed significantly between the two parents, while it did not differ in non mycorrhizal roots. Significant differences were found between poplar genotypes for all enzymatic activities measured on ECMs except for laccase activity. On the contrary, no significant differences were found between poplar genotypes for enzymatic activities of non-mycorrhizal root tips except for acid phosphatase activity.

Heritability values of enzymatic activities were similar for ectomycorrhizal and non-mycorrhizal root tips, except for glucuronidase in both types of roots and laccase which was not detected on non-mycorrhizal root tips. It is remarkable to find a high heritability value among the poplar genotypes for the enzymatic secretions of mycorrhizal roots, which are mainly due to fungal activity.

Several previous studies have demonstrated significant genetic variability within plants and/or fungal species for symbiotic capability in mycorrhizal interactions. Rosado *et al*. (1994) reported a high value of heritability for colonization of *Pinus ellilotii* by the ECMf *Pisolithus tinctorius,* and moderate heritability for the development of *P. tinctorius* extramatrical mycelium. *Eucalyptus grandies, E. globules, E. marginata* and *Pinus muricata* varied greatly in their growth response to different *Pisolithus* and *Rhizopogon* genotypes, respectively (Tonkin *et al.,* 1989; Burgess *et al.,* 1994; Thompson *et al.,* 1994; Piculell *et al.,* 2008). Tagu *et al.* (2003, 2005) have already shown that the ability of poplar to form ECMs is under its genetic control. Other studies with contrasting results have found that plant genotype can play a dominant role in controlling the associates soil microbial communities. Short-term experiment have either shown variations in mycorrhizal colonization, in microbial and in mycorrhizal communities (Gehring and Whitham, 2003; Kuske et al., 2003; Tan et al., 2003; Kasurinen et al., 2005) or few differences in arbuscular fungal and bacterial communities (Bever et al; 1996; Madritch and Hunter, 2002, 2003). To our knowledge,

our study is the first one reporting a genetic control of the host plant on some fungal functional traits. Here, the degree of fungal enzymatic secretion is under the control of poplar genome. These results strongly suggest the potential for poplar genome to direct the microbial-plant interaction and to create environments to which ECM fungi can respond.

In addition, we found a high positive heterosis for the capacity of poplar to form mycorrhizas ($h^2$ a = + 45 %). We also found positive heterosis for characters such as leaf maximum area, dry weight of roots and for five of the seven enzymatic activities of mycorrhizal roots. Heterosis for poplar hybrids is a well-known phenomenon (Stettler *et al.*, 1996; Li and Wu, 1997; Pearce *et al.*, 2004; Marron *et al.*, 2006). Heterosis is determined by non-mutually exclusive mechanisms, including genome-wide dominance complementation, locus-specific overdominance effects and epistasis, although the relative contribution of each of these mechanisms is still unclear (Lippman and Zamir, 2007). But it is also the first time that it is shown that plant heterosis could be expressed through the physiological activity of the fungal partner.

*Conclusion*

The genetic diversity in tree species can influence fluxes of nutrients and also interactions with soil microorganisms. Assessing tree genotype x environment interactions is a major challenge in functional ecology. In this paper, our data linked and quantified the general relationships between poplar plant genetics, ECM fungal infection, and physiological parameters. In the association *L. bicolor*/poplar, variations in plant and fungal responses in these controlled conditions illustrate the broad plasticity of the interaction. In this study, the role of poplar genetics in determining both poplar growth characteristics and fungal activities are clear.

## Acknowledgements

The first author was funded by grants of the French Ministry of Ecology and Sustainable Development; part of this research has been supported by the Biological Invasions program of the same ministry. This project was also supported by the European Network of Excellence EVOLTREE.

## References

Andersson A, Keskitalo J, Sjödin A, Bhalerao R, Sterky F, Wissel K, Tandre K, Aspeborg H, Moyle R, Ohmiya Y, Bhalerao R, Brunner A, Gustafsson P, Karlsson J, Lundeberg J, Nilsson O, Sandberg G, Strauss S, Sundberg B, Uhlen M, Jansson S, Nilsson P. 2004. A transcriptional timetable of autumn senescence. *Genome Biol* 5: R24.

Brendel O, Pot D, Plomion C, Rozenberg P, Guehl JM. 2002. Genetic parameters and QTL analysis of delta 13C and ring width in maritime pine. *Plant,Cell and Environment*. 25: 945-953.

Brunner AM, Busov VB, Strauss SH. 2004. Poplar genome sequence: functional genomics in an ecologically dominant plant species. *TRENDS in Plant Science* 9: 49-56.

Buée M, Courty PE, Mignot D & Garbaye J. 2007. Soil niche effect on species diversity and catabolic activities in an ectomycorrhizal fungal community. *Soil Biol Biochem* 39: 1947-1955.

Ceulemans, R. 1990. Genetic variation in functional and structural productivity determinants in poplar. Thesis Publishers, Amsterdam.

Conn C, Dighton J. 2000. Litter quality influences on decomposition, ectomycorrhizal community structure and mycorrhizal root surface acid phosphatase activity. *Soil Biol Biochem* 32: 489-496.

Courty PE, Pritsch K, Schloter M, Hartmann A & Garbaye J. 2005. Activity profiling of ectomycorrhiza communities in two forest soils using multiple enzymatic tests. *New Phytol* 167: 309-319.

Courty PE, Pouysegur R, Buée M & Garbaye J. 2006. Laccase and phosphatase activities of the dominant ectomycorrhizal types in a lowland oak forest. *Soil Biol Biochem* 38: 1219-1222.

Di Battista C, Selosse MA, Bouchard D, Stenstrom E, Le Tacon F. 1996. Variations in symbiotic efficiency, phenotypic characters and ploidy level among different isolates of the ectomycorrhizal basidiomycete *Laccaria bicolor* strain S238. *Mycological research* 100: 1315-1324.

Dighton J. 1983. Phosphatase production by mycorrhizal fungi. *Plant and soil* 71: 455-462.

Dillen SY, Marron N, Bastien C, Ricciotti L, Salani F, Sabatti M, Pinel MPC, Rae AM, Taylor G, Ceulemans R. 2007. Effectsof environment and progeny on biomass estimations of five hybrid poplar families grown at three contrasting sites across Europe. *Forest Ecology and Management* 252: 12-23.

Duponnois R, Garbaye J. 1991. Techniques for controlled synthesis of the Douglas fir-*Laccaria laccata* ectomycorrhizal symbiosis. *Annals of Forest Sciences* 48: 641-650.

Frey-Klett P, Pierrat JC, Garbaye J (1997) Location and survival of mycorrhiza helper *Pseudomonas fluorescens* during establishment of ectomycorrhizal symbiosis between *Laccaria bicolor* and Douglas fir. *Appl Environ Microbiol* 63: 139-144.

Khasa DP, Hambling B, Kernaghan G, Fung M, Ngimbi E. 2002. Genetic variability in salt tolerance of selected boreal woody seedlings. *Forest Ecology and Management* 165: 257-269.

Karst J, Jones MD, Turkington R. 2009. Ectomycorrhizal colonization and intraspecific variation in growth responses of lodgepole pine. *Plant Ecology* 200: 161-165.

Koide RT, Sharda JN, Herr JR, Malcolm GM. 2008. Ectomycorrhizal fungi and the biotrophy-saprotrophy continuum. *New Phytol* 178: 230-233.

Lefervre F, Goué-Mourier MC, Faivre-Rampant P, Villar M. 1998. A single gene cluster controls incompatibility and partial resistance to various *Melampsora larici-populina* races in

hybrid poplars. *Phytopathology* 88:156–163.

Marron N, Villar M, Dreyer E, Delay D, Boudouresque E, Petit JM, Delmotte FM, Guehl JM, Brignolas F. 2005. Diversity of leaf traits related to productivity in 31 *Populus deltoides-Populus nigra* clones. *Tree Physiol.* 25: 425-435.

Martin F, Kohler A, Duplessis S. 2007. Living in harmony in the wood underground: ectomycorrhizal genomics. *Current Opinion in Plant Biology* 10: 204-210.

Mather. K. 1955. The genetical basis of heterosis. Proc. R. Soc. London B, 144: 143-150.

Monje OA, Bugbee B. 1992. Inherent limitations of nondestructive chlorophyll meters: a comparison of two types of meters. *Hortscience* 27: 69-71.

McCaig, B., Meagher R.B. and Dean J.F.D. 2005. Gene structure and molecular analysis of the laccase-like multicopper oxidase (LMCO) gene family in *Arabidopsis thaliana*. *Planta* 221: 619-636.

Nygren CMR, Rosling A. 2009. Localisation of phosphomonoesterase activity in ectomycorrhizal fungi grown on different phosphorus sources. *Mycorrhiza* 19: 197-204.

R Development Core Team. 2007. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.

Rae AM, Robinson KM, Street NR, Taylor G. 2004. Morphological and physiological traits influencing biomass productivity in short rotation coppice poplar. *Canadian Journal of Forest Research* 34: 1488-1498.

Rae AM, Pinel MPC, Bastien C, Sabatti M, Street NR, Tucker J, Dixon C, Marron N, Dillen SY, Taylor G. 2007. QTL for yield in bioenergy Populus : identifying G×E interactions from growth at three contrasting sites. *Tree Genetics & Genomics* 4: 97-112.

Rosado, S.C.S., Kropp, B.R., Piche, Y., 1994. Genetics of ectomycorrhizal symbiosis. I. Host plant variability and heritability of ectomycorrhizal and root traits. *New Phytol.* 126: 105-110.

Shull GH. 1952. Beginning of the heterosis concept. *In Heterosis*. Edited by J. W. Gowen. Iowa State College Press, Ames. Iowa. pp. 14-48

Smith SE, Read DJ. 2008. Mycorrhizal Symbiosis 3rd Ed. Academic Press, London.

SPSS, 2003 SPSS Inc., SPSS Statistical Software CD-ROM Version 13.0 for Windows, SPSS Inc., Chicago Illinois, USA (2003).

Tagu D, Faivre Rampant P, Lapeyrie F, Frey-Klett P, Vion P, Villar M. 2001. Variation in the ability to form ectomycorrhizas in the F1 progeny of an interspecific poplar (Populus spp.) cross. *Mycorrhiza* 10: 237-240.

Tagu D, Bastien C, Faivre-Rampant P, Garbaye J, Vion P, Villar M, Martin F. 2005. Genetic analysis of phenotypic variation for ectomycorrhiza formation in interspecific F1 poplar full-sib family. *Mycorrhiza* 15: 87-91.

Tibbett, M., Sanders, F.E., and Cairney, J.W.G. 1998. The effect of temperature and inorganic phosphorus supply on growth and acid phosphatase production in arctic and temperate strains of ectomycorrhizal *Hebeloma* spp. in axenic culture. Mycological Research 102: 129-135.

Tonkin, C.M., Malajczuk, N., McComb, J.A., 1989. Ectomycorrhizal formation by micropropagated clones of *Eucalyptus marginata* inoculated with isolates of *Pisolithus tinctorius*. *New Phytol.* 111: 209-214.

Zhang X, Zang R, Li C. 2004. Population differences in physiological and morphological adaptations of *Populus davidiana* seedlings in response to progressive drought stress. *Plant Science* 166: 791-797.

**Tables and figures**

**Figure 1:** Percentage of the root colonization of the different Poplar clones ($n$ = 40). Genotypes are ranked in mean percentage of root colonization. Bars represent SE ($n$ = 8). Light grey correspond to *Populus deltoides* (female) and dark grey to *Populus trichocarpa* (male).

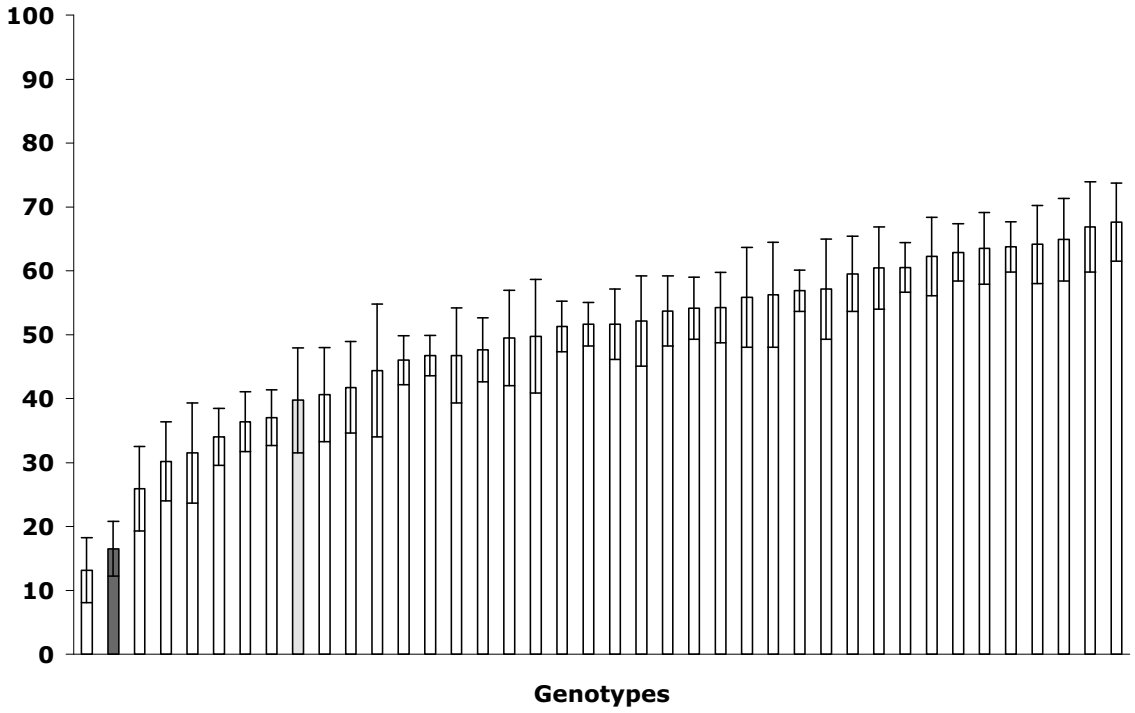


**Figure 2:** Enzymatic activity profiles measured on one root tip sampled on *P. trichocarpa* (dark grey line) and *P. deltoides* (light grey line) (*n* = 8 for each clone) either colonized (a) or non colonized (b) by *L. bicolor*. The polar graphs have been drawn from the relative activities calculated in per cent of the mean value of all root tips (either colonized or non-colonized by *L. bicolor*). (c) Traits measured on *P. trichocarpa* (dark grey line) and on *P. deltoides* (light grey line) (*n* = 8). The polar graphs have been drawn from the relative values calculated in per cent of the mean value measured on all poplar cuttings. The dotted line corresponds to the mean activity. As no laccase activity was measured on non mycorrhizal root tips. A statistical analysis was performed to highlight differences between the two parents (* $p<0.01$). Abbreviations for enzymes: pho, acid phosphatase; nag, *N*-acetyl-glucosaminidase; gls, β-glucosidase; cel, cellobiohydrolase; xyl, xylosidase; lac, laccase; glr, glucuronidase. Abbreviations for poplar traits: %, percentage of mycorrhizal colonization ; DW, dry weight; Chl; chorophyll content; LMA; Leaf Mass Area.

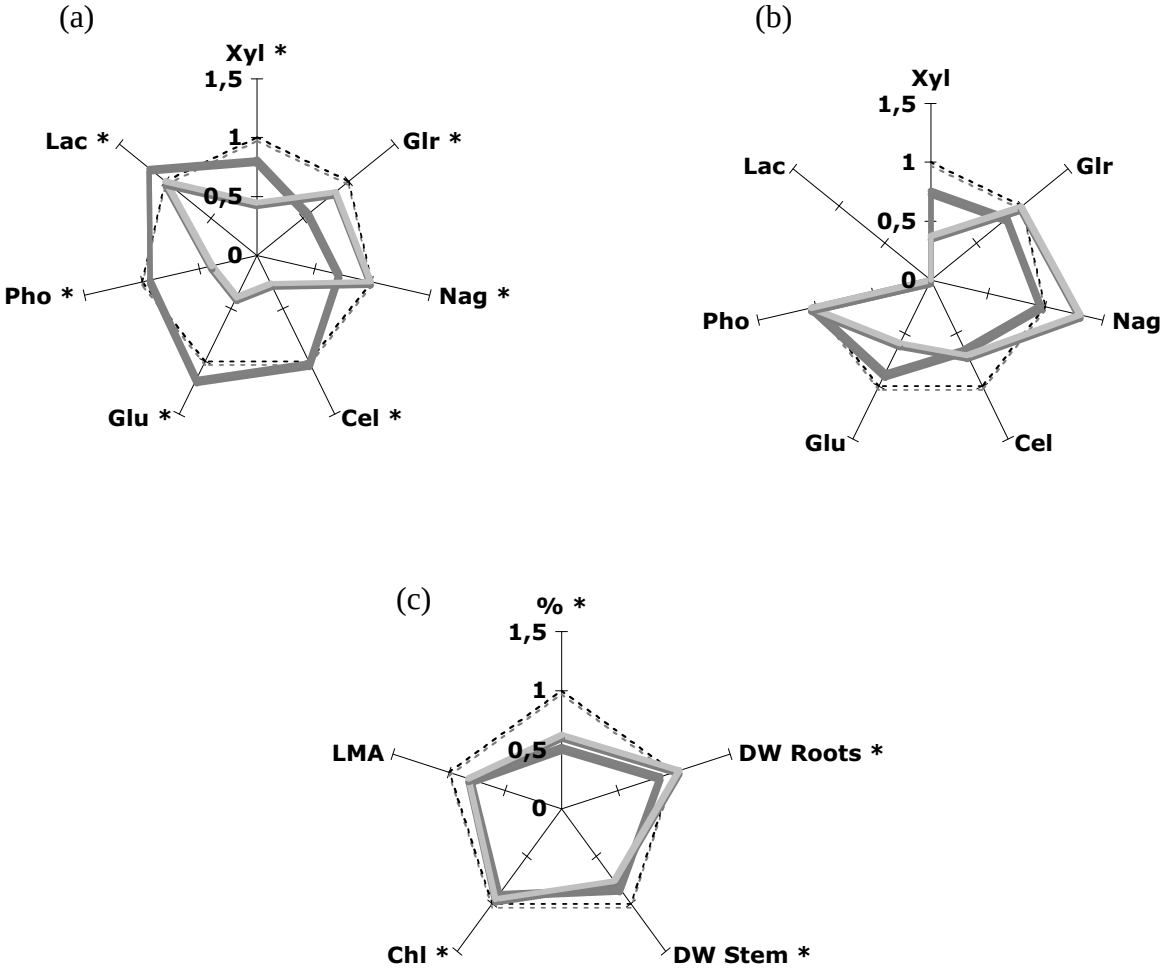


**Figure 3:** Values of the seven enzymatic activities successively measured on one root tip of the different poplar cuttings (40 clones, 8 replicates per clone), either colonized (a) or non-colonized (b), by *L. bicolor*. No laccase activity was detected on non mycorrhizal root tips. Genotypes are ranked by mean activity. Enzyme activities are expressed as pmol.mm$^{-2}$.min$^{-1}$ (Courty *et al*., 2005). Bars represent SE ($n$ = 8). Dark grey corresponds to *Populus trichocarpa* (male) and light grey to *Populus deltoides* (female).

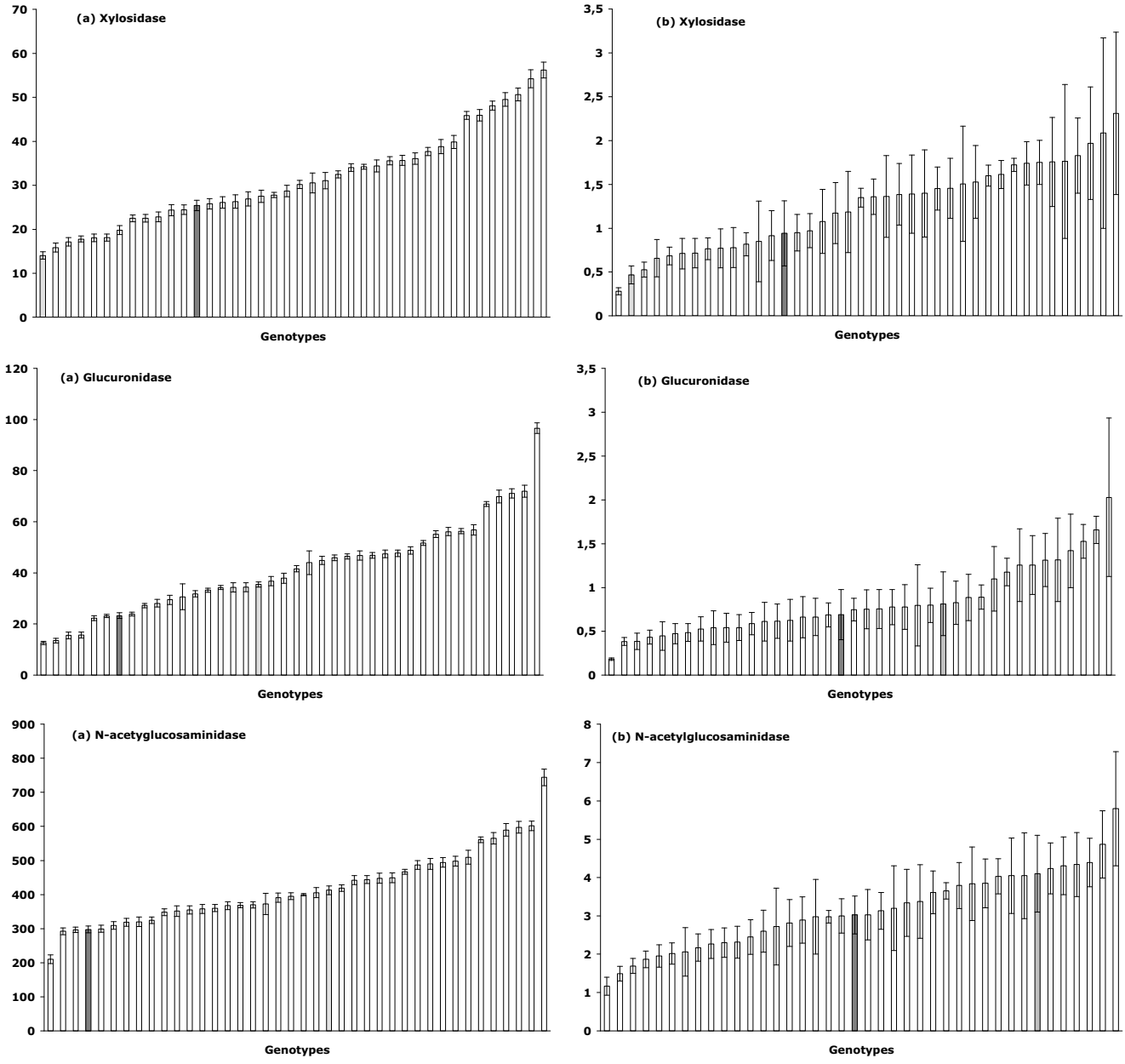
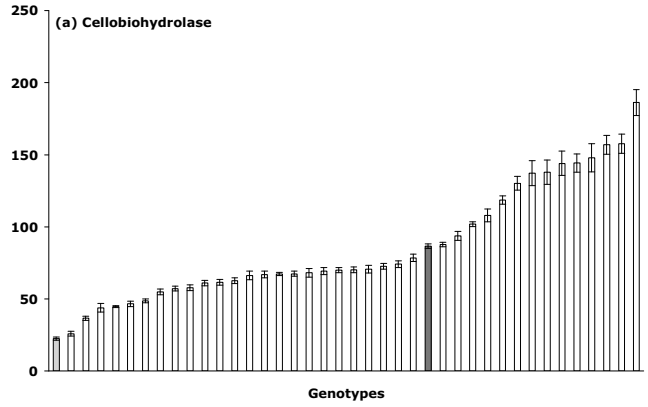
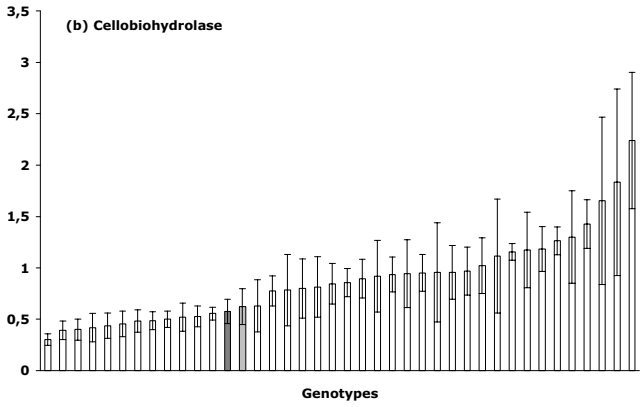
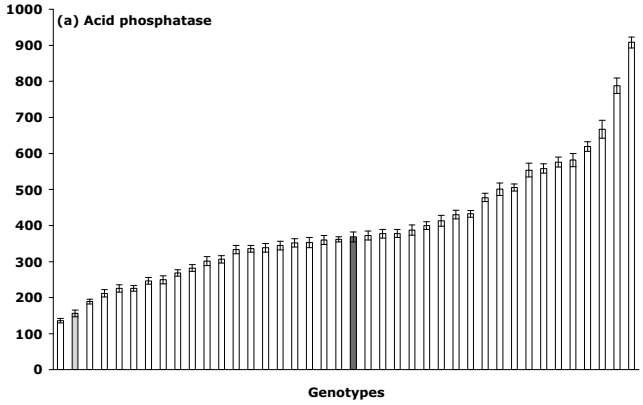
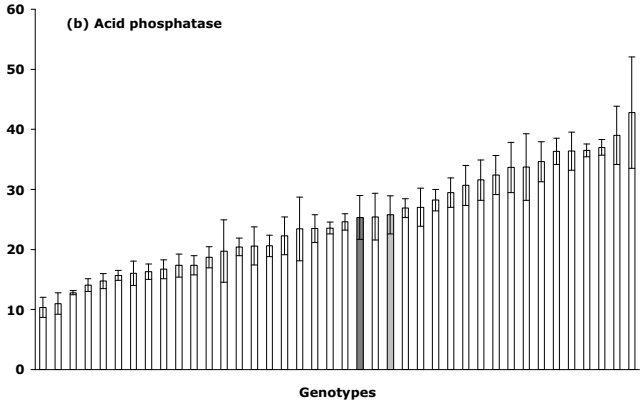


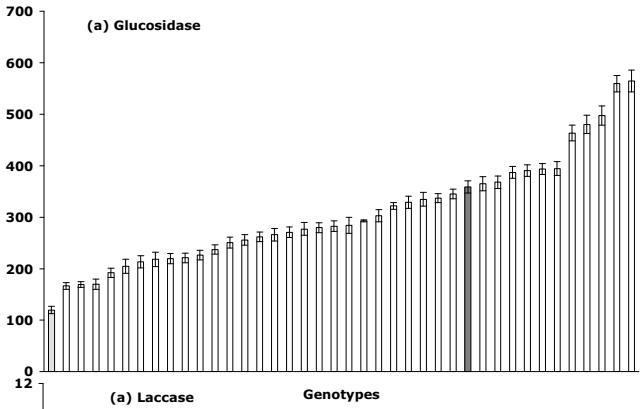
**Figure 3 (continued)**

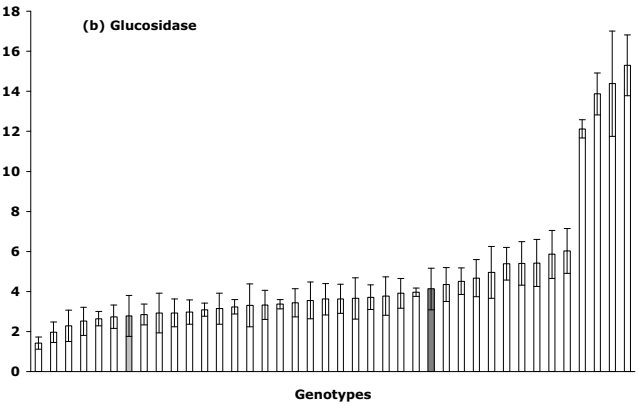


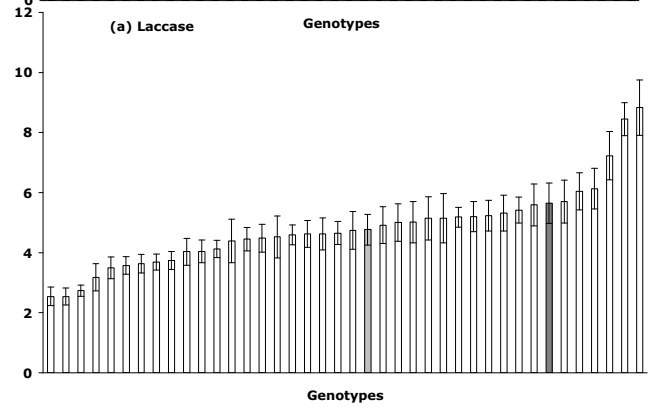



**Table 1** Trait variations among the two parents and the 38 progenies.

| Traits | F | P |
|---|---|---|
| Chlorophyll | 2.104 | <0.001 |
| Leaf maximum area | 1.661 | <0.001 |
| Stems (dry weight) | 8.105 | <0.001 |
| Roots (dry weight) | 0.768 | <0.001 |

**Table 2** Average enzymatic activities of ectomycorrhizal and non mycorrhizal root tips of poplar. Enzyme activities are expressed as pmol.mm$^{-2}$.min$^{-1}$ (Courty *et al.*, 2005). Mean and SE are given for each activity. Significant differences ($p < 0.001$) between mycorrhizal and non-mycorrhizal root tips are indicated by S.

| | n | Mycorrhizal root tips | Non mycorrhizal root tips | Ratio myc/non myc | $p < 0.001$ |
|---|---|---|---|---|---|
| Xylosidase | 317 | 31.32 ± 1.76 | 1.23 ± 0.07 | 25.5 | S |
| Glucuronidase | 317 | 41.41 ± 2.90 | 0.82 ± 0.06 | 50.5 | S |
| N-acetylhexosaminidase | 317 | 418.30 ± 17.26 | 3.14 ± 0.16 | 133.2 | S |
| Cellobiohydrolase | 317 | 85.07 ± 6.48 | 0.87 ± 0.06 | 97.8 | S |
| Glucosidase | 317 | 306.78 ± 16.93 | 4.68 ± 0.52 | 65.5 | S |
| Acid phosphatase | 317 | 396.65 ± 26.42 | 24.81 ± 1.35 | 16 | S |
| Laccase | 317 | 4.8 ± 0.2 | 0 | | nd |

**Table 3** Variation among 40 clones of poplar (two parents and the 38 progenies) in fungal ectomycorrhizal (a) and plant (b) physiological traits.

(a)

| Traits | Mycorrhizal root tips | | Non mycorrhizal root tips | |
|---|---|---|---|---|
| | *F* | *P* | *F* | *P* |
| Xylosidase | 18.69 | <0.001 | 0.51 | 1.0000 |
| Glucuronidase | 7.35 | <0.001 | 0.79 | 0.9285 |
| N-acetylhexosaminidase | 5.24 | <0.001 | 1.25 | 0.1295 |
| Cellobiohydrolase | 9.47 | <0.001 | 0.75 | 0.9599 |
| Glucosidase | 6.67 | <0.001 | 0.71 | 0.9510 |
| Acid phosphatase | 10.87 | <0.001 | 1.09 | <0.001 |
| Laccase | 0.74 | 0.3882 | nd | |

(b)

| Traits | | *F* | *P* |
|---|---|---|---|
| Chlorophyll | | 2.104 | <0.001 |
| Leaf maximum area | | 1.661 | <0.001 |
| Dry weight | Stems | 8.105 | <0.001 |
| | Roots | 0.768 | <0.001 |

**Table 4** Heritability ($h^2$) of poplar traits (a) and enzymatic activity of mycorrhizal and non-mycorrhizal root tips (b).

(a)

| | | n | $h^2$ |
|---|---|---|---|
| Mycorrhizal infection | | 40 | 0.45 ± 0.02 |
| Chlorophyll | | 40 | 0.45 ± 0.01 |
| Leaf Maximum Area | | 40 | 0.48 ± 0.01 |
| Dry weight | Stems | 40 | 0.43 ± 0.01 |
| | Roots | 40 | 0.21 ± 0.01 |
| | Leaves | 40 | 0.50 ± 0.01 |

(b)

| | n | Mycorrhizal root tips | Non-mycorrhizal root tips |
|---|---|---|---|
| Xylosidase | 40 | 0.16 ± 0.01 | 0.19 ± 0.01 |
| Glucuronidase | 40 | 0.04 ± 0.01 | 0.04 ± 0.01 |
| N-acetylhexosaminidase | 40 | 0.42 ± 0.01 | 0.40 ± 0.01 |
| Cellobiohydrolase | 40 | 0.33 ± 0.02 | 0.31 ± 0.02 |
| Glucosidase | 40 | 0.36 ± 0.01 | 0.34 ± 0.01 |
| Acid phosphatase | 40 | 0.41 ± 0.01 | 0.40 ± 0.01 |
| Laccase | 40 | 0.29 ± 0.01 | 0 |

**Table 5** Heterosis calculated as the ratio between the average of the hybrids and the best parent (a) or between the average of the two parents (b)

| Character | | Heterosis | |
|---|---|---|---|
| | | a | b |
| Mycorrhizal colonization | | + 25 | + 75 |
| Chlorophyll | | - 21 | - 7 |
| Leaf maximum area | | + 46 | + 51 |
| Dry weight | Stems | - 17 | - 14 |
| | Roots | + 1 | + 11 |
| Activities of mycorrhizal root tips | Xylosidase | + 26 | + 62 |
| | Glucuronidase | + 19 | + 44 |
| | N-acetylhexosaminidase | + 22 | + 19 |
| | Cellobiohydrolase | + 0.2 | + 59 |
| | Glucosidase | - 13 | + 30 |
| | Acid phosphatase | + 9 | + 54 |
| | Laccase | - 15 | - 8 |
| Activities of non mycorrhizal root tips | Xylosidase | + 34 | + 103 |
| | Glucuronidase | + 19 | + 82 |
| | N-acetylhexosaminidase | + 3 | + 11 |
| | Cellobiohydrolase | + 53 | + 75 |
| | Glucosidase | + 15 | + 38 |
| | Acid phosphatase | - 4 | - 3 |
| | Laccase | 0 | 0 |

**Table 6** Correlation matrix (Pearson Correlation r) between poplar parameters, enzymatic activities and percentage of mycorrhizal colonization. Correlation is significant for $p < 0.01$ (light grey cells).
Abbreviations: %, percentage of mycorrhizal colonization; Chl, Chlorophyl (g/m$^2$); DW, Dry Weight (g); LMA, Leaf maximum area (m$^2$); M, mycorrhized root tips; NM, non-mycorrhized root tips; Pho, acid phosphatase; Nag, *N*-acetyl-glucosaminidase; Gls, β-glucosidase; Cel, cellobiohydrolase; Xyl, xylosidase; Lac, laccase; Glr, glucuronidase.

| | % | Chl | DW Stem | DW Roots | LMA | M Xyl | M Glr | M Nag | M Cel | M Glus | M Pho | M Lac | NM Xyl | NM Glr | NM Nag | NM Cel | NM Glus | NM Pho |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| % | 1 | -0,09 | 0,20 | 0,20 | 0,06 | 0,22 | 0,14 | 0,13 | 0,07 | 0,13 | 0,04 | 0,05 | 0,05 | 0,13 | 0,02 | -0,02 | 0,06 | 0,13 |
| Chl | | 1 | -0,12 | -0,12 | -0,01 | -0,19 | -0,01 | -0,07 | -0,18 | -0,18 | -0,05 | 0,01 | -0,00 | 0,02 | 0,08 | 0,08 | 0,00 | 0,11 |
| DW Stem | | | 1 | 0,42 | -0,28 | -0,05 | -0,04 | 0,04 | -0,07 | -0,13 | -0,02 | -0,01 | -0,01 | 0,09 | 0,12 | 0,00 | -0,02 | 0,01 |
| DW Roots | | | | 1 | -0,20 | 0,05 | 0,04 | 0,06 | -0,01 | 0,02 | -0,03 | 0,06 | -0,07 | 0,00 | -0,09 | -0,09 | -0,10 | -0,02 |
| LMA | | | | | 1 | 0,14 | 0,09 | 0,14 | 0,06 | 0,10 | 0,12 | 0,13 | -0,08 | -0,13 | -0,04 | -0,07 | -0,06 | 0,04 |
| M Xyl | | | | | | 1 | 0,12 | 0,41 | 0,64 | 0,56 | 0,25 | 0,25 | 0,07 | 0,02 | -0,03 | 0,11 | -0,03 | 0,09 |
| M Glr | | | | | | | 1 | 0,27 | 0,18 | 0,15 | 0,30 | 0,15 | 0,01 | 0,07 | -0,12 | -0,13 | -0,07 | 0,15 |
| M Nag | | | | | | | | 1 | 0,35 | 0,33 | 0,47 | 0,31 | -0,11 | -0,08 | -0,04 | -0,01 | -0,19 | 0,32 |
| M Cel | | | | | | | | | 1 | 0,69 | 0,22 | 0,16 | 0,06 | -0,04 | 0,02 | 0,05 | -0,06 | 0,05 |
| M Gls | | | | | | | | | | 1 | 0,25 | 0,27 | 0,01 | -0,05 | -0,07 | 0,06 | -0,05 | 0,09 |
| M Pho | | | | | | | | | | | 1 | 0,29 | 0,01 | 0,08 | -0,01 | 0,02 | 0,01 | 0,36 |
| M Lac | | | | | | | | | | | | 1 | -0,12 | -0,05 | -0,14 | -0,06 | -0,10 | 0,21 |
| NM Xyl | | | | | | | | | | | | | 1 | 0,47 | 0,27 | 0,43 | 0,35 | -0,03 |
| NM Glr | | | | | | | | | | | | | | 1 | 0,25 | 0,39 | 0,29 | 0,04 |
| NM Nag | | | | | | | | | | | | | | | 1 | 0,24 | 0,31 | -0,03 |
| NM Cel | | | | | | | | | | | | | | | | 1 | 0,28 | -0,03 |
| NM Gls | | | | | | | | | | | | | | | | | 1 | 0,07 |
| NM Pho | | | | | | | | | | | | | | | | | | 1 |

## Article 7. Identification of quantitative trait loci affecting ectomycorrhizal symbiosis in an interspecific F1 poplar cross and differential expression of genes in ectomycorrhizas of the two parents*: Populus deltoides* and *Populus trichocarpa*

Jessy Labbé, Véronique Jorge, Annegret Kohler, Patrice Vion, Benoît Marçais, Catherine Bastien, Gerald A. Tuskan, Francis Martin and François Le Tacon

(à soumettre à *Theoretical and Applied Genetics*)

# Identification of quantitative trait loci affecting ectomycorrhizal symbiosis in an interspecific F1 poplar cross and differential expression of genes in ectomycorrhizas of the two parents*: Populus deltoides* and *Populus trichocarpa*

Jessy Labbé[1,*], Véronique Jorge[2], Annegret Kohler[1], Patrice Vion[1], Benoît Marçais[1], Catherine Bastien[2], Gerald A. Tuskan[3], Francis Martin[1] and François Le Tacon[1]

[1] UMR 1136, INRA-Nancy Université, Interactions Arbres/Microorganismes, INRA-Nancy, 54280 Champenoux, France.

[2] Unité Amélioration Génétique et Physiologie Forestière, INRA-Orléans, F-45166 Olivet Cedex, France.

[3] Environmental Sciences Division, Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, TN 37831-6422, USA.

* corresponding author

Email: labbe@nancy.inra.fr

**Summary:**

A *Populus deltoides* x *Populus trichocarpa* F1 progeny was analyzed for QTLs affecting ectomycorrhizal development and for characterization of gene networks involved in this symbiosis. This progeny was evaluated for its ability to form ectomycorrhiza with the basidiomycete *Laccaria bicolor*. The percentage of mycorrhizal root tips was determined on the root systems of 300 progeny genotypes and their two parents. A first QTL analysis (map v1) highligthed 4 significant QTLs, one (Myc_d1, linkage group LGg_d) on the *P. deltoides* and three on the *P. trichocarpa* map (Myc_t1, LGi_t; Myc_t2, LGk_t; Myc_t3, LGXI_t). Two QTLs positioned on the genetic map of *P. trichocarpa* could be anchored on the *P. trichocarpa* genome sequence. A second QTL analysis, performed with a different map (v2), identified three putative QTLs (MycI_d, LG_Ia d; MycII_d, LG_II d; MycVI_d, LG_VI d) on the *P. deltoides* map. On the *P. trichocarpa* map, two putative QTLs (MycIII_t, LG_IIIb t; MycXI_t, LG_XIa t) were also detected. A NimbleGen whole genome microarray analyses was performed by generating cDNA from RNA extracts of ectomycorrhizal root tips from the parental genotypes *P. trichocarpa* and *P. deltoides*. Among the 1543 differentially expressed genes (p-value < 0.05 and ≥ 5.0 change in transcript level) having different transcript levels in mycorrhizas of the two parents, 81 transcripts were located in the QTL genomic regions: 14 in Myc_t1 = MycXI_t, 7 in MycIII_t, 30 in MycI_d, 20 in MycII_d, and 10 in MycVI_d.

Among these 81 transcripts, 52 were up-regulated in *P. deltoides* of which 5 of top 20 of up-regulated genes; 29 were up-regulated in *P. trichocarpa* of which 2 of top 20. The gene showing the highest upregulation in *P. trichocarpa* mycorrhizas compared to *P. deltoides* mycorrhizas was lacated within the QTL MycIII_t on the linkage group LG_IIIb t and codes for a EREBP-4 protein involved in ethylene and auxin responses.

**Key words:** Quantitative trait loci, poplar, symbiosis, ectomycorrhiza

## Introduction

The majority of terrestrial plants live in association with symbiotic fungi which facilitate their access to soil mineral nutrients. The ectomycorrhizal (ECM) symbiosis is the most common association in forest under boreal and temperate climates. ECM symbioses are complex biological systems implying numerous interactions between the two partners and with their environments. These interactions also depend on the specific combination tree and fungus involved (Peterson and Bonfante, 1994; Nehls *et al.*, 2001; Martin *et al*, 2003; Martin *et al*, 2009). Ectomycorrhizal establishment depends as well on both trophic and climatic conditions (Smith and Read, 1997), as genetic traits of the partners (Rosado *et al.*, 1994; Tagu *et al.*, 2001).

The development and functioning of ectomycorrhizas involve the expression of numerous genes encoding for example for developmental proteins, enzymes or transporters both in the fungi and their partners (Martin *et al.*, 2007). The genetic predisposition and molecular adaptation needed for mycorrhizal establishment probably involve the combined action of numerous gene networks (Martin *et al*, 2009). One putative quantitative trait loci (QTL) related to ectomycorrhizal formation was located on a *P. trichocarpa* linkage group containing QTLs involved in the pathogenic interaction with the fungus *Melampsora larici-populina*, the causal agent of leaf rust (Tagu *et al.*, 2005). The recent release of genomic resources, such as the *Laccaria bicolor* genome (Martin *et al.*, 2008) and the *Populus trichocarpa* genome (Tuskan *et al*, 2006), will ease the identification of genetic factors involved in these complex interactions.

Here, we aimed to confirm and extend the previous QTL analysis (Tagu *et al.*, 2005) by positioning QTLs involved in root colonization by *L. bicolor* on two improved *Populus* genetic maps, using an extended progeny (300 F1 individuals). We then anchor the QTL markers on the *P. trichocarpa* genomic sequence and characterize genes that could be involved in the ectomycorrhizal establishment. Using a whole genome oligoarray of *P. trichocarpa* (Tuskan *et al.*, 2006) we compared the differential gene expression between *P. trichocarpa* and *P. deltoides* ECMs. The final objective was to determine whether these genes showing a differential expression between the parents exhibiting a low ectomycorrhizal infection (*P. deltoides*) and the parent exhibiting a high ectomycorrhizal infection (*P. trichocarpa*) located in putative QTLs related to ectomycorrhizal colonization.

## Material and methods

### *Biological materials*

Poplar material consisted of cloned 300 F1 individuals from an interspecific *P. deltoides* (female clone from Illinois, no. 73028-62) and *P. trichocarpa* (male clone from Washington, no. 101-74) controlled cross (family 54B, Pop2) (Lefèvre *et al.*, 1998; Tagu *et al.*, 2001; Tagu *et al.*, 2005). We have tested the two parents and the 300 progeny for ectomycorrhizal formation by inoculating them with *Laccaria bicolor* S238N (Di Battista *et al.*, 1996; Tagu *et al.*, 2001). The 300 F1 genotypes were chosen at random among the 336 genotypes used for the construction of the updated genetic maps (Jorge *et al.*, 2005). Moreover, these 300 genotypes comprised the 90 genotypes used for the construction of the first Pop2 genetic map (Faivre-Rampant *et al.*, 1999). The *L. bicolor* S238N fungal strain, issued from the INRA-Nancy collection of ectomycorrhizal fungi, was maintained on Pachlewski's medium (Pachlewski and Pachlewska 1974). This model fungal strain was chosen for its ability to form ectomycorrhizas with poplar and for the avaibility of its genome and transcriptome (Tagu *et al.*, 2001; Peter *et al.*, 2003; Martin *et al.*, 2008). The mycelium was produced on a peat-vermiculite nutrient mixture and grown in glass jars during 2 months in the dark at 25°C and kept at 4°C before use (Duponnois and Garbaye, 1991).

### *Inoculation*

Cuttings of one internode of each of the 300 poplar progeny and the two parents were rooted and individually inoculated at the same time, in 1-l pots containing a mixture of fungal inoculum (1:9 vol/vol) and calcinated attapulgite (Oil Dri US Special, Damolin, Denmark ; http://www.damolin.dk) during three and a half months, in a greenhouse at INRA-Nancy during spring (48° 41' 37" N, 6° 11' 05" E; glasshouse temperature 15-28°C, photoperiod 12h). Plantlets were watered twice daily during all the time of the experimentation until measurements, and weekly received a nutrient solution (like describe in Frey-klett *et al.*, 1997). In order to take in account the environmental heterogeneity of the greenhouse, eight replicates were done for each poplar genotype and were randomly distributed in eight blocks. Each block contained one pot of each 300 progeny and the two parents.

***Measurements***

The percentage of mycorrhizal infection was determined three and half months after inoculation by eight observers; it lasted 2 months. Each root system was rinsed with tap water, cut in 1-cm pieces and analyzed under a dissecting microscope. For each root system, 100 apices were randomly examined and assessed as mycorrhizal or non-mycorrhizal. Up to 25,000 mycorrhized root tips were harvested.

***Statistical analysis***

The percentage of mycorrhizal colonization was transformed by arcsin $\sqrt{X/100}$ function prior to variance analysis (ANOVA). The following mixed linear model was applied on an individual basis to detect significant differences in mycorrhizal colonization among the clones:

$Y_{ijk} = \mu + B_i + G_j + O_k + \varepsilon_{ijk}$

where $\mu$ is the overall mean, $B$ is the block effect (fixed), $G$ is the genotype effect (random), $O$ is the observer effect (fixed) and $\varepsilon$ is the random residual error. Restricted maximum likelihood estimates of genetic, block and residual variance components ($\sigma^2_G$, $\sigma^2_B$ and $\sigma^2_\varepsilon$) were computed, and for each trait, individual broad sense heritability ($h^2$) was estimated as follows:

$h^2 = \sigma^2_G / (\sigma^2_G + \sigma^2_\varepsilon /n)$ where $n$ is the average number of replicates per genotype. Standard deviations (SD) of $h^2$ were derived from classic estimation of SD for a ratio $x/y$ where $x = \sigma^2_G$ and $y = \sigma^2_G + \sigma^2_\varepsilon /n$. All analyses were performed with the statistical programs JMP 5.0 (SAS Institute Inc., Cary, NC, USA) and R version 2.9.1 (R. Development Core Team, 2003, www.R-project.org).

The genetic coefficient of variation ($CV_G$) was used (Cornelius, 1994) to compare the relative amounts of genetic variation of traits with different means: $CV_G = \sqrt{(\sigma^2_G / \mu)}$

***Molecular data and QTL detection***

Poplar genotyping was conducted using RFLP, STS, RAPD, and microsatellite markers in a first subset of 90 genotypes (Jorge *et al*., 2005). Microsatellite and AFLP genotyping was then extended to 253 supplementary genotypes (Jorge *et al*., unpublished). RFLP and STS genotyping was conducted as described by Bradshaw and Stettler (1993). RFLP markers were coded by the letter “P” followed by a 3- to 4-digit

number (*e.g.* P1273). RAPD genotyping was performed according to Villar *et al.* (1996). RAPD markers were coded by the Operon kit primer name followed by the molecular weight of the polymorphic band (*e.g.* M02-1150). AFLP genotyping was performed as described in Jorge *et al.* (2005). AFLP markers were named after the code of the EcoRI/MseI combination from the kit followed by the band ranking on the gel (e.g. E5M5-7). The SSR primers came from two different sources: (1) the International *Populus* Genome Consortium, SSR named "PMGC" (http://www.ornl.gov/sci/ipgc/ssr_resource.htm), "ORPM" (Tuskan *et al.*, 2004) and "WPMS" (Van der Schoot *et al.*, 2000) and (2) SSR named "ai", "bi" and "bu" (Jorge *et al.*, in prep.) developed from public EST databases. Different labeling techniques were successively used for SSR genotyping: primers were labeled with γ-[$^{33}$P] ATP, with forward fluorochrome-labelled primer or the M13 tailing strategy (Schuelke, 2000). The first two techniques were described elsewhere (Jorge *et al.*, 2005). The third one follows: PCR reaction were performed in a total volume of 10 µl containing 10 mM Tris-HCL, pH 8.3; 50 mM KCl; 1.5 mM $MgCl_2$; 200 µM each of dATP, dCTP, dGTP and dTTP; 0.2 unit *Taq* Polymerase (Invitrogen); 5 pmoles reverse primer; 0.5 pmoles M13-tailed forward primer; 5 pmoles M13-labelled primer; and 40 ng DNA. PCR were conducted with 42 cycles of a 30-s denaturation at 94°C, 30-s annealing at 55°C or 52°C and 30-s extension at 72°C. Each forward primer was 5'-tailed with the M13 forward consensus sequence. The M13-tailed forward primers were then used in combination with a standard M13 primer labelled with fluorescent dye (6FAM, HEX or NED) at its 5'-end. The amplicons for each SSR marker were separately produced, diluted and pooled post-PCR by three-color multiplexes (6FAM, HEX, and NED) for polymorphism screening. Microsatellite polymorphisms were visualized using an ABI PRISM 3100 genetic analyzer (PE Applied Biosystems, Foster City, Calif.). SSR allele lengths were recorded by GeneScan and Genotyper or Genemapper softwares.

Markers significantly deviating from Mendelian segregation ratios [1:1] (*i.e.* $P_{>\chi^2} < 1\%$) were eliminated from the linkage analysis. Linkage maps were constructed using Mapmaker version 3.0b (Lander & Botstein, 1989) and were based on the pseudo-testcross strategy, which led to the construction of two parental maps (Grattapaglia & Sederoff, 1994). Steps for map construction were the same than Jorge *et al.* (2005) except "Error detection" option was set up on.

***QTL analysis***

Data used for QTL detection were corrected of the observer effect using a logistic regression. QTLs were determined using MultiQTL 2.4 (http://www.multiqtl.com/; MultiQTL Ltd, Institute of Evolution, Haïfa University, Haïfa, Israel). The option 'marker restoration' was used to reduce the effect of missing information. The Kosambi mapping function was chosen for recalculation of maps on genotypic data. Single trait analysis was performed using a combination of interval mapping approach and multiple interval mapping. The entire genome was first scanned using the one QTL model and then using the two-linked QTL model. Permutation tests (1000 runs), comparing hypotheses $H_1$ (there is one QTL in the chromosome), and $H_0$ (no QTL in the chromosome) were run to obtain chromosome-wise statistical significance. In a second step, the genome was scanned for QTLs assuming a two-linked QTL model. For chromosomes for which a single QTL was already detected, permutation tests (1000 runs) were run to compare the hypotheses $H_2$ (two-linked QTLs in the chromosome) versus $H_1$. Subsequently, when $p_{(H2vsH1)}<0.05$, permutations were run to compare $H_2$ vs. $H_0$. A two-linked QTL model was only accepted, if the two intervals were not adjacent. To speed up calculation time, permutations for the two-linked QTL models were only conducted with 1000 runs when the p-value was <0.1 after 100 runs. In a last step, multiple interval mapping was performed including all the significant intervals mapping QTLs (single and 2 linked QTLs), and threshold was 0.05. For the remaining significant QTLs, permutations were run per chromosome, using thresholds (0.05) per chromosome. Bootstrap analysis was performed to estimate the 95% confidence interval. Two QTL analysis were performed, one with the genetic map v1 (*P. deltoides* and *P. trichocarpa* maps; Jorge *et al.*, 2005), and another one with the genetic map v2 (*P. deltoides* and *P. trichocarpa* maps; Jorge *et al.*, unpublished data).

***RNA extraction and cDNA synthesis***

Ectomycorrhizal root tips from *P. trichocarpa* and *P. deltoides* inoculated with *L. bicolor* S238N for three and half months in the greenhouse were harvested and snap frozen in liquid nitrogen. Total RNA was extracted from three biological replicates (3 x 2 ECMs) of each poplar genotype using the RNeasy micro kit (ref.74004, QIAGEN Courtaboeuf, France) following manufacturer's instructions. An in-column digestion step with DNAse I (ref.79254, QIAGEN Courtaboeuf, France) was included of the RNA purification. RNA quality was verified using Experion High sens Capillar gels (BIO-RAD, France). cDNAs for NimbleGen microarrays was synthesized using the

Super Smart cDNA Synthesis kit (ref.635000, Clontech-Takara Bio Saint-Germain-en-Laye, France).

***NimbleGen microarrays***

The *Populus* whole-genome expression array version 2.0 (S. DiFazio, A. Brunner, P. Dharmawardhana, and K. Munn, unpublished data) manufactured by NimbleGen Systems Limited (Madison, WI) contains in duplicates three independent, non-identical, 60-mer probes per whole gene model plus control probes and labeling controls. Included in the microarray are 65,965 probe sets corresponding to 55,970 gene models predicted on the *P. trichocarpa* genome sequence version 1.0 and 9,995 aspen cDNA sequences (*Populus tremula*, *Populus tremuloides*, and *P. tremula x P. tremuloides*). The Populus version 2.0 oligoarray is fully described in the platform Gene Expression Omnibus (GEO) at NCBI (http://www.ncbi.nlm.nih.gov/geo). NimbleGen whole genome microarray analyses were performed in triplicate as per manufacturer instructions. Expression data were processed in the following way: to assure a high specificity, all independent 60mer oligos for the 55,970 genes were blasted against the *Populus* genome v1.1 (Tuskan *et al*, 2006) available on http://genome.jgipsf.org/Poptr1_1/Poptr1_1.home.html and only probes with less than 10% homology to other gene models than the gene model they were designed for were retained for the further analysis. Due to this stringent filtering about 30% of the genes (16,667) were excluded since all three independent probes failed the given specifity. Fluorescence data were normalized between all different arrays using ARRAYSTAR software (DNASTAR, Madison, USA). Average expression levels from all three biological repetitions were calculated for each gene from the specific independent probes and were used for further analyses. A Student's T-Test with FDR (Benjamini-Hochberg) multiple testing corrections was applied on the data using ARRAYSTAR software (DNASTAR). Transcripts with a significant p-value ($<0.05$) $\geq$ 5.0-fold change in transcript level were considered as differentially expressed between *P. trichocarpa* ECM and *P. deltoides* ECM. The complete expression dataset is available as series (accession number……) at the Gene Expression Omnibus at NCBI (http://www.ncbi.nlm.nih.gov/geo/).

## Results

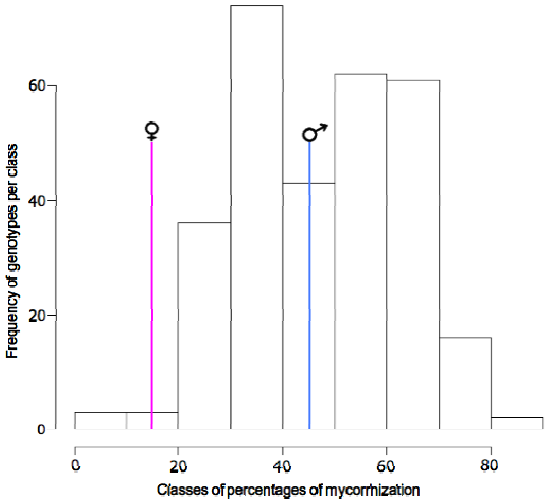
### *Mycorrhizal development*

Three and a half months after inoculation, root systems of the 300 progenies and two parental clones were exclusively colonized by the ectomycorrhiome *L. bicolor*. No contaminant ectomycorrhizal fungi such as *Thelephora terrestris*, were observed on the roots. Progenies and parental clones could be ranged following their degree of mycorrhizal infection (Figure 1). The two parents greatly differed in their mycorrhizal capacity-rate; *P. trichocarpa* exhibited a colonization rate of 45 % ± 2.4, whereas *P. deltoides* showed a rate of 15 % ± 3.2. The percentages of ECM colonization of the different genotypes varied from 8 % (hybrid clone N˚661300230) to 88 % (hybrid clone N˚661300592), with an average of 47 %. The genetic variation in the inter-specific family was assessed by a linear mixed model (Table 1). This variance analysis showed a significant genotype and observer effects. No block effect was detected. The broad sense heritability of the percentage of mycorrhizal infection was 0.40 ($CV_G$= 20.82 %) before correction of the observer effect while after correction it was 0.38 ($CV_G$= 22.94 %).

**Table 1 ANOVA for the percentage of mycorrhization (% Myc) in the interspecific family (54B) issued from a cross between *P. deltoides* and *P. trichocarpa*.**

| Trait | Effect | *df* | F value | P |
|---|---|---|---|---|
| % Myc | Genotype | 302 | 7.47 | < .0001 |
| | Block | 8 | 0.2726 | < 0.2732 |
| | Observer | 8 | 5.38 | < .0001 |

**Figure 1 Distribution of *Populus deltoides* (female) and *P. trichocarpa* (male) and their F1 progeny following their root colonization by *Laccaria bicolor***. The bars represent the number of genotypes distributed in each class according to their percentages of mycorhization. The percentage of *P. deltoides* is identified by the pink line while that of *P. trichocarpa* is identified by the blue line.



***QTL detection***

Before QTL detection, the observer effect was corrected using a logistic regression. On the genetic *P. deltoides* map v1 (Jorge *et al.*, 2005), a single putative QTL (Myc_d1, linkage group LGg_d), explaining 10.9 % of the variation of the mycorrhizal colonization, was detected (LOD score=3.59; Table 2a). This QTL was co-localized with three amplified fragment length polymorphism (AFLP) markers, which could not be anchored onto the poplar genomic sequence (Figure 2). On the genetic *P. trichocarpa* map v1 (Jorge *et al.*, 2005), three QTLs named Myc_t1, Myc_t2 and Myc_t3 respectively explaining 28.3 %, 5.2 %, 2.2 % of the mycorrhizal colonization trait were detected on the respective linkage groups LGi_t (LOD score=4.0), LGk_t (LOD score=1.3) and LGXI_t (LOD score=1.1) of the *P. trichocarpa* parent (Table 2a). No co-localization with components of partial resistance to *Melampsora larici-populina* was found (Tagu *et al.*, 2005). The QTL Myc_t1 was co-localized with two markers, an AFLP (E1M7.9) and a single nucleotide polymorphism (SNP) marker corresponding to a gene coding for the cinnamoyl-Coa reductase, CCR2 (ID: 172900). QTL Myc_t2 was co-localized with a microsatellite marker corresponding to a poplar genomic region of unknown function. QTL Myc_t3 was not co-localized with markers corresponding to a coding region of the poplar genome (Figure 2).

The use of an updated genetic map comprising additional markers (v2; Jorge *et al.*, unpublished data) allowed the positioning of these two Myc_t1 markers on *P. trichocarpa* LG XI of the *P. trichocarpa* and to anchor the Myc_t1 QTL onto the poplar genome using the microsatellites PMGC634 and PMGC2499, which frame E1M7.9 and CCR2 (Figure 2). Size of this QTL is 5.1 megabases (Mb).

The newly updated genetic maps (v2) also allowed the detection of three additional putative QTLs, named MycI_d, MycII_d and MycVI_d, respectively, explaining 2.3 %, 3.6 %, and 2.3 % of the variation of the mycorrhizal colonization and positioned on the linkage groups LG_Ia d, LG_II d and LG_VI d of *P. deltoides* parent (with LOD scores of 0.39, 0.118 and 0.085; Table 2b, Figure 3a, 3b and 3c). These three putative QTLs were co-localized with several microsatellite markers anchored to the genomic sequence (Tuskan *et al,* 2006; Figure 3a, 3b and 3c). MycI_d, MycII_d and MycVI_d respectively displayed a coverage of 17.2 (Mb) on the poplar chromosome I, 16.7 Mb on the chromosome II and 6.6 Mb on the chromosome III (Figure 3a, 3b and 3c).

Two other additional putative QTLs named MycIII_t and MycXI_t, respectively explaining 3 % and 1.8 % of the variation of the mycorrhizal development were detected on the respective linkage groups LG_IIIb t and LG_XIa t of the *P. trichocarpa* parent (Table 2b, Figure 3d). The putative QTL MycIII_t, which covered a region of 2.4 Mb on the chromosome III (Figure 3d), was co-localized with several microsatellite markers allowing its anchoring on the *Populus* genome v1.1(Tuskan *et al*, 2006). The QTL MycXI_t was co-localized with a SNP marker (SNPT1) and a microsatellite marker (PMGC2011) not positioned on a well defined region of the physical map (Figure 3e, Table 2b). No co-localization with components of partial resistance to *Melampsora larici-populina* was found.

***Analysis of microarrays transcript profiling***

We have identified 3679 transcripts corresponding to the putative QTLs Myc_t1 (264) MycI_d (1359), MycII_d (1528), MycVI_d (486) and MycIII_t (242). No significant transcripts differentially expressed are found for the QTL MycXI_t (Table 3g). The marker SNPT1 linked to this QTL belongs to a gene of unknown function, for which the expression is not significantly different. By considering a change in gene expression $\geq 5$ between *P. trichocarpa* ECM and *P. deltoides* ECM, we found 384 transcripts placed on the different QTL (Myc_t1: 27, MycI_d: 123, MycII_d: 154, MycVI_d: 51 and MycIII_d: 29) on a total of 8546 for the whole genome. In a second step, on the five QTLs, we have selected 81 transcripts displaying a significant p-value $<0.05$ and a gene expression $\geq 5$ between the mycorrhizas of the two parents (1453 for the whole genome; Table 3a, 3b, 3c, 3d, 3e, 3f and 3g). Four transcripts belonging to the putative detected QTLs were among the ten most up-regulated transcripts between *P. trichocarpa* ECM and *P. deltoides* ECM (Table 3a). On these 81 transcripts, 50 corresponded to known functions and 31 to unknown functions.

**Discussion**

The present study confirmed the previous results of Rosaldo *et al.* (1994) and Tagu *et al.* (2001 and 2005), who respectively shown that in *Pinus elliotii* and *P. trichocarpa* x *P. deltoides* hybrids the ability to form ectomycorrhizas was partly under the genetic control of the host.

On the same progeny from the 54B family (*P. deltoides* female clone from Illinois, no. 73028-62 and *P. trichocarpa* male clone from Washington, no. 101-74) but with less

clones than in this study (146 against 300), Tagu *et al.* (2001-2005) found that the heritability of the mycorrhizal development was comprised between 0.49 (Tagu et al., 2001) and 0.09 (Tagu et al. 2005). With a larger progeny (300 clones), our statistical analyses also have shown a significant genotype effect on ectomycorrhizal formation. The broad sense heritability of the percentage of mycorrhizal infection was 0.38 ($CV_G$= 22.94 %), indicating moderate genetic control of the trait and a large involvement of environmental factors. As in our previous studies (Tagu *et al.* 2001, 2005), we found a large difference in the response of the two parents to *L. bicolor* colonization (45% for *P. trichocarpa* and 15% for *P. deltoides*). The mean of the progeny (47 % against 35% in Tagu *et al.*, 2005) was not significantly different from that of the *P. trichocarpa* parent (45 %). We could hypothesize that the genes involved in this trait would be dominant and inherited from the *P. trichocarpa* parent. Some clones, such as 661300592, showed a higher percentage of mycorrhizal infection than that of their parents. This result is probably connected with the large genetic distance existing between the two parents, which could induce the formation of heterozygotes exhibiting heterosis for this trait. This fact suggests that the ability to form ectomycorrhizas in poplar could be improved by breeding.

In the present study, an initial QTL analysis (carried out using genetic maps v1) identified four QTLs, one on the *P. deltoides* map and three on the *P. trichocarpa* map. The putative QTLs positioned on the v1 genetic map of *P. deltoides* could not be anchored onto the *P. trichocarpa* genome sequence (Tuskan *et al*, 2006). From the three QTLs positioned on the genetic map of *P. trichocarpa*, two could be anchored onto the genome assembly. The QTL Myc_t2 was anchored in a region containing only genes of unknown function, while the QTL Myc_t1, explaining 28.3 % of the ectomycorrhizal infection, was linked to a SNP marker located in the CCR2 gene-coding region (ID: 172900). The existence of such a QTL reinforce the assumption of the dominance of the trait inherited from the *P. trichocarpa* parent.

No significant QTL was detected with the updated mapping data, while by using the same phenotypic data set we have found four significant QTLs as reported above. The updating of the genetic map induced undesirable effects. Indeed, to integrate the new markers, it was necessary to remove several pairs of markers linked to the four QTLs of the v1 map (Jorge *et al.*, 2005). As long as the genetic map of *P. deltoides* and *P. trichocarpa* will not be saturated and stabilized, there will remain difficult to definitively validate the QTLs.

Despite their lack of significance, thanks to microsatellite markers, three of these QTLs were positioned on the genetic map of *P. deltoides* (MycI_d, MycII_d, MycVI_d) and two on the genetic map of *P. trichocarpa* (MycIII_t and MycXI_t). It was possible to anchor the QTL MycXI_t onto the *P. trichocarpa* genome, but not to delimit it on the chromosome XI due to the fact that the corresponding region is still split into several scaffolds. However, MycXI_t is linked to a SNP marker (SNPT1) located in a gene of unknown function.

The corresponding genomic regions to QTL spread through several megabases. In order to determine the candidate genes which could play a role in the mycorrhizal development, we have selected genes differentially expressed in between *P. trichocarpa* and *P. deltoides* ECM and located in the QTL genomic regions. Among the 1543 differentially expressed genes (p-value <0.05 and fold-change ≥ 5 in transcript level) between the mycorrhizas of the two parents, 81 transcripts were located in the QTL genomic regions: 14 in Myc_t1, 7 in MycIII_t, 30 in MycI_d, 20 in MycII-d, and 10 in MycVI_d.

Among these 81 transcripts, 52 were up-regulated in *P. deltoides* and 29 were up-regulated in *P. trichocarpa*. We can hypothesize that the genes whose transcripts are up-regulated in *P. trichocarpa* play a positive role in ectomycorrhizal development, while the genes whose transcripts are up-regulated in *P. deltoides* play a negative role. The most up-regulated gene in *P. trichocarpa* mycorhizas compared to *P. deltoides* mycorrhizas belongs to the QTL MycIII_t on the chromosome III and codes for a protein EREBP-4. The proteins of this class seem to be important ethylene-responsive binding protein transcription factor (Leubner-Metzger *et al.*, 1998, Solano *et al.*, 1998; Ohta *et al.*, 2000). Hirota *et al.* (2002) shown that the auxin-regulated AP2/EREBP *PUCHI* gene, encoding a putative APETALA2/ethylene-responsive element binding protein transcription factor, is required for the coordinated pattern of cell divisions during lateral root formation in *Arabidopsis thaliana*. It is well known that phytohormones are involved in ectomycorrhizal development (Slankis, 1950, 1973; Gay *et al.* 1994; Karabaghli-Degron *et al.*, 1998; Barker and Tagu, 1998; Tagu *et al*., 2003). For example, Rupp and Mudge (1995) showed that ethylene and auxin induce mycorrhiza-like roots on *Pinus mugo*. In the *Eucalyptus globulus-Pisolithus microcarpus* ectomycorrhiza, Tagu *et al*. (2003) showed that many plant genes are up-regulated during the mycorrhizal development and that the expression of the *EgHypar* gene, exhibiting homology with plant auxin-induced glutathione-*S*-transferases (GST)

was up-regulated not only in the ectomycorrhiza, but also by exogenous application of auxin and hypaphorine. The up-regulation of this gene, which codes for a protein EREBP-4, fits perfectly well with the previous knowledge on the involvement of hormonal interactions in ectomycorhizal establishment. We are conscious that this kind of study presents severe limitations. Indeed, the QTL approach requires a stable and saturated genetic map. In the future, we hope to be able to use better genetic maps allowing to definitively localize significant QTLs involved in mycorrhizal establishment.

On the another hand, the comparison of the gene expression between the mycorrhizas of the two parents three or four months after fungal inoculation probably integrates only a part of the interactions which took place at the moment of mycorrhizal establishment (early recognition with exchange of signals, attachment of the mycelium on the root, internal colonization). Moreover several months after inoculation, the comparison of the mycorrhizas of the two parents also integrates processes related to the functioning of the association.

Nevertheless, despite the imperfections of the two approaches, the combination of both provides an exhaustive list of genes, which could be involved into mycorrhizal establishment or functionning and which could be further confirmed or infirmed. The interactions involved in mycorrhizal establishment and functioning seem extremely complex: For the whole genome, we have identified 8546 transcripts significantly and differentially expressed between *P. trichocarpa* and *P. deltoides* ECMs, including 3679 localized into the putative QTLs Myc_t1 (264), MycI_d (1359), MycII_d (1528), MycVI_d (486) and MycIII_t (242). By considering a change in transcript level more than 5, in the different QTLs, there were still 81 transcripts displaying a significant p-value (<0.05) between the mycorrhizas of the two parents (1453 for the whole genome).

**Acknowledgments**

This project was supported by grants from the European Commission project ENERGYPOPLAR and EVOLTREE (to F.M). JL was supported by a scholarship from the INRA and Région Lorraine. We thank Christine Delaruelle, Béatrice Palin, Judith Richter, Verónica Pereda, Saskia Reinhart, Anne Delaruelle, Simon Duchêne and Jean-Louis Churin for assistance in inoculation and in measuring the colonization rates. We also thank Denis Tagu for his help and fruitful discussions.

**References**


Barker SJ, Tagu D, Delp G. 1998. Regulation of root and fungal morphogenesis in mycorrhizal symbiosis. *Plant Physiology* 116: 1201-1207.

Bradshaw HD, Stettler RF. 1993. Molecular genetics of growth and development in *Populus*. I. Triploidy in hybrid poplars. *Theor Appl Genet* 86: 301 –307

Cervera MT, Storme V, Ivens B, Gusmao J, Liu BH, Hostyn V, VanSlycken J, Van Montagu M, Boerjan W. 2001. Dense genetic linkage maps of three *Populus* species (*Populus deltoides, P. nigra, P. trochocarpa*) based on AFLP and microsatellite markers. *Genetics* 158:787-809.

Cornelius J. 1994. Heritabilities and additive genetic coefficient of variation in forest trees. *Can J For Res* 24:372-379.

Di Battista C, Selosse MA, Bouchard D, Stenström E, Le Tacon F. 1996. Variations in symbiotic efficiency, phenotypic characters and ploidy level among different isolates of the ectomycorrhizal basidiomycete *Laccaria bicolor* strain S238. *Mycological Research* 100: 1315-1324.

Duponnois R, Garbaye J. 1991. Mycorrhization helper bacteria associated with Douglas fir-*Laccaria laccata* symbiosis: effects in aseptic and glasshouse conditions. *Ann. Sc. For.* 48: 239-251.

Faivre-Rampant P, Bastien C, Augustin S, Breton V, Delplanque A, Mourier MC, Kertadikara A, Laurans F, Lefèvre F, Lesage MC, Menard M, Pinon J, Saintagne C, Valadon A, Villar M, Prat D. 1999. Locating genomic regions involved in pest resistance in poplars. Proceedings of the International Poplar Symposium II, IUFRO, Orléans, France. *IUFRO,* Orléans p31.

Frey-Klett P, Pierrat JC, Garbaye J. 1997. Location and survival of mycorrhiza helper *Pseudomonas fluorescens* during establishment ofectomycorrhizal symbiosis between *Laccaria bicolor* and Douglas fir. *Appl Environ Microbiol* 63:139-144.

Gay G, Normand L, Marmeisse R, Sotta B, Debaud JC. 1994. Auxin overproducer mutants of *Hebeloma cylindrosporum* Romagnesi have increased mycorrhizal activity. *New Phytologist* 128: 645±657.

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.

Hirota A, Kato T, Fukaki H, Aida M and Tasaka M. 2002. The Auxin-Regulated AP2/EREBP Gene PUCHI Is Required for Morphogenesis in the Early Lateral Root Primordium of Arabidopsis. *The Plant Cell* 19: 2156–2168.

Jorge V, Dowkiw A, Faivre-Rampant P and Bastien C. 2005. Genetic architecture of qualitative and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar: geneticmapping and QTL detection. *New Phytol* 167: 113-127.

Karabaghli-Degron C , Sotta B, Bonnet M , Gay G & Le Tacon F. The auxin transport inhibitor 2, 3, 5-triiodobenzoic acid (TIBA) inhibits the stimulation of in vitro lateral root formation and the colonization of the tap-root cortex of Norway spruce (*Picea abies*) seedlings by the ectomycorrhizal fungus Laccaria bicolor. *New Phytologist* 140 (4): 723-733.

Lander ES, Botstein D. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185–199.

Lefèvre F, Goué-Mounier Mc, Faivre-Rampant P, Villar M. 1998. A single gene cluster controls incompatibility and partial resistance to various *Melampsora larici-populina* racesin hybrid poplars. *Phytopathology* 88: 156-163.

Leubner-Metzger G, Petruzzelli L, Waldvogel R, Vögeli-Lange R and Meins F. 1998. Ethylene-responsive element binding protein (EREBP) expression and the transcriptional regulation of class I-1, 3-glucanase during tobacco seed germination. *Plant Molecular Biology* **38:** 785–795.

Martin F. 2001. Frontiers in molecular mycorrhizal research genes, loci dots and spins. *New Phytologist* 150: 499-507.

Martin F, Duplessis S, Ditengou F, Lagrange H, Voiblet C, Lapeyrie F. 2003. Developmental cross talking in the ectomycorrhizal symbiosis: signals and communication genes. *New Phytologist* 151: 145-154.

Martin F, Lammers P, Tuskan GA, DiFazio SP, Podila GK,. 2004. Mycorrhizal symbionts of *Populus* to be sequenced by the United States Department of Energy's Joint Genome Institute. *Mycorrhiza* 14: 63-64.

Martin F, Kohler A & Dupessis S. 2007. Living in harmony in the wood underground: ectomycorrhizal genomics. *Current Opinion in Plant Biology* 10:204–210.

Martin F, Aerts A, Ahrén D, Brun A, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V, Salamov A, Shapiro HJ, Wuyts J, Blaudez D, Buée M, Brokstein P, Canbäck B, Cohen D, Courty PE, Coutinho PM, Danchin EGJ, Delaruelle C, Detter JC, Deveau A, DiFazio S, Duplessis S, Fraissinet-Tachet L, Lucic E, Frey-Klett P,

Fourrey C, Feussner I, Gay G, Grimwood J, Hoegger PJ, Jain P, Kilaru S, Labbé J, Lin YC, Legué V, Le Tacon F, Marmeisse R, Melayah D, Montanini B, Muratet M, Nehls U, Niculita-Hirzel H, Oudot-Le Secq MP, Peter M, Quesneville H, Rajashekar B, Reich M, Rouhier N, Schmutz J, Yin T, Chalot M, Henrissat B, Kües U, Lucas S, Van de Peer Y, Podila G, Polle A, Pukkila PJ, Richardson PM, Rouzé P, Sanders IR, Stajich JE, Tunlid A, Tuskan G, Grigoriev IV. 2008. Symbiosis insights from the genome of the mycorrhizal basidiomycete *Laccaria bicolor*. *Nature* 452: 88-92.

Martin F & Nehls U. 2009. Harnessing ectomycorrhizal genomics for ecological insights. *Current Opinion in Plant Biology* 12: 508–515.

Nehls U, Mikolajewski S, Magel E, Hampp R. 2001. Carbohydrate metabolism in ectomycorrhizas: gene expression, monosaccharide transport and metabolic control. *New Phytologist* 150: 533-541.

Ohta M, Ohme-Takagi M and Shinshi H. 2000. Three ethylene-responsive transcription factors in tobacco with distinct transactivation functions. *The Plant Journal* 22: 29-38.

Pachlewski R, Pchlevska J. 1974. Studies on symbiotic properties of mycorrhizal fungi of pine (*Pinus sylvestris* L.) with theaid of the method of mycorrhizal synthesis in pure cultures on agar. *PhD Thesis*, Forest Institute, Warsaw.

Peter M, Courty PE, Kohler A, Delaruelle C, Martin D, Tagu D, Frey-Klett P, Duplessis S, Chalot M, Podila G, Martin F. 2003. Analysis of expressed sequence tags from the ectomycorrhizal basidiomycete *Laccaria bicolor* and *Pisolithus microcarpus*. *New Phytologist* 159: 117-129.

Peterson L, Bonfante P. 1994. Comparative structure of vesicular-arbuscularmycorrhizas and ectomycorrhizas. In: Robson AD, Abbort LK, Malajczuk N (eds) Management of mycorhizas in agriculture, horticulture and forestery. Kluwer, Dordrecht, pp 79-88.

Rosado SCS, Kropp BR, Piché Y. 1994. Genetics of ectomycorrhizal symbiosis. I. Host plant variability and heritability of ectomycorrhizal root traits. *New Phytol* 126: 105-110.

Rupp L.A. and W. Mudge. 1985. Ethephon and auxin induce mycorrhiza-like changes in the morphology of root organ cultures of mugo pine. *Physiol. Plant* 64: 316-322.

Schuelke, M. 2000. An economic method for the fluorescent labelling of PCR fragments.

*Nature Biotechnology* 18: 233-234.

Solano R and Ecker JR. 1998.Ethylene gas: perception, signaling and response. *Current Opinion in Plant Biology* 1: 393-398.

Smith SE, Reads DJ (eds). 1997. Mycorrhizal symbiosis. *Academic Press,* San diego.

Slankis V. 1950. Effect of α-naphthalene-acetic acid on dichotomous branching of isolated roots of *Pinus silvestris*. *Physiologia Plantarum* 3: 40-44.

Tagu D, Faivre-Rampant P, Lapeyrie, Frey-klett P, Vion P, Villar M. 2001. Variation in the ability to form ectomycorrhizas in the F1 progeny of an interspecific poplar (*Populus spp.*) cross. *Mycorrhiza* 10: 237-240.

Tagu D, Bastien c, Faivre-Rampant P, Garbaye J, Vion P, Villar M, Martin F. 2005. Genetic analysis of phenotypic variation for ectomycorrhiza formation an interspecific F1 poplar full-sib family. *Mycorrhiza* 15: 87-91.

Trappe JM. 1977. Selection of fungi for ectomycorrhizal inoculation in nurseries. *Annu Rev Phytopathol* 15: 203-222.

Tuskan GA, Gunter LE, Yang ZMK, Yin TM, Sewell MM, DiFazio SP. 2004. Characterization of microsatellites revealed by genomic sequencing of *Populus trichocarpa*. *Canadian Journal of Forest Research* **34**: 85–93.

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, dePamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D. 2006. The genome of black cottonwood, *Populus trichocarpa*. *Science* 313: 1596-1604.

Van der Schoot J, Pospiskova M, Vosman B, Smulders MJM. 2000. Development and characterization of microsatellite markers in black poplar (*Populus nigra L.*). *Theor Appl Genet* 101: 317–322.

Villar M, Lefèvre F, Bradshaw HD, Teissier du Cros E. 1996. Molecular genetics of rust resistance in poplars (*Melampsora larici-populina Kleb*/*Populus sp*.) by bulked segregant analysis in a 2 x 2 factorial mating design. *Genetics* 143: 531-536.

**Table 2 QTLs detected by composite interval mapping in the full-sib *Populus deltoides* x *Populus trichocarpa* $F_1$ progeny for mycorrhizal infection evaluated in green house conditions.**

[a] *Populus deltoides* (d) and *Populus trichocarpa* (t) linkage groups
[b] Nearest marker anchored to the *Populus* genome from the maximum LOD value
[c] $P < 0.05$
[d] Maximum LOD (logarithm of odd ratio)
[e] In percentage of mycorrhizal infection
[f] Proportion of explained variability by the putative QTL

**Table 2a** QTL detection on the genetic linkage map v1 (Jorge *et al.*, 2005)

| Trait | LG[a] | Marker[b] | Significance[c] (on 1000 runs) | Position (cM) | LOD[d] | Mean[e] | PEV[f] |
|---|---|---|---|---|---|---|---|
| Myc_d1 | LGg_d | | 0.018 | 9.079 | 1.393 | 43.31 | 0.109 |
| | | M10.1600 | | 0 | 1.13 | | |
| | | G14.1050 | | 3 | 1.13 | | |
| | | G9.1600 | | 6.1 | 1.13 | | |
| Myc_t1 | LGi_t | | 0.002 | 50.82 | 4 | 45.95 | 0.283 |
| | | CCR2 | | 0 | 0,8 | | |
| | | E1M7.9 | | 19.2 | 1.6 | | |
| | | P1312 | | 19.4 | 2.4 | | |
| Myc_t2 | LGk_t | | 0.024 | 0 | 1.338 | 45.91 | 0.052 |
| | | E3M1.2* | | 0 | 1.34 | | |
| | | E3M1.3 | | 14.5 | 1.33 | | |
| Myc_t3 | LGXI_t | | 0.048 | 24.7 | 1.108 | 48.14 | 0.022 |
| | | P219 | | 0 | 0.537 | | |
| | | K19.550 | | 14.9 | 0.559 | | |
| | | PMGC333 | | 15 | 0.582 | | |

*distorted ($\chi^2 = 0.019$, segregation 57:85)

**Table 2b** QTL detection on the genetic linkage map v1 (Jorge *et al.*, unpublished data)

| Trait | LG[a] | Marker[b] | Significance[c] (on 1000 runs) | Position (cM) | LOD[d] | Mean[e] | PEV[f] |
|---|---|---|---|---|---|---|---|
| MycI_d | LG_Ia d | | 0.39 | 342.92 | 0.958 | 47.63 | 0.023 |
| | | PMGC2098 | | 178.5 | 0.0226 | | |
| | | bu831219 | | 278.4 | 0.00781 | | |
| | | PMGC2852 | | 289.9 | 0.0044 | | |
| MycII_d | LG_II d | | 0.116 | 54.93 | 1.366 | 47.75 | 0.036 |
| | | PMGC667 | | 35.3 | 0.412 | | |
| | | PMGC2088.1 | | 193.9 | 0.35 | | |
| | | PMGC2418 | | 219.3 | 0.338 | | |
| MycVI_d | LG_VI d | | 0.085 | 142.49 | 1.471 | 47.87 | 0.023 |
| | | PMGC2578 | | 92 | 0.14 | | |
| | | ORPM190 | | 171.3 | 0.284 | | |
| | | PMGC2328 | | 207.3 | 0.31 | | |
| MycIII_t | LG_IIIb t | | 0.152 | 51.17 | 1.088 | 47.48 | 0.03 |
| | | PMGC486 | | 18.3 | 0.0684 | | |
| | | PMGC2274 | | 45.1 | 0.0366 | | |
| MycXI_t | LG_XIa t | | 0.160 | 160.17 | 1.118 | 47.69 | 0.018 |
| | | PMGC2011 | | 129 | 0.00568 | | |
| | | SNPT1 | | 160.3 | 0.0131 | | |

**Table 3 Microarrays transcript profiling corresponding to the regions of putative detected QTLs.** Only transcripts with a significant p-value (<0.05) and ≥ 5.0 change in transcript level were considered as significantly differentially expressed between *P. trichocarpa* ECMs and *P. deltoides* ECM.

**Table 3a** Transcripts corresponding to the most up- and down-regulated genes, in the whole genome. Those corresponding to the detected QTLs appear in green and in pink.

| Top of up/down regulated gene | Protein ID | Expression ratio del/trich | Expression ratio trich/del | P value | Putative function |
|---|---|---|---|---|---|
| 1 | 8383 | 86,6 | 0,0 | 0,0011 | epoxide hydrolase |
| 2 | 412601 | 59,0 | 0,0 | 0,00424 | epoxide hydrolase |
| 3 | 830358 | 58,2 | 0,0 | 0,0000936 | putative flavonol reductase |
| 4 | 755638 | 54,8 | 0,0 | 0,000929 | glycosyl hydrolase family 3 |
| 5 | 753870 | 54,4 | 0,0 | 0,0000115 | Ubiquitin ligase |
| 6 | 569526 | 53,3 | 0,0 | 0,00000124 | Unknown |
| 7 | 654260 | 53,0 | 0,0 | 0,000641 | Steroid binding protein |
| 8 | 679320 | 50,7 | 0,0 | 0,000000548 | Unknown |
| 9 | 769294 | 50,6 | 0,0 | 0,0102 | Plant lipid transfer protein |
| 10 | 772380 | 49,2 | 0,0 | 0,00000123 | Ubiquitin ligase |
| 11 | 755637 | 45,1 | 0,0 | 0,00222 | Glycoside hydrolase |
| 12 | 271871 | 45,1 | 0,0 | 0,000891 | Plant lipid transfer protein |
| 13 | 774414 | 44,4 | 0,0 | 0,0000331 | UDP-glucosyltransferase |
| 14 | 423586 | 44,2 | 0,0 | 0,000321 | Expansin |
| 15 | 594950 | 44,1 | 0,0 | 0,00000997 | alcohol dehydrogenase |
| 16 | 201272 | 42,8 | 0,0 | 0,00111 | isoflavone reductase |
| 17 | 595584 | 42,2 | 0,0 | 0,000157 | disease resistance protein RPM1 |
| 18 | 595584 | 40,4 | 0,0 | 0,00000089 | acyltransferase |
| 19 | 416408 | 39,7 | 0,0 | 0,0000221 | nodulin-like protein |
| 20 | 766543 | 38,3 | 0,0 | 0,000477 | Leucine rich repeat |
| 21 | 580472 | 35,1 | 0,0 | 0,000101 | Unknown |
| 1 | 259627 | 0,0 | 321,7 | 0,0000994 | plastocyanin |
| 2 | 767919 | 0,0 | 262,1 | 0,000000399 | Cinnamoyl Alcohol Dehydrogenase |
| 3 | 788513 | 0,0 | 190,5 | 0,00000041 | Unknown |
| 4 | 649437 | 0,0 | 134,5 | 0,00000239 | Unknown |
| 5 | 669706 | 0,0 | 113,5 | 0,00000132 | Unknown |
| 6 | 675468 | 0,0 | 113,0 | 0,00000306 | Unknown |
| 7 | 800187 | 0,0 | 97,5 | 0,00000658 | protein EREBP-4 |
| 8 | 829444 | 0,0 | 86,8 | 0,0092 | putative flavonol reductase |
| 9 | 670561 | 0,0 | 82,2 | 0,000000977 | Unknown |
| 10 | 278262 | 0,0 | 81,2 | 0,00141 | methyltransferase |
| 11 | 727226 | 0,0 | 81,0 | 0,000000498 | Auxin Efflux Carrier |
| 12 | 579578 | 0,0 | 78,6 | 0,0126 | Serine/threonine protein kinase |
| 13 | 253773 | 0,0 | 71,1 | 0,00326 | cyclin |
| 14 | 561449 | 0,0 | 70,2 | 0,0259 | Glucose/ribitol dehydrogenase |
| 15 | 751341 | 0,0 | 69,3 | 0,000192 | Unknown |

**Table 3b** Transcripts corresponding to the QTL Myc_t1 (on LGi_t map v1, LG_Ib t map v2)

| Protein ID | Expression ratio delt/trich | Expression ratio trich/delt | P value | Putative function |
|---|---|---|---|---|
| 548125 | 23.9 | 0.04 | 0.00028 | potassium transporter AtKT1p |
| 750943 | 17.0 | 0.05 | 0.000333 | gag-pol polyprotein |
| 750935 | 8.31 | 0.12 | 0.000259 | Unknown |
| 640058 | 8.08 | 0.12 | 0.000219 | WRKY-DNA binding protein |
| 750915 | 6.80 | 0.14 | 0.0471 | Unknown |
| 548126 | 6.60 | 0.15 | 0.0278 | potassium transporter AtKT1p |
| 179061 | 6.3 | 0.15 | 0.00257 | cytochrome p450 |
| 179380 | 5.24 | 0.19 | 0.000351 | FKBP-type isomerase |
| 639958 | 0.02 | 43.8 | 3.98E-07 | Unknown |
| 797239 | 0.06 | 14.2 | 0.00121 | Lipolytic Enzyme |
| 179173 | 0.08 | 11.8 | 0.0246 | MADS-box protein AGL3 |
| 814871 | 0.15 | 6.56 | 0.00785 | chalcone synthase |
| 708629 | 0.17 | 5.57 | 0.00761 | putative phloem-specific lectin |
| 640148 | 0.18 | 5.38 | 0.00146 | Unknown |

**Table 3c** Transcripts corresponding to the QTL MycI d (on LG_Ia d)

| Protein ID | Expression ratio delt/trich | Expression ratio trich/delt | P value | Putative function |
|---|---|---|---|---|
| 641531 | 17.1 | 0.05 | 0.00725 | glucosyltransferase |
| 752790 | 16.8 | 0.05 | 0.00178 | ABA-responsive protein |
| 174800 | 16.7 | 0.05 | 0.000147 | putative phosphate transporter |
| 752836 | 16.1 | 0.06 | 0.000339 | F-box protein |
| 549720 | 15.6 | 0.06 | 0.0016 | Unknown |
| 549955 | 12.2 | 0.08 | 0.00582 | Unknown |
| 752786 | 11.8 | 0.08 | 0.00444 | Unknown |
| 174968 | 9.36 | 0.10 | 0.00152 | dehydrodolichyl diphosphate synthase |
| 706543 | 9.07 | 0.11 | 0.00123 | wall-associated kinase 2 |
| 641991 | 9.03 | 0.11 | 0.0017 | Unknown |
| 171534 | 8.03 | 0.12 | 0.000872 | protein kinase |
| 174544 | 7.46 | 0.13 | 0.00651 | MADS-box protein |
| 751815 | 7.29 | 0.13 | 0.0000522 | peptide transporter PTR2-B |
| 751865 | 7.17 | 0.13 | 0.0309 | Unknown |
| 174972 | 7.10 | 0.14 | 0.0135 | UDP-glucose glucosyltransferase |
| 549334 | 6.61 | 0.15 | 0.00059 | cinnamyl-alcohol dehydrogenase |
| 751814 | 6.03 | 0.16 | 0.000189 | peptide transporter PTR2-B SP |
| 174671 | 5.89 | 0.16 | 0.0395 | Unknown |
| 829783 | 5.64 | 0.17 | 0.0566 | Enoyl-Coa hydratase |
| 172616 | 5.40 | 0.18 | 0.00242 | GRAS transcription factor |
| 172038 | 5.05 | 0.19 | 0.00273 | membrane protein |
| 549698 | 0.03 | 26.5 | 0.0000719 | Unknown |
| 549181 | 0.05 | 18.4 | 0.000185 | Unknown |
| 172100 | 0.06 | 14.9 | 0.0000291 | Multicopper oxydase |
| 640672 | 0.08 | 12.3 | 0.0000256 | Unknown |
| 549673 | 0.09 | 10.7 | 0.0000362 | Unknown |
| 549969 | 0.09 | 10.3 | 0.000209 | Anaphylatoxin/fibulin |
| 174093 | 0.10 | 9.86 | 0.000787 | Chitinase |
| 815170 | 0.11 | 8.69 | 0.000335 | Chitinase 1 precursor |
| 705727 | 0.17 | 5.55 | 0.000532 | Ribosomal protein L37e |

**Table 3d** Transcripts corresponding to the QTL MycII d (on LG_II d)

| Protein ID | Expression ratio delt/trich | Expression ratio trich/delt | P value | Putative function |
|---|---|---|---|---|
| 412601 | 59.0 | 0.01 | 0.00424 | epoxide hydrolases |
| 830358 | 58.2 | 0.01 | 0.0000936 | flavonol reductase |
| 755638 | 54.8 | 0.01 | 0.000929 | glycosyl hydrolase family 3 |
| 552266 | 16.7 | 0.05 | 0.000535 | Ion transport protein ($K^+$) |
| 830240 | 13.1 | 0.07 | 0.00000295 | Multi antimicrobial extrusion protein MatE |
| 754447 | 11.7 | 0.08 | 0.00696 | Unknown |
| 409930 | 11.1 | 0.08 | 0.00000811 | NADH dehydrogenase |
| 754777 | 9.33 | 0.10 | 0.00008 | PPR repeat |
| 644137 | 8.67 | 0.11 | 0.0000591 | Cinamoyl-Coa reductase |
| 411764 | 7.69 | 0.12 | 0.000121 | Calcium-binding EF-hand |
| 816402 | 7.28 | 0.13 | 0.00144 | phosphoenolpyruvate carboxykinase |
| 754607 | 7.08 | 0.14 | 0.00553 | GRAS transcription factor |
| 644752 | 6.06 | 0.16 | 0.00811 | Unknown |
| 755810 | 0.03 | 32.5 | 0.0000103 | Unknown |
| 551488 | 0.08 | 12.0 | 0.0000149 | flavonol synthase (FLS) |
| 412508 | 0.08 | 11.2 | 0.000042 | UDP-glucose glucosyltransferase |
| 552582 | 0.09 | 10.8 | 0.0000349 | Serine/threonine phosphatase 2C (PP2C) |
| 816547 | 0.13 | 7.4 | 0.00000997 | t-complex polypeptide 1 |
| 411729 | 0.14 | 6.75 | 0.0000719 | riboflavin biosynthesis protein ribG |
| 643641 | 0.15 | 6.58 | 0.0000675 | transcription factor with AP2 domain |

**Table 3e** Transcripts corresponding to the QTL MycVI d (on LG_VI d)

| Protein ID | Expression ratio delt/trich | Expression ratio trich/delt | P value | Putative function |
|---|---|---|---|---|
| 416408 | 39.7 | 0.02 | 0.0000221 | nodulin-like protein |
| 762951 | 22.2 | 0.04 | 0.0111 | protein kinase |
| 819386 | 8.98 | 0.11 | 0.000868 | Adenine-specific DNA methylase |
| 561528 | 7.26 | 0.13 | 0.00673 | Unknown |
| 653743 | 6.84 | 0.14 | 0.000219 | Unknown |
| 416253 | 6.10 | 0.16 | 0.0117 | disease resistance response protein |
| 561474 | 0.12 | 7.99 | 0.000061 | Unknown |
| 561172 | 0.14 | 6.89 | 0.000534 | calcium-dependent protein kinase |
| 417451 | 0.14 | 6.88 | 0.00065 | NAC-domain protein |
| 561340 | 0.14 | 6.82 | 0.0000332 | Unknown |

**Table 3f** Transcripts corresponding to the QTL MycIII_t (on LG_IIIb t)

| Protein ID | Expression ratio delt/trich | Expression ratio trich/delt | P value | Putative function |
|---|---|---|---|---|
| 413017 | 21.6 | 0.0 | 0.000655 | peptide transporter |
| 830999 | 8.17 | 0.12 | 0.00013 | Unknown |
| 758242 | 7.22 | 0.13 | 0.000139 | Methionine lyase |
| 647351 | 5.59 | 0.17 | 0.000158 | PPR repeat |
| 800187 | 0.01 | 97.5 | 0.00000658 | protein EREBP-4 |
| 758186 | 0.07 | 13.9 | 0.0131 | steroid sulfotransferase |
| 555048 | 0.12 | 7.82 | 0.00152 | serine threonine-specific kinase |

**Table 3g** Transcripts corresponding to the QTL MycXI_t (on LG_Xia t)

| Protein ID | Expression ratio delt/trich | Expression ratio trich/delt | P value | Putative function |
|---|---|---|---|---|
| 234401 | 1.20 | 0.83 | 0.766 | Unknown |

**Figure 2 Graphical representation of QTL detected by permutation test on the corresponding linkage group of the map v1 (Jorge *et al.*, 2005), correspondence on the updated map when this is possible and corresponding transcripts, up- (red) and down- (green) regulated identified**. The genetic distances are given in centiMorgan (cM); markers in red are linked to the corresponding QTL; markers in pink allow the anchoring to the *Populus* genome v1.1.

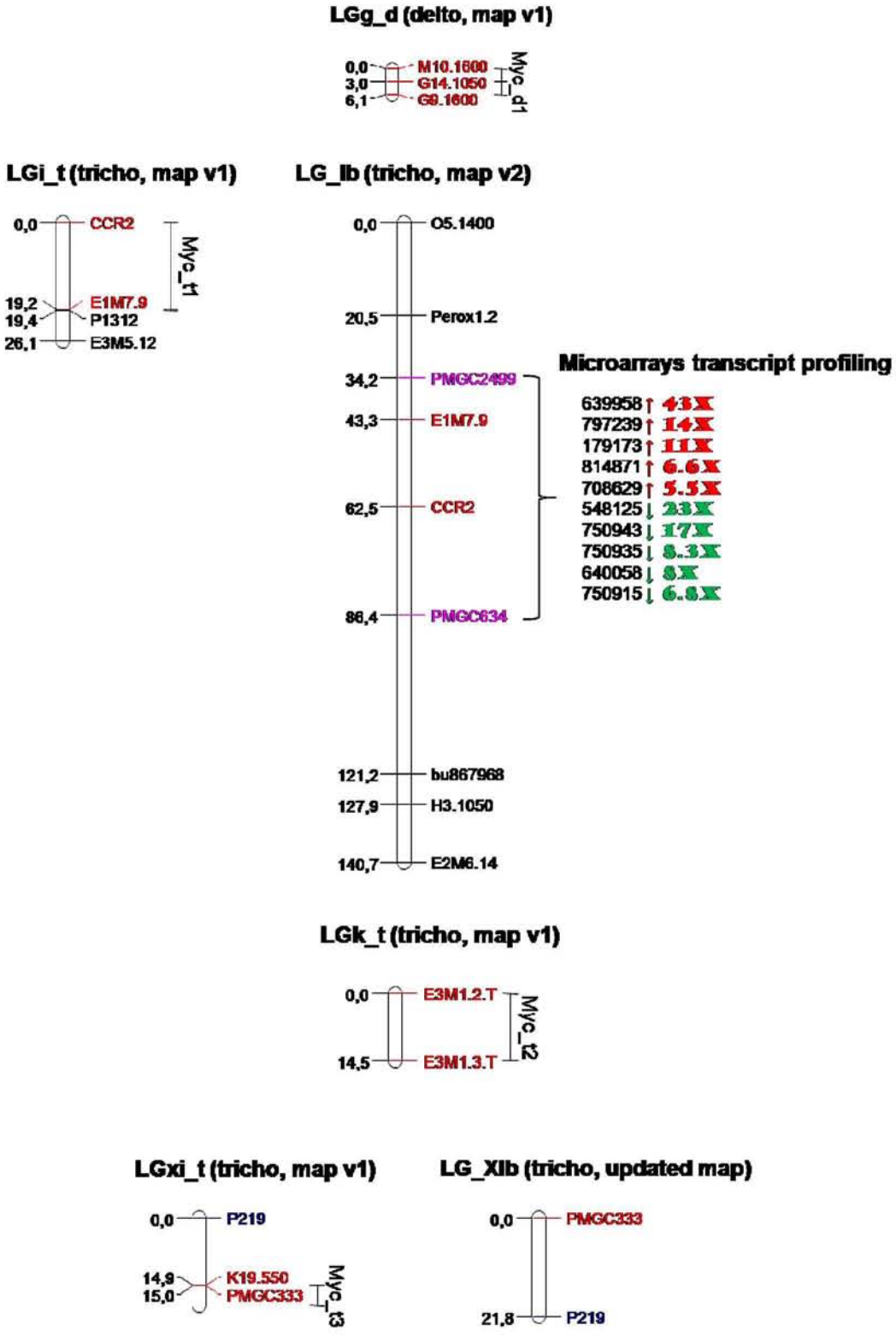


**Figure 3 Graphical representation of QTL detected by permutation test on the corresponding linkage group, anchoring on the *Populus* genome v1.1 and corresponding transcripts, up- (red) and down- (green) regulated identified**. The genetic distances are given in centiMorgan (cM) and physical distances in base pairs (bp).

**Figure 3a**

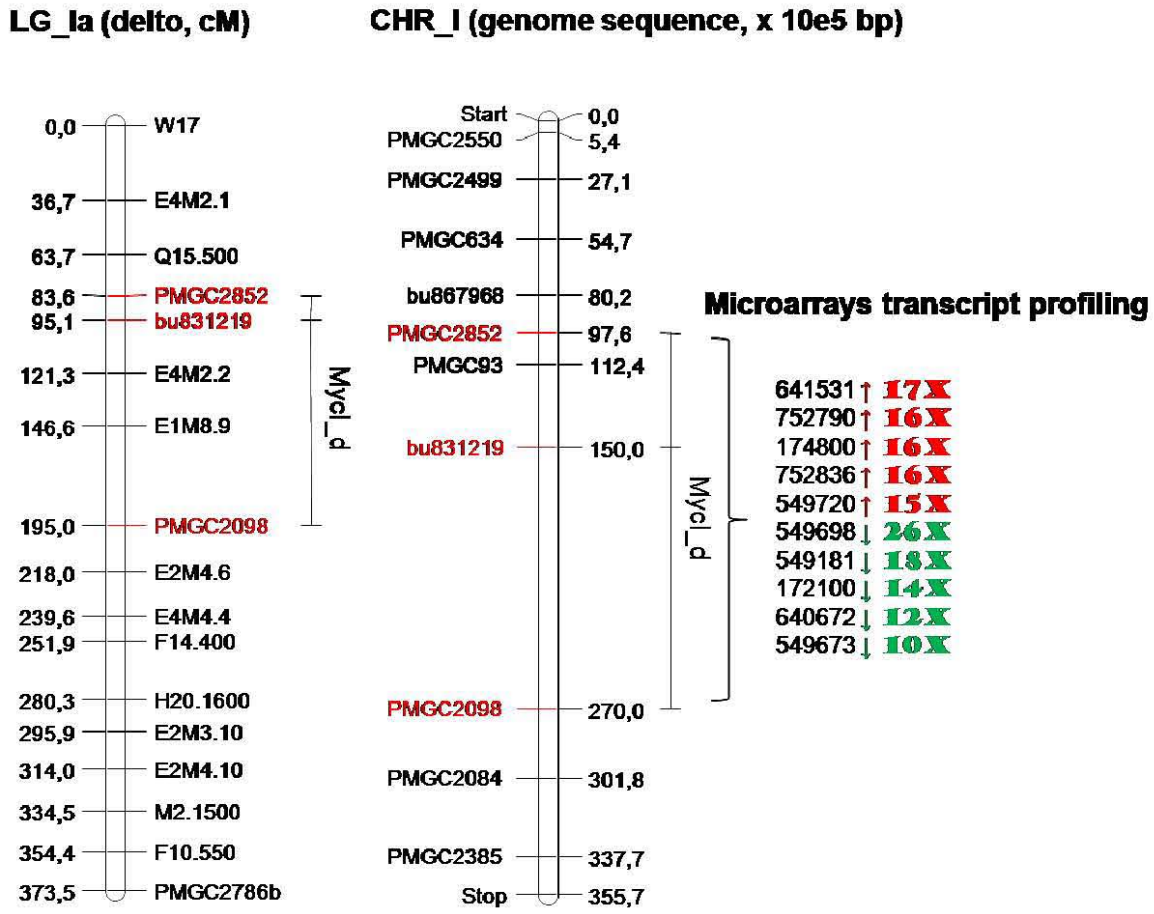


**Figure 3b**

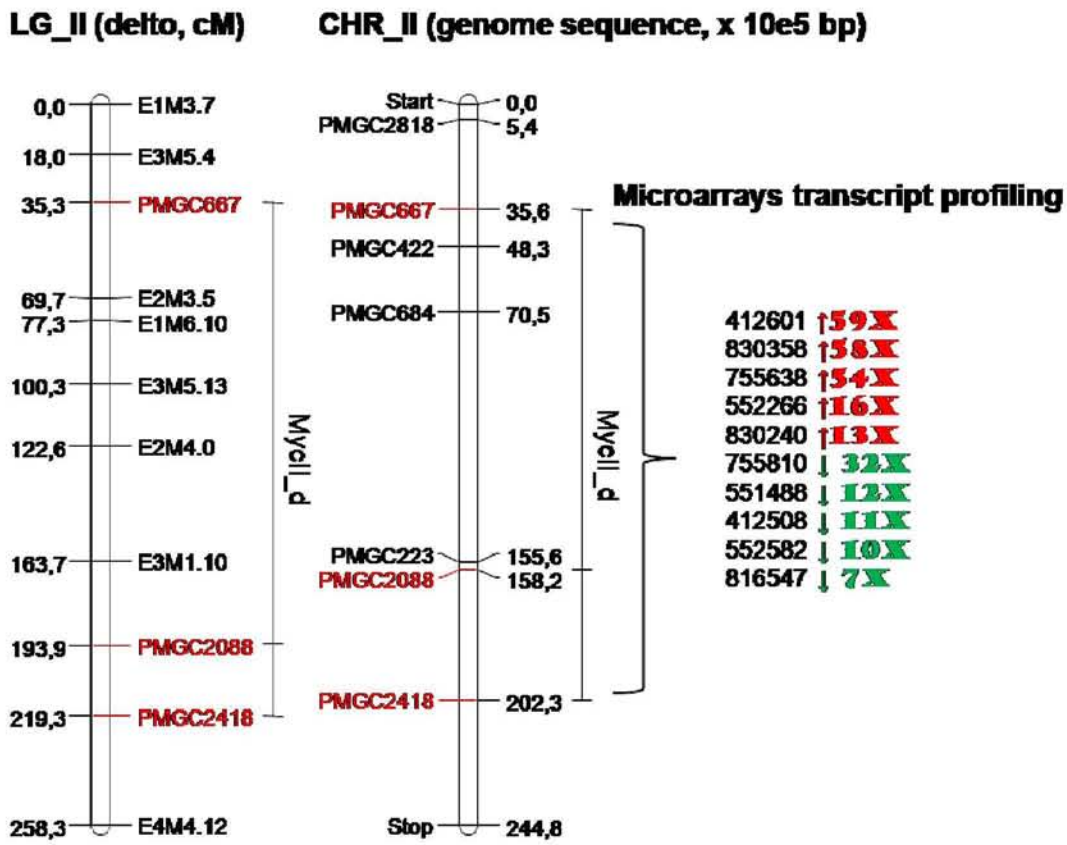

LG_II (delto, cM)
CHR_II (genome sequence, x 10e5 bp)
0,0 E1M3.7
18,0 E3M5.4
35,3 PMGC667
69,7 E2M3.5
77,3 E1M6.10
100,3 E3M5.13
122,6 E2M4.0
163,7 E3M1.10
193,9 PMGC2088
219,3 PMGC2418
258,3 E4M4.12
Mycll_d
Start 0,0
PMGC2818 5,4
PMGC667 35,6
PMGC422 48,3
PMGC684 70,5
PMGC223 155,6
PMGC2088 158,2
PMGC2418 202,3
Stop 244,8
Mycll_d
Microarrays transcript profiling
412601 ↑59X
830358 ↑58X
755638 ↑54X
552266 ↑16X
830240 ↑13X
755810 ↓ 32X
551488 ↓ 12X
412508 ↓ 11X
552582 ↓ 10X
816547 ↓ 7X

**Figure 3c**

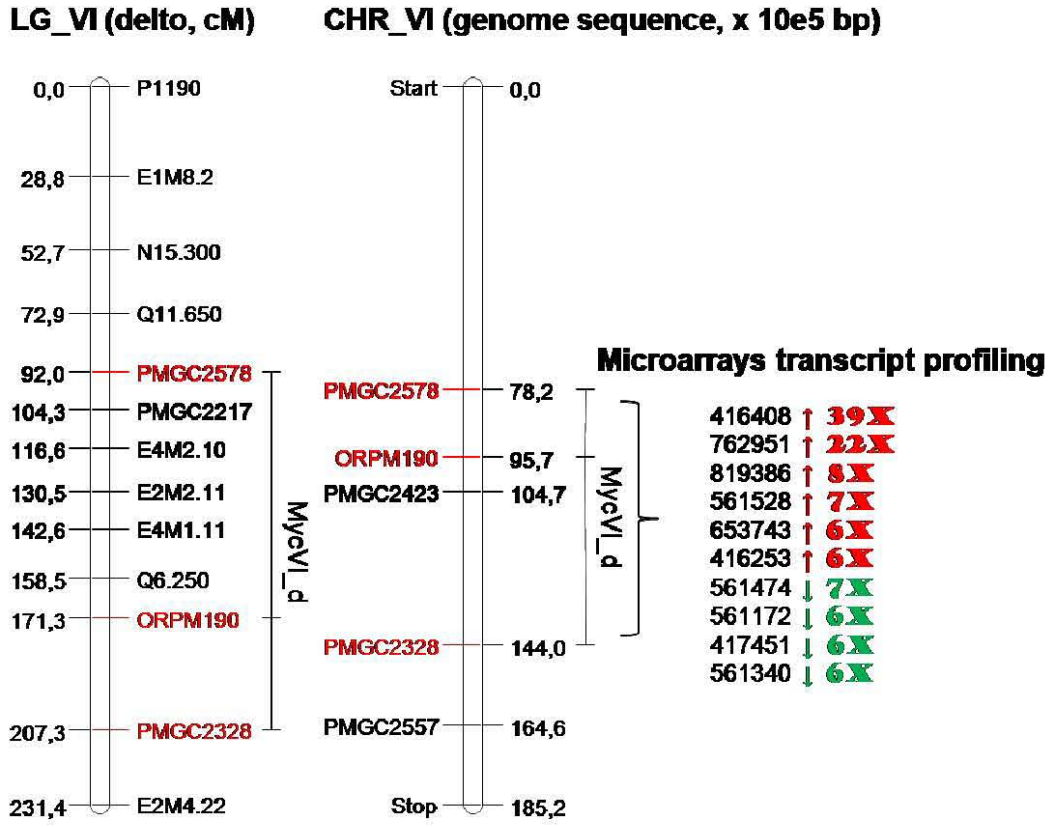

LG_VI (delto, cM)
CHR_VI (genome sequence, x 10e5 bp)
0,0 P1190
28,8 E1M8.2
52,7 N15.300
72,9 Q11.650
92,0 PMGC2578
104,3 PMGC2217
116,6 E4M2.10
130,5 E2M2.11
142,6 E4M1.11
158,5 Q6.250
171,3 ORPM190
207,3 PMGC2328
231,4 E2M4.22
MycVI_d
Start 0,0
PMGC2578 78,2
ORPM190 95,7
PMGC2423 104,7
PMGC2328 144,0
PMGC2557 164,6
Stop 185,2
MycVI_d
Microarrays transcript profiling
416408 ↑ 39X
762951 ↑ 22X
819386 ↑ 8X
561528 ↑ 7X
653743 ↑ 6X
416253 ↑ 6X
561474 ↓ 7X
561172 ↓ 6X
417451 ↓ 6X
561340 ↓ 6X

**Figure 3d**

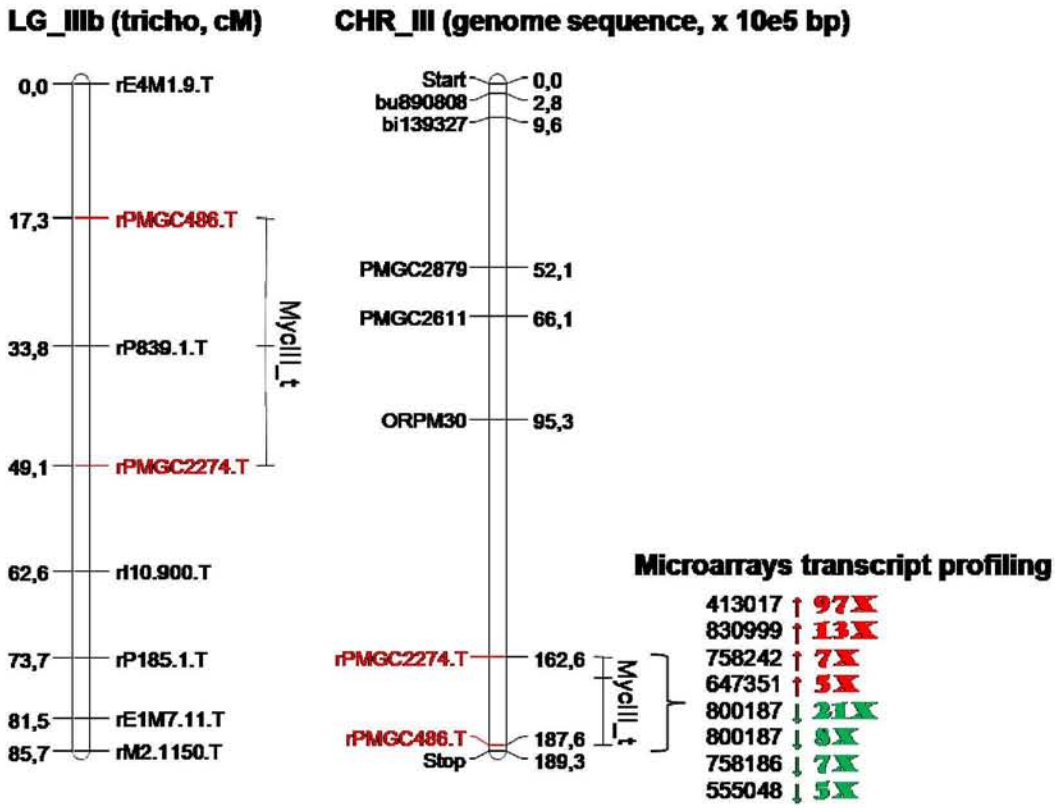
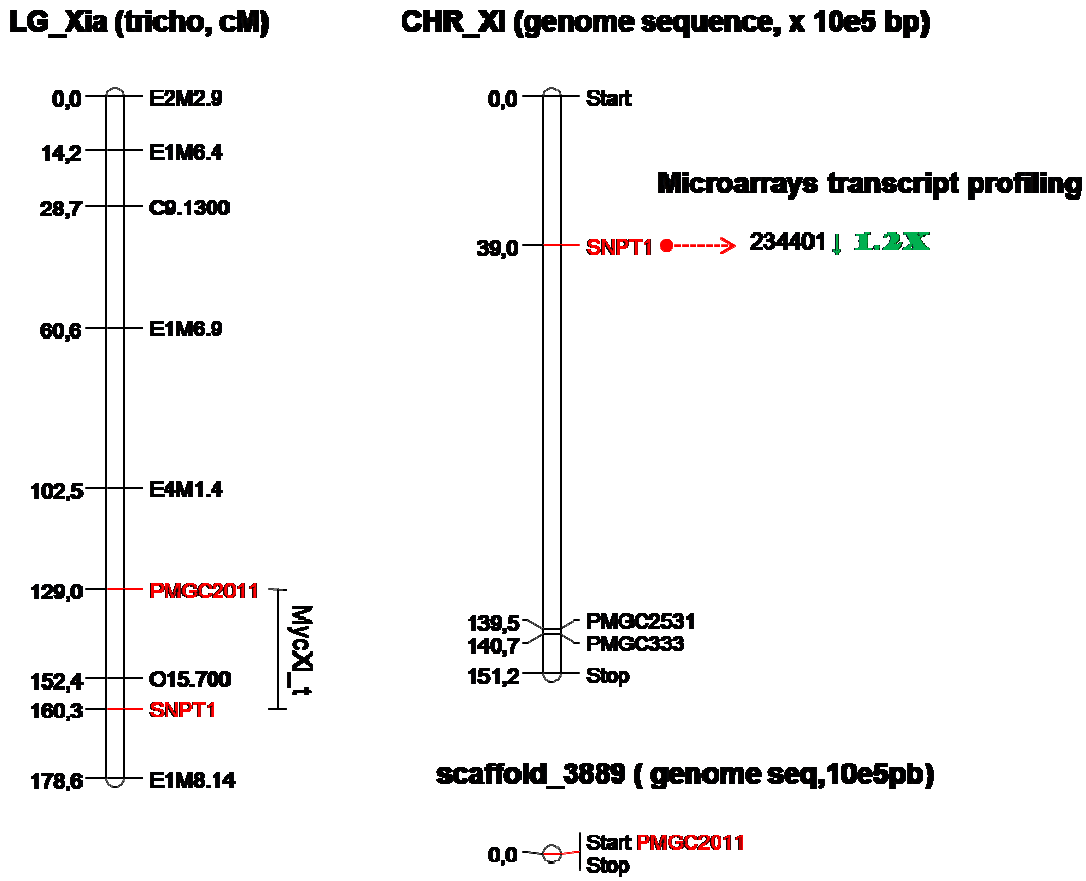

LG_IIIb (tricho, cM)
CHR_III (genome sequence, x 10e5 bp)
0,0 rE4M1.9.T
17,3 rPMGC486.T
33,8 rP839.1.T
49,1 rPMGC2274.T
62,6 rI10.900.T
73,7 rP185.1.T
81,5 rE1M7.11.T
85,7 rM2.1150.T
MycIII_t
Start 0,0
bu890808 2,8
bi139327 9,6
PMGC2879 52,1
PMGC2611 66,1
ORPM30 95,3
rPMGC2274.T 162,6
rPMGC486.T 187,6
Stop 189,3
MycIII_t
Microarrays transcript profiling
413017 ↑ 97X
830999 ↑ 13X
758242 ↑ 7X
647351 ↑ 5X
800187 ↓ 21X
800187 ↓ 8X
758186 ↓ 7X
555048 ↓ 5X

**Figure 3e**


LG_Xia (tricho, cM)
0,0
E2M2.9
14,2
E1M6.4
28,7
C9.1300
60,6
E1M6.9
102,5
E4M1.4
129,0
PMGC2011
152,4
O15.700
160,3
SNPT1
178,6
E1M8.14
MycXI_1
CHR_XI (genome sequence, x 10e5 bp)
0,0
Start
Microarrays transcript profiling
39,0
SNPT1
234401
L2X
139,5
PMGC2531
140,7
PMGC333
151,2
Stop
scaffold_3889 ( genome seq,10e5pb)
0,0
Start PMGC2011
Stop

## 4. Conclusions et perspectives

Depuis quelques années les avancés technologiques et la création de consortia internationaux ont permis la mise en place et la réalisation de plusieurs projets de séquençage de génomes. A ce jour, la disponibilité de génomes de champignons symbiotiques, saprotrophes et pathogènes, ainsi que celui de *Populus trichocarpa*, nous donne les moyens d'approfondir nos connaissances des processus de colonisation et d'interactions entre ces champignons et le peuplier, processus, essentiels aux fonctions remplies dans les cycles du carbone et de l'azote (Martin & Nehls, 2009).

Ce travail de thèse avait pour objectifs de participer à la caractérisation et au décryptage du génome de *L. bicolor* puis à la recherche des gènes impliqués dans la formation de l'ectomycorhize chez les deux partenaires de la symbiose *L. bicolor* – peuplier. Notre contribution à la connaissance du génome de *L. bicolor*, a principalement consisté à construire une carte génétique ancrée à la séquence génomique, puis à identifier et à étudier la distribution des séquences répétées (SSRs et éléments transposables). Nous avons également identifié des régions du génome du peuplier impliquées dans l'établissement de l'ectomycorhize et à la quantification de leurs effets en couplant une stratégie de détection de QTL et une stratégie d'expression différentielle de gènes.

Nous nous proposons dans cette conclusion de faire le point de nos acquis, d'en discuter et de préciser les principales questions qui restent en suspens avant de proposer des perspectives de recherche ouvertes par cette étude.

## Caractérisation et décryptage du génome de *L. bicolor*

*Laccaria bicolor* est le premier champignon ectomycorhizien dont le génome a été séquencé et publié (Martin *et al.*, 2008). A ce jour, le génome nucléaire haploïde a une taille de 60 Mb et contient 19 100 gènes dont beaucoup appartiennent à des familles multi-géniques probablement impliquées dans le développement et le fonctionnement de la symbiose (Martin *et al.*, 2008 ; Martin & Tunlid, 2008), et dans lesquelles des duplications en tandem sont apparues. Comparé à d'autres champignons, il possède un très grand nombre de gènes dont certains groupes ont subi une expansion, notamment ceux codant des GTPases, des protéases, des transporteurs. Ces propriétés révèlent un double rôle, celui de saprotrophe et celui de symbiote (Martin *et al.*, 2008 ; Martin & Nehls, 2009). A la suite du passage à l'état symbiotique, il y a eu une perte massive de gènes de dégradation des parois cellulaires de la plante hôte; de ce fait *L. bicolor* est maintenant incapable d'utiliser les sucres des polysaccharides pariétaux de la plante-hôte comme la cellulose et les pectines (Marmeisse R, données non publiées ; Martin & Nehls, 2009).

Le génome nucléaire de *L. bicolor* est plus grand que celui des génomes fongiques précédemment publiés. Cependant, il n'y a aucune évidence de duplications chromosomiques à grande échelle. Sachant qu'environ trente pour cent du génome code pour des protéines, cette taille est en partie expliquée par la grande proportion de séquences répétées et d'éléments transposables (TEs). En effet, dans ce travail, nous avons montré que le génome de *L. bicolor* contient 32 % de séquences répétées dont 8 % de microsatellites (ou simple sequence repeats, SSRs) et 24 % d'éléments transposables ; mais nos dernières investigations révèlent la présence de 26 % de répétitions minisatellites, portant le nombre de répétitions dans ce

génome à 58 %. Comparé à d'autres basidiomycètes, *L. bicolor* possède la plus grande densité de SSRs, qui par ailleurs, ne sont pas distribués au hasard dans les différentes régions génomiques. Comme notre étude le confirme, l'abondance des SSRs dans les génomes fongiques n'est ni inversement ni directement proportionnel à la taille du génome (Karaoglu *et al*., 2005). Le fait que la séquence génomique de *L. bicolor* ait une taille importante ne serait pas directement lié à ce type de séquences répétées mais à la combinaison de ces répétitions et des TEs. En effet, nous avons montré que la densité de SSRs est plus importantes dans les TEs que dans les autres régions du génome, particulièrement aux extrémités 5' et 3' de ces derniers. Ainsi, ces SSRs résulteraient de la mutation suivie de l'expansion du même proto-SSRs, répandu par la suite, à travers le génome comme un composant d'un élément transposable actif (Ramsay *et al.,* 1999 *;* Metzgar *et al*., 2000). Ils serviraient à la fois de signature "cicatrice" du passage de cet élément dans une région génomique et de site d'insertion potentiel pour d'autres types d'éléments mobiles (Metzgar *et al*., 2000).

Dans cette étude, nous avons également montré la présence de SSRs dans les régions géniques. Les possibles effets délétères expliquent probablement la faible densité de SSRs identifiés dans ces régions codant des protéines (Li *et al.,* 2004*)*. D'autre part, la présence de SSRs polymorphes dans les exons, les introns ou les régions non transcrites (Untranslated Regions UTR) pourrait changer la structure de la protéine et /ou l'expression du gène et alors jouer un rôle dans les mécanismes d'adaptation, de survie et d'évolution des organismes. La plupart des gènes de *L. bicolor* ayant des microsatellites de type tri- ou hexa-nucléotidiques dans leurs exons codent des protéines inconnues. Cependant, certains de ces gènes exprimés pendant la symbiose codent pour des homologues de glycosyl-transférases et pour des protéines contenant des domaines de type WD40, fermeture-éclair riche

en leucine et doigt de zinc (Martin *et al.*, 2008). Il reste à déterminer si ces protéines interviennent dans les interactions symbiotiques. Selon nos résultats, ces SSRs pourraient jouer un rôle dans le codage de la structure répétée de quelques domaines protéiques impliqués dans les processus d'échange de signaux et auraient ainsi un rôle clef dans l'évolution vers le style de vie symbiotique de ce clitocybe laqué bicolore.

D'autre part, en plus de la richesse en éléments mobiles chez ce symbiote fongique, nous avons mis en évidence une grande diversité des éléments transposables en y caractérisant 172 familles dont 40 possèdent des éléments intacts. La majorité des 172 familles appartiennent à la classe I, ayant pour principaux représentants des LTR rétrotransposons de type *Copia* et *Gypsy*. Nous avons également caractérisé des rétrovirus *ERV*, des *LINEs* et des *LARDs*. Une grande proportion d'éléments de classe II ont aussi été identifiés tels que des *TIRs*, un *Helitron* et des *MITEs*. Tout comme nous l'avons vu précédemment avec les microsatellites, les TEs sont aussi aléatoirement distribués à travers le génome, emboîtés (ou nichés) les uns dans les autres ou regroupés en « clusters ». Par ailleurs, l'emboîtement (ou nid) des TEs résulterait des potentiels sites d'insertion que formeraient les SSRs bordant les éléments mobiles (Metzgar *et al.*, 2000). La présence de tels nids et/ou groupes de TEs prédomine dans les régions télomèriques. L'ADN répété pouvant promouvoir des événements de recombinaison ectopique délétères, la sélection négative serait réduite dans les régions faiblement recombinantes et permettrait donc d'accumuler ces répétitions. Ainsi, l'insertion de TEs pourrait déclencher un processus fugitif qui fournirait une cible (pouvant être les SSRs bordant des TEs) dans laquelle d'autres éléments pourraient s'insérer sans conséquences délétères (Walbot *et al.*, 2001). Il y aurait alors par ce biais une accumulation de répétitions dans des zones faiblement polymorphes réparties à travers le génome, ce qui expliquerait les

difficultés rencontrées lors de l'assemblage de ce dernier. De plus, ceci est en adéquation avec le fait que les 78 marqueurs microsatellites développés dans la descendance de *L. bicolor* étaient plus polymorphes dans les régions inter-géniques à l'extérieur des TEs.

Ainsi de tels groupes de TEs seraient des acteurs majeurs dans la genèse de nouveaux gènes et/ou dans l'adaptation de gènes existants, tant au mode de vie saprotrophique que symbiotique du clitocybe laqué bicolore.

Des régions génomiques d'un ancêtre commun entre *L. bicolor* et le basidiomycète saprotrophe *Coprinopsis cinerea* ont été identifiées (J Stajich, F Martin *et al.*, résultats non publiés ; Martin & Nehls, 2009). Ces régions semblent s'être développées dans une mosaïque de blocs synténiques isolés par des régions fortement divergentes suite à l'insertion et à la prolifération différentielles de TEs. Or, nous avons mis en évidence que ces deux champignons contenaient des éléments rétrotransposables phylogénétiquement proches. De plus, l'estimation de l'âge d'insertion de ces familles de LTR rétrotransposons a révélé que *C. cinerea* les a acquises sur deux périodes (la plus ancienne se situant entre 10 et 59 millions d'années) alors que *L. bicolor* les a acquises sur trois périodes (la plus ancienne entre 10 et 57 millions d'années) dont une récente (entre aujourd'hui et 500 millions d'années) expliquant notamment son plus grand nombre de familles de *Copia* par rapport à *C. cinerea*. Ces résultats corroborent l'hypothèse précédente et montrent que le génome de *L. bicolor* est caractérisé par une grande plasticité, lui facilitant vraisemblablement le passage d'un mode de vie saprotrophique à un mode de vie symbiotique.

La caractérisation de l'ensemble de ces répétitions nous a permis de repérer les régions génomiques denses et nuisibles au bon assemblage du génome et de cibler celles permettant le développement de marqueurs

moléculaires. Nous avons ainsi pu établir la carte génétique de *L. bicolor* et aider à l'assemblage de son génome pour lequel les assembleurs JAZZ (Chapman *et al.*, données non publiées) et ARACHNE (Batzoglou *et al.,* 2002) donnaient des « scaffolds » de tailles et de compositions différentes. A ce jour, la carte génétique est composée de 326 marqueurs de type RAPD (Random Amplified Polymorphism DNA), AFLP (Amplified Fragment-Length Polymorphism), SSR et SNP (Single Nucleotide Polymorphism) répartis sur 10 groupes de liaisons. Cette carte génétique a une longueur génétique totale de 880 cM avec une distance génétique moyenne de 2,70 cM entre marqueurs adjacents. Elle est ancrée à la séquence génomique de *L. bicolor* par les marqueurs RAPD, SSRS et SNP, et aligne un total de 47,7 Mb. La carte obtenue à ce jour a réduit le nombre de groupes de liaison initial obtenus par Doudrick *et al.* (1995) de 15 à 10. Elle a permis l'amélioration de l'assemblage de la version 1 du génome et aidé à la mise à jour d'une version 2. Elle devrait être proche de la saturation, état où le nombre de groupes de liaison est égal au nombre de chromosomes. Cependant, le comptage cytologique des chromosomes de *Laccaria montana*, une espèce proche de *L. bicolor*, a abouti à 9 chromosomes haploïdes (Mueller *et al.*, 1993). De ce fait, il est nécessaire d'envisager un comptage cytologique précis ainsi qu'un caryotypage en champ pulsé de *L. bicolor* S238N. Si cette carte n'est pas saturée, le génotypage de marqueurs supplémentaires est à poursuivre jusqu'à saturation. De plus, pour envisager de futures détections de QTL éventuellement suivies de clonages positionnels, il conviendrait d'améliorer la résolution de la dite carte afin de réduire la distance génétique entre marqueurs adjacents à un minimum de 1 cM. Cela nécessite donc d'augmenter le nombre d'individus dans la descendance (111 à ce jour). Nous avons localisé sur cette carte génétique, l'ADN ribosomal dont la taille du locus est de 0,5 Mb (basé sur un nombre de copies = 50 pour le génome

haploïde) à l'extrémité distale du LG 7 (Martin *et al.,* 1999). Nous avons également localisé les loci de compatibilité sexuelle dans des régions différentes *MATa s*ur le LG 1 *et MATb* sur le LG 8. A la différence des travaux de Doudrick *et al.* (1995), ces loci nous apparaissent non liés, ce qui pourrait être expliqué par la meilleure saturation de la présente carte.

Nous avons confirmé ce résultat en caractérisant l'organisation des gènes de ces deux loci *MATa* (homéo-domaine composé de gènes codant pour des facteurs de transcription) et *MATb* (composé de gènes codant pour des récepteurs de phéromones). De plus, nous avons montré la conservation des séquences des domaines fonctionnels des protéines de *L. bicolor* et de *C. cinerea* impliquées dans la compatibilité sexuelle, et cela malgré plus de 100 millions d'années de divergence (Berbee et Taylor, 2001). Nous avons également montré que ces loci sont conservés dans l'ordre des Agaricales et intégrés dans des régions génomiques évoluant différemment. En effet, chez *L. bicolor* le locus *MATa* est localisé dans une région à faible taux de recombinaison et à faible densité en TEs et où les gènes sont donc soumis à une forte pression de sélection. Au contraire, le locus *MATb* est situé dans une région soumise à une faible pression de sélection, où des évènements tels que les duplications, les translocations et les insertions de TEs sont fréquents en raison de l'existence de longues distances inter-géniques. Ainsi la région *MATb* évolue plus rapidement que la région *MATa*. Ces évènements ressemblent à ceux qui ont permis la formation des régions de compatibilité sexuelle chez *Cryptococcus* (Fraser *et al.*, 2004) et des chromosomes sexuels chez les plantes et les animaux (Fraser & Heitman, 2004). Ceci corrobore les hypothèses précédentes sur l'importance de la distribution des nombreux éléments répétés caractérisant le génome de *L. bicolor*.

En perspective sur ce point, la caractérisation approfondie des divers

types de TE peut être un élément nécessaire tant à la compréhension de certaines fonctions et régulations qu'à l'évolution du génome de ce symbiote fongique.

## Recherche des gènes impliqués dans la formation des ectomycorhizes chez les deux partenaires

Les tentatives de quantification de l'établissement de la symbiose ectomycorhizienne par des souches monocaryotiques issues de *L. bicolor* S238N n'ont pas encore aboutis à des résultats satisfaisants. Cependant nous avons obtenu dans quelques cas une faible colonisation de racines de peuplier par des homocaryons de *L. bicolor* (de 0 à 30 %). De ce fait, aucune détection de QTL impliqué dans la mycorhization chez *L. bicolor* n'a pu être réalisée au cours de cette thèse. La mycorhization du peuplier par des monocaryons de *L. bicolor* S238N nécessiterait des conditions très strictes (non connues) ou dépendrait du parent d'origine. En effet, Kropp *et al.*, (1987) ont montré que la capacité des homocaryons de *L. bicolor* à former des mycorhizes avec *Pinus strobus* dépendait bien du parent (capacité à former des mycorhizes par les homocaryons variant de 0 à 22% dans une descendance et de 30 à 80% pour une autre). De même, Umata (1999) a mis en évidence chez *Laccaria laccata,* espèce proche de *L. bicolor*, l'incapacité de former des ectomycorhizes par certains monocaryons issus d'un même basidiome, alors que d'autres le pouvaient. D'autre part, il n'a détecté aucune variabilité dans la capacité à former des mycorhizes parmi des monocaryons d'*Hebeloma cylindrosporum*, qui eux peuvent facilement former des mycorhizes (Debaud *et al*., 1987). De plus, d'après les travaux de Kropp et Anderson (1994), certaines espèces fongiques exigent des informations supplémentaires à celles portées par un génome haploïde pour être capable de coloniser un hôte et d'établir une symbiose

fonctionnelle. Ainsi la dicaryotisation augmenterait les capacités de colonisation et pourrait même être indispensable chez certaines espèces ou certaines souches.

Afin d'optimiser les futures détections de QTL impliqués dans l'établissement de la symbiose chez *L. bicolor*, nous devrons utiliser des dicaryons reconstitués à partir des monocaryons composant l'actuelle descendance. Trente-sept dicaryons sont d'ores et déjà disponibles et maintenus en culture au laboratoire Interactions Arbres/Micro-organismes du centre INRA de Nancy. Néanmoins, une détection précise de QTL nécessitera l'augmentation du nombre de dicaryons reconstitués (quelques centaines) ; d'autre part, une augmentation du nombre de monocaryons dans la descendance est nécessaire pour affiner la carte génétique de *L. bicolor*.

Pour identifier les loci pouvant intervenir dans l'établissement de la symbiose chez le peuplier, nous avons inoculé 300 hybrides interaméricains *P. deltoides* x *P. trichocarpa* de la famille 54B (contenant les 146 hybrides utilisés dans l'étude de Tagu *et al.,* 2005) par la souche dicaryotique S238N de *L. bicolor.* Nous avons tout d'abord confirmé les résultats de Rosado *et al.* (1994) et de Tagu *et al.* (2001 et 2005), montrant que la formation d'ectomycorhizes était sous contrôle génétique de l'hôte. Cependant, l'héritabilité au sens large pour le pourcentage d'infection mycorhizienne égale à 0,38, indique un contrôle génétique modéré du caractère et une grande participation de facteurs exogènes. Par ailleurs nous avons analysé 40 des 300 génotypes utilisés et montré que la colonisation des racines de peuplier par *L. bicolor* modifie complètement leur capacité à excréter des enzymes impliquées dans la dégradation de la matière organique ou la mobilisation du phosphore organique. Nous

avons pu montrer que le niveau d'excrétion de ces enzymes principalement produites par le champignon associé est sous le contrôle génétique de l'hôte.

Comme dans les études précédentes de Tagu *et al.*, (2001, 2005), nous avons trouvé une grande différence de réponse à la colonisation des deux parents par *L. bicolor* (45 % pour *P. trichocarpa* et 15 % pour *P. deltoides*). La moyenne des descendants (47 % contre 35 % dans Tagu *et al.,* 2005) n'était pas significativement différente de celle du parent *P. trichocarpa* (45 %). Nous émettons l'hypothèse que les gènes impliqués dans ce caractère seraient dominants et hérités du parent *P. trichocarpa*. Quelques clones ayant montré un pourcentage d'infection mycorhizienne plus élevé que celui des parents suggèrent que la capacité à former des ectomycorhizes chez le peuplier peut être améliorée.

Nous avons réalisé deux analyses QTL. La première, réalisée avec les cartes génétiques v1 de Jorge *et al.,* 2005, a révélé quatre QTL statistiquement significatifs. L'un est localisé sur la carte génétique de *P. deltoides* et ne peut pas être ancré sur la séquence génomique du peuplier (Tuskan *et al.,* 2006). Les trois autres sont localisés sur la carte génétique de *P. trichocarpa* et deux, Myc_t2 et Myc_t1 ont pu être ancrés au génome du peuplier (Tuskan *et al.,* 2006). Le QTL Myc_t2 est ancré dans une région de fonction inconnue, tandis que Myc_t1, expliquant 28.3 % de la colonisation ectomycorhizienne, est lié à un marqueur SNP situé dans un gène codant une cinamoyl-coa réductase (CCR), enzyme intervenant dans la synthèse de lignines.

Dans la seconde analyse, utilisant les cartes génétiques de Jorge *et al.*, 2005 mises à jour, aucun QTL statistiquement significatif n'a été identifié alors qu'avec le même jeu de données phénotypiques nous trouvions

quatre QTL significatifs sur les cartes v1 (Jorge *et al.,* 2005). En effet, pour intégrer les nouveaux marqueurs, il a été nécessaire d'exclure plusieurs paires de marqueurs, dont certains étaient liés aux quatre QTL localisés sur les cartes v1 de Jorge *et al*., 2005. Tant que les cartes génétiques de *P. deltoides* et de *P. trichocarpa* ne seront pas saturées et stabilisées, la validation de QTL impliqués dans l'infection ectomycorhizienne restera difficile. De plus, la détection de tous les QTL impliqués dans ce caractère nécessiterait d'améliorer le jeu de données phénotypiques actuel en augmentant le nombre d'hybrides phénotypés.

Néanmoins, trois des cinq QTL non significatifs statistiquement de la deuxième analyse ont été placés sur la carte génétique de *P. deltoides* (MycI_d, MycII_d, MycVI_d) et deux sur la carte génétique de *P. trichocarpa* (MycIII_t et MycXI_t). Seul le QTL MycXI_t ancré sur le chromosome XI du génome du peuplier (Tuskan *et al.,* 2006), par un seul marqueur SNP, n'a pas pu avoir une région génomique correspondante délimitée. Cependant ce marqueur SNP (SNPT1) est situé dans un gène de fonction inconnue.

Les régions génomiques correspondant aux autres QTL sont réparties sur plusieurs mégabases. Par une approche d'expression différentielle de gènes, nous avons tenté d'identifier les gènes potentiels de ces régions pouvant jouer un rôle dans la formation des ectomycorhizes. Nous avons identifié 81 transcrits (sur- ou sous exprimés au moins 5 fois et ayant une P-value < 0,05) dont 52 sont surexprimés chez *P. deltoides* et 29 chez *P. trichocarpa*. Parmi ceux-ci, le gène le plus surexprimé chez *P. trichocarpa* code pour une protéine EREBP-4, facteur de transcription sensible à l'éthylène (Leubner-Metzger *et al.,* 1998, Solano *et al.,* 1998; Ohta *et al.,* 2000). Or, Il est bien connu que les phytohormones sont impliquées dans

le développement ectomycorhizien (Slankis, 1950 ; Gay *et al.,* 1994; Karabaghli-Degron *et al.,* 1998; Barker & Tagu, 1998; Tagu *et al.,* 2003). En effet, Rupp & Mudge (1995) ont montré que l'éthylène et l'auxine induisent la mycorhization des racines de *Pinus mugo*. Ce facteur de transcription sensible à l'éthylène pourrait jouer un rôle important dans l'établissement de la symbiose. De plus l'inoculation de peupliers par *L. bicolor* a confirmé le rôle de l'auxine fongique dans l'augmentation du nombre de racines chez la plante-hôte (Richter *et al.,* 2009). Cela a induit chez la plante, une augmentation du nombre de transcrits codant pour des transporteurs d'auxine et des gènes de réponse auxine/éthylène (Richter *et al.,* 2009).

Cependant, nous avons conscience que cette étude présente d'importantes limitations. En effet, comme nous l'avons discuté précédemment, l'approche QTL exige une carte génétique stable et saturée mais également un jeu de données phénotypiques comportant un grand nombre d'hybrides. Nous espérons pouvoir utiliser de meilleures cartes et un plus grand jeu de données phénotypiques afin de localiser définitivement des QTL significatifs impliqués dans l'établissement de l'ectomycorhize. D'autre part, la comparaison de l'expression des gènes entre les ectomycorhizes des deux parents, plusieurs mois après l'inoculation fongique, n'intègre probablement qu'une partie des interactions existantes lors de l'établissement de la mycorhize (reconnaissance avec l'échange de signaux, l'attachement du mycélium sur la racine puis la colonisation interne) et/ou intègre en plus des processus liés au fonctionnement de la symbiose. Ces interactions semblent extrêmement complexes.

Cependant, malgré les imperfections des deux approches, cette combinaison nous a permis de fournir une liste de gènes, qui pourraient être impliqués dans l'établissement et/ou le fonctionnement de la mycorhize.

Les perspectives ouvertes par ce travail sont nombreuses dans l'objectif de déterminer les gènes des deux partenaires qui peuvent intervenir dans l'établissement puis le fonctionnement de la symbiose.

Peuplier :

1. Il apparaît nécessaire de confirmer la détection QTL et l'identification transcriptomique des gènes de peuplier ici proposés comme potentiels acteurs dans l'établissement de la symbiose. Ce travail complémentaire nécessite une saturation et une stabilisation des cartes génétiques utilisées dans cette étude mais aussi sans doute le phénotypage d'un plus grand nombre de clones de peuplier.

Une autre alternative possible serait de quantifier l'infection mycorhizienne par *L. bicolor* dans une autre descendance pour laquelle des cartes génétiques saturées sont d'ores et déjà disponibles telle celle issue du croisement (*P. trichocarpa* x *P. deltoides*) x *P. deltoides* et utilisée par Yin *et al.*, 2004. Dans cette hypothèse, il faudrait d'abord vérifier que l'aptitude à la symbiose est suffisamment différente entre les parents et la descendance.

2. Une nouvelle combinaison QTL/approche transcriptomique nécessiterait cette fois de prendre en compte des racines de peuplier non mycorhizés et mycorhizés à différents stades de colonisation, de façon à faire la part entre les différences d'expression génique caractérisant la mise en place de la symbiose puis son fonctionnement.

*L. bicolor* :

1. Le caryotypage en champs pulsé associé à la saturation de la carte génétique par l'utilisation de marqueurs complémentaires devrait permettre de définir le nombre exact de chromosomes.

2. Il serait nécessaire dans l'avenir d'augmenter le nombre de monocaryons et de dicaryons reconstitués de *L. bicolor* S238N, afin d'envisager une quantification et une détection des QTL impliqués dans l'établissement de la symbiose chez ce champignon.

3. Il conviendrait maintenant de continuer à analyser dans les régions géniques, l'insertion des séquences et des éléments répétés, afin de mieux comprendre leurs rôles possibles dans l'évolution de ce génome d'un stade saprophyte vers un stade symbiotique.

## Bibliographie de l'introduction et conclusions


Armstrong JL, Fowles NL & Rygiewicz PT. 1989. Restriction fragment length polymorphisms distinguish ectomycorrhizal fungi. *Plant and Soil* 116: 1-7.

Balasubramanian S, Kim SJ & Podila G. 2002. Differential expression of a malate synthase gene during the preinfection stage of symbiosis in the ectomycorrhizal fungus *Laccaria bicolor*. *New Phytologist* 154 (2): 517-527.

Barker SJ, Tagu D & Delp G. 1998a. Regulation of root and fungal morphogenesis in mycorhizal symbioses. *Plant Physiology* 116 (4): 1201-1207.

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, 1987. Mapmaker an interactive computer package for constructing primary genetic linkage maps of experimental and natural population. *Genomics* 1: 174-181.

Batzoglou S, Jaffe DB, Stanley K, Butler J, Gnerre S, Mauceli E, Berger B, Mesirov JP, Lander ES. 2002. ARACHNE: a whole-genome shotgun assembler. *Genome Research* 12: 177–189.

Beguiristain T, Cote R, Rubini P, Jay-Allemand C and Lapeyrie F. Hypaphorine accumulation in hyphae of the ectomycorrhizal fungus, *Pisolithus tinctorius*. *Phytochemistry* 40 (4): 1089-1091.

Berbee ML, Taylor JW. 2001. Fungal molecular evolution: gene trees and geologic time. *The Mycota: a comprehensive treatise on fungi as experimental systems for basic and applied research. Systematics and Evolution, Part B. Volume VII.* Berlin Heidelberg, Germany: Springer-Verlag, p. 229–245.

Bigelow HE. 1985. North American Species of Clitocybe. Part II. J. Cramer, Berlin, p. 190.

Courty PE, Pritsch K, Schloter M, Hartmann A, Garbaye J. 2005. Activity profiling of ectomycorrhiza communities in two forest soils using multiple enzymatic tests. *New* Phytologist 167: 309-319.

Kropp BR, McAfee, BJ and Fortin JA. 1987. Variable loss of ectomycorrhizal ability in monokaryotic and dikaryotic cultures of *Laccaria bicolor*. *Can. J. Bot*. 65(3): 500–504.

Kropp BR & Anderson AJ. 1994. Molecular and genetic approaches to understanding variability in mycorrhizal formation and functioning. In "Mycorrhizae and Plant Health", Pfleger FL, Linderman RG, eds, APS Press, St. Paul, Minnesota, 309-336.

Burdon RD. 2001. Genetic diversity and disease resistance: some considerations for research, breeding, and deployment. *Can. J. For. Res.* 31: 596-606.

Burn RM & Honkala BH. 1990. Silvics of North America: 1. Conifers; 2. Hardwoods. Vol. 2. Agriculture Handbook 654. Washington DC, USA: US Department of Agriculture, p.877.

Buscot F, Munch JC, Charosset JY, Gardes M, Nehls U, Hampp R. 2000. Recent advances in exploring physiology and biodiversity of ectomycorrhizas highlight the functioning of these symbioses in ecosystems. *New phytologist* 24: 601-614.

Cervera MT, Plomion C, Malpica C. 2000. Molecular markers and genome mapping in woody plants. *In Molecular Biology of Woody Plants*. Edited by SM Jain and SC Minocha 1: 375-394.

Cervera MT, Storme V, Ivens B, Gusmao J, Liu BH, Hostyn V, Slycken JV, Montagu MV & Boerjan W. 2001. Dense genetic linkage map of three populous species (*P. deltoides, P. nigra, P. trichocarpa*) based on AFLP an d microsatellite markers. *Genetics* 158: 787-809.

Cripps CL and Miller OK Jr. 1993. Ectomycorrhizal fungi associated with aspen on three sites in the north-central Rocky Mountains. *Can. J. Bot* 71: 1414-1420.

Chalot M & Brun A. 1998. Physiology of organic nitrogen acquisition by ectomycorrhizal fungi and ectomycorrhizas. *FEMS Microbiol. Rev*. 22: 21- 44.

Debaud JC and Gay G. 1987. In vitro Fruiting Under Controlled Conditions of the Ectomycorrhizal Fungus *Hebeloma cylindrosporum* Associated with *Pinus pinaster*. *New Phytologist* 105 (3): 429-435.

Debaud JC, Gay G, Prevost A, Lei J and Dexheimer J. 1988. Ectomycorrhizal Ability of Genetically Different Homokaryotic and Dikaryotic Mycelia of Hebeloma cylindrosporum. *New Phytologist* 108 (3): 323-328.

Debaud JC, Marmeisse R & Gay G. 1997. Genetics and molecular biology of the fungal partner in the ectomycorrhizal symbiosis *Hebeloma cylindrosporum* x *Pinus pinaster*. *In the Mycota*, volume V, part B, pp. 95-115. Carroll, GC and Tuzynski P (eds), springer, Berlin.

DeBellis T, Kernaghan G, Bradley R, Widden P. 2006. Relationships between stand composition and ectomycorrhizal community structure in boreal mixed-wood forests. *Microbial Ecology* 52: 114–126.

Dell, B., & N. Malajczuck, 1997**.** L'inoculation des Eucalyptus introduits en Asie avec des champignons ectomycorhiziens australiens en vue d'augmenter la productivité des plantations. *Revue forestière Française* XLIX (no spec. 1997): 174-184.

Di Battista C, Selosse MA, Bouchard D, Stenström E, & Le Tacon F. 1996. Variation in symbiotic efficiency, phenotypic characters and ploidy level among

different isolates of the ectomycorrhizal basidiomycete *Laccaria bicolor* S238N. *Mycological Research* 100: 1315-1324.

Ditengou FA, Béguiristain T & Lapeyrie F. 2000. Root hair elongation is inhibited by hypaphorine, the indole alkaloid from the ectomycorrhizal fungus *Pisolithus tinctorius*, and restored by indole-3- acetic acid. *Planta* 211 (5): 722-728.

Ditengou FA, Raudaskoski M, Lapeyrie F. 2003. Hypaphorine, an IAA antagonist delivered by the ectomycorrhizal fungus *Pisolithus tinctorius*, induces reorganization of actin and microtubule cytoskeleton in *Eucalyptus globulus ssp bicostata* root hairs. *Planta* 17: 45-49.

Doudrick RL, Raffle VL, Nelson NL, & Furnier GR. 1995. Genetic analysis of homocaryons from a basidiome of *Laccaria bicolor* using random amplified polymorphic DNA (RAPD) markers. *Mycological Research* 99: 1361-1366.

Drost DR, Novaes E, Boaventura-Novaes C, Benedict C, Brown RC, Yin T, Tuskan GA and Kirst M. 2009. A microarray-based genotyping and genetic mapping approach for highly heterozygous outcrossing species enables localization of a large fraction of the unassembled *Populus trichocarpa* genome sequence. *The Plant Journal* 58: 1054–1067.

Dulbecco P, Sales C, et Vincent MH. 1995. Le peuplier, l'une des premières essences feuillues récoltées en France: ses emplois se diversifient. *La Forêt Privée* 225: 33-42.

Duplessis S, Sorin C, Voiblet C, Palin B, Martin F, Tagu D. 2001. Cloning and expression analysis of a new hydrophobin cDNA from the ectomycorrhizal basidiomycete *Pisolithus*. *Curr Genet* 39: 335-339.

Duplessis S, Courty PE, Tagu D, Martin F. 2005. Transcript patterns associated with ectomycorrhiza development in *Eucalyptus globulus* and *Pisolithus microcarpus. New Phytologist* 165: 599-611.

Eckenwalder JE. 1996. Systematics and evolution of *Populus*. Biology of *Populus* and its implications for management and conservation. *NRC Research Press* 3: 7-32.

Esser K & Kuehenen R. 1967. Genetics of fungi. Springer Verlag, Heidelberg, p. 500.

Fraser JA, Diezmann S, Subaran RL, Allen A, Lengeler KB, Dietrich FS, Heitman J. 2004. Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. *Plos Biology* 2: e384

Fraser JA, Heitman J. 2004. Evolution of fungal sex chromosomes. *Molecular Microbiology* 51: 299–306.

Frey P, Frey-Klett P, Garbaye J, Berge O & Heulin T. 1997. Metabolic and genotypic fingerprinting of fluorescent pseudomonas associated with Douglas fir-*Laccaria bicolor* mycorrhizosphere. *Applied and environmental Microbiology* 63: 1852-1860.

Frey-Klett P, Churin JL, Pierrat JC, Garbaye J. 1997. Dose effect in the dual inoculation of an ectomycorrhizal fungus and a mycorrhiza helper bacterium in two forest nurseries. *Soil Biology and Biochemistry* 31: 1555-1562.

Fries N. 1983. Spore germination homing reaction, and intersterility groups in *Laccaria laccata*. *Mycologia* 75: 221-227.

Fries N, Serck-Hanssen K, Dimberg L.H, Theander O. 1987. Abietic acid, an activator of basidiospore germination in ectomycorrhizal species of the genus Suillus (Boletaceae). *Plant physiology and biochemistry* 11(4): 360-363

Gardes M, Wong KKY & Fortin JA. 1990. Interactions between monokaryotic and dikaryotic isolates of *Laccaria bicolor* on roots of *Pinus banksiana*. *Symbiosis* 8: 233-250.

Gay G, Normand L, Marmeisse R, Sotta B & Debaud JC. 1992. Auxin overproducer mutants of *Hebeloma cylindrosporum* Romagnesi have increased mycorhizal activity. *New Phytologist* 128 (4): 645-657.

Gehring CA & Whitman TG. 1994. Interactions between above ground herbivores and the mycorrhizal mutualists of plants. *Trends in Ecology and Evolution* 9: 251-255.

Hampp R, Ecke M, Schaeffer C, Wallenda T, Wingler A, Kottke I and Sundberg B. 1996. Axenic mycorrhization of wild type and transgenic hybrid aspen expressing T-DNA indoleacetic acid-biosynthetic genes. *Trees-Structure and Function* 11: 59-64.

Harr B, Todorova J and Schlotter J. 2002. Mismatch repair-driven mutational bias in *D. melanogaster*. *Mol. Cel.* 10: 199-205.

Henrion B, Di Battista C, Bouchard D, Vairelles D, Thompson BD, Le Tacon F & Martin F. 1994. Monitoring the persistence of *Laccaria bicolor* as an ectomycorrhizal symbiot of nursery-grown Douglas fir by PCR of the rDNA intergenic spacer. *Molecular ecology* 3: 571-580.

Hibbet DS & Donoghue MJ. 1996. Implication of phylogenetic studies for conservation of genetic diversity in Shiitake murshrooms. *Conservation Biology* 10: 1321-1327.

Hiremath ST, Balasubramanian S, Zheng J and Podila GK. 2006. Symbiosis-regulated expression of an acetyl-CoA acetyltransferase gene in the ectomycorrhizal fungus *Laccaria bicolor*. *Canadian Journal of Botany* 84: 1405-1416.

Högberg P, Nordgren A, Buchmann N, Taylor A F S, Ekblad A, Högberg M N, Nyberg G, Ottosson-Löfvenius M & Read D J, 2001. Large-scale forest girdling shows that current photosynthesis drives soil respiration. *Nature* 411: 789-792.

Horan DP & Chilvert GA. 1990. Chemotropism – the key to ectomycorrhizal formation? *New Phytologist* 116 (2): 297-301.

Jackson S, Rounsley S and Purugganan M. 2006. Comparative Sequencing of Plant Genomes: Choices to Make. *Plant Cell* 18: 1100-1104.

Johansson T, Le Quéré A, Ahren D, Söderström B, Erlandsson R, Lundeberg J, Uhlen M, Tunlid A. 2004. Transcriptional responses of Paxillus involutus and Betula pendula during formation of ectomycorrhizal root tissue. *Molecular Plant–Microbe Interactions* 17: 202–215.

Jokipii S, Haggman H, Brader G, Kallio PT and Niemi K. 2008. Endogenous *PttHb1* and *PttTrHb*, and heterologous Vitreoscilla vhb haemoglobin gene expression in hybrid aspen roots with ectomycorrhizal interaction. *Journal of Experimental Botany* 59: 2449-2459.

Jorge V, Dowkiw A, Faivre-Rampant P and Bastien C. 2005. Genetic architecture of qualitative and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar: genetic mapping and QTL detection. *New Phytol* 167: 113-127.

Kaldorf M, Renker C, Fladung M, Buscot F. 2004. *Mycorrhiza* 14: 295–306.

Karabaghli-Degron C, Sotta B, Bonnet M, Gay G and Le Tacon G. 1998. The Auxin Transport Inhibitor 2,3,5-Triidobenzoic Acid (TIBA) Inhibits the Stimulation of In vitro Lateral Root Formation and the Colonization of the Tap-Root Cortex of Norway Spruce (Picea abies) Seedlings by the Ectomycorrhizal Fungus Laccaria bicolor. *New Phytologist* 140 (4): 723-733.

Karaoglu H, Lee CMY, Meyer W. 2005. Survey of simple sequence repeats in completed fungal genomes. *Molecular Biology and Evolution* 444 22: 639-649.

Kawano T, Kawano N, Hosoya H, Lapeyrie F. 2001. Fungal auxin antagonist, hypaphorine, competitively inhibits IAA-dependant superoxide generation by horseradish peroxidase. *Biochemical and Biophysical Research Communication* 288: 546-551.

Kelleher CT, Chiu R, Shin H, Bosdet IE, Krzywinski MI, Fjell CD, WilkinJ, Yin T, DiFazio SP, Ali J, Asano JK, Chan S, Cloutier A, Girn N, Leach S, Lee D, Mathewson CA, Olson T, O'Connor K, Prabhu A, Smailus DE, Stott JM, Tsai M, Wye NH, Yang GS, Zhuang J, Holt RA, Putnam NH, Vrebalov J, Giovannoni JJ, Grimwood J, Schmutz Rokhsar D, Jones SJM, Marra MA, Tuskan GA, Bohlmann JR, Ellis BE, Ritland K, Douglas CJ and Schein .JE. 2007. A physical map of the highly heterozygous *Populus* genome: integration with the genome sequence and genetic map and analysis of haplotype variation. *The Plant Journal* 50: 1063–1078.

Kim SJ, Hiremath ST & Podila GK. 1999. Cloning and identification of symbiosis-regulated genes from the ectomycorrhizal *Laccaria bicolor*. *Mycological Research* 103 (2): 168-172.

Krpata D, Peintner U, Langer I, Fitz WJ and Schweiger P. 2008. Ectomycorrhizal communities associated with Populus tremula growing on a heavy metal contaminated site. *Mycological Research* 112: 1069-1079.

Kropp BR, & Fortin JA. 1988. The incompatibility system and relative etcomycorrhizal performance of monokaryons and reconstituted dikaryons of *Laccaria bicolor*. *CanadianJjournal of Botany*. 66: 289-294.

Kropp BR. and Anderson AJ. 1994. Molecular and genetic approaches to understanding variability in mycorrhizal formation and function. In: Mycorrhizae and plant health, (ed. by F. L. Pfleger and R. G. Linderman), p. 309-336. The American Phytopathological Society, St. Paul, Minnesota, USA.

Lacourt I, Duplessis S, Abba S, Bonfant P, Martin F. 2002. Isolation and characterization of differentially expressed genes in mycelium and fruit body of *Tuber borchii*. *Applied & environmental Microbiology* 68: 4574-4582.

Lamhamedi MS, Fortin Sa, Kope HH & Kropp BR. 1990. Genetic variation in ectomycorrhiza formation by *Pisolithus arhizus* on *Pinus pinaster* and *Pinus baksiana*. *The New Phyt* 115: 689-697.

Le Tacon F, Alvarez IF, Bouchard D, Henrion B, Jackson RM, Luff S, Parlade JI, Pera J, Stenström E, Villeneuve N & Walker C. 1992. Variation in field response of forest trees to nursery ectomycorrhizal inoculation in Europe. *In Mycorrhizas in ecosystems*. CAB International, Wallingford, p. 119-134.

Le Tacon, F., D. Mousain, J. Garbaye, D. Bouchard, J.L. Churin, C. Argillier, J.-M. Amirault & B. Généré, 1997. Mycorhizes, pépinières et plantations forestières en France. *Revue forestière Française* XLIX (no spec. 1997) : 131-154.

Lefèvre F, Goué-Mourier MC, Faivre-Rampant P, Villar M. 1998. A single gene cluster controls incompatibility and partial resistance to various *Melampsora larici-populina* races in hybrid poplars. *Phytopathology* 88: 156–163.

LePage BA, Currah RS, Stockey RA & Rothwell GW. 1997. Fossil ectomycorrhizae from the middle Eocene. *American Journal of Botaby* 84 (3): 410-412.

Le Quéré A, Wright D, Söderström B, Tunlid A, Johansson T. 2005. Global patterns of gene regulation associated with the development of ectomycorrhiza between birch (Betula pendula Roth.) and Paxillus involutus (Batsch) Fr. *Mol Plant Microbe Interact.* 18: 659-673.

Leubner-Metzger G, Petruzzelli L, Waldvogel R, Vögeli-Lange R and Meins F. 1998. Ethylene-responsive element binding protein (EREBP) expression and the transcriptional regulation of class I-1,3-glucanase during tobacco seed germination. *Plant Molecular Biology* 38: 785–795.

Li YC, Korol AB, Fahima T, Nevo E. 2004. Microsatellites within genes: Structure, function, and evolution. *Mol Biol Evol* 21: 991-1007.

Linton *et al.*. 2001. Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.

Lodge DJ. 2000. Ecto-or arbuscular mycorrhizas-which are best? *New Phytologist* 146: 353-354.

Olivier, J.-M., J. Guimberteau, J. Rondet & M. Mamoun, 1997. Vers l'inoculation contrôlée des cèpes et bolets comestibles ? *Revue forestière Française* XLIX (no spec.): 222-234.

Pachlewski R & Pchlevska J. 1974. Studies on symbiotic properties of mycorrhizal fungi of pine (*Pinus sylvestris* L.) with the aid of the method of mycorrhizal synthesis in pure cultures on agar. *PhD Thesis*, Forest Institute, Warsaw.

Malloch DW, Pirozynski KA & Raven PH. 1980. Ecological and evolutionary significance of mycorhizal symbioses in vascular plants. *Proceeding of the national Academy of Sciences USA* 77: 2113-2118.

Mankel A, Krause K, Genenger M, Kost G, Kothe E. 2000. A hydrophobin accumulated in the Hartig net of ectomycorrhiza formed between *Tricholoma terreum* and its compatible host tree is missing in an incompatible association. *Journal of Applied Botany* 74: 95–99.

Mankel A, Krause K and Kothe E. 2002. Identification of a Hydrophobin Gene That is Developmentally Regulated in the Ectomycorrhizal Fungus *Tricholoma terreum. Appl Environ Microbiol*. 68 (3): 1408-1413.

Martin F & Hilbert JL. 1991. Morphological, biochemical and molecular changes during ectomycorrhiza development. *Experimentia* 47 (4): 321-331.

Martin F, Selosse MA, Le Tacon F. 1999. The nuclear rDNA intergenic spacer of the ectomycorrhizal basidiomycete *Laccaria bicolor*: structural analysis and allelic polymorphism. *Microbiology* 145: 1605–1611

Martin F, Duplessis S, Ditengou F, Lagrange H, Voiblet C, Lapeyrie F. 2001. Developmental cross talking in the ectomycorrhizal symbiosis: signals and communication genes. *New Phytologist* 151: 145-154.

Martin F, Plassard C. 2001. Nitrogen assimilation by ectomycorrhizal symbiosis. In : *Nitrogen Assimilation by Plants*. Physiological, Biochemical and Molecular Aspect. (JF Morot-Gaudry, ed) Science publisher, Inc. Enfield, NH, USA, p. 169-184.

Martin F. 2006. Fair trade in the underworld: the ectomycorrhizal symbiosis. In: The Mycota Vol. III, 2nd edition (eds: RJ Howard & NAR Gow), Springer-Verlag Berlin Heidelberg 1994, 2006.

Martin F. 2007. Fair Trade in the Underworld: the Ectomycorrhizal Symbiosis. Biology of the Fungal Cell, 2nd Edition *The Mycota VIII.*

Martin F et Selosse MA. 2008. The *Laccaria* genome : a symbiont blueprint decoded. *New Phytologist* 180 : 296-310.

Martin F & Tunlid A. 2008. The ectomycorrhizal symbiosis: a marriage of convenience. In The Mycota. Vol. V, Plant Relationships, edn 2. Edited by Deising HB p.237-257.

Martin F & Nehls U. 2009. Harnessing ectomycorrhizal genomics for ecological insights. *Current Opinion in Plant Biology* 12: 508–515.

Marx DH, Ruehle JL, Kenny DS, Cordell CE, Riffle JW, Molina RJ, Pawuk WH, Navratil S, Tinus RW & Goodwin OC.1982. Commercial vegetative inoculums of *Pisolithus tinctorius* and inoculation techniques for development of ectomycorrhizae on container-grown tree seedlings. *Forest Science* 28: 373-400.

Matheny PB, Curtis JM, Hofstetter V, Aime MC, Moncalvo JM, Ge ZW, Yang ZL, Slot JC, Ammirati JF, Baroni TJ, Bougher NL, Hughes KW, Lodge DJ, Kerrigan RW, Seidl MT, Aanen DK, DeNitis M, Daniele GM, Desjardin DE, Kropp BR, Norvell LL, Parker A, Vellinga EC, Vilgalys R and Hibbett DS. 2007. Major clades of *Agaricales*: a multi-locus phylogenetic overview, *Mycologia* 98: 984–997.

Metzgar D, Bytof J, Wills C. 2000. Selection against frame-shift mutations limits microsatellite expansion in coding DNA. *Genome Res* 10: 72-80.

Molina R, Massicotte H & Trappe JM. 1992. Specificity phenomena in mycorrhizal symbioses : community-ecological consequences and pratical implications. *In Mycorrhizal functioning* – an integrated plant-fungal process (édité par M.F. Allen). Chapman and Hall, New York, p. 357-417.

Morel M, Jacob C, Kohler A, Johansson T, Martin F, Chalot M, Brun M. 2005. Identification of differentially-expressed genes in extra radical mycelium versus ectomycorrhizal roots in *Paxillus involutus/Betula pendula* ectomycorrhizal symbiosis. *Applied & Environmental Microbiology* 71: 382-391.

Mueller GM & Gardes. 1991. Intra- and interspecific relations within *Laccaria bicolor sensu lato*. *Mycological Research* 95: 592-601.

Mueller GM. 1992. Systematics of *Laccaria* (Agaricales) in the continental United states and Canada, with discussions on extralimital taxa and descriptions of extant types. *Fieldiana* 30: 1-158.

Mueller GJ, Mueller GM, Shih LH, Ammirati JF. 1993. Cytological studies in *Laccaria* (Agaricales). *American Journal of Botany* 80: 316–321.

Nehls U, Mikolajewski S, Ecke M & Hampp R. 1999. Identification and expression analysis of two fungal cDNAs regulated by ectomycorrhiza and fruit body formation. *New Phytologist* 144 (1): 195-202.

Nehls U, Grunze N, Willmann M, Reich M, Kuester H. 2007. Sugar for my honey: Carbohydrate partitioning in ectomycorhizal symbiosis. *Phytchemistry* 68: 82-91.

Nelson CD & Tauer CG. 1987. Genetic variation in juvenile characters of *Populus deltoides* Bartr. From Southern Great Plains. *Silv. Genet.* 36: 216-221.

Newcombe G, Martin F & Kohler A. 2009. Defense and nutrient mutualisms in *Populus*. *Plant Genetics and Genomics in Press*.

Nicolson TH. 1967. Vesicular-arbuscular mycorhiza a universal plant symbiosis. *Scientific Progress* (Oxford) 55: 561-581.

Ohta M, Ohme-Takagi M and Shinshi H. 2000. Three ethylene- responsive transcription factors in tobacco with distinct transactivation functions. *The Plant Journal* 22 : 29-38.

Pachlewski R, Pachlewska J. 1974. Studies on symbiotic properties of mycorrhizal fungi of pine (Pinus sylvestris L.) with the aid of the method of mycorrhizal synthesis in pure cultures on agar. PhD thesis. Forest Research Institute of Warsaw, Warsaw.

Peter M, Courty PE, Kohler A, Delaruelle C, Martin D, Tagu D, Frey-Klett P, Duplessis S, Chalot M, Podila G, Martin F. 2003. Analysis of expressed sequence tags from the ectomycorrhizal basidiomycete *Laccaria bicolor* and *Pisolithus microcarpus*. *New Phytologist* 159: 117-129.

Piotrowski JS, Lekberg Y,Harner MJ, Ramsey PW and Rillig MC. 2008. Dynamics of mycorrhizae during development of riparian forests along an unregulated river. *Ecography* 31: 245-253.

Pourtet J. 1951. Matériels pour faciliter la récolte des graines et l'élagage. *Revue Forestière Française* N° 6, p. 451.

Raffle VL, Anderson NA, Furnier GR & Doudrick RI. 1995. Variation in mating competence and random amplified polymorphic DNA in *Laccaria bicolor* (Agaricales) associated with three tree host species. *Canadian Journal of Botany* 73: 884-890.

Ramsay L, Macaulay M, Cardle L, Morgante M, Ivanissevich S, Maestri E, Powell W, Waugh R. 1999. Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. *Plant Journal* 17: 415-425.

Read DJ & Perez-Moreno J. 2003. Mycorrhizas and nutrient cycling in ecosystems - a journey towards relevance? *New Phytol* 157: 475–492.

Richter J, Kohler A, Morin E, Palme K, Martin F, Ditengou F and Legué V. 2009. The ectomycorrhizal fungus *L. bicolor* stimulates lateral root formation in *A. thaliana* and *P. tremula x P. alba* through auxin signaling pathways. *Plant Physiology*: submitted.

Rincón A, Gérard J, Dexheimer J, and Le Tacon F. 2001. Effect of an auxin transport inhibitor on aggregation and attachment processes during ectomycorrhiza formation between *Laccaria bicolor* S238N and *Picea abies*. *Can. J. Bot.* 79(10): 1152–1160.

Rincón A, Priha O, Sotta B, Bonnet M and Le Tacon F. 2003. Comparative effects of auxin transport inhibitors on rhizogenesis and mycorrhizal establishment of spruce seedlings inoculated with *Laccaria bicolor*. *Tree Physiology* 23(11): 785-791.

Rosado SCS, Kropp BR, Piche Y. 1994. Genetics of ectomycorrhizal symbiosis. II. Fungal variability and heritability of ectomycorrhizal traits. *New Phytol.* 126: 111-117.

Rupp LA & Mudge KW. 1995. Ethephon and auxin induce mycorrhiza-like changes in the morphology of root organ cultures of Mugo pine. *Physiologia Plantarum* 64: 316-322.

Salzer P, Hebe G & Hager A. 1997. Cleavage of chitinous elicitors from the ectomycorrhizal fungus *Hebeloma crustuliniforme* by host chitinases prevents induction of $K^+$ and $Cl^-$ release, extracellular alkalinization and $H_2O_2$ synthesis of *Picea abies* cells. *Planta* 203 (4): 470-479.

Sapp J. 1994. Evolution by Association. A History of Symbiosis. Oxford University Press, New York, Oxford, p 255.

Slankis V. 1950. Effect of α-Naphthaleneacetic Acid on Dichotomous Branching of Isolated Roots of Pinus silvestris. *Physiologia Plantarum* 3 (1): 40-44.

Schreiner RP, & Koide RT. 1993. Antifungal compounds from the roots of mycotrophic and non-mycotrophic plant species. *New Phytologist* 123: 99-103.

Seddas P, Gianinazzi-Pearson V, Schoefs B, Küster H and Wipf D. 2009. Communication and Signaling in the Plant–Fungus Symbiosis: The Mycorrhiza. *Plant-Environment Interactions* 10: 45-71.

Selosse MA, & Le Tacon F, 1998. The land flora: a prototroph-fungus partnership? *Trends in ecology and Evolution* 13: 15-20.

Selosse MA, Bouchard D, Martin F, Le Tacon F, 2000. Survival after outplanting and effect of two *Laccaria bicolor* strains inoculated on Douglas-Fir. *Can J For Res* 30: 360-371.

Settler RF, Zsuffa L & Wu R. 1996. The role of hybridization in the genetic manipulation of *Populus* and its implications for management and conservation. *NRC Research Press*. 87-112.

Shepperd WD, Brinkley D, Bartos DL, Stohlgren TJ and Eskew LG, compilers, 2001. Sustaining Aspen in Western Landscapes: Symposium proceedings, June 13-15, Grand Junction, Colorado. p 459.

Sjödin A, Street NR, Sandberg G, Gustafsson P and Jansson S. 2009. PopGenIE : The *Populus* Genome Integrative Explorer. A new tool for exploring the *Populus* genome. *New Phytologist* 182 (4): 1013 – 1025

Smith SE & Read DJ : Mycorrhizal Symbiosis. edn 3. Academic Press; 2008.

A balanced, lucid overview of the past, present and future of mycorrhizal research.

Smith SE & Read DJ. 1996. Mycorrhizal Symbiosis 2nd edn (Academic, London, 1996).

Solano R and Ecker JR. 1998. Ethylene gas: perception, signaling and response. *Current Opinion in Plant Biology* 1: 393-398.

Southern EM. 1975. *Detection of specific sequences among DNA fragments separated by gel*
*electrophoresis. J Mol Biol* **98** (3): p. 503-17.

Stettler FS & Fenn RC. 1988. *P. deltoides* x *P. trichocarpa* hubrids for short rotation culture : variation patterns and 4-year field performance. *Can. J. For. Res.* 18: 745-753.

Stettler FS, Zsuffa L & Wu R. 1996.The role of hybridization in the genetic manipulation of *Populus* and its implications for management and conservation. Edité par Stettler RF, Bradshaw HDJ, Heilman PE & Hinckley TM. Ottawa ON, Canada: NRC Research Press p. 542.

Street NS, Skogström O, Sjödin A, Tucker J, Rodríguez-Acosta M, Nilsson P, Jansson S and Taylor G. 2006. The genetics and genomics of the drought response in *Populus. The Plant Journal* 48: 321-341.

Tagu D & Martin F. 1996. Molecular analysis of cell wall proteins expressed during the early steps of ectomycorrhiza development. *New Phytologist* 133 (1): 73-85.

Tagu D, De Bellis R, Balestrini R, De Vries OMH, Piccoli G, Stocchi V, Bonfante P, Martin F. 2001. Immunolocalization of hydrophobin HYDPt-1 from the ectomycorrhizal basidiomycete *Pisolithus tinctorius* during colonization of *Eucalyptus globulus* roots. *New Phytologist* 149: 127-135.

Tagu D, Lapeyrie F, Martin F. 2002. The ectomycorrhizal symbiosis: genetics and development. *Plant & Soil* 244: 97-105.

Tagu D, Palin B, Balestrini R, Gelhaye E, Lapeyrie F, Jacquot JP, Sautiere PE, Bonfante P, Martin F. 2003. Characterization of a symbiosis- and auxin- regulated glutathion S-transferase from *Eucalyptus globulus* roots. *Plant physiology & Biochemistry* 41: 611-618.

Thielges BA. 1985. Breeding poplars for disease resistance. *F.A.O. Forestry papers* Vol. 56 Rome, Italy.

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, dePamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K Goodstein

D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D. 2006. The genome of black cottonwood, *Populus trichocarpa*. *Science* 313: 1596-1604.

Umata H. 1999. Germination and growth of *Erythrorchis ochobiensis* (Orchidaceae) accelerated by monokaryons and dikaryons of *Lenzites betulinus* and *Trametes hirsute*. *Mycoscience* 40: 367-371.

Venter *et al.* 2001. The sequence of human genome. *Science* 292: 1838-1851.

Viart M. 1999. Un demi-siècle de populiculture en France, deuxième partie : les changements intervenus de 1947 à 1997 et leurs conséquences. *La Forêt Privée* 246: 27-42.

Voiblet C, Martin F. 2000. Identifying symbiosis-regulated genes in *Eucalyptus globulus-Pisolithus tinctorius* ectomycorhiza using suppression substractive hybridization and cDNA arrays. In: *Biology of Plant-Microbe Interactions* (PGM de Wit, T Bisseling, W Stiekema, eds), International Society for molecular Plant-Microbe interactions, St Paul, Minnesota, p. 208-213.

Voiblet C, Duplessis S, Encelot N, Martin F. 2001. Identification of symbiosis-regulated genes in *Eucalyptus globulus-Pisolithus tinctorius* ectomycorrhiza by differential hybridization of arrayed sDNAs. *The Plant Journal* 25: 181-191.

VanOoijen JW, Voorrips RE. 2001. JoinMap 3.0, software for the calculation of genetic linkage maps. *Plant Research International*, Wageningen, The Netherlands. Website: http://www.plant.wageningen-ur.nl.

Walbot V, Petrov DA. 2001. Gene galaxies in the maize genome. *Proc Natl Acad Sci USA* 98: 8163-8164.

Weiss MS, Milkolajewski S, Peipp H, Schmitt U,Schmidt J, Wray V & Strack D. 1997. Tissue-specific and development-dependent accumulation of phenylpropanoids in larch mycorrhizas. *Plant physiology* 114: 15-27.

Williams BD and Johnston RS. 1984. Natural Establishment of Aspen from Seed on a Phosphate Mine Dump. *Journal of Range Management* 37 (6): 521-522.

Xu JR, Peng YL, Dickman MB, Sharon A. 2006. The dawn of fungal pathogen genomics. *Annu Rev Phytopathol* 44: 337-366.

Yin TM, Difazio SP, Gunter Lee, Riemenschneider D, Tuskan GA. 2004. Large-scale heterospecific segregation distortion in *Populus* by a dense genetic map. *Theor Appl Genet* 109: 451-463.

Ying CC & Bagley WT. 1976. Genetic variation of eastern cottonwood in an eastern Nebraska provenance study. *Silv. Genet.* 25: 67-73.

# Annexes

**Liste des génotypes de peuplier utilisés dans l'expérience 1 – 2007 :**

661300228
661300229
661300244
661300249
661300258
661300262
661300271
661300279
661300286
661300303
661300315
661300316
661300324
661300331
661300338
661300353
661300367
661300373
661300377
661300400
661300406
661300416
661300417
661300446
661300453
661300454
661300463
661300466
661300474
661300483
661300489
661300504
661300506
661300519
661300526
661300534
661300535
661300566
661300570
661300577
101-74
73028-62

**Liste des génotypes de peuplier utilisés dans l'expérience 2 – 2008 :**

| | | | |
|---|---|---|---|
| 661300306 | 661300364 | 661300419 | 101-74 |
| 661300307 | 661300365 | 661300420 | 73028-62 |
| 661300308 | 661300366 | 661300422 | 661300237 |
| 661300309 | 661300367 | 661300423 | 661300238 |
| 661300311 | 661300368 | 661300424 | 661300240 |
| 661300312 | 661300369 | 661300425 | 661300241 |
| 661300313 | 661300370 | 661300426 | 661300242 |
| 661300314 | 661300371 | 661300427 | 661300243 |
| 661300315 | 661300372 | 661300428 | 661300244 |
| 661300316 | 661300373 | 661300429 | 661300246 |
| 661300317 | 661300374 | 661300430 | 661300247 |
| 661300318 | 661300375 | 661300431 | 661300248 |
| 661300319 | 661300376 | 661300432 | 661300249 |
| 661300321 | 661300377 | 661300433 | 661300250 |
| 661300322 | 661300378 | 661300436 | 661300251 |
| 661300323 | 661300379 | 661300437 | 661300252 |
| 661300324 | 661300380 | 661300438 | 661300254 |
| 661300326 | 661300381 | 661300439 | 661300255 |
| 661300327 | 661300383 | 661300440 | 661300256 |
| 661300328 | 661300384 | 661300441 | 661300258 |
| 661300329 | 661300385 | 661300442 | 661300260 |
| 661300330 | 661300386 | 661300443 | 661300262 |
| 661300331 | 661300387 | 661300444 | 661300263 |
| 661300332 | 661300388 | 661300445 | 661300267 |
| 661300333 | 661300389 | 661300446 | 661300270 |
| 661300334 | 661300390 | 661300447 | 661300271 |
| 661300335 | 661300391 | 661300448 | 661300272 |
| 661300336 | 661300392 | 661300449 | 661300275 |
| 661300337 | 661300393 | 661300450 | 661300276 |
| 661300338 | 661300394 | 661300451 | 661300277 |
| 661300339 | 661300395 | 661300453 | 661300278 |
| 661300340 | 661300396 | 661300454 | 661300279 |
| 661300341 | 661300397 | 661300455 | 661300280 |
| 661300342 | 661300398 | 661300456 | 661300281 |
| 661300343 | 661300399 | 661300457 | 661300282 |
| 661300344 | 661300400 | 661300458 | 661300283 |
| 661300345 | 661300401 | 661300459 | 661300284 |
| 661300346 | 661300402 | 661300460 | 661300286 |
| 661300347 | 661300403 | 661300461 | 661300287 |
| 661300348 | 661300405 | 661300462 | 661300288 |
| 661300349 | 661300406 | 661300463 | 661300289 |
| 661300350 | 661300407 | 661300464 | 661300292 |
| 661300351 | 661300408 | 661300465 | 661300293 |
| 661300352 | 661300409 | 661300466 | 661300295 |
| 661300353 | 661300410 | 661300467 | 661300297 |
| 661300355 | 661300411 | 661300468 | 661300298 |
| 661300356 | 661300413 | 661300469 | 661300299 |
| 661300357 | 661300414 | 661300470 | 661300301 |
| 661300358 | 661300415 | 661300471 | 661300302 |
| 661300360 | 661300416 | 661300472 | 661300303 |
| 661300361 | 661300417 | 661300473 | 661300304 |

661300362
661300475
661300476
661300477
661300478
661300479
661300480
661300481
661300482
661300483
661300484
661300485
661300486
661300487
661300488
661300489
661300490
661300491
661300492
661300493
661300494
661300495
661300496
661300497
661300498
661300499
661300500
661300501
661300502
661300503
661300504
661300505
661300506
661300507
661300508
661300509
661300510
661300511
661300512
661300513
661300514
661300515
661300516
661300517
661300518
661300519
661300520
661300521
661300522
661300523
661300524
661300525
661300526
661300418
661300527
661300528
661300529
661300530
661300531
661300532
661300533
661300534
661300535
661300536
661300537
661300538
661300539
661300540
661300541
661300542
661300543
661300544
661300545
661300546
661300547
661300548
661300549
661300550
661300551
661300552
661300553
661300554
661300555
661300556
661300557
661300558
661300559
661300560
661300561
661300562
661300563
661300564
661300565
661300566
661300474

**Contribution to the study of the structure and the polymorphism of the genome of the ectomycorrhizal basidiomycete *Laccaria bicolor* (Maire) Orton and identification of QTLs of mycorrhization in two Poplars, *Populus trichocarpa* Torr. & A. Gray ex Hook. and *Populus deltoides* (Bartr.) Marsh.**

The mycorrhizal symbioses between fungi and roots concern 95 % of the plant species. Social trees of boreal and temperate forests form a particular type of root association with fungi: the ectomycorrhizal symbiosis. Ectomycorrhizas play a major role in tree hydromineral nutrition, nutrient cycles and primary production. However, their complexity have so far prevented from deciphering their precise function and role. The recent availability of the genome of the ectomycorrhizal fungus *Laccaria bicolor* and that of the host-tree *Populus trichocarpa* provides an unprecedented opportunity to decipher the key components of development and functioning of this symbiosis. The aims of this study were to participate to the characterization and deciphering of the genome of *L. bicolor*, and to determine the genes involved in the formation of ectomycorrhizas in both partners. Firstly, in order to facilitate the assembly of the genomic sequence of *L. bicolor*, we have identified the repeated sequences and generated a genetic map. On the 60 Mb of this genome, 8 % are microsatellite sequences and 24 % transposable elements. A genetic map was built from 111 monokaryons issued from *L. bicolor* S238N. This map includes 326 markers (8 RAPD, 243 AFLP, 59 SSR and 14 SNP) distributed on 10 linkage groups anchored onto the genomic sequence of *L. bicolor*. Secondly, we have identified the genes involved in the establishment of ectomycorrhizas in poplar by combining QTL detection and DNA microarrays. We targeted 81 genes which can be involved in the establishment and/or the functioning of the symbiosis.

**Keywords:** ectomycorrhizal symbiosis, genome organization, genetic mapping and QTL detection

**Contribution à l'étude de la structure et du polymorphisme du génome du basidiomycète ectomycorhizien *Laccaria bicolor* (Maire) Orton et identification de QTLs de mycorhization chez les Peupliers, *Populus trichocarpa* Torr. & A. Gray ex Hook. et *Populus deltoides* (Bartr.) Marsh.**

Les symbioses mycorhiziennes, entre champignons et racines de plantes, concernent 95 % des espèces végétales. Les arbres sociaux des forêts boréales et tempérées forment avec les champignons un type particulier d'association : la symbiose ectomycorrhizienne. Les ectomycorrhizes jouent un rôle essentiel dans la nutrition hydrominérale des arbres, le cycle des éléments minéraux et la production primaire. Cependant, leur complexité n'a pas permis à ce jour de déchiffrer leurs rôles et leurs fonctions précises. La récente disponibilité du génome du champignon ectomycorhizien *Laccaria bicolor* et de celui de l'arbre hôte *Populus trichocarpa* fournit une occasion inégalée d'approfondir nos connaissances du développement et du fonctionnement de cette symbiose. Les objectifs de cette étude ont donc été de participer à la caractérisation et au décryptage du génome de *L. bicolor* puis à la recherche des gènes impliqués dans la formation des ectomycorhizes chez les deux partenaires. Dans un premier temps et afin d'aider à l'assemblage de la séquence génomique de *L. bicolor*, nous avons identifié les séquences répétées et construit une carte génétique. Sur les 60 Mb de ce génome, nous avons mis en évidence 8 % de séquences microsatellitaires et 24 % d'éléments transposables. Une carte génétique a été construite à partir de 111 monocaryons issus de *L. bicolor* S238N. Cette carte comprend 326 marqueurs (8 RAPD, 243 AFLP, 59 SSR et 14 SNP) répartis sur 10 groupes de liaison ancrés à la séquence génomique de *L. bicolor*. Dans un second temps, nous avons tenté d'identifier les gènes impliqués dans l'établissement des ectomycorhizes chez le peuplier en combinant une approche de détection par QTL et par puces à ADN. Nous avons ciblé 81 gènes potentiellement impliqués dans l'établissement et/ou le fonctionnement de la symbiose.

**Mots clés :** Symbiose ectomycorhizienne, organisation du génome, cartographie génétique et détection de QTL