



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>



Faculté des Sciences et Technologies,
Ecole Doctorale Ressources Procédés Produits Environnement (RP2E)

THESE

Présentée pour l'obtention du titre de
Docteur de l'Université Henri Poincaré, Nancy I
En Biologie Végétale et Forestière

Par **Emilie TISSERANT**

Analyse bio-informatique du transcriptome des champignons mycorhiziens *Tuber melanosporum* et *Glomus intraradices*

Membres du Jury :

Rapporteurs :

Dr. Elodie Gaulin, Maître de Conférences, Université Paul Sabatier, Toulouse
Dr. Franck Panabières, Directeur de Recherches, INRA Sophia Antipolis

Examineurs :

Pr. Christophe Roux, Professeur, Université Paul Sabatier, Toulouse
Pr. Pierre Leblond, Professeur, Université Henri Poincaré, Nancy
Dr. Marie-Dominique Devignes, Chargée de recherche au LORIA, Nancy
Dr. Francis Martin, Directeur de Recherches, INRA de Nancy (Directeur de thèse)

Remerciement

Je tiens tout d'abord à remercier Elodie Gaulin, Franck Panabières, Pierre Leblond, Christophe Roux et Marie-Dominique Devignes, qui me font l'honneur de juger ce travail.

Je tiens à exprimer ma vive reconnaissance à Francis pour m'avoir accueillie au sein du laboratoire, pour la confiance et la patience qu'il m'a accordée, ainsi que pour sa passion pour la recherche qu'il a su partager.

Un grand merci à Manue, pour son soutien bio-informatique qui a été essentiel à ce travail, mais surtout pour son soutien moral et sa sympathie.

Je remercie Annegret pour ses précieux conseils et son enthousiasme qu'elle a su me transmettre.

Je tiens à remercier Sébastien sans qui je n'aurais pas pu réalisé cette thèse et avec qui j'ai apprécié travailler sur le coté sombre de la symbiose.

Je tiens à adresser mes plus sincères remerciements à tous les membres du laboratoire pour leur accueil chaleureux et leur gentillesse. Un merci particulier à Adeline, dont la compagnie m'a permis de surmonter les moments difficiles.

Enfin, Je remercie ma famille et mes amis, et plus particulièrement Julien, pour m'avoir soutenue et supportée durant ces trois années.

Sommaire

Introduction	1
Chapitre I : Analyse bibliographique	5
I - Contexte biologique.....	5
1 - La symbiose mycorhizienne	5
1.1 - Développement de la symbiose mycorhizienne arbusculaire.....	7
1.2 - Développement de la symbiose ectomycorhizienne.....	11
2 - Génomique des symbioses mycorhiziennes.....	12
2.1 - Les génomes des champignons ectomycorhiziens	13
2.2 - La génomique chez les champignons mycorhiziens arbusculaires	15
3 - Transcriptomique des symbioses mycorhiziennes	16
3.1 - Les échanges de nutriments	18
3.3 - Echapper au système de surveillance de la plante hôte	24
II - Contexte génomique et bio-informatique	25
1 - La génomique.....	25
2 - La transcriptomique	27
2.1 - Séquençage d'ESTs.....	27
2.2 - Puce à ADN	28
2.3 - Application des NGS à la transcriptomique	29
3 - Les nouvelles technologies de séquençage	31
3.1 - Roche/454.....	31
3.2 - Illumina/Solexa.....	33
3.3 - ABI SOLiD Applied Biosystem.....	33
4 - Les outils bio-informatiques pour l'analyse des transcriptomes	35
4-1 - Assemblage <i>de novo</i> des données NGS	35
4-2 - Alignement à un génome de référence.....	41
Chapitre II : Analyse globale du génome et du transcriptome de <i>T. melanosporum</i>	45
Chapitre III : Analyse globale du transcriptome de <i>Glomus intraradices</i>.....	85
Chapitre IV : Discussion générale	115
I - Le transcriptome de <i>Tuber melanosporum</i> par RNA-Seq	115
II - Caractérisation du transcriptome de <i>Glomus intraradices</i>	119
Chapitre V : Conclusions et perspectives.....	123
Liste des publications	127
Références bibliographiques	129

Introduction

Les plantes constituent la base de la productivité dans les écosystèmes naturels et agricoles. En outre, elles sont des sources potentielles de médicaments et d'énergie renouvelable ainsi que les moteurs clés de la séquestration du carbone. Leur capacité à s'associer avec différents partenaires pour former des associations mutualistes, ou symbioses, leur permet d'exploiter de nombreuses niches écologiques. Ainsi pour conquérir le milieu terrestre, certains végétaux s'associèrent, il y a plusieurs centaines de millions d'années, à divers organismes (Selosse et Le Tacon, 1998). Certaines algues s'associent avec des champignons pour former les lichens. D'autres plantes, telles que les légumineuses forment au niveau de leurs racines des nodosités abritant des bactéries du genre *Rhizobium* capables de fixer l'azote atmosphérique. La symbiose mycorhizienne, associant certains champignons du sol aux racines de près de 95% des végétaux terrestres vasculaires, est une des formes les plus fréquentes (Smith et Read, 2008). Cette association avec les champignons mycorhiziens permet à la plante d'exploiter efficacement les ressources minérales du sol, limitantes dans la plupart des écosystèmes naturels. Ces interactions mutualistes à bénéfices réciproques reposent sur un échange équitable de nutriments, les plantes offrant à leurs symbiotes un apport constant de carbone *via* le sucrose issu de l'activité photosynthétique. En promouvant la croissance des plantes *via* une nutrition améliorée, ces associations mycorhiziennes jouent donc un rôle clé dans la productivité des écosystèmes. Elles contribuent ainsi au cycle du carbone dans les écosystèmes terrestres (Smith et Read, 2008). Ce rôle majeur dans le bon fonctionnement des écosystèmes justifie pleinement l'étude de leur développement et fonctionnement, ainsi que leur évolution.

Les approches génétiques et moléculaires ont démontré que le développement et le fonctionnement de cette symbiose étaient sous le contrôle génétique des deux partenaires (Martin et al, 2007). Ce contrôle semble s'exercer dès le processus de reconnaissance entre les partenaires, lors de la mise en place des structures symbiotiques et enfin, lors de la réorganisation conjointe des métabolismes de la plante-hôte et du champignon. Il apparaît que les mécanismes moléculaires sous-jacents sont basés sur l'expression coordonnée de réseaux de gènes, eux même orchestrés par des gènes maîtres et des molécules de signalisation (Martin & Nehls, 2009 ; Veneault-Fourrey et Martin, 2011). L'étude de l'expression des gènes et de leur régulation est donc une base importante pour l'étude de la symbiose mycorhizienne.

Récemment, le séquençage de génomes de plusieurs plantes et champignons impliqués dans la symbiose a donné lieu à des avancées significatives dans la compréhension des interactions mycorhiziennes (Tuskan et al, 2006 ; Martin et al, 2008a ; Martin et al, 2010a). A travers la génomique comparative, ces séquences permettent d'identifier des traits moléculaires spécifiques de la symbiose mycorhizienne, tels que des répertoires de gènes spécifiques, des éléments structuraux et des régions régulatrices conservés. Ces séquences génomiques permettent également d'étudier les patrons d'expression génique à l'échelle du génome et donc de mettre à jour des dizaines de gènes régulés par la symbiose ainsi que des réseaux de gènes à expression coordonnée. Ces approches ont récemment permis l'identification de gènes nécessaires au développement de la symbiose (par exemple MiSSP7 chez *L. bicolor* ; Plett et al, 2011). Par ailleurs, le développement récent des technologies de séquençage à haut débit (454, Illumina) permet maintenant d'envisager le séquençage massif du transcriptome de dizaines d'organismes non modèles et le développement de nouvelles méthodes de quantification numérique de l'expression des gènes à une échelle et une précision jamais atteinte (Marguerat, 2010). Ces technologies de séquençage offrent de nouvelles perspectives pour la compréhension des mécanismes moléculaires gouvernant les symbioses mycorhiziennes.

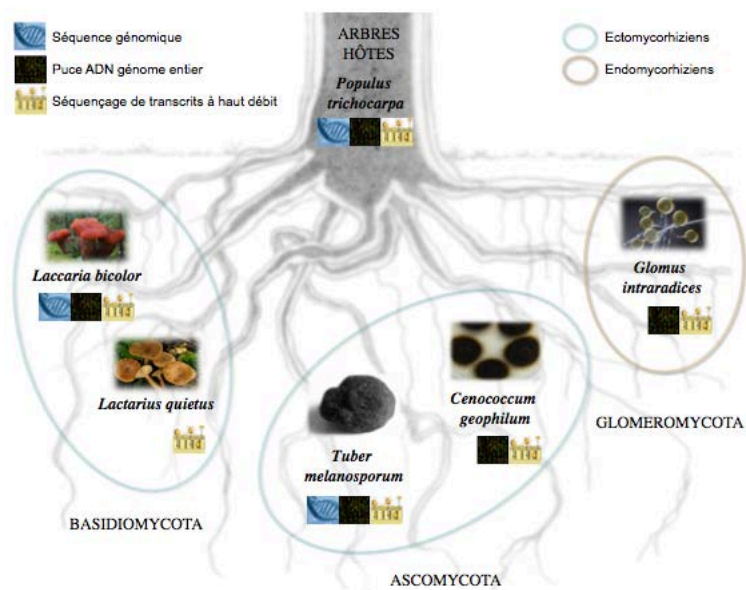


Figure 1 : Ressources génomiques et transcriptomiques acquises au sein du laboratoire 'Ecogénomique des Interactions' (UMR IaM).

Objectifs de la thèse

Le projet que j'ai réalisé au sein de l'équipe 'Ecogénomique des Interactions' de l'UMR Interactions Arbres/Micro-organismes s'inscrit dans le cadre de la compréhension des mécanismes moléculaires qui sous-tendent le développement et le fonctionnement des interactions symbiotiques mycorhiziennes. Ce projet de thèse vise à caractériser les patrons d'expression génique lors de la symbiose induite par plusieurs espèces de champignons mycorhiziens. Cette étude repose sur la caractérisation *in silico* de plusieurs transcriptomes d'espèces mycorhiziennes (*Tuber melanosporum*, *Glomus intraradices*, *Laccaria bicolor*, *Cenococcum geophilum* et *Lactarius quietus*).

D'un point de vue méthodologique, le principal objectif de ce travail est l'exploitation *in silico* des ressources transcriptomiques acquises sur les différentes symbioses mycorhiziennes (Figure 1). En effet, l'utilisation de méthodes et d'outils bio-informatiques optimisés est indispensable pour traiter la masse considérable de données générées par les approches de génomique et de transcriptomique à haut débit (plusieurs dizaines de millions de séquences). Il s'agit notamment de filtrer, stocker et faciliter la gestion de ces données dans des bases de données dédiées ; d'évaluer et de mettre en place les algorithmes et outils nécessaires à leur analyse rapide ; de développer des scripts permettant la manipulation et l'interrogation de ces données, ainsi que l'exécution et l'analyse automatique des résultats générés ; et enfin d'appliquer ces différents outils d'analyse et d'intégrer les données provenant de différentes sources pour en extraire des connaissances biologiques et identifier les gènes d'intérêt.

En ce qui concerne la biologie de ces systèmes, les objectifs de ce projet de thèse sont :

- L'analyse du transcriptome de la Truffe du Périgord (*T. melanosporum*) par séquençage de transcrits à très haut débit (RNA-Seq) afin d'améliorer l'annotation structurale de son répertoire de gènes, d'accéder à la complexité fonctionnelle de ses transcrits et de caractériser les modifications dans l'expression génique au cours des phases successives de son cycle de vie. Le **Chapitre II** regroupe dans trois publications l'ensemble des résultats obtenus sur le transcriptome de *T. melanosporum*.
- La caractérisation du transcriptome du symbiote endomycorhizien *Glomus intraradices* afin d'identifier son répertoire de gènes et d'identifier les transcrits induits lors de l'établissement de la symbiose. Le **Chapitre III** présente les résultats de ce travail.

Chapitre I : Analyse bibliographique

Chapitre I : Analyse bibliographique

I - Contexte biologique

1 - La symbiose mycorhizienne

La relation symbiotique mutualiste entre les racines des plantes et les champignons du sol, appelée symbiose mycorhizienne, est l'une des associations les plus prévalentes au sein des écosystèmes terrestres (Smith et Read, 2008). En effet, les champignons mycorhiziens forment des symbioses avec près de 95% des plantes terrestres, colonisant des milieux tels que les forêts boréales, tempérées et tropicales ainsi que les toundras, les prairies et de nombreuses terres cultivées (Read et Perez-Moreno, 2003). Lors de cette interaction, les réseaux d'hyphes fongiques, spécialisés dans l'absorption des minéraux du sol et de l'azote organique, agissent comme une extension du système racinaire, augmentant ainsi l'assimilation par la plante de nutriments essentiels tels que le phosphore, l'azote, le soufre et l'eau (Smith et Read, 2008). A travers cette symbiose, avec les champignons mutualistes, les espèces végétales sont en mesure d'acquérir des capacités métaboliques leur permettant l'utilisation de niches écologiques inaccessibles auparavant (Read et Perez-Moreno, 2003). En outre, les champignons protègent leurs hôtes contre les stress abiotiques (pollution par les métaux, sécheresse) et biotiques (infection par un pathogène) (Smith et Read, 2008). En contrepartie, les champignons obtiennent des sucres dérivés de la photosynthèse. Jusqu'à 20% des éléments carbonés issus des plantes terrestres sont alloués aux champignons mycorhiziens (Högberg et al, 2001). Ainsi, cette symbiose contribue fortement au budget global du cycle du carbone au sein des écosystèmes (Read et Perez-Moreno, 2003).

Selon la position phylogénétique de leurs partenaires et selon leurs structures symbiotiques, plusieurs types de mycorhizes sont définis. Les deux formes de symbiose mycorhizienne les plus représentées sont la symbiose mycorhizienne arbusculaire (AM) et la symbiose ectomycorhizienne (ECM). Lors de la symbiose ECM, le partenaire fongique colonise les espaces intercellulaires des racines (apoplaste), tandis que lors de la symbiose AM, une partie des hyphes se développe à l'intérieur des cellules végétales (Figure 2) (Smith et Read, 2008).

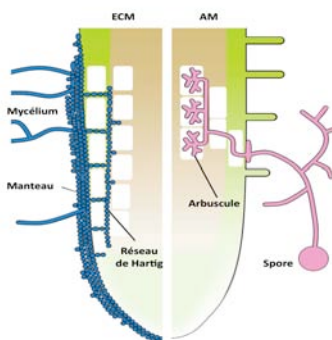


Figure 2 : Structures de colonisation racinaire de la symbiose ectomycorhizienne (bleu) et de la symbiose mycorhizienne arbusculaire (rose). Les hyphes des champignons ectomycorhiziens s'agregent autour de l'extrémité racinaire pour former le manteau et se développent entre les cellules épidermiques pour former le réseau de Hartig. Les hyphes des champignons mycorhiziens arbusculaires se développent entre et à travers les cellules végétales et pénètrent à l'intérieur des cellules corticales internes pour former des arbuscules. (Bonfante et Genre, 2010).

La symbiose AM est probablement la symbiose terrestre la plus répandue. En effet, celle-ci est associée à plus de 80% des espèces de plantes terrestres actuelles, essentiellement des espèces herbacées, et notamment des espèces cultivées importantes, telles que le blé, le riz, le maïs et le soja (Smith et Read, 2008). Cependant, il n'existe qu'environ 160 espèces de champignons formant des symbioses AM, toutes appartenant au phylum des Gloméromycètes, un clade monophylétique frère des Ascomycètes et des Basidiomycètes (Figure 3) (Schübler et al, 2001). Des données fossiles apportent la preuve de l'existence de cette symbiose il y a plus de 400 millions d'années, ce qui coïncide avec l'apparition des premières plantes terrestres (Selosse et Le Tacon, 1998). Il a donc été suggéré que les champignons AM auraient joué un rôle crucial dans la colonisation du milieu terrestre par les plantes. Ceci pourrait expliquer la distribution quasi ubiquitaire de cette symbiose AM dans le règne végétal ainsi qu'au sein des écosystèmes.

Figure 3 : Arbre phylogénétique représentant les Glomérormycètes au sein des autres lignées fongiques. Tous les champignons AM appartiennent au clade des Glomérormycètes et, mis à part *Geosiphon pyriformis*, tous les membres des Glomérormycètes forment une symbiose AM. Les clades non monophylétiques des Chytridiomycètes et des Zygomycètes sont représentés en vert et en bleu, respectivement. (Parniske, 2008).

La symbiose mycorhizienne joue un rôle clé dans la bioséquestration du carbone, le cycle des nutriments, la biodiversité végétale et la productivité des écosystèmes naturels et agricoles. Les symbiotes mutualistes mycorhiziens sont donc d'une importance majeure en tant que moteurs des fonctions écologiques et des processus évolutifs dans les écosystèmes terrestres.

1.1 - Développement de la symbiose mycorhizienne arbusculaire

Les champignons AM sont des biotrophes obligatoires qui dépendent de leur association avec les racines des plantes pour compléter leur cycle de vie. La colonisation des racines, vitale pour ces champignons, suit une série d'étapes distinctes. Les champignons AM existent dans le sol sous forme de spores. Le développement fongique commence par la germination des hyphes, à partir des spores, qui explorent le sol à la recherche d'une racine hôte. En l'absence d'hôte, la croissance des hyphes est limitée par la quantité de carbone stockée dans les spores sous forme de lipides qui alimentent les hyphes (Bécard et Piché, 1989 ; Bago et al, 2000). Les hyphes arrêtent alors leur croissance et rétractent leur masse cytoplasmique dans la spore, qui entre de nouveau en dormance (Bécard et al, 2004). Des cycles successifs de germination des spores peuvent se produire chez les champignons AM. La perception de la plante hôte par le mycélium se fait *via* des exsudats racinaires diffusés sur une courte distance et rapidement dégradés, qui ont récemment été identifiés comme étant des strigolactones (Akiyama et al, 2005). Ces composés induisent une réaction pré-symbiotique fongique caractérisée par une croissance des hyphes en continu, une augmentation de l'activité physiologique et mitochondriale ainsi qu'une ramification abondante des hyphes qui augmente ainsi les chances d'une rencontre avec l'hôte. En retour, les spores germées produisent des signaux diffusibles, appelés facteurs Myc, tels que des lipochitoooligosaccharides récemment identifiés, qui sont perçus par les racines des plantes même en l'absence de contact physique avec le champignon (Maillet et al, 2011). Ces signaux, qui stimulent la croissance et la ramification racinaires, activent chez la plante une voie de signalisation commune à celle déclenchée par la symbiose entre les bactéries rhizobium et les plantes légumineuses (Figure 4) (Kistner et al, 2005). La symbiose mycorhizienne arbusculaire étant apparue avant la symbiose rhizobium fixatrice d'azote, la voie de transduction du signal mycorhizienne pourrait avoir été recrutée par les bactéries fixatrices d'azote (Kistner et Parniske, 2002).

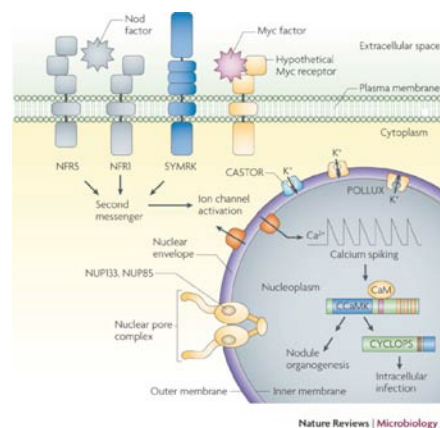
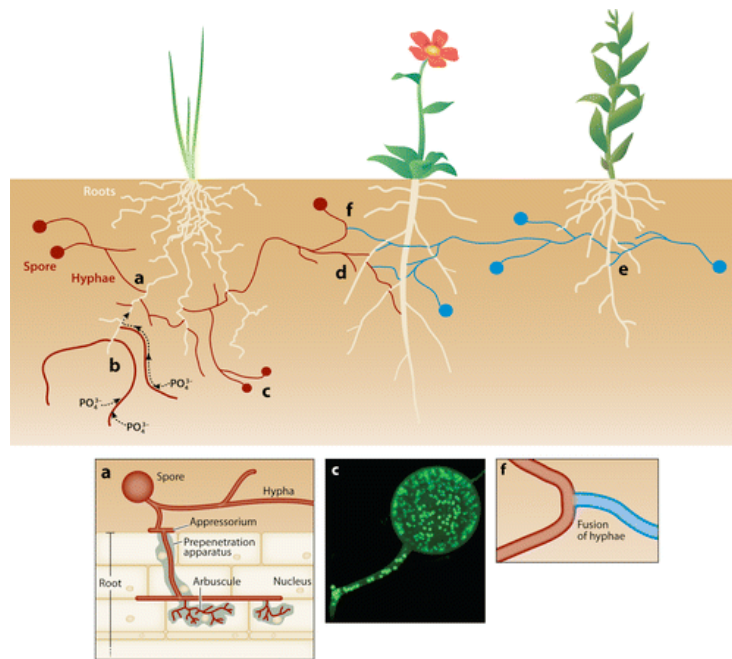


Figure 4 : Voie de signalisation commune aux symbioses rhizobiennes et mycorhiziennes arbusculaires. La perception des signaux dérivés des champignons AM et des bactéries rhizobiums déclenche une voie de transduction du signal, qui implique au moins sept gènes nécessaires pour le développement correct des deux symbioses (gènes SYM). Les facteurs Myc sont reconnus par la plante *via* un récepteur encore non caractérisé. La perception de ces facteurs induit des oscillations périodiques et transitoires de la concentration de calcium dans le cytoplasme de la plante. Un récepteur membranaire SYMRK détecte les signaux fongiques et les transmet au cytoplasme par la phosphorylation d'un substrat encore inconnu. Ce signal est rapidement transmis au noyau. Deux nucléoporines NUP85 et NUP133, ainsi que les deux protéines CASTOR et POLLUX codant des canaux de potassium localisés dans l'enveloppe nucléaire, sont nécessaires pour les oscillations calciques. Au sein du noyau, une protéine kinase calcium et calmoduline dépendante, CCaMk, forme un complexe avec un substrat de phosphorylation, CYCLOPS. Ce complexe pourrait être responsable du décodage des oscillations calciques et provoquer la régulation des gènes en aval. (Parniske, 2008)



 Sanders IR, Croll D. 2010.
Annu. Rev. Genet. 44:271–92

Figure 5 : Développement de la symbiose AM. (a) Les spores du champignon germent dans le sol. Les hyphes explorent le sol pour trouver une racine hôte. Le champignon pénètre dans la racine et, après avoir formé un appressorium, il colonise le cortex racinaire et forme des arbuscules à l'intérieur des cellules végétales. Ces structures permettent le transfert des nutriments entre les deux partenaires. (b) Le champignon produit alors des hyphes qui se propagent dans le sol et absorbent les nutriments pour les transférer à la plante. (c). Ces hyphes peuvent produire des spores contenant de nombreux noyaux. (d) Les hyphes peuvent également coloniser de nouvelles plantes, créant ainsi un réseau d'hyphes. (e) Des individus génétiquement différents peuvent coloniser les plantes voisines ou la même plante. (f) Les hyphes d'individus génétiquement différents peuvent fusionner. (Encart (c) : les noyaux sont marqués par fluorescence). (Sanders et Croll, 2010).

Après le premier contact physique entre les hyphes et la plante, le champignon forme un appressorium ou hyphopodium sur la surface de la racine par lequel il colonise l'espace intercellulaire du cortex racinaire (Genre et al, 2005). La cellule végétale forme une structure subcellulaire appelée « prepenetration apparatus » (PPA) qui prédétermine le chemin de la croissance de l'hyphe à travers la cellule végétale (Figure 5). C'est seulement après la différenciation de ce pont cytoplasmique que l'hyphe fongique pénètre dans la cellule hôte (Genre et al, 2005).

Ensuite, le champignon traverse les couches de cellules externes, se propage longitudinalement dans le cortex interne et forme des hyphes ramifiés dichotomiques à l'intérieur des cellules corticales, appelés arbuscules (Figure 5). Ces structures élaborées restent séparées du cytoplasme des cellules végétales par une extension de la membrane plasmique de l'hôte, la membrane periarbusculaire. Cette membrane suit les contours des branches des hyphes, conduisant à une augmentation de la surface de la membrane plasmique. Les arbuscules sont considérés comme étant le site principal d'échange des éléments nutritifs entre le champignon et la plante (Paszkowski, 2006 ; Bonfante et Genre, 2010). Les arbuscules sont des structures éphémères avec une durée de vie estimée à 4-5 jours. Les structures fongiques sont ensuite dégradées et la cellule hôte retrouve son ancienne organisation. Les cellules corticales sont alors en mesure de permettre une nouvelle pénétration fongique et la formation d'arbuscules (Paszkowski, 2006 ; Bonfante et Genre, 2010).

Le champignon développe également un mycélium extraradicalaire qui s'étend à l'extérieur de la racine. Ce mycélium acquière les nutriments dans le sol qui seront transférés à la plante. Le cycle de vie des champignons AM est complété par la formation de spores au niveau du mycélium extraradicales, qui peuvent entrer dans un autre processus de colonisation (Figure 5).

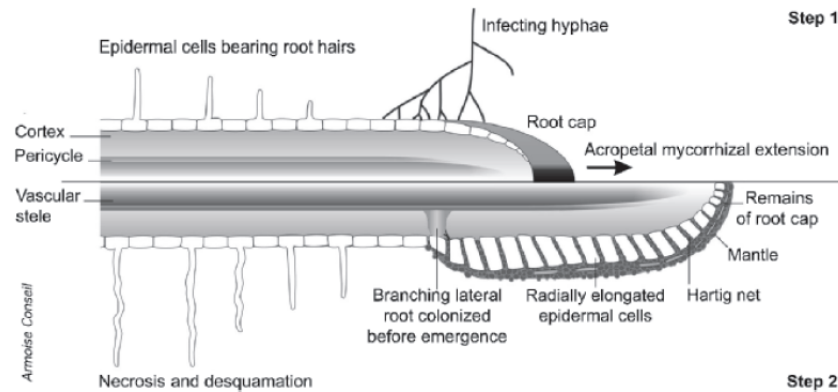


Figure 6 : Dynamique de la colonisation racinaire chez les champignons ectomycorhiziens. Les hyphes pénètrent dans la racine au niveau de la coiffe racinaire et se propagent pour atteindre les cellules végétales. Les cellules végétales épidermiques s'allongent radialement et le réseau de Hartig se développe entre ces cellules. (Martin et Tunlid, 2009).

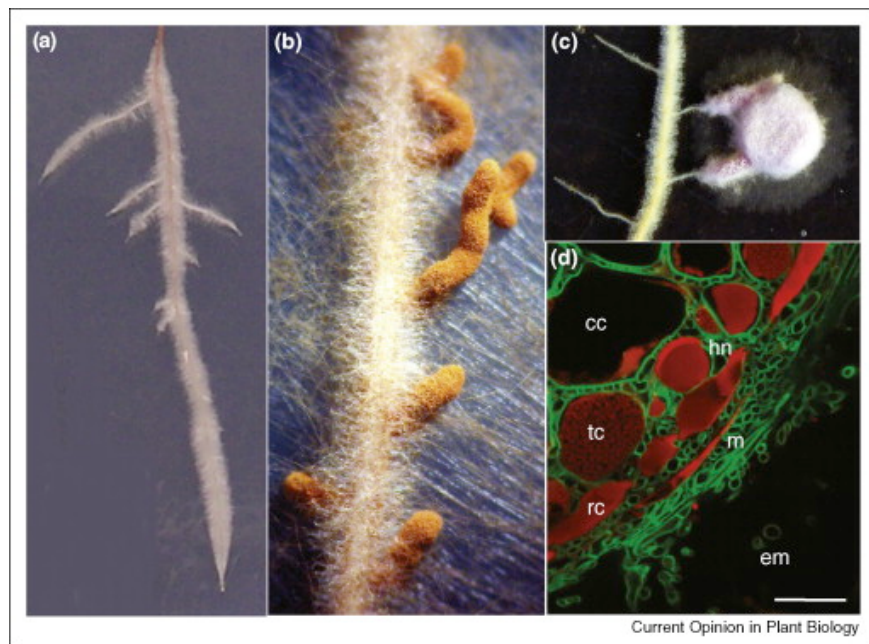


Figure 7 : Impact de la symbiose ectomycorhizienne sur la morphologie des racines. (a) Racines non mycorhizées d'*Eucalyptus globulus* (b) Racines latérales de *E. globulus* colonisées par le champignon ECM *Tomentella* sp. (c) Mycélium de l'ECM *Laccaria bicolor* colonisant deux racines latérales de *Populus tremula alba*. (d) Coupe transversale en microscopie électronique ? d'une ectomycorhize *Pseudotsuga menziesii*-*Laccaria bicolor* (mycélium extramatriciel (em), manteau (m) réseau de Hartig (HN) cellules rhizodermiques (rc), cellules taniques (tc), cellules corticales (cc) ; échelle, 10 mm). (Martin et Nehls, 2009).

1.2 - Développement de la symbiose ectomycorhizienne

Le développement des ECM se produit à travers une série d'événements morphologiques bien caractérisés. Les hyphes fongiques émergent à partir des spores du sol ou de racines préalablement mycorhizées et croissent dans la rhizosphère. La rencontre et la reconnaissance entre les deux symbiotes impliquent l'échange de signaux chimiques. Seules les grandes lignes de ce processus de signalisation sont définies. Dans le voisinage de la racine, la morphologie des champignons change, notamment avec une augmentation de la croissance et une ramification intense des hyphes. Une telle réponse est probablement déclenchée par les exsudats racinaires de l'hôte, tels que la rutine et la zéatine (Lagrange et al, 2001). Ceci suggère que le champignon perçoit la position de l'hôte grâce à la détection des signaux dérivés de l'hôte. En retour, les hyphes fongiques, présents dans la rhizosphère, stimulent la formation des racines latérales et la différenciation des cellules racinaires, à travers l'action de molécules sécrétées, telles que l'éthylène, l'auxine ou l'hypaphorine, fournissant ainsi un moyen d'accroître les sites de contact (Felten et al, 2009).

Les hyphes fongiques rentrent alors en contact avec la surface des racines, pénètrent dans la région de la coiffe racinaire et se propagent jusqu'à atteindre les cellules corticales vivantes (Horan et al, 1988). Ce contact provoque un bouleversement de la morphologie du champignon. L'apex des hyphes se gonfle formant une structure proche des structures d'infection des pathogènes, les appressoria. La plante subit également des changements morphologiques tels que l'allongement radial des cellules corticales. Les hyphes fongiques se différencient alors en différents tissus (Figure 6).

Les hyphes se multiplient pour former une gaine très dense autour de la racine, constituée d'une série de couches de plusieurs centaines de microns d'épaisseur, appelée le manteau (Figure 6 et 7). Les hyphes de cette structure sont enfermés dans une matrice de polysaccharides extracellulaires et de protéines (Dexheimer et Pargney, 1991). La fonction principale du manteau est le stockage des éléments nutritifs.

A partir du manteau, les hyphes progressent entre les cellules corticales au sein de l'espace apoplastique, conduisant à la formation d'une structure digitée complexe, de forme labyrinthique, appelée réseau de Hartig (Figure 6 et 7) (Bonfante, 2001). Les hyphes du réseau de Hartig sont coenocytiques et contiennent de nombreuses mitochondries, des corps lipidiques et un vaste réticulum endoplasmique, traduisant un état métabolique très actif (Kottke et oberwinkler, 1987). Les membranes abondantes de cette structure permettent l'échange de nutriments et de signaux entre les cellules végétales et les cellules fongiques. Aucune pénétration des hyphes au sein des cellules racinaires n'a été observée chez les champignons ECM, à l'exception des cellules sénescents.

D'autre part, un réseau d'hyphes extramatriciels s'étend alors dans le sol à partir des couches supérieures du manteau pour prospecter le sol et collecter des nutriments. Ces hyphes s'organisant parfois en structure de forme racinaire, appelées rhizomorphes. Ces structures sont également capables d'initier la formation de primordia pour le développement de carpophores qui assureront la dissémination des spores sexuées.

2 - Génomique des symbioses mycorhiziennes

D'un point de vue génomique, plusieurs mécanismes peuvent expliquer l'émergence de la symbiose, comme, par exemple, la présence de nouveaux gènes acquis par duplication ou transfert horizontal, la perte de gènes ou la régulation de l'expression génique (Martin et al, 2007). A travers la génomique comparative, le séquençage de génomes fongiques provenant de différents taxons et niches écologiques, peut donc permettre d'identifier ces mécanismes. Les premiers génomes de champignons mycorhiziens à avoir été séquencés sont ceux du basidiomycète ectomycorhizien *L. bicolor* (Martin et al, 2008a) et de l'ascomycète ectomycorhizien *T. melanosporum* (Martin et al, 2010a). Actuellement, plus d'une centaine de projets de séquençage de génomes d'espèces fongiques saprotrophes, parasites ou symbiotiques sont achevés ou en cours de réalisation. Ces projets ont été rendus possibles grâce aux programmes « Fungal Genome Initiative » du Broad Institute et « Community Sequencing Program » du JGI (United States Department of Energy Joint Genome Institute). Par ailleurs, dans le cadre du programme FGP (Fungal Genomic Program) qui vise à explorer la diversité génomique des champignons, 30 champignons saprotrophes et 25 champignons mycorhiziens sont en cours de séquençage (Figure 8) (Grigoriev et al. 2011, Martin et al, 2011). Ainsi, très récemment, plusieurs séquences de génomes de champignons ectomycorhiziens, tels que le basidiomycète *Paxillus involutus* et l'ascomycète *Oidiodendron maius*, sont maintenant disponibles.

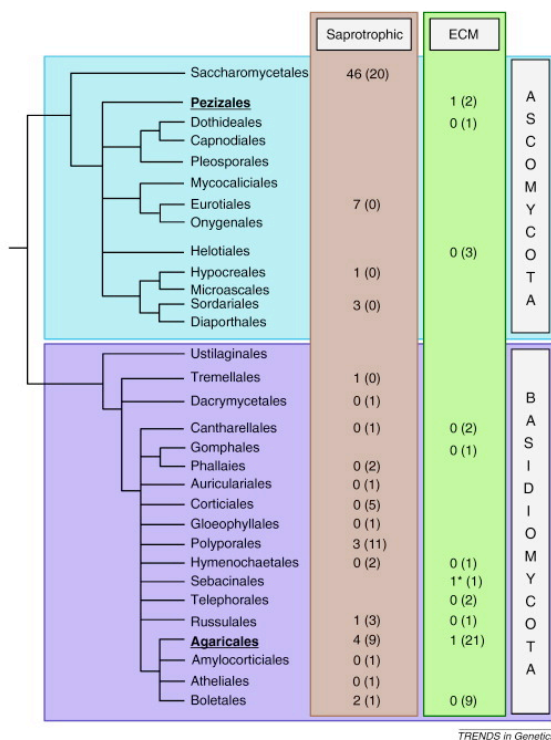


Figure 8 : Distribution phylogénétique et écologique des organismes fongiques, basidiomycètes (violet) et ascomycètes (bleu), saprotrophes (brun) et ECM (vert), séquencés (ou en cours de séquençage). * *Piriiformospora indica*, de l'ordre des Sebacinales, est un endophyte favorisant la croissance d'*Arabidopsis thaliana* et non un champignon mycorhizien. (Plett et Martin, 2010).

En ce qui concerne les partenaires végétaux de la symbiose, un seul génome d'arbre s'associant aux champignons ectomycorhiziens a actuellement été séquencé, celui de *Populus trichocarpa* (Tuskan et al, 2006). Plusieurs génomes de plantes formant les symbioses AM, tels que *M. truncatula* ou *O. sativa*, sont actuellement disponibles (International Rice Genome Sequencing Project, 2005).

2.1 - Les génomes des champignons ectomycorhiziens

Les séquences génomiques des deux champignons ECM, *L. bicolor* et *T. melanosporum*, ont permis d'identifier plusieurs traits moléculaires liés à la symbiose mycorhizienne. Par rapport à la plupart des génomes fongiques, ces deux champignons possèdent des génomes de taille relativement importante, 65 et 125 Mb, respectivement. Ceci s'explique principalement par une prolifération importante des éléments transposables (TE) qui représentent environ 20% et 60% des génomes de *L. bicolor* et *T. melanosporum*, respectivement. (Martin et al, 2008a ; Martin et al 2010). Ces éléments transposables sont distribués de manière uniforme à travers les deux génomes. En permettant des réarrangements génomiques, ils pourraient avoir joué un rôle fondamental dans la plasticité et l'évolution des génomes symbiotiques.

Alors que le répertoire de gènes prédits chez *L. bicolor* est l'un des plus importants chez les champignons filamenteux (20 000 gènes), celui de *T. melanosporum*, estimé à 7500 gènes, est relativement restreint. Une des caractéristiques marquantes du génome de *L. bicolor* est l'expansion des familles multigéniques issue d'événements de duplication (Figure 9). Les gènes de ces familles codent pour des protéines membranaires et pariétales, des protéines sécrétées et des protéines impliquées dans la transduction du signal (domaines kinases, GTPase et WD40). Cependant, de nombreuses familles sont spécifiques de ce champignon. L'expansion des familles multigéniques participant à la diversité génique, ces familles pourraient contenir des gènes importants pour l'adaptation au mode de vie symbiotique (Martin et al, 2008a). Toutefois, chez *T. melanosporum*, les familles multigéniques sont rares (Martin et al 2010). Le statut symbiotique ne semble donc pas imposer l'existence de familles multigéniques en expansion.

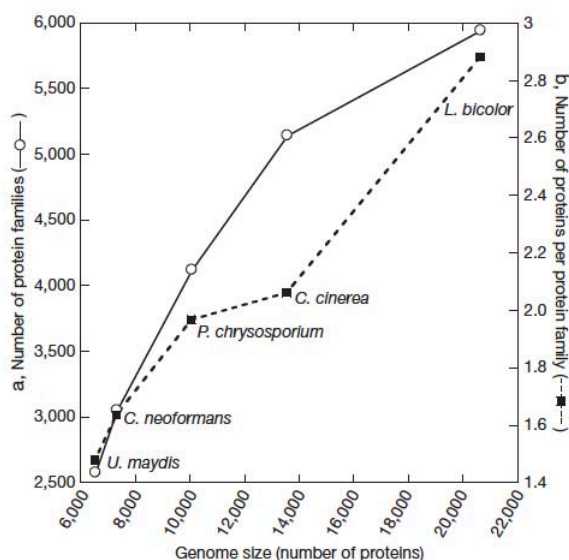


Figure 9 : Expansion des familles multigéniques chez *L. bicolor*. Relation entre la taille du génome et (a) le nombre de familles multigéniques, (b) la taille des familles multigéniques chez cinq basidiomycètes. (Martin et al, 2008a)

La comparaison des génomes de *L. bicolor* et *T. melanosporum* avec ceux de champignons saprotrophes et pathogènes révèle des pertes massives des gènes codant pour des enzymes dégradant les polymères de la paroi cellulaire des plantes (PCP) telles que la cellulose, la lignine, la pectine et le xylane (Figure 10) (Martin et al, 2008a ; Martin et al 2010). De même, les champignons ECM du genre *Amanita* ont également perdu plusieurs gènes codant pour des cellulases (Nagendrana et al, 2009). L'adaptation au mode de vie symbiotique conduirait donc à des pertes convergentes d'enzymes agissant sur les PCP, probablement pour éviter le déclenchement de réactions de défense de la plante. Fait intéressant, les génomes de deux pourritures brunes, *Postia placenta* (Martinez et al, 2009) et *Serpula lacrymans* (Eastwood et al, 2011) ont également un répertoire réduit de cellulases agissant sur les PCP. Des analyses phylogénétiques récentes indiquent que l'ancêtre des Agaricomycotina était probablement un champignon saprotrophe (Hibbett et Matheny, 2009). La perte de la machinerie enzymatique de dégradation de la cellulose et de la lignine, à partir d'un ancêtre de type pourriture blanche, aurait conduit à l'apparition des champignons de type pourriture brune. La perte de ces enzymes aurait alors facilité l'interaction entre les champignons saprotrophes, colonisant la litière, et les racines sans endommager ces dernières, donnant lieu à la symbiose mycorhizienne (Eastwood et al, 2011).

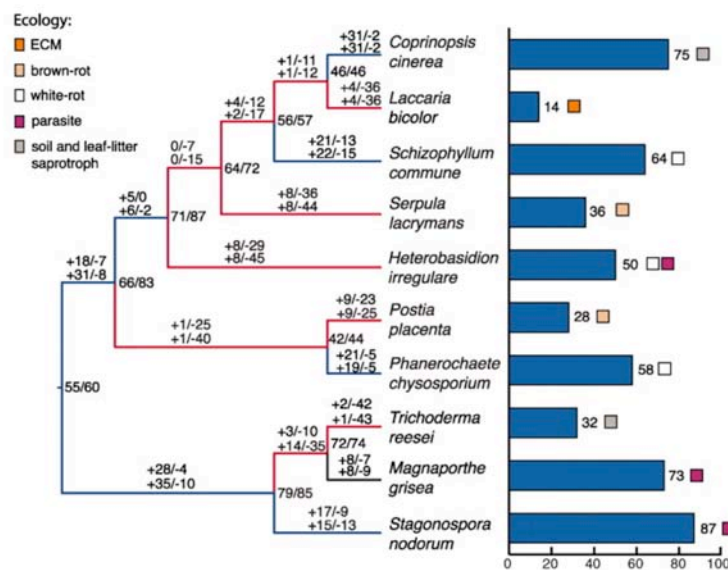


Figure 10 : Structure des duplications et des pertes de gènes codant pour les enzymes agissant sur la lignocellulose chez les Agaricomycètes. L'histogramme indique le nombre de copies dans chaque génome. Les lignées en expansion et en contraction sont signalées par les branches bleues et rouges, respectivement. Le nombre de copies estimé pour les espèces ancestrales est indiqué au niveau des nœuds de l'arbre phylogénétique. Le nombre de gains et de pertes de gènes est indiqué au niveau des branches de l'arbre phylogénétique. (Eastwood et al, 2011).

2.2 - La génomique chez les champignons mycorhiziens arbusculaires

Les Glomérormycètes sont des organismes coenocytiques, c'est à dire que leurs spores et leurs hyphes contiennent des centaines de noyaux au sein d'un même cytoplasme (Figure 5c) (Jany et Pawlowska, 2010). De plus, il existe une grande variabilité génétique au sein d'une même spore. Cette variabilité a été observée au sein de séquences d'ADNr (Sanders et al, 1995 ; Clapp et al, 2001; Rodriguez et al, 2004) et, plus tard, dans des régions codant des protéines telles que la tubuline β et la H⁺/ATPase (Corradi et al, 2004 ; Corradi et al, 2009). L'origine et la répartition de ce polymorphisme, au sein des différents noyaux d'une même spore, sont l'objet d'un débat en cours (Figure 11). Ce polymorphisme pourrait provenir d'une population de noyaux génétiquement différents au sein d'une même spore (Kuhn et al, 2001), ou d'une population de noyaux génétiquement identiques caractérisée par un polymorphisme intranucléaire (Pawlowska et Taylor, 2004).

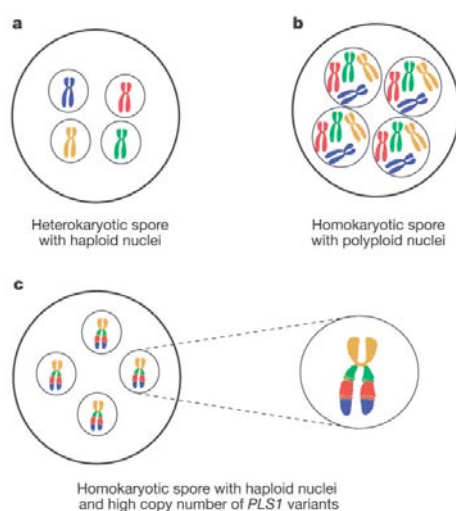


Figure 11 : Différentes hypothèses pour expliquer la variabilité génétique chez les champignons mycorhiziens arbusculaires. (a) Les spores contiennent des noyaux haploïdes génétiquement différents. (b) Les spores contiennent des noyaux polyploïdes génétiquement identiques. (c) Les spores contiennent des noyaux haploïdes génétiquement identiques mais plusieurs copies de mêmes gènes sont présentes dans chaque noyau en raison d'événements de duplication. (Hijri et Sanders, 2005).

Actuellement, aucune séquence génomique de champignon AM n'est disponible et l'organisation globale des génomes AM reste une énigme. La taille de ces génomes, généralement estimée par cytométrie de flux, est très variable selon les espèces, allant de 15 Mb (*G. intraradices*) à plus de 700 Mb (*Gigaspora margarita*) par noyau (Hijri et Sanders, 2004). En 2004, le champignon mycorhizien *G. intraradices* a été choisi par le JGI comme candidat pour le premier projet de séquençage de génome AM (Martin et al, 2004). Bien qu'une quantité considérable de séquences ait été acquise, l'assemblage de ces séquences s'est révélé problématique. En effet, la taille du génome assemblé est plusieurs fois supérieure à la taille estimée par la quantité d'ADN par noyau (Martin et al, 2008b). Les difficultés pour assembler ce génome sont probablement dues au niveau de polymorphisme élevé chez ce champignon. Des séquençages complémentaires sont en cours pour compléter ce génome. (Martin, 2008b). En raison du manque de données génomiques et transcriptomiques à grande échelle disponibles pour le partenaire fongique, la symbiose AM a principalement été étudiée sous l'angle du partenaire végétal.

3 - Transcriptomique des symbioses mycorhiziennes

Le développement et le fonctionnement de la symbiose mycorhizienne repose sur l'orchestration de processus complexes chez les deux partenaires, tels que l'induction de programmes d'organogenèse, la facilitation de la survie *in planta* du champignon face aux réponses de défense des plantes et l'augmentation de la capacité des flux d'éléments nutritifs. Les mécanismes moléculaires sous-jacents impliquent une coordination étroite de la régulation de l'expression des gènes chez le champignon et sa plante hôte.

Des études de transcriptomiques ont été réalisées pour un certain nombre d'associations ECM (*Pisolithus-Eucalyptus*, *Paxillus-Betula*, *Tuber-Tilia*, *Pinus-Laccaria*,) en utilisant plusieurs méthodes (séquençage d'ESTs, puces à ADNc, polymorphisme de longueur de fragment (AFLP), hybridations suppressive soustractive (SSH)). Plusieurs centaines de gènes exprimés préférentiellement dans les tissus symbiotiques ont été identifiés (Voiblet et al, 2001 ; Duplessis et al, 2005 ; Johansson et al, 2004 ; LeQueré et al, 2005 ; Wright et al, 2005 ; Morel et al, 2005 ; Menotta et al, 2004). Ces études ont confirmé que le développement des mycorhizes est accompagné par des changements dans les profils de transcription dès le contact entre les deux partenaires, avant la formation des ectomycorhizes fonctionnelles. De nombreuses fonctions cellulaires, communes aux différents systèmes ectomycorhiziens, sont impliquées dans le développement de la symbiose, notamment, la prolifération, la différenciation et la signalisation cellulaire, la synthèse de la paroi cellulaire et des matrices extracellulaires, la respiration, la synthèse des protéines, les réactions de stress, l'assimilation et le transport (Figure 12). Seulement une faible proportion des transcrits analysés (5-20%) est affectée par le développement de la symbiose et les variations des niveaux d'expression sont modérées. De plus, aucun gène spécifiquement induit par la symbiose n'a été détecté. Plus récemment, le premier profil transcriptomique à l'échelle du génome, de l'interaction de *L. bicolor* avec les racines de peuplier ou de Douglas, a cependant révélé des gènes très fortement régulés dans les tissus symbiotiques (>1000 fois), la plupart codant pour des protéines orphelines (Martin et al, 2008a). La majorité de ces gènes fortement régulés sont détectés au sein des autres tissus. De même, seulement 3% de l'ensemble des gènes sont régulés lors de l'interaction symbiotique (Martin et al, 2008a). Ces gènes sont essentiellement impliqués dans le métabolisme primaire, la respiration, la synthèse des protéines, la régulation de la morphologie, la formation des parois cellulaires, le transport et la signalisation cellulaire. Le faible nombre de gènes spécifiques des ectomycorhizes suggère que les programmes génétiques qui conduisent la symbiose sont régulés par l'activité de facteurs de transcription et/ou de voies de transduction, plutôt que par l'expression de jeu de gènes spécifiques. De plus, le nombre de transcrits régulé étant très faible, le contrôle traductionnel joue probablement un rôle important dans la régulation génique au cours du développement symbiotique.

Chez les champignons AM, les études d'expression génique au cours de la symbiose ont principalement été basées sur des approches ciblées, visant en particulier les gènes impliqués dans l'absorption des nutriments. Contrairement à la symbiose ECM, peu de gènes fongiques régulés lors de la symbiose AM ont été identifiés à ce jour (quelques dizaines). Les gènes identifiés comme étant régulés au cours de la symbiose sont essentiellement liés au métabolisme des nutriments, au transport membranaire, à la signalisation et à la réponse au stress oxydatif (Requena et al, 2003 ; Breuninger et Requena, 2004 ; Aono et al, 2004 ; Kuznetsova et al, 2010 ; Lanfranco et al, 2005 ; Govindarajulu et al, 2005 ; Gomez et al, 2009 ; Seddas et al, 2009).

Du côté du partenaire végétal, les transcriptomes des symbioses ECM (*Populus trichocarpa*) et AM (*Medicago truncatula* ; *Pisum sativum* ; *Oryza sativa* ; *Lotus japonicus*) ont été étudiés par l'analyse de puces génome entier. Plusieurs centaines de gènes induits au cours de ces symbioses ont été identifiés, mais ils ne représentent qu'une fraction limitée de l'ensemble des gènes de ces plantes (2% chez *Populus trichocarpa*). La plupart de ces gènes sont impliqués dans le métabolisme de l'auxine, les voies de signalisation du calcium, la régulation transcriptionnelle, les mécanismes de défense, les modifications des parois cellulaires et le transport des nutriments (Liu et al, 2004 ; Hohnjec et al, 2005 ; Grunwald et al, 2004 ; Guïmil et al, 2005 ; Guether et al, 2009a ; Gomez et al, 2009).

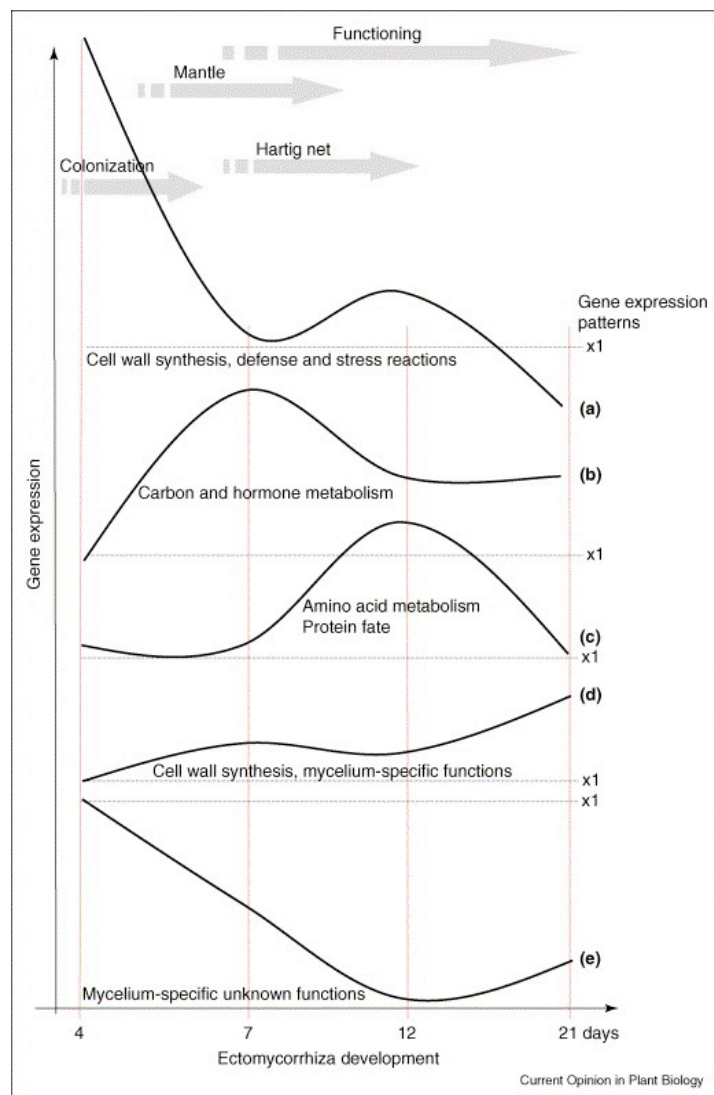


Figure 12 : Profils d'expression des gènes végétaux et fongiques régulés au cours du développement de la symbiose ectomycorhizienne. Ils ont été obtenus à partir des transcriptomes fongiques et racinaires étudiés pour les interactions *Betula-Paxillus* et *Eucalyptus-Pisolithus*, à l'aide de puces à ADNc. Différents groupes de gènes à expression coordonnée sont mis en jeu lors des différentes phases du développement de la symbiose (Martin et al, 2007).

3.1 - Les échanges de nutriments

Les échanges d'éléments nutritifs entre la plante et le champignon sont au cœur de la symbiose mycorhizienne. En effet, dans les sols des écosystèmes naturels, les nutriments essentiels pour la croissance des plantes sont fixés dans les couches organiques du sol. Ils ne sont donc disponibles qu'en quantité limitée pour les plantes. Ceci les rend écologiquement dépendantes de leur partenaire fongique (Smith and Read, 2008). D'autre part, les champignons AM, qui sont des biotrophes obligatoires, dépendent entièrement de la plante pour leur apport en carbone. A l'inverse, les champignons ECM ne sont pas dépendants de leur hôte pour leur besoin en carbone. Cependant, les sucres simples étant rares dans les sols forestiers, le mode de vie symbiotique leur permet d'avoir un accès privilégié aux sucres présents en quantité importantes dans les exsudats racinaires (Smith and Read, 2008). Pour établir et maintenir la symbiose, les deux partenaires doivent donc activer les transporteurs et les enzymes anaboliques et cataboliques nécessaires. D'une manière générale, l'établissement de la symbiose est, en effet, accompagné par des modifications de l'expression des gènes codant pour des transporteurs et des enzymes d'assimilation.

- Métabolisme de l'azote

L'azote inorganique absorbé par les hyphes extraradicales, est tout d'abord assimilé en acides aminés, probablement sous forme d'arginine chez les AM et de glutamine chez les ECM. Ces acides aminés sont ensuite transportés vers l'interface symbiotique, d'où l'azote peut être exporté vers la plante (Chalot et al, 2006 ; Govindarajulu, 2005). Le mécanisme de transfert de l'azote à l'interface symbiotique est encore mal compris. L'azote pourrait être libéré sous forme de composés organiques, tels que les acides aminés, ou sous forme d'ions ammonium (ou nitrate) (Figure 13) (Chalot et al, 2006).

Pour que le transfert de l'azote du champignon vers son hôte soit efficace, les transporteurs d'influx fongiques correspondant devraient être réprimés à l'interface symbiotique (Chalot et al, 2006). Chez *Amanita muscaria* et *Hebeloma cylindrosporum*, les transporteurs d'influx d'ammonium AMT sont fortement réprimés au sein des ECM (Willmann et al, 2007 ; Javelle et al, 2003). En revanche, chez *L. bicolor*, aucun des gènes de la famille AMT n'est réprimé au sein des ECM. Le transporteur LbAMT2.2 est même fortement induit par la symbiose (Martin et al, 2008a ; Lucic et al, 2008). De même, chez *G. intraradices* le transporteur d'influx d'ammonium GintAMT2 est induit au sein des structures intraracinaires (Perez-Tienda et al, 2011). Ceci suggère que les hyphes symbiotiques ont une capacité d'absorption de l'ammonium apoplastique renforcée. Alternativement, l'induction des gènes AMT au sein des structures symbiotiques pourrait être expliquée par le stockage de l'ammonium en excès au sein de vésicules intracellulaires via un système de transport AMT. Ces vésicules pourraient ensuite libérer l'ammonium à l'interface symbiotique par un mécanisme d'exocytose (Chalot et al, 2006). Le profil d'expression de ces gènes AMT pourrait également indiquer l'existence d'une compétition entre la plante et le champignon pour l'azote présent à l'interface symbiotique (Guether et al. 2009b).

L'activation des enzymes anaboliques et cataboliques appropriées est nécessaire pour maintenir le transfert de l'azote vers l'hôte. Chez *G. intraradices*, l'analyse de l'expression des gènes du métabolisme de l'azote, a montré que les gènes impliqués dans l'assimilation de l'ammonium et la synthèse de l'arginine sont préférentiellement exprimés dans le mycélium extraradical alors que les gènes impliqués dans la dégradation de l'arginine et la libération de l'ammonium sont plus fortement exprimés au sein du mycélium intraracinaire (Govindarajulu et al. 2005, Tian et al, 2010). De même, le niveau de transcription de l'uréase augmente chez *L.*

bicolor au cours de la symbiose (Martin et al, 2008a), ce qui est en faveur de la libération de l'ammonium pour son exportation à l'interface symbiotique.

Bien que la nature des formes d'azote échangées entre le mycobionte et sa plante hôte soit l'objet de controverses, des transporteurs d'ammonium à haute affinité sont induits au sein des apex racinaires de peuplier lors de la formation des ECM (Selle et al, 2005 ; Couturier et al, 2007). Plusieurs transporteurs de nitrate et d'ammonium sont également induits lors de la symbiose AM chez *M. truncatula* et *Lotus japonicus* (Hohnjec et al. 2005 ; Guether et al, 2009b). Fait intéressant, des souches de *L. bicolor* déficientes pour l'assimilation du nitrate, ont perdu la capacité à former des ECM, ce qui confirme le rôle clé joué par le transfert de l'azote dans le maintien de la symbiose (Kemppainen et al, 2009).

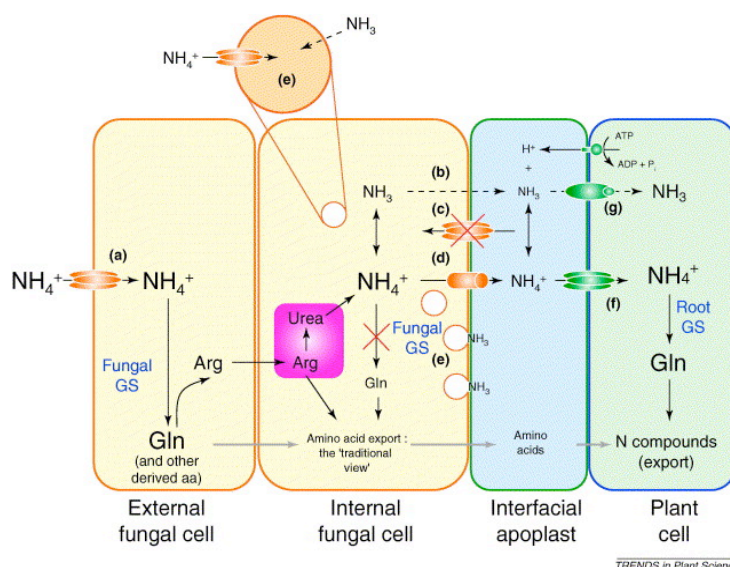


Figure 13 : Mécanismes possibles pour le transfert d'azote entre les partenaires symbiotiques. (A) L'ammonium du sol est absorbé à travers la membrane plasmique par des transporteurs fongiques AMT. Le transfert d'azote entre les cellules fongiques et la racine a lieu à l'interface apoplastique sous forme d'acides aminés (flèches grises), ou sous forme d'ammonium. (B) La concentration en nitrate dans le cytosol est plus élevée que dans l'apoplaste, ce qui génère un gradient favorable pour la diffusion passive du nitrate. Les concentrations apoplastiques doivent alors être maintenues à un faible niveau, par exemple, par un mécanisme de piégeage acide du nitrate. (C) Dans ce modèle, les transporteurs AMT devraient être réprimés à la surface de la membrane plasmique des hyphes symbiotiques par rapport aux hyphes extraradicales. (D) Un système d'efflux fongique, de type Ato, pourrait permettre la libération d'ammonium à l'interface apoplastique. (E) La diffusion passive du nitrate ou le transport actif d'ammonium, par un système AMT, dans des vésicules acidifiées permettrait de compartimenter les excès d'ammonium. Les vésicules chargées d'ammonium pourraient alors se déplacer vers la membrane symbiotique, via les microtubules, où elles pourraient fusionner avec la membrane plasmique pour libérer l'ammonium dans l'apoplaste interfaciales. (F) Le transfert de l'ammonium de l'apoplaste vers le cytoplasme de la plante hôte pourrait impliquer des transporteurs AMT (G) ou des systèmes voltage-dépendant. (Abréviations: Arg, arginine, Gln, glutamine, GS, glutamine synthétase). (Chalot et al, 2006).

- Métabolisme du phosphate

Le phosphate inorganique (Pi) dans les solutions du sol est facilement absorbé par les hyphes fongiques (Smith et Read, 1997). Les champignons stockent généralement l'excès en Pi sous forme de granules de polyphosphates qui sont ensuite transportées vers les racines de l'hôte (Figure 14). Les polyphosphates sont les seuls anions macromoléculaires dans la vacuole des champignons. Le potassium, l'arginine et le magnésium stabilisent alors les charges négatives de ces derniers (Martin et al 1985; Orlovich & Ashford 1993; Martin et al. . 1994; Bücking & Heyser 1999). La séquestration du phosphore dans la vacuole permet au champignon d'accumuler les éléments nutritifs tout en évitant la toxicité.

Plusieurs transporteurs de Pi à haute et faible affinité ont été identifiés dans les génomes de *L. bicolor* et de *H. cylindrosporum* (Martin et al 2008, 2010; Tatry et al 2009), mais les mécanismes moléculaires contrôlant l'absorption du Pi chez les champignons ECM sont jusqu'ici inconnus. Chez les champignons AM, trois transporteurs de phosphate à haute affinité ont été identifiés chez *G. versiforme*, *G. intraradices* et *G. mosseae*. Tous sont exprimés dans les hyphes extraradicales, alors que seul celui de *G. mosseae* montre également une expression significative dans les hyphes intraradicales (Benedetto et al. 2005, Balestrini et al, 2007).

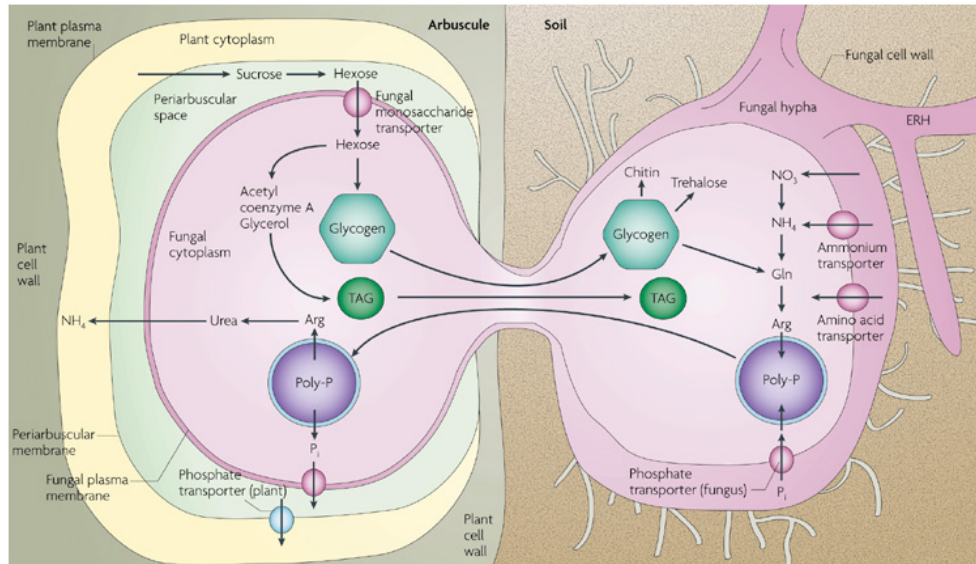
L'absorption du Pi par les cellules végétales doit être active car le gradient électrochimique ne permet pas l'absorption passive. Parmi les 13 transporteurs de Pi à haute et faible affinité identifiés chez *Populus trichocarpa*, un seul est préférentiellement exprimé dans les racines ectomycorhiziennes (Martin et al. 2007). Plusieurs transporteurs de phosphate chez *O. sativa*, *L. japonicus*, *M. truncatula*, *S. tuberosum*, et *L. esculentum* sont également induits lors de la symbiose AM (Javot et al, 2007). Il est alors intéressant de constater que certains de ces transporteurs semblent être impliqués dans la maintenance des arbuscules (Javot et al, 2007).

- Métabolisme du carbone

La capacité des champignons mycorrhiziens à dégrader la matière organique du sol étant très limitée, leur principale source de carbone est constituée par les photoassimilats de la plante hôte. La plante libère les photoassimilats dans l'espace apoplastique principalement sous la forme de saccharoses (Nehls et al, 2001). Cependant, chez les champignons mycorrhiziens, les sucres sont absorbés sous la forme d'hexoses (Nehls et al, 2007 ; Bago et al, 2002). Le séquençage du génome de *L. bicolor* a révélé l'absence de gènes codant pour des invertases, enzymes qui hydrolysent le saccharose en glucose et en fructose (Deveau et al, 2008). Le saccharose apoplastique serait alors clivé par une invertase végétale. *L. bicolor* dépendrait donc de l'activité invertase de son hôte (Nehls et al, 2007). L'absence d'invertase semble être une caractéristique partagée par la plupart des champignons basidiomycètes ECM (Parrent et al, 2009). La plante pourrait utiliser cette dépendance à l'activité invertase pour réguler la quantité de carbone allouée au champignon (Nehls et al, 2007).

Une quinzaine de gènes codant pour les transporteurs d'hexose ont été identifiés chez *L. bicolor* et la moitié d'entre eux sont induits au cours de la symbiose (Fajardo Lopez et al, 2008). Deux transporteurs d'hexoses ont également été caractérisés chez *Amanita muscaria* (Nehls et al, 1998 ; Nehls, 2004). Ceci suggère une forte augmentation de la capacité d'absorption des hexoses à l'interface symbiotique. A l'exception d'un gène chez le champignon non-AM *Geosiphon pyriforme*, (Schussler et al, 2006), aucun transporteur d'hexose responsable de l'absorption du glucose à partir de la cellule hôte n'a jusqu'à présent été caractérisé chez les Glomeromycètes.

Pour maintenir un puits en carbone fort vers le champignon, le glucose doit être rapidement converti à travers la glycolyse ou/et le stockage de métabolites fongiques, tels que le tréhalose, le glycogène ou les lipides (Figure 14) (Nehls, 2008 ; Bago et al, 2003). Ces composés sont ensuite transférés au mycélium extraradical. Lors de l'interaction *P. microcarpus/Eucalyptus*, les gènes codant pour la glycolyse et le cycle de l'acide tricarboxylique sont induits (Duplessis et al, 2005). Chez *A. muscaria*, les gènes codant pour des protéines impliquées dans la biosynthèse du tréhalose sont également induits à l'interface apoplastique (Fajardo Lopez 2007). De même, chez *L. bicolor*, tous les gènes codant pour la synthèse du tréhalose sont induits dans les tissus symbiotiques (Deveau et al, 2008).



Nature Reviews | Microbiology

Figure 14 : Flux métaboliques chez les champignons mycorhiziens à arbuscules. Le carbone d'origine végétale est tout d'abord libéré dans l'espace periarbusculaire, probablement sous la forme de saccharoses, puis clivé en hexoses et absorbé par les champignons AM. Dans le cytoplasme fongique, les hexoses sont convertis en granules de glycogène et en gouttelettes de triacylglycérol, qui permettent le transport sur de longues distances. Les nutriments acquis dans le sol par les champignons sont transportés sur une longue distance vers les arbuscules. Le phosphate est absorbé par les transporteurs de phosphate fongiques qui sont présents dans les hyphes extraradicales. Le phosphate est transporté vers la racine sous la forme de granules de polyphosphate. La charge négative de ces granules leur permet de transporter des ions métalliques et de l'arginine. Le phosphate est libéré à partir de granules de polyphosphate dans les arbuscules. Les transporteurs végétaux impliqués dans le transport du phosphate ont été clonés et caractérisés, tandis que les transporteurs de phosphate du champignon, responsables de la libération de phosphate, sont encore inconnus. L'azote du sol est absorbé par des transporteurs d'ammonium, de nitrate ou d'acides aminés. L'azote est principalement transporté sous forme d'arginine, puis libéré dans les arbuscules sous forme d'urée et transporté à la plante directement ou après clivage en ammonium. (Parniske, 2008)

Chez les champignons AM, le profil d'expression des gènes codant pour une acyl-coA déshydrogénase et une glycogène synthase sont cohérents avec la synthèse de triacylglycérols et de glycogène dans les hyphes intraradicales (Bago et al. 2002, 2003). En revanche, les gènes codant pour les enzymes du cycle glyoxylate sont activement exprimés dans le mycélium extraradical, permettant ainsi la conversion des lipides en glucides (Lammers et al, 2001).

Récemment, des transporteurs d'efflux de glucose, appelés SWEET, ont été identifiés chez les plantes. Certains de ces gènes sont induits par des effecteurs de bactéries pathogènes (Chen et al, 2010). Des orthologues des gènes SWEET identifiés chez le peuplier sont également induits lors du développement de la symbiose ECM. Des effecteurs de champignons ECM, non identifiés à ce jour, pourraient donc jouer un rôle actif dans le contrôle de l'efflux de glucose lié à la symbiose.

Un certain nombre de transporteurs et d'enzymes clefs impliqués dans le transfert des métabolites à l'interface symbiotiques ont été mis en évidence. Cependant, la fonction et la localisation précise de ces transporteurs dans les tissus fongiques, ainsi que la nature des nutriments échangés à l'interface apoplastique, restent largement incomprises. L'acquisition de données génomiques et transcriptomiques à l'échelle du génome, associée aux techniques de microdissection par capture laser (LCM), pourraient permettre d'étudier de façon précise l'expression des gènes au sein des différents tissus de la symbiose.

Il a récemment été montré que l'échange nutritionnel lors de la symbiose mycorhizienne est sous le contrôle de deux partenaires, stabilisant ainsi la coopération au sein de la symbiose mycorhizienne (Kiers et al, 2011). Les mécanismes moléculaires contrôlant le développement des structures de l'interaction et des échanges de nutriments pourraient impliquer des signaux de transduction, des systèmes d'évasion aux défenses de l'hôte ainsi que des molécules effectrices.

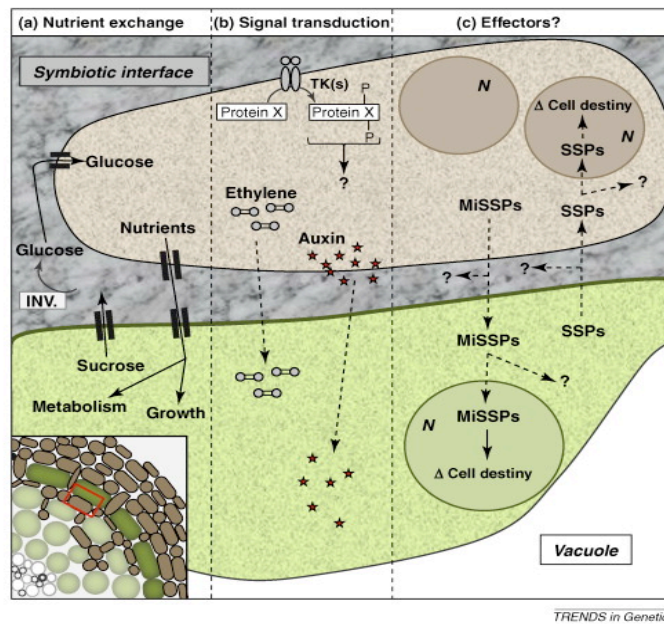


Figure 15 : Niveaux de contrôle dans l'interaction symbiotique entre les champignons ECM et les cellules végétales. Représentation schématique d'une coupe transversale du réseau de Hartig. (a) La principale caractéristique commune à toutes les interactions mutualistes ECM est l'échange réciproque de nutriments. Le saccharose, dérivé de la photosynthèse de la plante, est clivé par une invertase (INV) en glucose et en fructose dans l'interface symbiotique. Les transporteurs d'hexoses des hyphes absorbent ce glucose qui est utilisé par le champignon comme source primaire de carbone. En retour, le champignon libère des éléments nutritifs (azote, phosphore) dans l'espace apoplastique. Les transporteurs des cellules végétales absorbent ces nutriments où ils sont utilisés pour soutenir la croissance et le métabolisme des plantes. (b) Pour établir une symbiose, les champignons ECM doivent avoir développé de nouvelles voies de signalisation. Celles-ci pourraient être contrôlées par les tyrosines kinases (TKS) qui activent les cascades de signalisation dans les cellules fongiques. Des éléments diffusibles tels que l'éthylène et l'auxine ont également été impliqués dans le développement de la symbiose entre les deux partenaires. (c) Les MiSSPs sont une autre voie de contrôle chez les champignons ECM. Ces protéines sécrétées agissent comme effecteurs pour contrôler les cellules végétales et perturber leur système de défenses. Le partenaire végétal pourrait utiliser de tels effecteurs pour contrôler le développement des champignons dans la racine. (N, noyau ; la cellule fongique est représentée en marron et la cellule végétale en vert). (Plett et Martin, 2010)

3.3 - Echapper au système de surveillance de la plante hôte

Bien que certaines réponses à large spectre soient induites dans les racines de l'hôte colonisé par les champignons mycorrhiziens, le niveau de défense de la plante reste relativement faible. Cela indique que les champignons symbiotiques ont développé des stratégies pour éviter le déclenchement des réactions de défense de la plante hôte. Cette inhibition des défenses de la plante est cruciale dans l'établissement de la symbiose. Les analyses transcriptomiques des tissus ECM de *L. bicolor* ont identifié plusieurs gènes fongiques induits lors de la symbiose codant pour des protéines secrétées telles que des protéines à motif RGD (Arg-Gly-Asp) et des hydrophobines qui pourraient agir dans l'espace apoplastique pour détourner les protéases de la plante (Martin et al, 2008a). De plus, plusieurs membres de la famille des cliticypines sont parmi les transcrits les plus régulés lors de la symbiose. *L. bicolor* pourrait utiliser ces inhibiteurs de protéases pour contrer les protéases végétales sécrétées dans l'apoplaste (Martin et al, 2008a). Un gène codant pour la chitine déacétylase est également induit lors de l'association des plantes. Cette protéine qui modifie la chitine en chitosane pourrait permettre d'éviter la détection de la chitine par l'hôte (Martin et al, 2008a).

Les pathogènes fongiques, tels que *Ustilago maydis* et *Melampsora larici-populina*, utilisent des petites protéines secrétées effectrices pour manipuler les défenses des cellules végétales et coloniser les tissus végétaux. Ces protéines effectrices agissent dans l'espace apoplastique ou dans la cellule végétale pour modifier les voies de signalisation des plantes (Stergiopoulos et de Wit, 2009). Le séquençage des génomes et l'analyse des sécrétomes ont permis l'identification à grande échelle d'effecteurs candidats chez les champignons biotrophes (Choi et al, 2010). Chez *L. bicolor*, une douzaine de petites protéines secrétées, partageant certaines ressemblances avec les protéines effectrices des pathogènes, sont spécifiquement exprimées au sein des tissus symbiotiques (Figure 15) (Martin et al, 2008a). C'est le cas de l'effecteur MiSSP7 qui est le transcrit fongique le plus fortement induit lors de l'interaction de *L. bicolor* avec le peuplier et dont la protéine correspondante s'accumule au niveau du réseau de Hartig (Martin et al, 2008a). MiSSP7 est importé dans la cellule végétale par endocytose par le biais d'un motif RXLR et migre vers le noyau afin de modifier le transcriptome de la cellule végétale. Cette protéine est nécessaire pour l'établissement de la symbiose (Plett et al, 2011). Par ailleurs, il a été montré récemment que le champignon AM, *G. intraradices*, sécrète également une protéine effectrice, SP7, qui interagit dans le noyau de la plante avec le facteur de transcription ERF19 lié à la pathogenèse, neutralisant ainsi les systèmes de défenses de la plante (Kloppholz et al, 2011). La présence de ce type d'effecteur semble donc être importante pour les interactions symbiotiques mycorrhiziennes.

Lors de la symbiose entre les bactéries rhizobiums et les légumineuses, des petites protéines secrétées émises par la plante sont nécessaires pour le bon développement des nodules symbiotiques, notamment pour initier la différenciation des bactéries, limitant ainsi leur prolifération dans les tissus végétaux (Wang et al, 2010a). Bien qu'aucune petite protéine végétale sécrétée induite par l'interaction n'ait été détectée à ce jour, les plantes hôtes mycorrhiziennes pourraient également utiliser ce type de molécules pour contrôler la symbiose.

II - Contexte génomique et bio-informatique

1 - La génomique

Le génome est le plan génétique d'un individu ou d'une espèce. Il contient les informations nécessaires au développement et au fonctionnement d'un organisme vivant. Ces informations sont codées au sein de molécules d'acide désoxyribonucléique (ADN), à l'exception de certains virus dont le génome est porté par des molécules d'acide ribonucléique (ARN). L'ADN est composé de quatre bases nucléotidiques différentes, l'adénine (A), la guanine (G), la cytosine (C) et la thymine (T). Il contient en particulier des régions transcrites en molécules d'ARN messagers, elles même traduites en protéines (séquence d'acides aminés) qui assurent une fonction particulière au sein de la cellule. Ces régions sont appelées gènes codant des protéines. Il contient également des régions non codantes (non transcrites ou transcrites mais non traduites) dont certaines sont impliquées dans la régulation de nombreux processus biologiques. La connaissance de la séquence nucléotidique d'un génome permet d'en identifier les gènes, d'accéder à leurs organisations et de prédire leurs fonctions potentielles. Cette discipline, appelée génomique, a donné lieu à une multitude de découvertes en biologie cellulaire, en biologie du développement et en biologie de l'évolution.

L'essor de la génomique a été permis par le développement des techniques de séquençage nucléotidique. Ces techniques ont commencé à se développer dans les années 1970. Cependant, des méthodes de séquençage fiables, permettant d'obtenir des séquences d'ADN suffisamment longues, à savoir, les méthodes Sanger et Maxam-Gilbert, n'ont été disponibles qu'à partir de 1977 (Maxam et Gilbert, 1977 ; Sanger et al, 1977). L'automatisation de la méthode Sanger par capillaire a été la première percée technologique en terme de débit de séquençage. Elle a été désignée comme étant la première génération de technologie de séquençage à haut débit. Elle est devenue la méthode dominante durant plus de deux décennies pour les projets de séquençage de génome (Hutchison, 2007) et a conduit à un certain nombre d'accomplissements de grande ampleur, tels que le séquençage du génome humain en 2000 (Lander et al, 2001) et du génome d'*Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000). Cependant, la méthode Sanger reste lourde en terme de temps et de coût, et nécessite des investissements importants en terme d'infrastructure. Ainsi, le séquençage du génome humain a nécessité l'utilisation de plus d'une centaine de séquenceurs produisant plus de 175 000 lectures de séquences par jour durant 13 ans pour décoder les 3 milliards de bases d'ADN du génome humain (Venter et al, 2001). Cette méthode est donc restée hors de portée des simples laboratoires et les projets de séquençage de génome ont été restreints à un certain nombre d'organismes dit modèles.

Plusieurs technologies de séquençage de nouvelle génération (NGS) (ou technologies de séquençage de seconde génération) sont récemment apparues. Ces technologies permettent de générer rapidement des quantités massives de données de séquences à faible coût (Schuster, 2008 ; Hudson, 2008 ; Shendure et Ji, 2008). Elles ont ouvert la porte à un champ remarquable d'applications, à la fois pour les espèces modèles et non modèle. Une des plus vaste application des NGS est le reséquençage des génomes pour lesquels un génome de référence est déjà disponible. Le reséquençage de génome permet d'identifier les variations génétiques individuelles afin de comprendre les bases génétiques des différences phénotypiques. Le premier exemple d'un projet de re-séquençage de génome a été la détermination de la séquence du génome du Dr James Watson (Wheeler et al, 2008). Ainsi des projets visant à séquencer le génome de nombreux individus au sein d'une même espèce, telle que l'homme (1000 Genomes Project Consortium, 2010) ou *Arabidopsis thaliana* (Weigel et Mott, 2009), ont émergé. Le reséquençage de génomes peut apporter des informations sur la structure et l'histoire évolutive des populations. Il peut ainsi être appliqué à la génétique des populations, à l'écologie et à la biologie de l'évolution, par exemple, pour identifier des différences entre des souches fongiques

symbiotiques ou non symbiotiques. De plus, les technologies NGS permettent maintenant d'obtenir la séquence génomique d'espèces non modèles. Dans un premier temps, elles ont été utilisées en combinaison avec le séquençage Sanger, par exemple pour le séquençage du génome de *Vitis vinifera* (Velasco et al, 2007). Les deux premiers génomes d'eucaryotes séquencés entièrement par ces technologies sont le génome du Panda géant (Li et al, 2010) et le génome du champignon *Sordaria macrospora* (Nowrousian et al, 2010). De nombreux génomes sont actuellement séquencés avec ces technologies (Martinez et Nelson, 2010). La capacité à séquencer beaucoup d'organismes permet de réaliser des études comparatives et évolutives à grande échelle, ce qui était inimaginable il y a juste quelques années. Par ailleurs, le débit de séquençage des NGS permet également d'appréhender la biodiversité de manière plus vaste par des approches de métagénomique. En effet, le séquençage de communautés microbiennes entières permet d'accéder à l'étendue et à la composition de la biodiversité au sein de différents écosystèmes (Alvarez et al, 2009 ; Buée et al, 2009). Ces technologies sont également appliquées à d'autres questions biologiques qui étaient jusqu'alors traitées par d'autres approches. Par exemple, elles peuvent être utilisées pour étudier la distribution des protéines interagissant avec l'ADN (ChIP-Seq) ou la distribution de la méthylation (methyl-Seq) de l'ADN au sein d'un génome, qui sont deux processus gouvernant la régulation de l'expression des gènes (Johnson et al, 2007 ; Zemach et al, 2010). Enfin ces technologies sont appliquées à l'étude du transcriptome.

2 - La transcriptomique

Le transcriptome représente l'ensemble des ARN issus de l'expression du génome (séquences transcrites) dans un type cellulaire donné à un moment donné et dans une condition biologique précise. Il s'agit d'une combinaison d'ARN messagers codant pour des protéines (ARNm) et d'ARN non codants. L'un des principaux mécanismes initiant les processus adaptatifs dans une cellule est la modification des niveaux de transcription des gènes qui conduit à des changements dans la machinerie cellulaire. Les principaux objectifs de l'étude du transcriptome, appelée transcriptomique, sont de cataloguer tous les types de transcrits, de déterminer leurs structures, et de quantifier leurs niveaux d'expression pour un stade de développement ou une condition physiologique donnée. La transcriptomique permet ainsi d'avoir un aperçu du fonctionnement interne d'un génome dans un tissu ou une condition physiologique donnée. Cette discipline est donc fondamentale pour la compréhension des mécanismes gouvernant les processus biologiques. La transcriptomique peut s'étudier par différentes approches, notamment par des approches basées sur le séquençage d'ADN et des approches basées sur l'hybridation moléculaire.

2.1 - Séquençage d'ESTs

Les ARNm représentent les gènes exprimés au sein d'une cellule dans une condition donnée. Le séquençage de ces ARNm permet donc un accès direct à la part exprimée du génome pour cette condition. Toutefois, les ARNm ont un taux de dégradation rapide et ne peuvent pas être clonés directement. Afin de déterminer leurs séquences, des banques d'ADN complémentaire (ADNc) sont synthétisées à partir des ARNm extraits d'un échantillon de cellules ou de tissus d'un organisme. Ces séquences d'ADNc, appelées étiquettes de séquences exprimées (ESTs), ne représentent que des fragments des transcrits d'origine. Le séquençage Sanger de banques d'ESTs permet d'avoir une idée du contenu en gènes d'un génome. Cette approche a longtemps été la méthode de base pour la découverte de transcrits de référence (Boguski et Toltoshev, 1994 ; Gerhard et al, 2004). Elle a également été largement utilisée comme soutien aux grands projets de séquençage génomique, notamment pour la prédiction de la structure des gènes. Toutefois, le séquençage Sanger de banques d'ESTs reste une méthode lourde avec un débit relativement faible et ne peut généralement pas identifier les transcrits exprimés à des niveaux d'expression faibles. Cette approche est donc généralement non quantitative. De plus, elle comporte des biais de clonage qui affectent la représentation des séquences.

Plus récemment, d'autres méthodes basées sur le séquençage Sanger de courts fragments d'ADNc (9 à 14 pb) ont été développées, notamment l'analyse sérielle d'expression génique (SAGE) (Velculescu et al, 1995 ; Harbers et Carninci, 2005). Ces approches ont un débit plus élevé et permettent de quantifier l'abondance de plusieurs milliers de transcrits au sein d'une population d'ARN donnée. Toutefois, ces méthodes sont longues à mettre en œuvre et coûteuses. La mesure exhaustive des niveaux d'expression des ARNm de faible abondance n'est pas possible avec ces méthodes, en raison du coût et des contraintes technologiques. De plus, ces méthodes se limitent à l'évaluation des niveaux d'expression des gènes et ne fournissent aucune information sur les isoformes d'épissage ou la découverte de nouveaux gènes.

2.2 - Puce à ADN

Les niveaux d'expression des gènes peuvent également être étudiés par des approches basées sur l'hybridation moléculaire entre des séquences complémentaires d'ADN, appelées puces à ADN. Des dizaines de milliers de sondes oligonucléotidiques spécifiques des gènes de l'organisme étudié sont fixées par synthèse *in situ* sur un support solide. Afin d'étudier le niveau d'expression de l'ensemble des transcrits d'un échantillon donné, des ADNc sont synthétisés à partir de la population d'ARN de cet échantillon et sont marqués par un fluorochrome. Ces ADNc sont ensuite hybridés de manière spécifique aux sondes oligonucléotidiques. La quantité des molécules d'ADNc qui s'hybride aux sondes représente la valeur d'expression relative ou absolue des transcrits. Celle-ci est mesurée par l'intensité du signal de fluorescence. Ainsi, les puces à ADN permettent de suivre simultanément les niveaux d'expression de tous les gènes annotés dans une population cellulaire (Shalon et al, 1996 ; Schena et al, 1998). Suite aux progrès de ces technologies, les approches basées sur l'hybridation de séquences ont un débit élevé et sont relativement peu coûteuses. Elles sont donc devenues largement disponibles et applicables pour de nombreux laboratoires. Ainsi, depuis plus de 10 ans, les puces à ADN constituent l'approche la plus largement utilisée pour la quantification du transcriptome (Kammenga et al, 2007 ; Shalon et al, 1996 ; Schena et al, 1998). La capacité d'analyser l'ensemble des niveaux de l'expression des gènes a permis de générer des informations sur l'ensemble de la réponse cellulaire et d'élargir ainsi notre compréhension des processus globaux régulant l'expression des gènes.

Plus récemment, des puces à ADN, appelées "tilling arrays", ont été conçues pour mesurer systématiquement les niveaux d'expression à l'échelle de la séquence génomique entière, indépendamment de toute annotation de gènes. L'application de cette technique à plusieurs organismes différents a démontré que la complexité des transcriptomes avait été largement sous-estimée et que les ARN transcrits à partir de portions non codantes du génome jouaient un rôle fondamental (Bertone et al 2005 ; Kapranov et al, 2007).

Mais ces approches basées sur l'hybridation de séquences ont aussi plusieurs limitations (Bloom et al, 2009). Tout d'abord, l'application des puces a été principalement limitée à des espèces modèles avec des séquences de génome connues. De plus, la technologie des puces a une gamme dynamique (le ratio du plus petit au plus grand signal fluorescent) relativement limitée pour la détection de niveaux de transcription, en raison à la fois du bruit de fond et de la saturation du signal. De plus, avec un niveau de bruit de fond élevé et des problèmes d'hybridation croisée, la détection des ARNm de très faible abondance et la distinction entre séquences d'ARNm très proches sont rendues difficiles (Okoniewski et Miller, 2006 ; Royce et al, 2007 ; Casneuf et al, 2007). Par ailleurs, la comparaison des niveaux d'expression entre différentes expériences est souvent ardue et peut exiger des méthodes de normalisation complexes (Hinton et al, 2004).

2.3 - Application des NGS à la transcriptomique

Récemment, le développement des technologies de séquençage à haut débit a rendu possible la production de quantités massives de données de séquences avec un débit variant de centaines de milliers à plusieurs millions de séquences selon la plateforme utilisée. La capacité à générer rapidement et à moindre coût de grands ensembles de données de séquence d'ADNc permet la réalisation d'études de transcriptomique pour un large éventail d'organismes, de types cellulaires et de conditions environnementales. Cette approche d'analyse du transcriptome basée sur le séquençage direct à haut débit de banques d'ADNc est appelée RNA-Seq (Marguerat et Bähler, 2008 ; Wang et al, 2009). Selon les objectifs de l'étude transcriptomique, différentes stratégies peuvent être employées. Ces objectifs peuvent être de caractériser le transcriptome d'espèces non modèle, de déterminer la structure des transcrits ou de quantifier l'expression des gènes.

La grande profondeur de séquençage générée par les NGS est extrêmement puissante pour la quantification des transcrits (Mortazavi et al, 2008 ; Sultan et al, 2008). Si les banques d'ADNc sont séquencées avec une couverture élevée, les lectures séquencées peuvent être alignées au génome de référence de l'espèce et le nombre de lectures alignées à chaque transcrit peut être quantifié. Le raisonnement qui sous-tend cette approche est que la représentation des séquences issues du séquençage profond des ADNc est proportionnelle à la quantité d'ARN du gène dans l'échantillon initial. L'abondance de chaque transcrit au sein des différents échantillons peut ensuite être comparée. Si un nombre suffisant de lectures sont séquencées pour un échantillon, il devrait en théorie être possible de détecter et de quantifier tous les transcrits présents dans une population cellulaire (Mortazavi et al, 2008). Cette mesure numérique de l'expression est généralement moins touchée par les problèmes de bruit de fond élevé que les approches basées sur l'hybridation. Un certain nombre d'études ont démontré que cette approche est considérablement plus sensible pour la détection de transcrits rares (Cloonan et al, 2008 ; Sultan et al, 2008). Cette approche basée sur le séquençage permet également d'obtenir une gamme de mesure de l'expression beaucoup plus vaste que celle des puces à ADN (Marioni et al, 2008 ; Mortazavi et al, 2008 ; Sultan et al, 2008). Cette gamme, limitée en théorie uniquement par la profondeur de séquençage, est plus proche de la gamme d'abondance des transcrits dans une cellule. Ainsi, les niveaux d'expression mesurés par les puces à ADN et par RNA-Seq divergent pour les transcrits de faible ou de forte abondance (Wang et al, 2009). Les écarts entre les mesures d'expression obtenues par RNA-Seq et par puce à ADN ont été testés par qRT-PCR. Globalement les résultats obtenus sont en faveur des données RNA-seq (Marioni et al, 2008). De plus, les données d'expression issues de l'approche RNA-seq montrent peu de variations techniques et sont donc très répliquables (Mortazavi et al, 2008 ; Marioni et al, 2008).

La profondeur de séquençage accrue fournit également une résolution à l'échelle du nucléotide, ce qui permet l'identification des limites de la transcription sans la connaissance préalable des régions transcrites dans un génome (Mortazavi et al, 2008 ; Wilhelm et al, 2008). Cette approche peut donc être utilisée pour cartographier les limites de transcription, par exemple, les extrémités 5' et 3' des régions non traduites, ainsi que les limites des exons et des introns. Elle peut également permettre la découverte de nouveaux gènes et la détection d'épissage alternatif. En outre, cette approche peut aider à l'amélioration ou à la correction de l'annotation de la structure des transcrits du génome (Denoëud et al, 2008, Larsen et al, 2010). L'introduction de ce type de données dans un projet de séquençage de génome peut donc être extrêmement utile pour améliorer la prédiction automatique des gènes.

De nombreuses études ont utilisées cette approche pour analyser simultanément la structure des gènes et les profils d'expression à un degré de détail remarquable. Les premières études ont été tout d'abord menées sur des cellules souches de souris et d'homme ainsi que sur les levures

(Nagalakshmi et al, 2008 ; Wilhelm et al, 2008 ; Cloonan et al, 2008 ; Mortazavi et al, 2008 ; Sultan et al, 2008). Depuis ces études se sont multipliées et ont permis d'explorer le transcriptome de nombreux organismes, notamment de plantes et de champignons (Lister et al, 2008 ; Zhang et al, 2010 ; Wang et al, 2010b ; Larsen et al, 2011).

Les NGS peuvent également être utilisées pour caractériser le transcriptome d'organismes non modèles pour lesquels la séquence génomique n'a pas encore été déterminée. L'objectif est alors d'obtenir une description générale des gènes et des fonctions exprimés dans un tissu ou un stade de vie au sein d'un organisme. Les premières études décrivant le transcriptome d'une espèce non modèle par séquençage NGS ont été réalisées sur la guêpe *Polistes metricus* (Toth et al, 2007) et le papillon *Melitaea cinxia* (Vera et al, 2008). Depuis ces premières études, une multitude de séquençage de transcriptomes a été réalisé pour un certain nombre d'organismes non modèle. Pour de nombreuses études de caractérisation du transcriptome, il est préférable d'avoir une représentation aussi large que possible du transcriptome, notamment par la mise en commun de plusieurs tissus et/ou de stades de vie lors de la production des banques d'ADNc. Ces études de caractérisation des transcriptomes fournissent un point de départ important et une ressource précieuse pour des analyses plus approfondies (Ellegren, 2008). Par exemple, il est possible d'utiliser ce transcriptome de référence pour créer une puce à ADN pour des études fonctionnelles à grande échelle (Garcia-Reyero et al, 2008).

3 - Les nouvelles technologies de séquençage

Les technologies de séquençage de nouvelle génération (NGS) permettent à des millions de réactions de séquençage d'avoir lieu en même temps dans de petits volumes de réaction. De plus, elles ne nécessitent pas de clonage de l'ADN dans une banque de plasmides, ce qui, en plus d'économiser du temps de travail, a l'avantage d'éviter les biais de clonage. Toutefois, ces méthodes produisent des lectures de séquence plus courtes que celles générées par le séquençage Sanger par capillaire (650 à 1000 pb), ainsi que des taux d'erreurs plus élevés. Actuellement, les trois plateformes de séquençage NGS les plus populaires sont 454, Solexa et SOLID. Elles sont généralement divisées en deux catégories : les plateformes produisant de larges quantités de lectures courtes (Solexa et SOLID) et les plateformes produisant des lectures plus longues en moindre quantité (454). Il est à noter que chacune de ces plateformes est en constante évolution, ce qui entraîne une augmentation du débit, de la longueur et de la qualité des lectures de séquences.

3.1 - Roche/454

Le système de pyroséquençage Roche/454 a été la première technologie de séquençage de nouvelle génération introduite sur le marché en 2005 (Figure 16) (Margulies et al, 2005). Après fragmentation de l'ADN par nébulisation, les fragments sont sélectionnés par la taille puis chaque fragment est fixé individuellement à une nanobille grâce à des adaptateurs spécifiques. Ces billes sont placées dans une émulsion permettant l'amplification des fragments en parallèle par PCR (Polymerase Chain Reaction). Chaque bille est ensuite déposée dans un puits sur une plaque de picotitration et les réactions de séquençage ont lieu simultanément dans chacun des puits. Pendant le pyroséquençage, chaque cycle se compose de l'introduction d'un type unique de nucléotides. L'incorporation d'un nucléotide complémentaire par la polymérase, au sein du brin néosynthétisé dans l'un des puits, libère un groupe pyrophosphate. Ce groupe pyrophosphate va être converti en ATP par une ATP-sulfurylase à partir d'APS (adénosine 5'-phosphosulfate). Finalement, un signal lumineux sera produit par l'enzyme luciférase lors de l'oxydation de la luciférine en oxyluciférine. Cette réaction est rendue possible par la présence d'ATP. Le signal lumineux sera alors capté par une caméra à transfert de charge (CCD) capable de détecter la position de la lumière émise. On peut ainsi en déduire la séquence à partir d'une succession de pics de signaux lumineux (pyrogramme). Comme il n'y a pas de nucléotides terminaux qui empêchent plusieurs incorporations consécutives pour un cycle donné, la longueur des homopolymères doit être déduite de l'intensité du signal. Par conséquent, le type d'erreur le plus courant pour cette technologie sont les insertions, suivies par les délétions (Metzer, 2011). Ce système qui générait à l'origine des lectures de 100 pb, permet actuellement d'obtenir des lectures de 400 pb en moyenne, et la société Roche est maintenant sur le point de produire des séquences de 800 pb. Le dernier modèle commercialisé permet de générer plus d'un million de lectures de séquences.

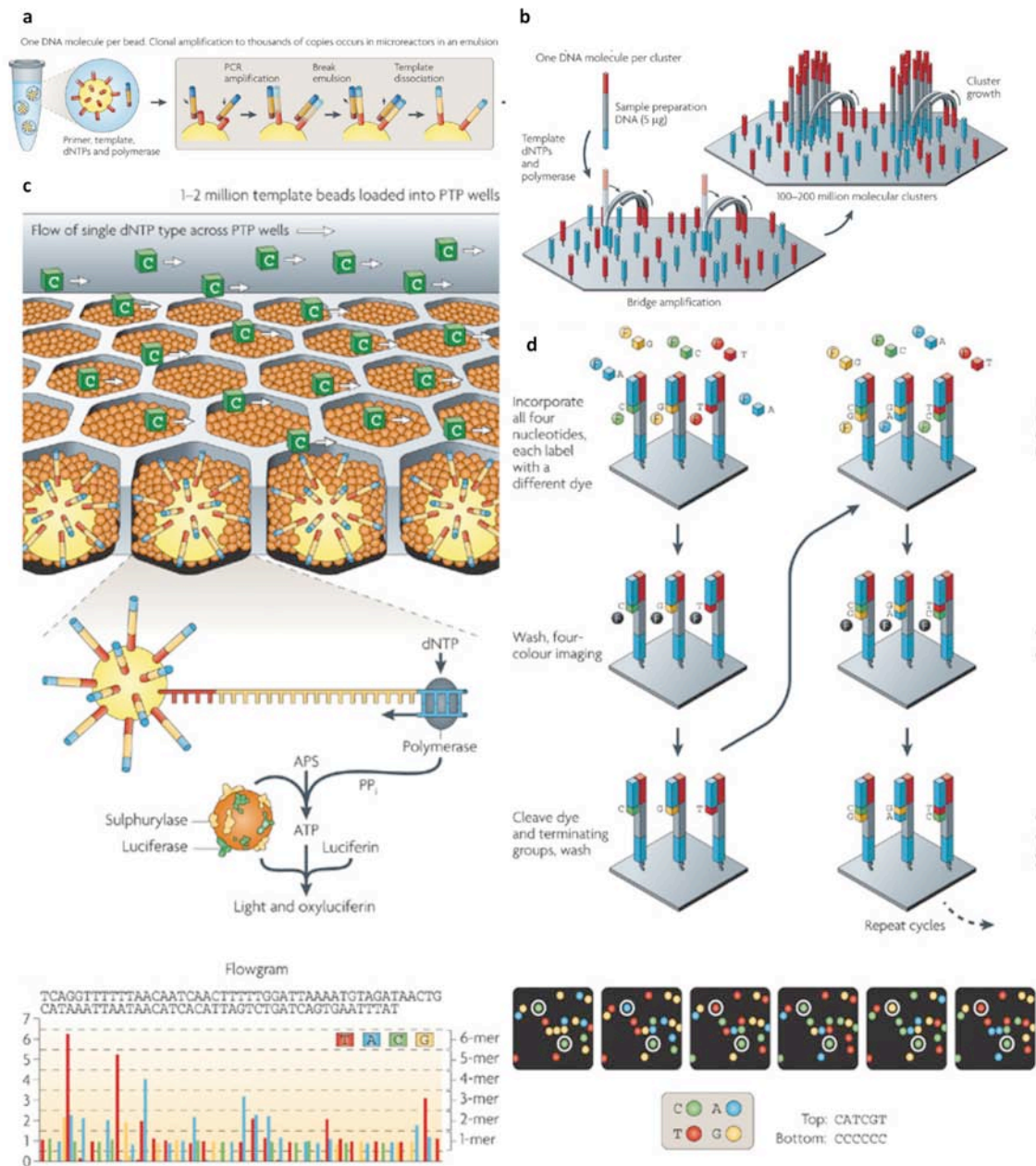


Figure 16 : Principe des technologies de séquençage de nouvelles générations 454/Roche et Illumina/Solexa. (a) Méthode d'amplification 454 : PCR en émulsion (emPCR). Chaque nanobille portant une molécule d'ADN simple brin est mise en émulsion au sein de gouttelettes (microréacteurs) en présence des réactifs de PCR pour permettre l'amplification clonale de chaque fragment. (b) Méthode d'amplification Illumina : Les fragments d'ADN simple brin sont immobilisés sur une plaque de verre. Après l'ajout de réactif de PCR, l'amplification par pont des brins immobilisés avec les amorces adjacentes est réalisée pour former des clusters denses de fragments d'ADN. (c) Méthode de séquençage 454 : Après amplification, les nanobilles porteuses d'ADN amplifié sont déposées dans des puits individuels, sur une plaque de picotitration en fibre optique. D'autres billes couplées avec la sulphurylase et la luciférase, ainsi que les réactifs de séquençage, sont ajoutés dans les puits. La lumière générée à partir de chaque puits subissant la réaction de pyroséquençage est détectée par une caméra CDD à haute résolution reliée à la lame en fibre optique. (d) Méthode de séquençage Illumina : Chaque cycle de séquençage est initié par l'ajout des quatre terminateurs marqués ainsi que des réactifs de séquençage. Après l'incorporation des nucléotides terminaux marqués, l'image de la fluorescence émise par chaque cluster est capturée par lecture optique. Chaque cycle se termine par une étape de clivage des fluorochromes et restauration du groupe 3'-OH. (Metzer, 2010)

3.2 - Illumina/Solexa

Solexa est la seconde technologie NGS apparue sur le marché (Figure 16) (Bentley et al, 2008). Les brins d'ADN sont fragmentés aléatoirement par nébulisation et des adaptateurs sont fixés à chaque extrémité des fragments. Pour cette technologie, l'amplification des fragments se fait en phase solide. Les fragments sont fixés arbitrairement sur une surface par hybridation des adaptateurs. Les extrémités libres des fragments immobilisés s'hybrident à une amorce complémentaire (amplification par ponts) et sont copiés par la polymérase. Ce processus d'amplification est ensuite répété, aboutissant à la formation de clusters de brins d'ADN. Plusieurs millions (100 à 200) de clusters sont ainsi amplifiés simultanément sur la surface. En fin d'amplification, chaque cluster contient jusqu'à 1000 molécules identiques. Cette haute densité de brins d'ADN facilite et augmente le débit de séquençage. Le séquençage des fragments amplifiés se fait *via* un procédé de séquençage par synthèse qui incorpore les quatre nucléotides, marqués par des fluorochromes différents, au sein de brins d'ADN immobilisés. Comme pour la technique de Sanger, cette technique est basée sur l'utilisation de terminateurs réversibles contenant un groupement de protection attaché au nucléotide qui termine la synthèse d'ADN. A chaque cycle de séquençage, les quatre terminateurs réversibles ainsi que des amorces et la polymérase sont ajoutés sur la plaque de verre. Les terminateurs réversibles sont alors incorporés dans les brins d'ADN. Après l'excitation laser, l'image de la fluorescence émise par chaque cluster est capturée par lecture optique. Chaque cycle se termine par le clivage du terminateur et la restauration du groupement fonctionnel du nucléotide incorporé, ce qui permet à l'ADN polymérase d'incorporer le prochain nucléotide lors du cycle suivant. Les cycles sont ainsi répétés pour déterminer la séquence de chacun des fragments, base par base. Cette technique reposant sur la chimie des terminateurs, les insertions et les délétions sont rares. Les substitutions sont le type d'erreurs le plus courant pour cette technologie. En raison de la diminution de la capacité de détection de la fluorescence à partir d'un certain nombre de cycles, le taux d'erreurs augmente à la fin des séquences. De plus, il semble que les séquences riches en AT et en GC soient sous-représentées avec cette technologie (Metzker, 2011). La très haute densité de la puce (plus de 100 millions de molécules par centimètre carré) permet de séquencer jusqu'à plusieurs dizaines de millions de lectures. Cette plateforme fournissait à l'origine des séquences courtes de 25 à 36 pb. Les systèmes actuels permettent d'obtenir des lectures de 100 pb.

3.3 - ABI SOLiD Applied Biosystem

Avec cette technologie, l'amplification est réalisée par émulsion PCR sur des billes qui sont ensuite fixées sur une plaque de verre. Par la suite, le système SOLiD utilise une technique de séquençage par ligation. Ainsi la détection de la base ne se fait plus par l'intégration des nucléotides *via* la polymérase mais par l'incorporation de sondes par une ligase, chacune des sondes représentant un dinucléotide. Bien qu'il y ait 16 paires de bases possibles, ce système n'utilise que quatre couleurs pour les représenter. Ces couleurs codent les transitions entre deux nucléotides successifs. Chaque base est donc codée deux fois par deux couleurs successives. Ce système permet une correction interne des erreurs lors de l'alignement. Les données de séquences sont codées sous la forme d'une succession de données de couleurs, appelée « color-space ». Ce type de données nécessite l'utilisation d'outils bio-informatiques spécifiques. Cette technologie fournie plusieurs centaines de millions de lectures courtes de 25 à 75 pb (Metzker, 2011).

Actuellement, des technologies de séquençage de troisième génération, telle que PacBio (Pacific Biosciences), sont en développement et commencent à devenir disponibles (Eid et al, 2009). Ces technologies ne requièrent pas l'amplification des molécules et séquentent directement les brins d'ADN en temps réel. Dans l'avenir, ces plateformes pourraient produire des lectures de plusieurs kilobases, laissant ainsi entrevoir la possibilité de séquencer un transcrit dans toute sa longueur en une seule lecture. De plus, ces technologies pourraient, dans un futur proche, permettre le séquençage direct des molécules d'ARN, sans conversion préalable en ADNc, éliminant ainsi les biais introduits lors de la synthèse des cDNA. Ces technologies ont un taux d'erreurs encore élevé par rapport aux autres plateformes (15 à 20%) mais elles sont en cours d'amélioration (Metzker, 2011).

4 – Les outils bio-informatiques pour l'analyse des transcriptomes

L'analyse des grands jeux de données générées par les NGS nécessite l'utilisation d'outils bio-informatiques afin de gérer et d'interpréter pleinement ces données de séquences. Selon les objectifs, le type de plateforme utilisé et les ressources génomiques disponibles pour l'organisme étudié, les séquences issues des NGS sont soit alignées sur un génome de référence ou assemblées *de novo*. Enfin, une autre étape importante pour l'analyse des transcriptomes est l'annotation fonctionnelle qui permet d'attribuer des fonctions biologiques potentielles aux différents transcrits.

4-1 - Assemblage *de novo* des données NGS

Pour de nombreux organismes, la séquence de référence du génome est incomplète ou indisponible. Dans ce cas, il est nécessaire de reconstruire les séquences des transcrits d'origine par une stratégie d'assemblage *de novo*. Cette stratégie utilise la redondance des lectures séquencées pour trouver des chevauchements entre celles-ci et les assembler en contigs. Un contig correspond à la séquence consensus d'un alignement multiple de lectures. Cette stratégie permet d'éliminer la redondance au sein des données et d'obtenir des séquences plus longues. La plupart des projets nécessitant un assemblage *de novo* ont utilisé des données issues de la plateforme 454 (Vera et al, 2008 ; Novaes et al, 2008 ; Meyer et al, 2009 ; Parchman et al, 2010).

Toutefois, reconstruire un transcriptome à partir de lectures NGS relativement courtes n'est pas trivial. La taille importante des jeux de données nécessite des systèmes informatiques ayant une mémoire importante et/ou de nombreux cœurs pour exécuter les algorithmes en parallèles (Zerbino et Birney, 2008 ; Birol et al, 2009). De plus, le taux d'erreur plus élevé des lectures NGS peut entraîner des assemblages fragmentés ou des erreurs d'assemblage (Martin et Wang, 2011). Par ailleurs, il existe des défis supplémentaires qui sont spécifiques à l'assemblage des données de transcriptome. Certains programmes dédiés à l'assemblage de génomes ne peuvent pas être directement appliqués à l'assemblage de transcriptomes. En effet, beaucoup d'assembleurs de génomes utilisent la distribution uniforme de la couverture de séquençage pour distinguer les régions répétitives du génome et pour calculer un ensemble de paramètres optimaux pour l'assemblage. Toutefois, le nombre de copies d'ARNm par gène dans une cellule peut varier d'un seul transcrit à plusieurs dizaines de milliers. Dans un jeu de données transcriptomiques, la couverture de séquençage peut donc s'échelonner sur plusieurs ordres de grandeur, selon l'abondance des transcrits. L'information de couverture ne peut donc pas être utilisée pour l'assemblage des transcriptomes (Zerbino et Birney, 2008). Ces variations de profondeur de séquençage conduisent également à une explosion des besoins en mémoire. Par ailleurs, la présence des UTRs, des variants de transcription, des variants d'épissage, des gènes paralogues et de polymorphisme complique l'assemblage des transcriptomes (Chevreux et al, 2004 ; Martin et Wang, 2011).

Les artefacts présents au sein des jeux de données brutes peuvent conduire à de mauvais assemblages. Il est donc nécessaire de les éliminer pour améliorer la précision et l'efficacité des calculs d'assemblage. Plusieurs types d'artefacts peuvent être retirés des données de séquences, notamment, les adaptateurs de séquençage, les fragments de vecteurs utilisés pour créer des banques d'ADNc et les lectures de faible complexité. Les séquences indésirables sur-représentées, telles que les séquences provenant de transposons ou d'ADNr ainsi que d'autres contaminants doivent également être éliminés. Il existe plusieurs outils pour nettoyer ces données de séquençage : SeqClean (<http://compbio.dfci.harvard.edu/tgi/software/>), SeqTrim (Falgueras et al, 2010) et TagDust (Lassmann et al, 2009).

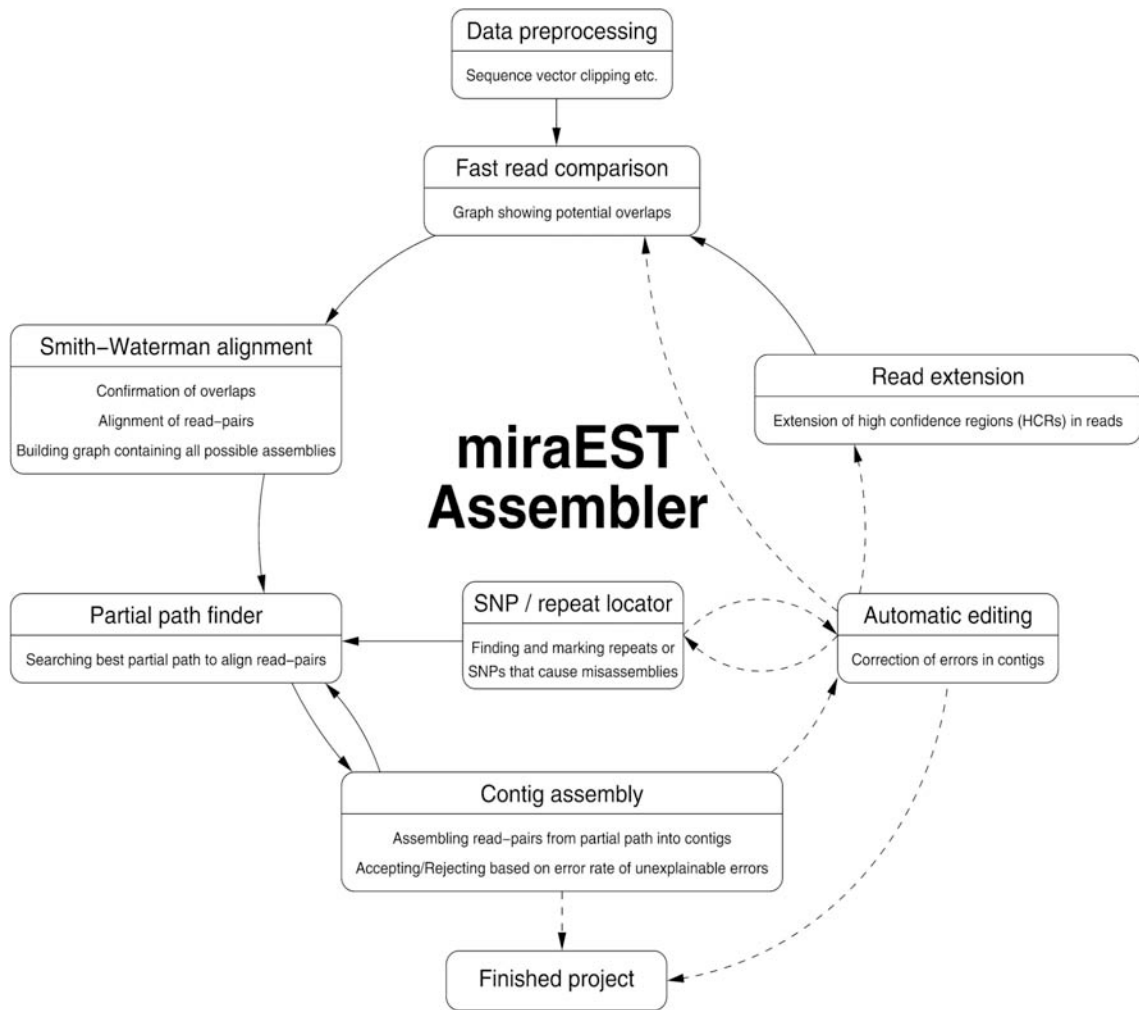


Figure 17 : Processus itératif d’assemblage de MIRA. Un algorithme rapide, SKIM, compare les régions de haute qualité des lectures deux à deux, afin de détecter les chevauchements potentiels. Un algorithme d’alignement Smith-Waterman confirme alors les chevauchements. Un graphe de chevauchement est ensuite construit. La résolution de ce graphe permet de trouver la meilleure manière d’aligner les lectures qui sont ensuite assemblées en contig. Les erreurs d’assemblage au sein des contigs sont corrigées. Les régions de haute qualité des lectures sont ensuite étendues. De manière réursive, les contigs sont de nouveaux traités par l’algorithme d’assemblage (Chevreux et al, 2004).

Les queues polyA, qui font partie des ARNm et donc des données séquencées, sont un autre problème pour l'assemblage des transcrits. Si elles sont conservées dans le jeu de données, les différents variants d'épissage au niveau de ces régions ne seront pas assemblés au sein d'un même contig. De plus, les homopolymères étant des sources d'erreurs pour la technologie 454, des erreurs au niveau de ces queues polyA peuvent entraver la construction correcte des contigs. C'est la raison pour laquelle la pratique la plus courante est d'éliminer les queues polyA du jeu de données de séquences.

Les erreurs de séquençage dans les lectures NGS peuvent être corrigées en utilisant les scores de qualité qui peuvent indiquer de possibles erreurs de séquençage. Alternativement, les erreurs de séquençage peuvent aussi être identifiées de façon empirique en utilisant les fréquences de chaque k-mer (oligomère de longueur k) dans l'ensemble de données. Etant donné que la même molécule d'ARN est séquencée plusieurs fois, les k-mers sans erreurs seront plus nombreux. En revanche, les k-mers avec des fréquences très faibles sont probablement des erreurs de séquençage ou des transcrits de faible abondance. Les lectures contenant ces erreurs peuvent être enlevées pour améliorer la qualité d'assemblage et diminuer la quantité de mémoire vive requise. Toutefois, cette approche peut éliminer les lectures issues de transcrits rares (Miller et al, 2010 ; Martin et al, 2010b).

La plupart des assembleurs actuels reposent sur une approche de graphe de chevauchement, appelée méthode OLC (« Overlap/Layout/Consensus »). Cette approche est typique des assembleurs de données Sanger (Miller et al, 2010). Toutes les lectures sont d'abord comparées deux à deux pour identifier les chevauchements. La construction d'un graphe de chevauchement représentant les lectures et leurs chevauchements permet ensuite d'organiser les lectures en contigs. Enfin, un alignement multiple des lectures détermine la séquence consensus (Miller et al, 2010). La longueur minimale du chevauchement et le pourcentage d'identité de celui-ci sont des paramètres qui affectent la robustesse de l'assemblage face aux erreurs de séquençage et à la faible couverture. L'augmentation des valeurs de ces paramètres conduit à des contigs plus précis mais plus courts. La méthode OLC a été adaptée aux données 454 pour de nombreux assembleurs tels que Newbler (Margulies et al, 2005), MIRA (Chevreux et al, 2004), SeqMan (Swindell et Plasterer, 1997), TGICL (Pertea et al, 2003).

MIRA et MiraSearchESTSNPs sont deux assembleurs qui ont été utilisés lors de ce travail de thèse. MIRA (Mimicking Intelligent Read Assembly) est un programme qui permet d'assembler *de novo* des séquences génomiques ou transcriptomiques selon le paramétrage utilisé. Il a été le premier assembleur en mesure de réaliser des assemblages hybrides combinant des données issues de plusieurs plateformes de séquençage. MIRA est un assembleur itératif, qui recherche les erreurs d'assemblage après chaque itération et utilise les informations sur les séquences qui les ont causées pour résoudre les conflits d'assemblage dans les itérations suivantes (Figure 17) (Chevreux et al, 2004). Les routines internes de MIRA prennent en compte les scores de qualité. Ainsi, MIRA utilise dans un premier temps les régions de bonne qualité au sein des lectures, puis étend progressivement les alignements. MIRA fonctionne comme un assembleur et non comme un algorithme de « clustering », de ce fait les lectures seront assemblées en contigs de manière à reconstituer les séquences transcrites d'origine. Les séquences composées de différentes structures d'exons ou de large indels ne seront donc pas assemblées ensemble. MiraSearchESTSNPs est un pipeline automatisé issu de MIRA, dont le but est d'assembler des transcrits issus de différentes souches d'un organisme donné. Les séquences qui diffèrent par un polymorphisme d'un seul nucléotide (SNP) seront tout d'abord traitées comme des transcrits différents. Les SNP peuvent ainsi être assignés à leur séquence d'ARNm correspondante au sein du transcriptome. La dernière étape de ce processus assemble alors les transcrits contenant la même séquence exonique en une seule séquence consensus. MiraSearchESTSNPs peut également être utilisé avec des données provenant d'une seule souche. Dans ce cas, les transcrits seront assemblés selon leur allèle respectif lorsque ceux-ci contiennent des mutations.

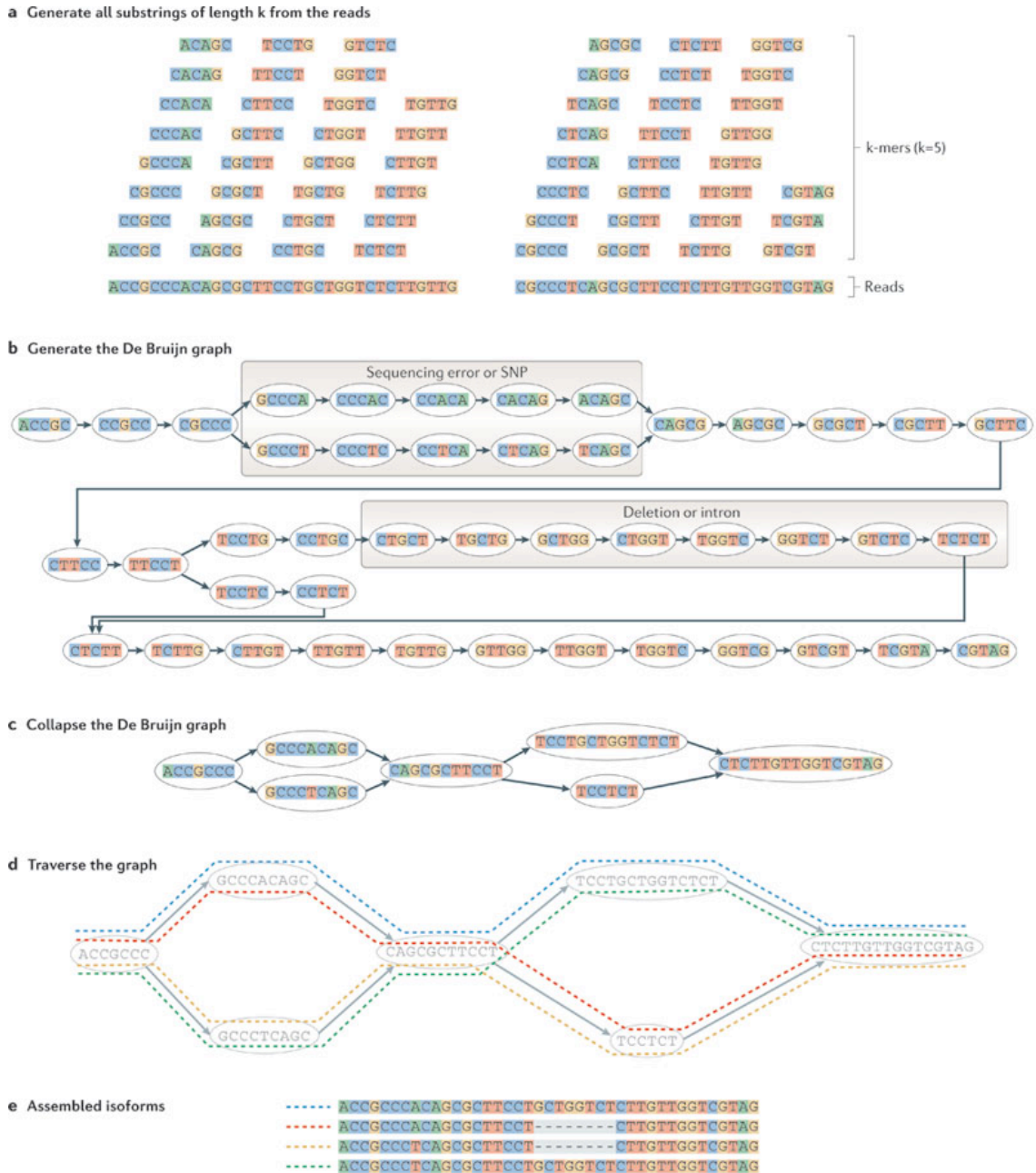


Figure 18 : Stratégie d'assemblage *de novo* de transcriptome basée sur l'approche de graphe de Bruijn. Toutes les sous-chaînes de longueur k (k-mer) possibles sont générées à partir de chaque lecture. Chaque k-mer représente un nœud dans le graphe et deux nœuds seront connectés lorsque le décalage d'un nucléotide au sein d'un k-mer conduit à un chevauchement parfait (de taille k-1) entre deux k-mer. Les chaînes de nœuds dans le graphe sont ensuite fusionnées si deux nœuds successifs ne sont connectés que par un seul chemin. Les différents chemins traversant le graphe sont alors utilisés pour assembler les différentes isoformes. (Martin et Wang, 2011)

Avec l'arrivée des plateformes de type Illumina ou SOLiD, aucun assembleur basé sur la méthode OCL n'a été en mesure de traiter les quantités massives de données de séquences courtes générées. Récemment une nouvelle approche basée sur le graphe de Bruijn a été développée pour gérer efficacement ce type de données (Figure 18) (Martin et Wang, 2011). Cette méthode repose sur la notion de k-mers qui correspondent à des sous-chaînes de lectures de longueur k. Intuitivement, deux lectures ayant une similitude de séquence partagent des sous-chaînes identiques dans leurs régions chevauchantes. La détection rapide de k-mers identiques entre plusieurs lectures permet d'identifier implicitement les chevauchements parfaits, sans aucun calcul d'alignement deux à deux, ce qui réduit considérablement le coût de calcul (Miller et al, 2010). Récemment, l'assemblage *de novo* d'un transcriptome d'une espèce, dont le génome est connu, a permis de montrer que les lectures courtes de type Solexa/SOLID peuvent être utilisées pour l'assemblage de transcriptomes d'organismes non-modèle (Gibbons et al, 2009). Une poignée d'assembleurs basés sur cette approche permettent maintenant de reconstruire *de novo* les transcrits et leurs isoformes à partir de lectures courtes : Rnnotator (Martin et al, 2010b), Multiple-k (Surget-Groba et Montoya-Burgos, 2010) et Trans-ABYSS (Robertson et al, 2010). La qualité de l'assemblage dépend notamment de la longueur de k-mer choisie. Dans le cas des transcriptomes, pour lesquels la couverture n'est pas uniforme, l'utilisation d'une valeur de k-mer plus grande conduit à un meilleur assemblage des transcrits fortement exprimés ; alors que l'assemblage des transcrits faiblement exprimés sera de meilleure qualité si la valeur de k-mer est plus faible (Zerbino et Birney, 2008). Pour résoudre ce problème, certains assembleurs construisent plusieurs assemblages avec différentes valeurs de k-mer et fusionnent ensuite les contigs résultants, ce qui aboutit à un assemblage de meilleure qualité (Martin et al, 2010b). La plupart des assembleurs basés sur cette approche sont conçus pour gérer les lectures courtes (25 à 100 pb) de taille uniforme. Ils ne sont donc pas en mesure de traiter les données de séquence de type 454.

Alors qu'il existe des normes pour évaluer la qualité des assemblages de génomes, telles que le N50, les critères pour apprécier la qualité des assemblages de transcriptomes n'ont pas été établis. Les assemblages peuvent être comparés en utilisant des paramètres standard tels que le nombre total de lectures utilisées dans l'assemblage, le nombre de contigs générés ainsi que la longueur des contigs. Toutefois, ces statistiques n'apportent pas d'information sur la manière dont l'assemblage reflète le transcriptome. D'autres critères ont récemment été proposés (Martin et al, 2010b). Par exemple, la précision d'un assemblage peut être mesurée en calculant le pourcentage de contigs s'alignant au génome de référence. L'exhaustivité d'un assemblage peut être estimée en calculant le pourcentage de gènes annotés couverts par les contigs assemblés. La contiguïté d'un assemblage peut être reflétée par le pourcentage de gènes entièrement couverts par un seul contig. Enfin, la redondance d'un assemblage peut également être estimée en calculant le nombre de contigs couvrant une même région du génome de référence. Toutefois, ces mesures nécessitent qu'un génome de référence soit disponible pour l'organisme étudié. Alternativement, différents assemblages peuvent être comparés en utilisant des jeux de données de séquences provenant d'organismes différents (Papanicolaou et al, 2009).

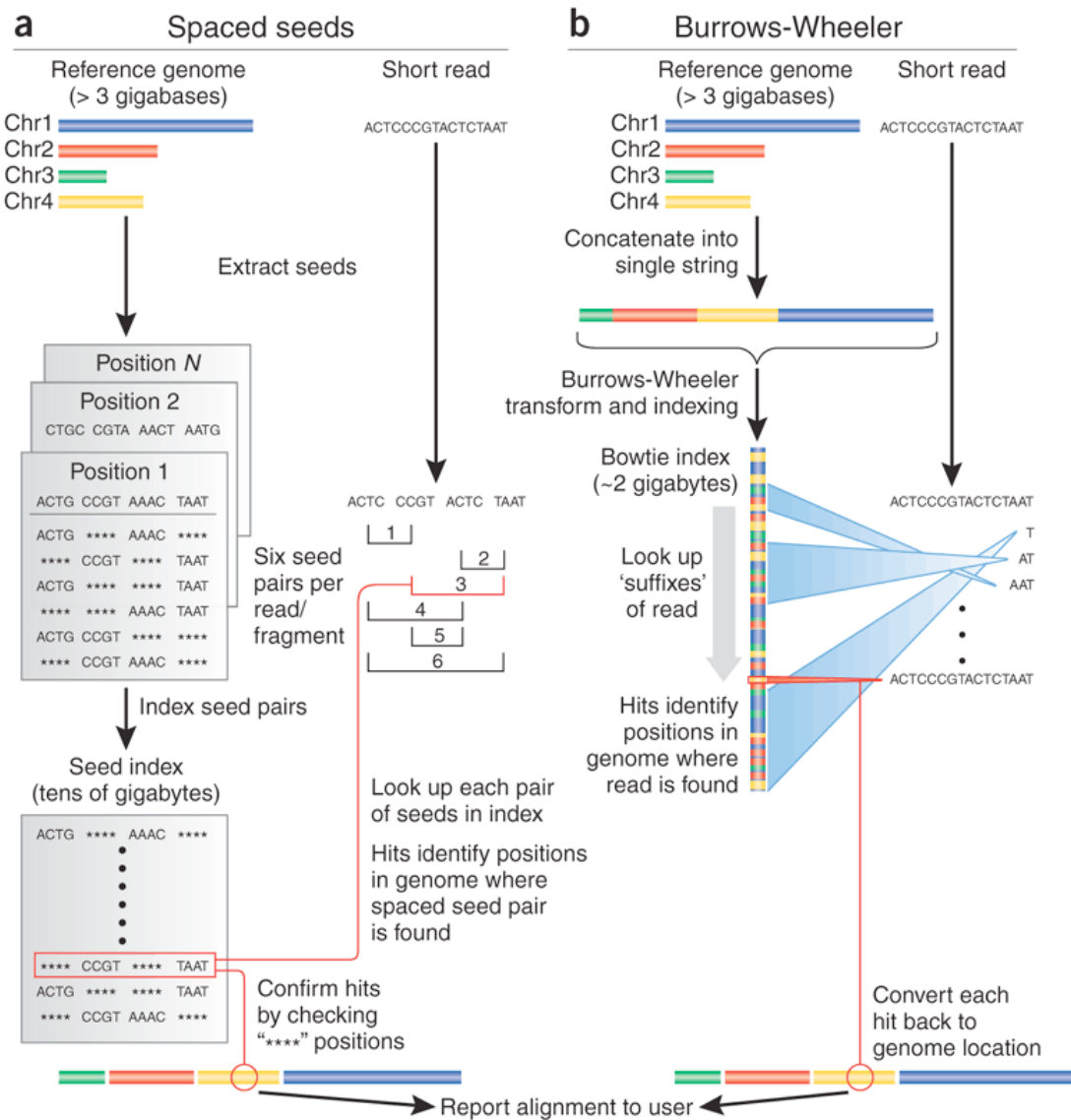


Figure 19 : Deux approches algorithmiques pour l'alignement de lectures courtes à un génome de référence. (a) Les algorithmes basés sur l'indexation des k-mer, par exemple Maq. (b) Les algorithmes basés sur la transformée de Burrows-Wheeler, par exemple Bowtie. (Trapnell et Salzberg, 2009).

4-2 - Alignement à un génome de référence

Lorsqu'une séquence génomique de référence est à disposition, une première étape dans l'analyse des données RNA-Seq est l'alignement des lectures au génome de référence. Les programmes précédemment développés pour les lectures Sanger, tels que BLAST (Altschul et al, 1997) ou BLAT (Kent, 2002) ne sont pas adaptés aux données NGS, car ils ne gèrent ni le grand volume de données, ni la courte taille des lectures générées (Trapnell et Salzberg, 2009). Par conséquent, un certain nombre de nouveaux algorithmes adaptés à l'alignement des lectures courtes NGS de type Solexa ont été développés ces dernières années ; par exemple, SOAP (Li et al, 2008a), Maq (Li et al, 2008b), Bowtie (Langmead et al, 2009), BFAST (Homer et al, 2009), BWA (Li et Durbin, 2009). Ces programmes ne permettent pas toujours l'utilisation de lectures plus longues de type 454.

En pratique, l'alignement de millions de séquences à un génome de grande taille fait appel à des algorithmes extrêmement efficaces et optimisés dans leur utilisation de la mémoire et des processeurs (méthodes de parallélisation). Ces programmes utilisent une stratégie appelée «indexation», qui leur permet de trouver rapidement des séquences courtes au sein de la séquence génomique. Deux méthodes différentes sont généralement utilisées. La première est basée sur l'indexation de k-mers (Figure 19) (Trapnell et Salzberg, 2009). Pour cette stratégie, le génome comme les lectures sont divisés en segments de longueur égale, appelés graines. Les différentes positions de ces graines dans le génome sont stockées dans une table de hachage. Cette structure de données permet une association clé-élément, ici la clé est la graine, l'élément représente les différentes positions de cette graine dans le génome. Ensuite, pour chaque lecture, les différentes graines sont recherchées dans la table donnant ainsi accès aux positions potentielles de chaque graine dans le génome. Afin de permettre les mésappariements, les graines sont comparées au génome de référence par paires. En effet, chaque lecture étant divisée en quatre graines, deux mésappariements laisseront au moins deux graines correspondre parfaitement au génome de référence. Ainsi, en comparant les paires de graines au génome de référence, il est possible de parcourir la liste des emplacements où la lecture entière pourrait être alignée en permettant au plus deux mésappariements. La seconde méthode utilise la transformée de Burrows-Wheeler (BWT) qui permet de stocker une représentation économe en mémoire du génome de référence (Figure 19) (Trapnell et Salzberg, 2009). Les lectures sont alignées caractère par caractère contre cette représentation. Chaque caractère successivement aligné permet de parcourir la liste des positions pour lesquelles la lecture pourrait être alignée. Si aucune position ne correspond, l'algorithme retourne au caractère précédent, réalise une substitution et reprend la recherche. La méthode BWT permet ainsi d'aborder le problème de l'alignement en résolvant tout d'abord le problème simple de l'alignement d'un seul caractère, et s'appuyant ensuite sur cette solution pour résoudre le problème plus complexe de l'alignement de deux caractères, et ainsi de suite, jusqu'à ce que la lecture entière ait été alignée. Les algorithmes d'alignement basés sur la méthode BWT sont généralement plus rapides que les algorithmes basés sur les tables de hachage, car il requière moins de mémoire.

Certains algorithmes, tels que Bowtie et Maq, utilisent les données de qualité des bases pour aligner chaque lecture. Ceci est particulièrement important pour ce type de données qui combinent des lectures de longueurs courtes et des taux d'erreurs élevés. Ces algorithmes peuvent ainsi diminuer la pénalité attribuée pour les mésappariements due à des erreurs de séquençage, augmentant ainsi la précision des alignements. L'alignement des lectures provenant d'éléments répétitifs, notamment de gènes paralogues ou de domaines répétés est un autre problème. Une lecture pourra alors être associée de manière équivalente à différentes localisations au sein du génome. Le comportement des programmes d'alignement pour résoudre ce problème peut varier, certains pouvant choisir d'exclure complètement de telles lectures, d'autres de déclarer plusieurs emplacements possibles pour une même lecture. Bowtie permet de paramétrer ce comportement.

Une des principales applications d'une expérience de RNA-Seq est la quantification du niveau d'expression des gènes. L'approche la plus simple pour mesurer le niveau d'expression d'un gène est de compter le nombre de lectures s'alignant à celui-ci. Toutefois le nombre de lecture s'alignant à un gène dépend également de la longueur de l'ARNm. Une solution simple est de normaliser le nombre de lectures par la longueur de l'ARNm et par le nombre total de lectures obtenues pour un échantillon. Les valeurs d'expression peuvent ainsi être définies comme le nombre de lectures alignées par kilobase par million de lectures alignées (RPKM) (Mortazavi et al, 2008) ; c'est la méthode la plus utilisée. Toutefois cette forme de normalisation globale peut être biaisée par la composition de la population d'ARN échantillonnée. Ainsi une proportion relativement faible de gènes fortement et différenciellement exprimés dans une condition expérimentale peut « consommer » une partie des données de séquences et conduire à un sous-échantillonnage des autres gènes. Cet artefact d'échantillonnage peut biaiser l'analyse des gènes différenciellement exprimés entre deux conditions expérimentales. Ce problème peut être atténué par l'exclusion de ces gènes lors de la normalisation des valeurs d'expression. De nouvelles méthodes de normalisation apparaissent donc pour traiter ce type de données (Bullard, 2010 ; Robinson et Oshlack, 2010). Le cas des lectures provenant de régions répétées peut également introduire des biais dans les mesures d'expression. Par exemple, l'exclusion de ces lectures conduit à une sous-estimation de l'expression des gènes paralogues (Mortazavi et al, 2008 ; Pepke et al, 2009). Dans ce cas, les niveaux d'expression peuvent être normalisés en ne prenant en compte que les oligomères uniques contenus dans chacun des exons (Sultan et al, 2008). Alternativement, les lectures répétées peuvent être assignées proportionnellement au nombre de lectures non répétées détectées au même loci (Mortazavi et al, 2008). Toutefois, ces méthodes ont tout de même un impact sur la quantification de l'expression, avec peu d'effet sur les paralogues les plus divergents et de plus grands effets sur les paralogues les plus proches. Enfin, ces méthodes de quantification reposent sur l'annotation du génome étudié et dépendent donc de la qualité de celle-ci.

Un autre problème d'alignement, spécifique des séquences transcriptomiques, provient du fait que certaines lectures s'étendent à travers les jonctions d'épissage à l'extrémité des exons et ne peuvent donc pas être alignées directement à la séquence génomique (Figure 20). Cette difficulté peut être levée en concaténant les exons adjacents connus, pour créer une base de données de jonctions. Les lectures qui ne s'alignent pas avec le génome sont alors alignées à ces jonctions (Mortazavi et al, 2008). En utilisant, toutes les combinaisons entre les différents exons, il est possible d'identifier des événements d'épissage alternatif (Zhang et al, 2010). Toutefois, cette méthode dépend de la qualité de l'annotation du génome analysé et se limite aux exons connus. Récemment, un certain nombre d'algorithmes ont été développés pour identifier ces jonctions d'épissage *ab initio*. Par exemple, TopHat utilise dans un premier temps Bowtie pour aligner les lectures au génome de référence. Les séquences génomiques couvertes par ces lectures et flanquées de sites d'épissage sont alors concaténées pour former une base de données de jonctions d'épissage potentielles. Enfin les lectures qui ne s'alignent pas au génome sont alignées à ces jonctions putatives pour identifier les sites d'épissage (Trapnell et al, 2009).

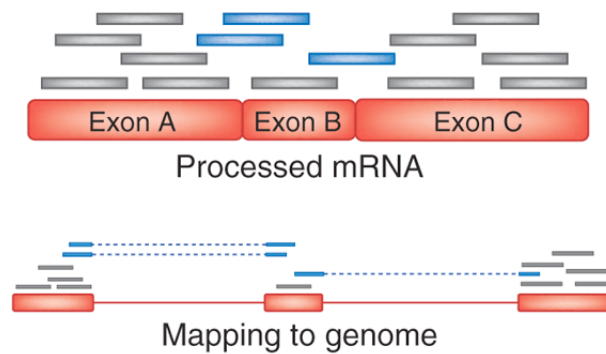


Figure 20 : Les lectures RNA-Seq sont obtenues à partir des ARNm. Les programmes d'alignement, tels que Bowtie et Maq, permettront d'aligner les lectures grises au génome de référence mais échoueront à aligner les lectures traversant les jonctions exoniques (lectures bleues). Des programmes, tels que TopHat, peuvent permettre d'aligner ces lectures au génome de référence. (Trapnell et Salzberg, 2009).

Actuellement, de nouveaux programmes, tels que G-Mo.R-Se (Denoeud et al, 2008), Cufflinks (Trapnell et al, 2010) et Scripture (Guttman et al, 2010), permettent la construction de modèles de gènes à partir d'alignements RNA-Seq sans connaissance a priori sur l'annotation du génome. Ces méthodes permettent l'identification de nouveaux exons, de formes alternatives d'épissage ainsi que de nouveaux loci. Tous ces programmes sont globalement basés sur une approche de graphes. Par exemple, G-Mo.R-Se construit tout d'abord des exons candidats directement à partir de la position des lectures alignées dans le génome (Figure 21). Toutes les jonctions possibles entre ces exons sont ensuite testées avec les lectures non alignées. Finalement une approche par graphes est utilisée pour chaîner les exons à travers les jonctions validées (Denoeud et al, 2008). Cufflinks et Scripture, plus récents, permettent de reconstruire de novo les différentes isoformes transcrites à partir des lectures alignées et d'estimer leurs abondances. Tout d'abord, les lectures sont alignées sur un génome de référence en utilisant un programme d'alignement permettant l'épissage, tel que TopHat. Deuxièmement, les lectures chevauchantes de chaque locus sont regroupées pour construire un graphe représentant toutes les isoformes possibles. L'étape finale consiste à parcourir le graphe pour résoudre les isoformes individuelles (Trapnell et al, 2010 ; Guttman et al, 2010). Toutefois, ces stratégies dépendent de la qualité du génome de référence et de la profondeur de couverture. Ainsi certains transcrits faiblement exprimés peuvent être incomplets.

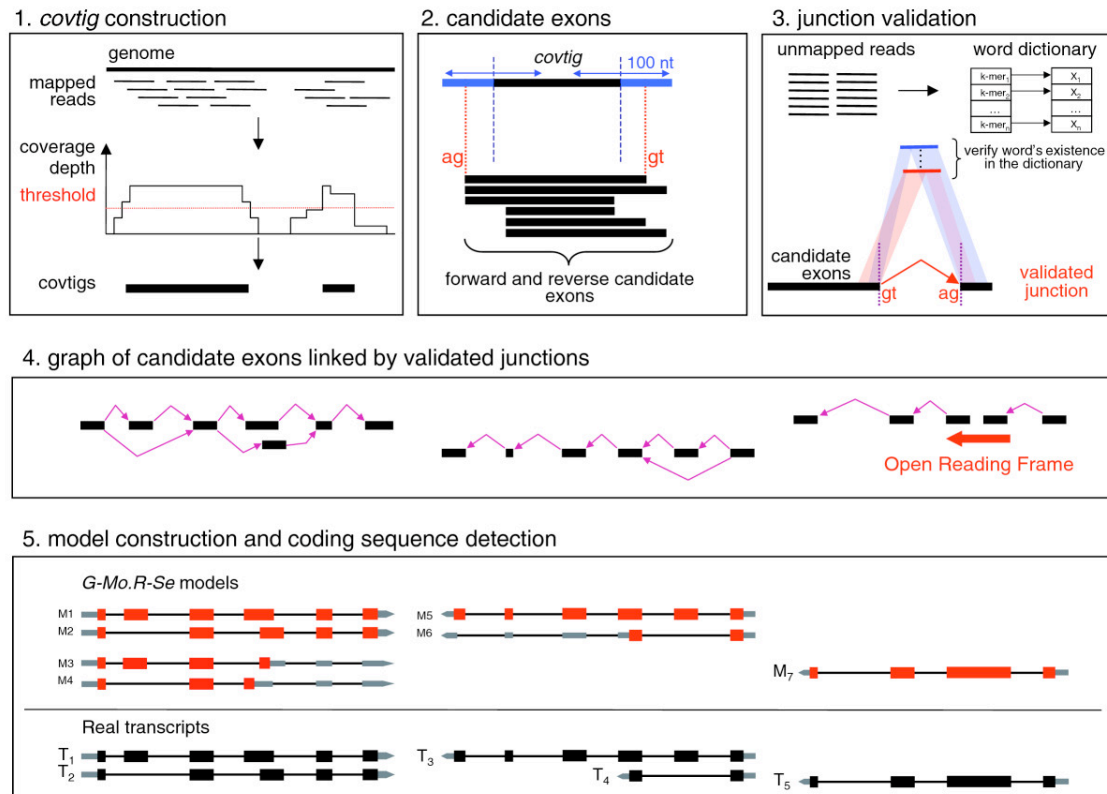


Figure 21 : Méthode utilisée par G-Mo.R-Se pour construire des modèles de gènes à partir des lectures RNA-Seq. Etape 1 : L'alignement des lectures au génome de référence permet de définir des covtigs (« coverage contigs »). Ils sont construits à partir d'un certain seuil de profondeur de couverture. Etape 2 : Des sites d'épissage donneurs et accepteurs sont recherchés dans un intervalle de 100 pb autour de chaque extrémité des covtigs. Les sites d'épissage détectés permettent de créer une liste d'exons candidats pour chaque covtig et d'orienter ces exons. Etape 3 : Les jonctions entre deux exons candidats sont validées en utilisant un dictionnaire de mots construits à partir des lectures non alignées. Etape 4 : Un graphe est ensuite créé. Dans ce graphe, les nœuds correspondent aux exons candidats et les connexions entre les nœuds représentent les jonctions validées. Etape 5 : Tous les chemins possibles sont extraits. Chacun de ces chemins représente alors un transcrit prédit. Des régions codantes sont ensuite recherchées au sein de chacun des transcrits. (Denoeud et al, 2008),

Chapitre II : Analyse globale du génome et du transcriptome de *T. melanosporum*

Chapitre II : Analyse globale du génome et du transcriptome de *T. melanosporum*

L'acquisition de séquences génomiques constitue une étape indispensable pour la découverte des gènes, l'étude de leur évolution et la construction de profils de transcription à l'échelle du génome entier. Les séquences génomiques de champignons mycorhiziens sont donc un atout majeur pour identifier les gènes dont le rôle pourrait être essentiel durant le processus de symbiose. L'exploration de la séquence du génome du basidiomycète ectomycorhizien *L. bicolor* a permis une avancée considérable dans la compréhension des bases moléculaires de l'interaction en identifiant plusieurs facteurs régulant le développement et le fonctionnement de la symbiose ectomycorhizienne. Cependant, les mécanismes moléculaires identifiés chez *L. bicolor* pourraient ne pas être partagés par toutes les espèces ectomycorhiziennes. Il est donc essentiel d'identifier les mécanismes moléculaires mis en jeu lors de la symbiose au sein de différents taxons. En 2006, notre laboratoire et le Genoscope ont initié le séquençage du génome de l'ascomycète ectomycorhizien *T. melanosporum*. La truffe noire du Périgord est un champignon endémique des sols calcaires du sud de l'Europe, qui a besoin de son partenaire végétal pour compléter son cycle de vie et produire des fructifications hypogées, que l'on appelle truffes.

Afin d'accéder à la complexité fonctionnelle du transcriptome de *T. melanosporum* et d'explorer les modifications de l'expression génique au cours des phases successives de son cycle de vie, notamment lors de la formation de la symbiose, j'ai appliqué le séquençage d'ARN à haut-débit Illumina (RNA-Seq) au transcriptome de *T. melanosporum* pour différents stades développementaux majeurs : le mycélium libre, les fructifications et les ectomycorhizes.

Le manuscrit intitulé « **Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis** » présente les résultats de l'analyse du génome de *T. melanosporum*. J'ai contribué à cette étude en réalisant l'analyse des données RNA-Seq. Ce travail a été utilisé pour d'autres analyses détaillées du transcriptome de *T. melanosporum* (Ceccaroli et al, 2011 ; Montanini et al, 2011 ; Rubini et al, 2011).

Le manuscrit intitulé « **Deep RNA sequencing improved the structural annotation of the *Tuber melanosporum* transcriptome** » correspond à l'utilisation de ces données RNA-Seq pour améliorer l'annotation structurale des modèles de gènes et pour l'identification de mécanismes transcriptionnels (épissage alternatif, μ ORF) chez *Tuber melanosporum*.

Le manuscrit intitulé « **Comprehensive transcriptome analysis of the Perigord black truffle unravels gene expression alterations during symbiosis and fruiting body formation** », actuellement en cours de préparation, décrit l'analyse détaillée de l'expression des gènes de *T. melanosporum* par RNA-Seq au cours de la formation de la symbiose et des fructifications, afin de fournir une image globale des profils d'expression et d'identifier les processus qui déclenche et conditionne la formation de la symbiose. Lors de cette analyse, nous avons également comparé les profils d'expression obtenus par RNA-Seq et puces à ADN.

Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis

Francis Martin, Annegret Kohler, Claude Murat, Raffaella Balestrini, Pedro M. Coutinho, Olivier Jaillon, Barbara Montanini, Emmanuelle Morin, Benjamin Noel, Riccardo Percudani, Bettina Porcel, Andrea Rubini, Antonella Amicucci, Joelle Amselem, Véronique Anthouard, Sergio Arcioni, François Artiguenave, Jean-Marc Aury, Paola Ballario, Angelo Bolchi, Andrea Brenna, Annick Brun, Marc Buée, Brandi Cantarel, Gérard Chevalier, Arnaud Couloux, Corinne Da Silva, France Denoeud, Sébastien Duplessis, Stefano Ghignone, Benoît Hilselberger, Mirco Iotti, Benoît Marçais, Antonietta Mello, Michele Miranda, Giovanni Pacioni, Hadi Quesneville, Claudia Riccioni, Roberta Ruotolo, Richard Splivallo, Vilberto Stocchi, Emilie Tisserant, Arturo Roberto Viscomi, Alessandra Zambonelli, Elisa Zampieri, Bernard Henrissat, Marc-Henri Lebrun, Francesco Paolocci, Paola Bonfante, Simone Ottonello et Patrick Wincker.

Article publié dans la revue Nature (Avril 2010)

LETTERS

Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis

Francis Martin¹, Annegret Kohler¹, Claude Murat¹, Raffaella Balestrini², Pedro M. Coutinho³, Olivier Jaillon⁴⁻⁶, Barbara Montanini⁷, Emmanuelle Morin¹, Benjamin Noel⁴⁻⁶, Riccardo Percudani⁷, Bettina Porcel⁴⁻⁶, Andrea Rubini⁸, Antonella Amicucci⁹, Joelle Amselem¹⁰, Véronique Anthouard⁴⁻⁶, Sergio Arcioni⁸, François Artiguenave⁴⁻⁶, Jean-Marc Aury⁴⁻⁶, Paola Ballario¹¹, Angelo Bolchi⁷, Andrea Brenna¹¹, Annick Brun¹, Marc Buée¹, Brandi Cantarel³, Gérard Chevalier¹², Arnaud Couloux⁴⁻⁶, Corinne Da Silva⁴⁻⁶, France Denoeud⁴⁻⁶, Sébastien Duplessis¹, Stefano Ghignone², Benoît Hilselberger^{1,10}, Mirco Iotti¹³, Benoît Marçais¹, Antonietta Mello², Michele Miranda¹⁴, Giovanni Pacioni¹⁵, Hadi Quesneville¹⁰, Claudia Riccioni⁸, Roberta Ruotolo⁷, Richard Splivallo¹⁶, Vilberto Stocchi⁹, Emilie Tisserant¹, Arturo Roberto Viscomi⁷, Alessandra Zambonelli¹³, Elisa Zampieri², Bernard Henrissat³, Marc-Henri Lebrun¹⁷, Francesco Paolocci⁸, Paola Bonfante², Simone Ottonello⁷ & Patrick Wincker⁴⁻⁶

The Périgord black truffle (*Tuber melanosporum* Vittad.) and the Piedmont white truffle dominate today's truffle market^{1,2}. The hypogaeous fruiting body of *T. melanosporum* is a gastronomic delicacy produced by an ectomycorrhizal symbiont³ endemic to calcareous soils in southern Europe. The worldwide demand for this truffle has fuelled intense efforts at cultivation. Identification of processes that condition and trigger fruit body and symbiosis formation, ultimately leading to efficient crop production, will be facilitated by a thorough analysis of truffle genomic traits. In the ectomycorrhizal *Laccaria bicolor*, the expansion of gene families may have acted as a 'symbiosis toolbox'⁴. This feature may however reflect evolution of this particular taxon and not a general trait shared by all ectomycorrhizal species⁵. To get a better understanding of the biology and evolution of the ectomycorrhizal symbiosis, we report here the sequence of the haploid genome of *T. melanosporum*, which at ~125 megabases is the largest and most complex fungal genome sequenced so far. This expansion results from a proliferation of transposable elements accounting for ~58% of the genome. In contrast, this genome only contains ~7,500 protein-coding genes with very rare multigene families. It lacks large sets of carbohydrate cleaving enzymes, but a few of them involved in degradation of plant cell walls are induced in symbiotic tissues. The latter feature and the upregulation of genes encoding for lipases and multicopper oxidases suggest that *T. melanosporum* degrades its host cell walls during colonization. Symbiosis induces an increased expression of carbohydrate and amino acid transporters in both *L. bicolor* and *T. melanosporum*, but the comparison of genomic traits in the two ectomycorrhizal fungi showed that genetic predispositions for symbiosis—the 'symbiosis toolbox'—evolved along different ways in ascomycetes and basidiomycetes.

The 125-megabase (Mb) genome of *T. melanosporum* is the largest sequenced fungal genome to date⁶, but no evidence for whole-genome duplication or large scale dispersed segmental duplications was observed (Supplementary Table 1 and Supplementary Information section 2). The approximately fourfold larger size of the truffle genome compared with other sequenced ascomycetes is accounted for by multi-copy transposable elements (TE) which constitute about 58% of the assembled genome (Fig. 1, Supplementary Figs 5, 6 and 8, Supplementary Information section 3). Estimated insertion times suggest a major wave of retrotransposition at <5 million years ago (Supplementary Fig. 7). TEs are not uniformly spread across the genome, but are clustered in gene-poor or gene-lacking regions (Fig. 1 and Supplementary Fig. 8). The expansion of regions between blocks of protein-coding genes results from an increased density of TEs. The proliferation of TEs within the truffle genome may result from its low effective population size⁷ during postglaciation migrations⁸ (Supplementary Information section 2.5).

The predicted proteome is in the lower range of sequenced filamentous fungi⁶, as only 7,496 protein-coding genes were identified (Supplementary Information section 4). They are mainly located in TE-poor regions and the gene density is heterogeneous when compared with that of other ascomycetes (Fig. 1, Supplementary Figs 8 and 9). Among the predicted proteins, only 3,970, 5,596 and 5,644 showed significant sequence similarity to proteins from *Saccharomyces cerevisiae*, *Neurospora crassa* and *Aspergillus niger*, respectively (Supplementary Fig. 10). This agrees with the predicted ancient separation (>450 Myr ago) of the Pezizomycetes from the other ancestral fungal lineages (Supplementary Fig. 4)⁹. Of the ~5,650 *T. melanosporum* genes that have an orthologue, very few show conservation of neighbouring orthologues (synteny) in at least one of the other species (Supplementary Fig. 11, Supplementary

¹INRA, UMR 1136, INRA-Nancy Université, Interactions Arbres/Microorganismes, 54280 Champenoux, France. ²Istituto per la Protezione delle Piante del CNR, sez. di Torino and Dipartimento di Biologia Vegetale, Università degli Studi di Torino, Viale Mattioli, 25, 10125 Torino, Italy. ³Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS-Universités Aix-Marseille I & II, 13288 Marseille, France. ⁴CEA, IG, Genoscope, 2 rue Gaston Crémieux CP5702, F-91057 Evry, France. ⁵CNRS, UMR 8030, 2 rue Gaston Crémieux, CP5706, F-91057 Evry, France. ⁶Université d'Evry, F-91057 Evry, France. ⁷Dipartimento di Biochimica e Biologia Molecolare, Università degli Studi di Parma, Viale G.P. Usberti 23/A, 43100 Parma, Italy. ⁸CNR-IGV Istituto di Genetica Vegetale, Unità Organizzativa di Supporto di Perugia, via Madonna Alta, 130, 06128 Perugia, Italy. ⁹Dipartimento di Scienze Biomolecolari, Università degli Studi di Urbino, Via Saffi 2 - 61029 Urbino (PU), Italy. ¹⁰INRA, Unité de Recherche Génomique Info, Route de Saint-Cyr, 78000 Versailles, France. ¹¹Dipartimento di Genetica e Biologia Molecolare & IBPM (CNR), Università La Sapienza, Roma, Piazzale, A. Moro 5, 00185 Roma, Italy. ¹²INRA, UMR Amélioration et Santé des Plantes, INRA-Université Blaise Pascal, INRA - Clermont-Theix, 63122 Saint-Genès-Champanelle, France. ¹³Dipartimento di Protezione e Valorizzazione Agroalimentare, Università degli Studi di Bologna, 40 126 Bologna, Italy. ¹⁴Dipartimento di Biologia di Base ed Applicata, ¹⁵Dipartimento di Scienze Ambientali, Università degli Studi dell'Aquila, Via Vetoio Coppito 1 - 67100 L'Aquila, Italy. ¹⁶University of Goettingen, Molecular Phytopathology and Mycotoxin Research, Grisebachstrasse 6, D-37077 Goettingen, Germany. ¹⁷INRA, UMR BIOGER-CPP, INRA-Grignon, av Lucien Brétignières - 78850 Thiverval Grignon, France.

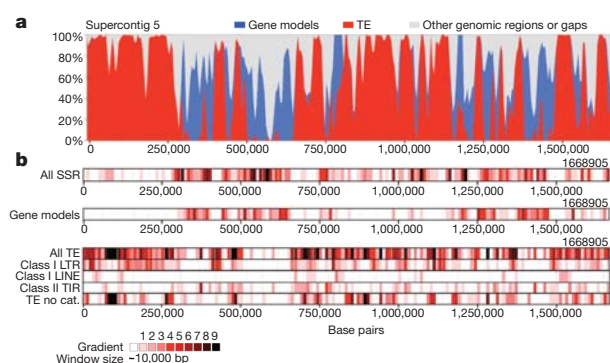


Figure 1 | Genomic landscape of *T. melanosporum*. **a**, The area chart quantifies the distribution of transposable elements (TE) and protein-coding genes (Gene models) along supercontig 5. The y axis represents the percentage of base pairs corresponding to TE (red), genes (blue), and other regions and gaps (white) in 10,000-bp sliding windows. **b**, Heat maps display the distribution of selected elements, including simple sequence repeats (SSR), gene models, all TE, long terminal repeat retrotransposons (class I LTR), long interspersed elements (class I LINE), terminal inverted repeats (class II TIR), and unknown TE classes (TE no cat.). Abundance of TE, protein-coding genes and other sequences is represented by a colour scale from 0 (white) to ≥ 9 occurrences (black) per 10 kbp window.

Information section 5.2). The *T. melanosporum* genome shows a structural organization strikingly different from other sequenced ascomycetes; the largest syntenic region (with *Coccidioides immitis*) only contains 99 genes with 39 orthologues (Supplementary Fig. 12). TE proliferation probably facilitated genome rearrangements. Some regions of meso-synteny were however detected, suggesting that *T. melanosporum* could be used for assessing the genome organization of ancestral ascomycete clades.

Expression of most predicted genes was detected in free-living mycelium, ectomycorrhizal (ECM) root tips and/or fruiting body

by custom oligoarrays, expressed-sequence-tag pyrosequencing and Illumina RNA-Seq (Supplementary Information sections 2.4 and 8, Supplementary Table 2, Supplementary Fig. 26). Only a low proportion of transcripts (7.6%) is differentially expressed (fold-ratio ≥ 4.0 , $P < 0.05$) in either ectomycorrhiza or fruiting body by comparison to free-living mycelium (Table 1, Supplementary Table 4). Only 61 transcripts unique to ectomycorrhiza, fruiting body or free-living mycelium were detected (Supplementary Table 5). A few transcripts coding for a H-type lectin, an arabinogalactan protein, a LysM-domain containing protein, major facilitator superfamily (MFS) transporters, laccase/tyrosinase, a lipase and polysaccharide-degrading enzymes are strikingly enriched ($> 1,000$ -fold) in symbiotic tissues (Table 1). They may play a role in adhesion to host cells, detoxication of plant defence metabolites, nutrient exchange, and colonization of root apoplast through the deconstruction of cell walls.

A process that is crucial to the success of ECM interactions is the mutualistic exchange of nutrients between the micro symbiont and its host plant. A comparison with other fungi revealed that the total number of predicted transporters is lower in *T. melanosporum* (381 members) compared with *L. bicolor* (491 members) as well as with saprotrophic and pathogenic ascomycetes (481–781 members) (Supplementary Table 26). However, 64 predicted membrane transporters showed an upregulated expression in truffle ectomycorrhizas, suggesting increased fluxes of carbohydrates, oligopeptides, amino acids and polyamines at the symbiotic interface (Supplementary Table 27). PFAM classification of fungal genes induced in symbiotic tissues of either *L. bicolor* or *T. melanosporum* ECM root tips revealed strikingly divergent fungal symbiotic proteomes (Supplementary Fig. 15). However, the PFAM categories corresponding to the MFS transporters (PFAM00083), aquaporin-related major intrinsic proteins (PFAM00230) and amino acid permeases (PFAM000324) were among the most strongly overrepresented in genes that were transcriptionally upregulated in both *L. bicolor* and *T. melanosporum* ectomycorrhizas.

Orthologous genes (that is, reciprocal best hits, BLASTP e-value $\leq 10^{-5}$) significantly induced in the symbiosis represent only 1.5% and

Table 1 | The most highly upregulated transcripts in *T. melanosporum*/Corylus avellana ECM root tips

SEQ_ID	ECM level	FB level	FLM level	ECM/FLM ratio	FB/FLM ratio	Definition	Size (AA)	Location	TMD
GSTUMT00012772001	13,262	731	1	13,262	731	H-type lectin	279	C	0
GSTUMT00012792001	11,423	250	1	11,423	250	Fasciclin-like arabinogalactan protein	414	S	0
GSTUMT00012437001	19,542	180	2	9,796	90	Lipase/esterase	346	S	0
GSTUMT00009894001	18,205	2	2	7,721	1	Cytochrome P450	396	M	0
GSTUMT00008973001	10,866	2	2	6,213	1	Endoglucanase GH5	342	C	0
GSTUMT00008992001	19,305	9	4	4,538	2	Laccase	586	S	0
GSTUMT00006890001	5,436	3,787	2	2,636	1,836	Sporulation-induced protein	481	C	0
GSTUMT00010076001	2,063	1	1	2,063	1	Tyrosinase	604	C	0
GSTUMT00003538001	3,588	2,818	2	2,050	1,610	FAD oxidoreductase	564	C	0
GSTUMT00012780001	6,542	44	4	1,745	12	LysM domain protein	87	S	0
GSTUMT00009016001	6,737	3,257	5	1,432	692	DUF1479-domain protein	379	C	0
GSTUMT00005760001	12,966	989	9	1,401	107	Major facilitator superfamily (MFS) permease*	142	S	2
GSTUMT00007927001	1,073	593	1	1,073	593	Hypothetical protein	306	C	0
GSTUMT00008954001	19,493	35	18	1,066	2	MFS permease	496	C	10
GSTUMT00000763001	902	2	1	902	2	Cytochrome P450	397	M	0
GSTUMT00002130001	1,534	1	2	833	1	β -Glucan synthesis-associated protein SKN1	575	M	1
GSTUMT00006579001	14,241	8,701	22	662	404	DUF1479-domain protein	426	M	0
GSTUMT00010279001	7,848	1,926	12	651	160	DUF2235-domain protein	403	C	0
GSTUMT00000499001	49,524	634	77	644	8	Phosphatidylserine decarboxylase	316	C	0
GSTUMT00012667001	1,151	7	2	575	3	Hypothetical protein	883	C	2
GSTUMT00009500001	2,529	2,846	6	403	453	Tuber-specific protein	86	M	0

Upregulation in ectomycorrhizal root tips and fruiting body is assessed by comparing transcript profiles to those from free-living mycelium. Ectomycorrhizal root tips were sampled from five-month-old common hazel (*Corylus avellana* L.) plantlets. Values are the means of seven, four and five biological replicates for free-living mycelium (FLM), ectomycorrhizal root tips (ECM) and fruiting body (FB), respectively. Based on the statistical analysis, a gene was considered significantly upregulated if it met two criteria: (1) t-test P value < 0.05 (ArrayStar, DNASTAR); (2) mycorrhiza or fruiting body versus free-living mycelium fold change ≥ 4 ; 571 genes (7.6% of the total gene repertoire) showed an upregulated expression. The highest signal intensity value observed on these arrays was 65,189 arbitrary units. Signals below the cut-off values (100 arbitrary units) were assigned a signal intensity value of 1. Bold font emphasises upregulation in symbiosis. AA, amino acid; C, predicted as cytosolic; S, predicted as secreted; M, predicted mitochondrial protein; TMD, no. of transmembrane domain.

* Truncated sequence. See Supplementary Information section 8 for details.

Table 2 | Orthologous symbiosis upregulated genes of *L. bicolor* and *T. melanosporum*

<i>Tuber</i> gene models	<i>Lb</i> gene models (protein ID)	Percentage identity	Putative function	<i>Tm</i> -Hazel ECM/FLM ratio	<i>Lb</i> -poplar ECM/FLM ratio	<i>Lb</i> -Douglas fir ECM/FLM ratio
GSTUMT00008973001	303005	61	Cellulase CBM1-GH5	6,213	4	6
GSTUMT00008992001	229839	31	Laccase	4,538	8	19
GSTUMT00002130001	239682	38	β -Glucan synthesis-associated protein	833	6	4
GSTUMT00005583001	314483	45	MFS permease	55	6	8
GSTUMT00005380001	306393	39	DUF1996	54	12	22
GSTUMT00005614001	249535	33	Sugar (and other) transporter	35	5	8
GSTUMT00010098001	300838	41	Amino acid permease	33	18	13
GSTUMT00012011001	247019	36	GH28 pectinase	29	5	6
GSTUMT00001268001	308894	35	Hypothetical protein	23	5	5
GSTUMT00009221001	187852	37	MFS permease	7	8	6
GSTUMT00003668001	256686	53	Sugar (and other) transporter	6	12,548	16,648
GSTUMT00000861001	186401	45	Sulphate transporter	6	6	5
GSTUMT00010105001	162486	43	Hypothetical protein	5	7	5
GSTUMT00007281001	152257	40	Amino oxidase	5	5	4

Table shows genes upregulated in ectomycorrhizas (ECM) compared to free-living mycelia (FLM). ECM of *T. melanosporum* were sampled on roots of five-month-old common hazel (*Corylus avellana* L.) plantlets, whereas ECM of *L. bicolor* were sampled on roots of nine-month-old Douglas fir (*Pseudotsuga menziesii*) seedlings or three-month old poplar (*P. trichocarpa*) plantlets⁴. Orthologous genes have been identified by BLASTP (reciprocal best hits, e-value $\leq 10^{-5}$). Transcript profiling was performed on FLM and ECM root tips. Values are the means of seven and four biological replicates, respectively. Based on the statistical analysis, a gene was considered significantly upregulated if it met two criteria: (1) t-test *P* value < 0.05 (ArrayStar, DNASTAR); (2) ECM versus FLM fold change ≥ 4 . Bold font emphasises upregulation in *Tuber* symbiosis. *Lb*, *L. bicolor*; *Tm*, *T. melanosporum*; MFS, major facilitator superfamily. See Supplementary Information section 8 for details.

4.1% of the ectomycorrhiza-upregulated genes in both *T. melanosporum* and *L. bicolor*, respectively. Most of these rare transcripts code for membrane transporters involved in sugar, amino acid or sulphate uptake (Table 2). This transcriptome trait appears to be a hallmark of the mycorrhizal symbiosis. The resulting increased nutrient flux probably explains the beneficial effect of the symbionts on the growth of their host seedlings (Supplementary Information section 1 and Supplementary Fig. 3). Other overrepresented PFAM categories displayed different patterns in the two symbionts. None of the effector-like small secreted MiSSP proteins specifically expressed in *L. bicolor* ectomycorrhizas⁴ were detected among ectomycorrhiza-regulated *T. melanosporum* transcripts.

One of the most striking characteristics of the *T. melanosporum* genome is the almost complete absence of highly similar gene pairs. Of the predicted 7,496 protein-coding genes, only seven pairs share $>90\%$ amino-acid identity in their coding sequence, whereas 30 pairs share $>80\%$ identity (Supplementary Information section 5.3, Fig. 2). This feature was also observed in the ascomycetous saprotroph *N. crassa*¹⁰. In striking contrast to the ECM *L. bicolor*⁴, multigene families in *T. melanosporum* are limited in number and comprise only 19% of the predicted proteome; most families have only two members (Supplementary Fig. 13). The rate of gene family gain is much lower than the rate of gene loss and among the 11,234 gene families found in ascomycetes, 5,695 appear to be missing in *T. melanosporum* (Supplementary Information section 5.4, Fig. 2). This compact gene coding space may reflect the genome organization of an ascomycete common ancestor, as the Pezizomycetes clade is the earliest diverging lineage within the Pezizomycotina (Supplementary Fig. 4). By comparison to other ascomycetes, gene families predicted to encode metabolite transporters (for example, amino acid and sugar permeases) and secondary metabolism enzymes (such as polyketide synthases and cytochrome P450s) are much smaller. Only 465 genes encoded by expanding gene families of *L. bicolor* are also found in the *T. melanosporum* genome (BLASTP, e-value $\leq 10^{-5}$) and 154 orthologues are shared between expanding gene families of both symbionts. None of them is differentially expressed in ectomycorrhizas. Differences in gene family expansion, in particular dynamic repertoires of genes encoding symbiosis-regulated effector-like proteins and sugar-cleaving enzymes (see below), are probably responsible for different symbiotic traits between *T. melanosporum* and *L. bicolor*, such as altered host specificity. The compact genome of *T. melanosporum* might be a product of selection for specialization; this is because genome expansion, as observed in *L. bicolor*, is probably driven by selection on the symbiont to exploit a diversity of encountered substrates provided by multiple potential hosts and by their diverse soils^{4,5}.

The volatiles released by truffles are attractive to rodents and truffle flies¹¹, which disperse their spores, but also to humans who consider this elusive mushroom a delicacy. *T. melanosporum* is the first sequenced fungus producing highly flavoured hypogeous fruiting bodies (Supplementary Information section 6.4, Fig. 3). Genomic signatures of the long-standing ($>2,000$ -year-old) reputation of the

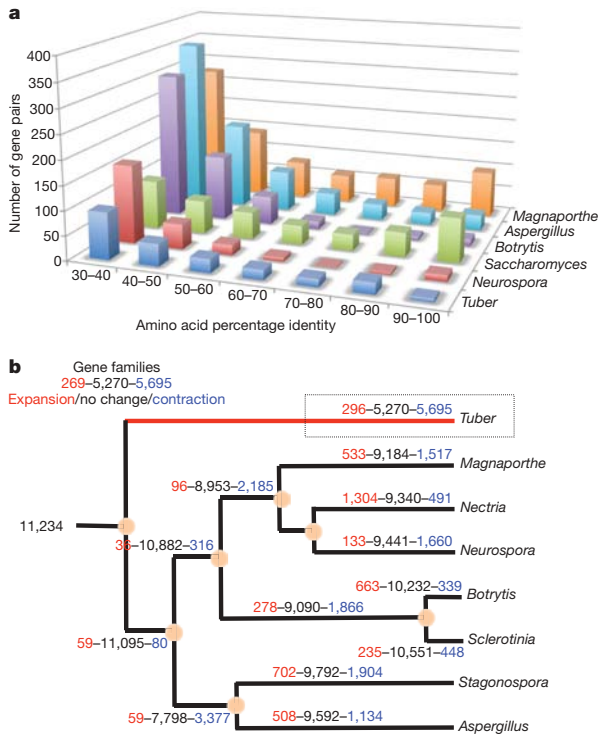


Figure 2 | Genome redundancy in the truffle genome. a, The percentage of amino-acid identity of the top-scoring self-matches for protein-coding genes in *T. melanosporum*, *Saccharomyces cerevisiae*, *Aspergillus nidulans*, *Neurospora crassa*, *Magnaporthe grisea*, and *Botrytis cinerea*. For each fungus, the protein-coding regions for each gene were compared with those of every other gene in the same genome using BLASTX. **b**, The figure represents the total number of protein families in each species or node. The numerals on branches show numbers of expanded (left, red), unchanged (middle, black) or contracted (right, blue) protein families along lineages by comparison to the putative pan-proteome.

black truffle as a gastronomic delicacy are its extremely low allergenic potential (Supplementary Fig. 18), coupled with the lack of key mycotoxin biosynthetic enzymes (Supplementary Information section 6.2, Supplementary Table 14), and the preferential overexpression of various flavour-related enzymes in the fruiting body (Supplementary Figs 19–21). Among the latter are specific subsets of sulphur assimilation and S-amino acid interconversion enzymes. These include cystathionine lyases known to promote the side-formation of methyl sulphide volatiles abundant in truffles¹² as well as various enzymes involved in amino acid degradation through the Ehrlich pathway which are giving rise to known truffle volatiles and flavours, for example, 2-methyl-1-butanal (Fig. 3, Supplementary Information section 6.4, Supplementary Figs 20 and 21). Also notable, given the subterranean habitat of this fungus, is the presence of various putative light-sensing components (Supplementary Information section 6.6), which might be involved in light avoidance mechanisms and/or in the control of seasonal developmental variations, especially those related to fruiting body formation and sexual reproduction.

The analysis of genes implicated in the mating process, including pheromone response, meiosis and fruiting body development showed that most sex-related components identified in other ascomycetes are also present in *T. melanosporum* (Supplementary Table 11). Sexual reproduction in ascomycete filamentous fungi is partly controlled by two different mating-type (*MAT*) genes that establish sexual compatibility¹³: one *MAT* gene codes for a protein with an α -box domain, whereas the other encodes a high mobility group (HMG) DNA binding protein (Supplementary Information section 6.5). It was widely believed that *T. melanosporum* was a homothallic or even an exclusively selfing species¹⁴. The sequenced Mel28 strain contains the HMG locus, and the opposite linked *MAT α* locus was identified in another natural isolate (Supplementary Fig. 22), confirming recent hints that *T. melanosporum* is heterothallic and thus an obligate outcrossing species¹⁵. This result has major implications for

truffle cultivation, which will be improved by the use of host plants harbouring truffle strains of opposite mating types. In most ascomycetes, the genomic regions flanking the *MAT* locus show an extended conservation¹³, but there is no synteny of the *MAT* loci between *T. melanosporum* and other sequenced fungi (Supplementary Fig. 23).

To determine whether *T. melanosporum* sugar-cleaving capabilities resemble those of other fungi, we have undertaken a comparison of the glycoside hydrolase (GH) and polysaccharide lyase (PL) repertoires¹⁶ of 18 completely sequenced fungi (Fig. 4). As expected for a symbiotic fungus living in the root apoplast, *T. melanosporum* has a relatively small number of GH-encoding genes (91 members; Supplementary Tables 23 and 24); much fewer than phytopathogens (for example, *Magnaporthe grisea*, *Fusarium graminearum*) and saprotrophs (for example, *N. crassa*, *Podospora anserina*). The *T. melanosporum* GH repertoire bears some similarity with that of *L. bicolor*⁴, especially a reduced spectrum of enzymes targeting the plant cell wall compared to saprobes, culminating in both fungi with the absence of cellulases from families GH6 and GH7. There are however significant differences in the spectrum of enzymes present in these two symbiotic fungi. For instance, *T. melanosporum* has hemicellulases from families GH10 and GH43, whereas *L. bicolor* has none. Similarly, *T. melanosporum* has a family GH45 cellulase that is absent from the *L. bicolor* genome. Other differences include different strategies to cleave pectin: whereas *L. bicolor* utilizes six hydrolytic GH28 pectinases, *T. melanosporum* has only two, but these are complemented by three pectin lyases and a pectin methyltransferase that are missing in *L. bicolor*. Both fungi have a set of proteins, few in number, bearing cellulose-binding domains, but differences appear here too: the single cellulose-binding CBM1 motif of *L. bicolor* is appended to a GH5 endoglucanase, whereas *T. melanosporum* has two CBM1 motifs attached to a GH61 enzyme and to a protein of unknown function. GH61 enzymes have been reported to display weak cellulolytic activity¹⁷.

An extended comparison with other sequenced fungi (Fig. 4) shows that *T. melanosporum* clusters neither with *L. bicolor* nor with saprotrophic ascomycetes, most probably because of its limited overall number of GHs and PLs that make it closer to yeasts and fungi that do not interact with plants, but rather with animals (*Cryptococcus neoformans*, *Malassezia globosa*). Differences between the enzyme repertoires of *T. melanosporum* and *L. bicolor* suggest differences in the mode of interaction of the two symbionts with their respective host plants. A striking difference is the presence of an invertase gene in *T. melanosporum*, whereas *L. bicolor* has none and is therefore completely dependent on its host for its provision of glucose⁵. In contrast, *T. melanosporum* could access and hydrolyse the plant-derived sucrose. This would suggest that although both fungi develop symbiotic relationships with plants, *T. melanosporum* is probably less dependent than *L. bicolor*. The overall pattern of induction of genes coding for enzymes acting on polysaccharides is similar in both *L. bicolor* and *T. melanosporum* symbiotic transcriptomes, although a larger number of carbohydrate-cleaving enzyme transcripts are up-regulated for some families—for example, GH16 (β -1,6-glucanases), GH18 (chitinases) and GT20 (α , α -trehalose-phosphatase) in *L. bicolor* (Supplementary Table 25 and Supplementary Fig. 24). Intriguingly, a GH5 cellulase and a GH28 pectinase are among the rare transcripts that are highly upregulated in both *L. bicolor* and *T. melanosporum* ectomycorrhizas, suggesting that they play a key role in the symbiosis. On the other hand, the β -glucan synthesis-associated protein present in both ectomycorrhizas is involved in fungal cell wall remodelling¹⁶ and may play a role in the alteration of cell wall surface during infection to conceal the hyphae from the host.

The ability to establish ECM symbioses is a widespread characteristic of various ascomycetes and basidiomycetes³. The truffle genome reveals features of an ancestral fungal lineage that diverged from other lineages >450 Myr ago⁹. Despite their similar symbiotic structures and similar beneficial effects on plant growth, the ascomycete *T. melanosporum* and the basidiomycete *L. bicolor* encode strikingly

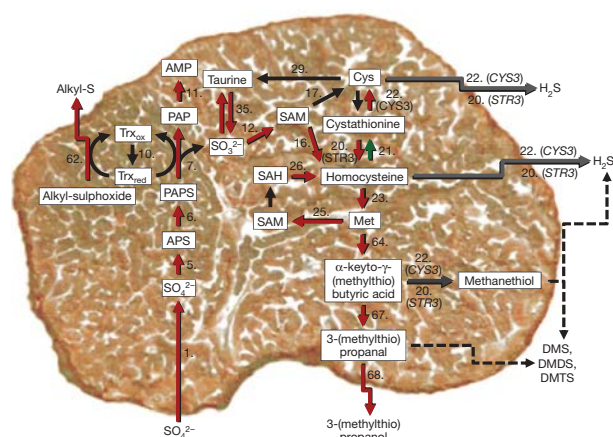
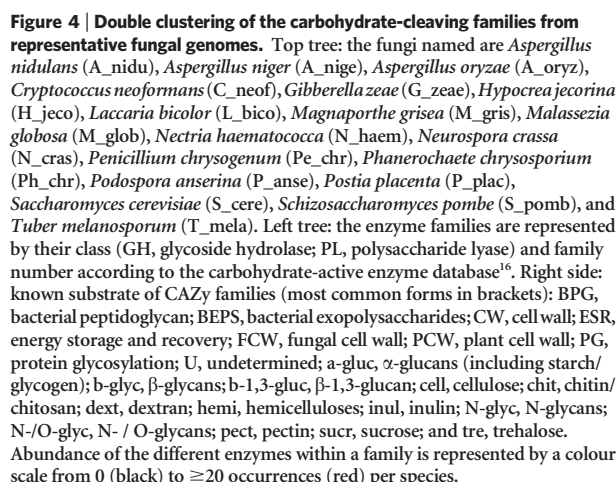


Figure 3 | Outline of sulphur metabolism in *T. melanosporum* fruiting body. Numbers identify enzymes and gene models as specified in Supplementary Table 15. Reactions (arrows) catalysed by enzymes whose mRNAs are upregulated in fruiting bodies are shown in red; mRNAs downregulated by at least twofold in fruiting bodies, or whose expression levels differ by less than twofold compared to free-living mycelia and ectomycorrhizas are represented by green and black arrows, respectively. Off-pathway cystathionine- γ -lyase (no. 22)- and cystathionine β -lyase (no. 20)-supported reactions, and spontaneous (non-enzymatic) breakdown reactions are indicated by grey and dashed arrows, respectively. APS, adenosine phosphosulphate; PAPS, phosphoadenosine phosphosulphate; PAP, phosphoadenosine phosphate; Trx, thioredoxin; SAM, S-adenosylmethionine; SAH, S-adenosylhomocysteine; DMS, dimethylsulphide; DMDS, dimethyl-disulphide; DMTS, dimethyl-trisulphide.



1. Mello, A., Murat, C. & Bonfante, P. Truffles: much more than a prized and local fungal delicacy. *FEMS Microbiol. Lett.* **260**, 1-8 (2006).
2. Hall, I. R., Brown, G. T. & Zambonelli, A. *Taming the Truffle. The History, Lore, and Science of the Ultimate Mushroom* (Timber Press, Portland, 2007).
3. Smith, S. E. & Read, D. J. *Mycorrhizal Symbiosis* 3rd edn (Academic, 2008).
4. Martin, F. *et al.* The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **452**, 88-92 (2008).
5. Martin, F. & Slesosse, M. *et al.* The *Laccaria* genome: a symbiont blueprint decoded. *New Phytol.* **180**, 296-310 (2008).
6. 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).
7. Lynch, M. & Conery, J. S. The origins of genomic complexity. *Science* **302**, 1401-1404 (2003).
8. Murat, C. *et al.* Polymorphism at the ribosomal DNA ITS and its relation to postglacial re-colonization routes of the Perigord truffle *Tuber melanosporum*. *New Phytol.* **164**, 401-411 (2004).
9. Taylor, J. W. & Berbee, M. L. Dating divergences in the fungal tree of life: review and new analyses. *Mycologia* **98**, 838-849 (2006).
10. Galagan, J. E. & Selker, E. U. RIP: the evolutionary cost of genome defense. *Trends Genet.* **20**, 417-423 (2004).
11. Maser, C., Claridge, A. W. & Trappe, J. M. *Trees, Truffles and Beasts* (Rutgers Univ. Press, 2008).
12. Liu, M., Nauta, A., Francke, C. & Siezen, R. J. Comparative genomics of enzymes in flavor-forming pathways from amino acids in lactic acid bacteria. *Appl. Environ. Microbiol.* **74**, 4590-4600 (2008).
13. Fraser, J. A. & Heitman, J. Evolution of fungal sex chromosomes. *Mol. Microbiol.* **51**, 299-306 (2004).
14. Bertault, G., Raymond, M., Berthomieu, A., Callot, G. & Fernandez, D. Trifling variation in truffles. *Nature* **394**, 734 (1998).
15. Riccioni, C. *et al.* *Tuber melanosporum* outcrosses: analysis of the genetic diversity within and among its natural populations under this new scenario. *New Phytol.* **180**, 466-478 (2008).

16. Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37**, D233–D238 (2009).
17. Karkehabadi, S. *et al.* The first structure of a glycoside hydrolase family 61 member, Cel61B from *Hypocrea jecorina*, at 1.6 Å resolution. *J. Mol. Biol.* **383**, 144–154 (2008).
18. Hibbett, D. S. & Matheny, P. B. Relative ages of ectomycorrhizal mushrooms and their plant hosts. *BMC Biol.* **7**, 13 (2009).
19. Soanes, D. M. *et al.* Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. *PLoS One* **3**, e2300 (2008).

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

Acknowledgements We thank the late L. Rioussset and C. Dupré for providing the Mel28 isolate, and acknowledge F. Le Tacon and J. Weissenbach for continuous support. The genome sequencing of *T. melanosporum* was funded by the Génomscope, Institut de Génétique, CEA, and Agence Nationale de la Recherche (ANR). Genome annotation and transcriptome analysis were supported by INRA, the European FP6 Network of Excellence EVOLTREE, Région Lorraine, the ANR FungEffector project, Fondazione Cariparma, Compagnia di San Paolo-Torino, the Italian Ministry of Education, University and Research (MIUR), Regione Umbria and Istituto Pasteur Fondazione Cenci Bolognetti. We thank D. Hibbett and J. Heitman for comments on an early draft of the manuscript, and J. Plett for a critical reading of the paper.

Author Contributions B.H., M.-H.L., F.P., P. Bonfante, S.O. and P.W. contributed equally to this work as senior authors; A.K., C.M., R.B., P.M.C., O.J., B.M., E.M., B.N., R.P., B.P. and A.R. contributed equally to this work as second authors. F.M. initiated the project and coordinated the genome annotation, data analysis and manuscript preparation; P.W. coordinated the sequencing and automated annotation at Genoscope. F.M. and S.O. wrote the manuscript with input from P. Bonfante. R.B., P.M.C., B.H., O.J., A.K., M.H.L., B.M., E.M., C.M., B.N., F.P., R.P., A.R. and P.W. also made substantial contributions (listed in alphabetical order). All other authors are members of the genome sequencing consortium and contributed annotation, analyses or data throughout the project, and are listed in alphabetical order. We also thank A. Bonfigli, M. Buffalini, S. Colafarina, T. Flutre, S. Kamal, P. Ceccaroli, C. Roux, R. Saltarelli, S. von Pall di Tolna and O. Zarivi for their assistance in annotation.

Author Information Genome assemblies together with predicted gene models and annotations have been deposited at DNA Data Bank of Japan/European Molecular Biology Laboratory/GenBank under the project accession numbers CABJ01000001–CABJ01004455 (whole genome shotgun sequencing data) and FN429986–FN430383 (scaffolds and annotations). The complete expression dataset is available as series (accession number GSE17529) at the Gene Expression Omnibus at NCBI. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.M. (fmartin@nancy.inra.fr).

Deep RNA sequencing improved the structural annotation of the *Tuber melanosporum* transcriptome

E. Tisserant, C. Da Silva, A. Kohler, E. Morin, P. Wincker et F. Martin.

Article publié dans la revue New Phytologist (Novembre 2010)

Methods

Deep RNA sequencing improved the structural annotation of the *Tuber melanosporum* transcriptome

E. Tisserant¹, C. Da Silva², A. Kohler¹, E. Morin¹, P. Wincker² and F. Martin¹

¹INRA, UMR INRA/Nancy Université 'Interactions Arbres/Micro-Organismes', INRA-Nancy, 54280 Champenoux, France; ²CEA, IG, Genoscope, 2 rue Gaston Crémieux CP5702, F-91057 Evry, France

Summary

Author for correspondence:
Francis Martin
Tel: +33 383 39 40 80
Email: fmartin@nancy.inra.fr

Received: 4 November 2010
Accepted: 20 November 2010

New Phytologist (2011) 189: 883–891
doi: 10.1111/j.1469-8137.2010.03597.x

Key words: ectomycorrhiza, fruiting body, high-throughput Illumina RNA-sequencing (RNA-Seq), transcriptome, *Tuber*.

- The functional complexity of the *Tuber melanosporum* transcriptome has not yet been fully elucidated. Here, we applied high-throughput Illumina RNA-sequencing (RNA-Seq) to the transcriptome of *T. melanosporum* at different major developmental stages, that is free-living mycelium, fruiting body and ectomycorrhiza.
- Sequencing of cDNA libraries generated a total of c. 24 million sequence reads representing > 882 Mb of sequence data. To construct a coverage signal profile across the genome, all reads were then aligned to the reference genome assembly of *T. melanosporum* Mel28.
- We were able to identify a substantial number of novel transcripts, antisense transcripts, new exons, untranslated regions (UTRs), alternative upstream initiation codons and upstream open reading frames.
- This RNA-Seq analysis allowed us to improve the genome annotation. It also provided us with a genome-wide view of the transcriptional and post-transcriptional mechanisms generating an increased number of transcript isoforms during major developmental transitions in *T. melanosporum*.

Introduction

The Périgord black truffle (*Tuber melanosporum*) is endemic to calcareous soils in southern Europe and is found in mutualistic symbiosis with roots of deciduous trees, mostly oak (*Quercus* spp.) and hazelnut (*Corylus avellana*) trees. This ectomycorrhizal ascomycete requires a host tree to complete its life cycle and to produce hypogeous fruiting bodies (FBs) (ascosporangia), so-called truffles (Mello *et al.*, 2006). This ectomycorrhizal symbiosis is of major ecological importance in woodland ecosystems (Mello *et al.*, 2006). However, it has been suggested that *T. melanosporum* may act as a biotroph with an antagonistic nature in the later stages of the interaction (G. Chevalier, INRA, Clermont-Ferrand, France, pers. comm.). The identification of the developmental processes that condition and trigger symbiosis and FB formation, ultimately leading to sustainable truffle crop management, will be facilitated by a thorough analysis of the genomic traits, physiology and ecology of the mycobiont (Ceccaroli

et al., 2011; Martin *et al.*, 2010; Montanini *et al.*, 2011; Murat *et al.*, 2011; Rubini *et al.*, 2011; Splivallo *et al.*, 2011). The host plant, environment and community of soil organisms with which *T. melanosporum* interacts affect the function of its genome (Martin *et al.*, 2010), but our current knowledge on the gene, transcript and protein networks involved in this complex web of interactions is scarce.

The available draft genome sequence provides access to the c. 7500-gene repertoire of *T. melanosporum*, allowing the genome-wide changes in gene expression to be monitored during the successive stages of the truffle life cycle (Martin *et al.*, 2010). The transcriptome of *T. melanosporum* has been interrogated using customized genome-wide exon oligoarrays (Martin *et al.*, 2010). Both ectomycorrhiza and FB development induce alterations in gene expression; 7.6% of the total gene repertoire shows an upregulated expression in these life cycle stages. These changes in gene expression affect major cellular processes, such as mating (Rubini *et al.*,

2011) and transcription (Montanini *et al.*, 2011), and various metabolic pathways (Ceccaroli *et al.*, 2011).

The interrogation of the *T. melanosporum* transcriptome in the latter study was greatly facilitated by the use of customized exon oligoarrays, which quantify transcript abundance by hybridization. However, all expression arrays are based on *a priori* assumptions regarding the expressed parts of the genome. This limitation, together with the restricted dynamic range of signal intensity measurements (1–1000), may hamper the potential of microarrays for unbiased studies of transcription and the discovery of novel transcripts. Recently, the development of next-generation Illumina sequencing of steady-state RNA populations (RNA-Seq) has provided unprecedented details about the RNA landscape within a cell (Mortazavi *et al.*, 2008; Nagalakshmi *et al.*, 2008; Sultan *et al.*, 2008; Wilhelm *et al.*, 2008). This cost-effective sequencing-by-synthesis produces millions of short cDNA reads which are mapped to a reference genome to obtain a quantitative measure of the transcript expression levels at single-nucleotide resolution, as well as transcript structure information. RNA-Seq provides access to rare transcripts (thanks to a wide dynamic range from 1 to > 1 000 000), more accurate quantification of abundant transcripts (above the signal saturation point of arrays), novel gene structures, alternative splicing (AS) and allele-specific expression information (Marioni *et al.*, 2008; Mortazavi *et al.*, 2008; Nagalakshmi *et al.*, 2008; Sultan *et al.*, 2008; Wilhelm *et al.*, 2008; Larsen *et al.*, 2010). Recently, this approach has been used to study the transcriptome of the ectomycorrhizal basidiomycete *Laccaria bicolor* (Larsen *et al.*, 2010) and the mutualistic endophyte *Epichloë festucae* (Eaton *et al.*, 2010). In addition to transcript profiling, RNA-Seq can support the annotation of draft genome sequences through read mapping and gene model reconstruction (Denoeud *et al.*, 2008; Sultan *et al.*, 2008; Wilhelm *et al.*, 2008; Larsen *et al.*, 2010; Trapnell *et al.*, 2010). Although several gene finder programs have been used in *T. melanosporum* genome annotation, highly accurate prediction of the coding regions of many genes is still challenging and can be alleviated by RNA-Seq.

In this study, we report the results from an RNA-Seq analysis of the *T. melanosporum* transcriptome in saprobic free-living mycelium (FLM), ectomycorrhizal root tips (ECM) and truffle using the Illumina platform. We were able to detect previously uncharacterized transcripts and exons, untranslated regions (UTRs) that extended *in-silico* gene models, AS, alternative upstream initiation codons (μ ATG) and upstream open reading frames (μ ORFs). We were also able to assess the impact of developmental stages on the generation of alternative mRNA isoforms. This genomic resource provides a global view of the transcriptome complexity of an ectomycorrhizal symbiont that has been used for further detailed analyses of the *T. melanosporum*

transcriptome (Ceccaroli *et al.*, 2011; Montanini *et al.*, 2011; Rubini *et al.*, 2011).

Materials and Methods

Biological materials

FLM of *T. melanosporum* (Vittad.) Mel28 was grown on 1% malt agar (Cristomalt-D, Difal, Villefranche-sur-Saône, France) for 5 wk before harvesting (seven biological replicates). The Mel28 strain used in this study is the same as the reference genome strain. ECM were sampled (four biological replicates of 100 ECM each) from 5-month-old Common Hazel (*Corylus avellana* L.) plantlets inoculated by a mycelium slurry produced from an FB of *T. melanosporum* harvested in Meuse (France) by Dr Gérard Chevalier. Inoculated plants were grown in the glasshouse facilities of the AGRI-TRUFFE Company (Saint-Maixant, France). Five different FBs of *T. melanosporum* were collected at different maturation stages below hazel trees or oak trees at different locations (Auvergne, Meuse and Dordogne, France, and Piceno, Italy).

Construction of cDNA library for RNA-Seq and Illumina sequencing

FLM, FB and ECM samples (100 mg fresh weight) were snap frozen in liquid nitrogen and RNA extraction was carried out using the RNeasy Plant Mini Kit including a DNase treatment (Qiagen, Cat. No. 74904). RNA quality and integrity were checked before cDNA synthesis using the Bio-Rad Experion analyzer and Experion RNA StdSens analysis kit (Bio-Rad, Hercules, CA, USA). Total RNA preparations (four biological replicates for ECM, five for FB and seven for FLM) were primed by Oligo-dT primer and amplified using the SMART PCR cDNA Synthesis Kit (Takara Bio Europe/Clontech, Saint-Germain-en-Laye, France), according to the manufacturer's instructions, and used either for hybridizations to NimbleGen oligoarrays (for details, see Martin *et al.*, 2010) or pooled for Illumina sequencing.

Cluster generation was performed after applying each DNA sample to the individual lanes of the Illumina Genome Analyser II flowcell according to the manufacturer's recommendations (Illumina, San Diego, CA, USA). Three lanes of massively parallel sequencing by synthesis were performed and processed into SCARF files containing millions of 36-bp sequencing reads and raw quality scores using the Illumina Genome Analyser II and the Illumina Pipeline Software according to the manufacturer's instructions. The raw Illumina sequences were submitted to the National Center for Biotechnology Information under accession number ERP000346.

Mapping reads to the reference genome and annotated genes

The alignments of Illumina reads to the genome were used to map reads to genomic features. The genomic location of each alignment may correspond to features such as genes, tRNA, rRNA or transposable elements. The *T. melanosporum* Mel28 genomic sequence and gene models were taken from the Genoscope Tuber Genome Browser (<http://www.genoscope.cns.fr/tuber>) using publicly available data files: http://www.Tuber_melanosporum_v1.0_unmasked.fa.gz and http://www.Tuber_melanosporum_v1.0_annot.gff.gz (Martin *et al.*, 2010). Genome assemblies together with predicted gene models and annotations for *T. melanosporum* are also available at the DNA Data Bank of Japan/European Molecular Biology Laboratory/GenBank under project accession numbers CABJ01000001–CABJ01004455 (whole genome shotgun sequencing data) and FN429986–FN430383 (scaffolds and annotations). In version 1.0 of the annotation, the *T. melanosporum* gene repertoire contains 7496 best gene models. *Populus trichocarpa* gene models were downloaded from the Joint Genome Institute Phytozome 5.0 website (<http://www.phytozome.net/poplar.php>).

After removing Illumina reads containing sequencing adapters, the remaining 24 509 204 read FASTA files were aligned to the *T. melanosporum* Mel28 genome sequence (Table 1) and gene models using BOWTIE software (Langmead *et al.*, 2009). This software is capable of producing alignments of the input cDNA fragments that were aligned to the genome. The default mapping criteria used by BOWTIE were as follows: alignments may have no more than two mismatches in the first 28 bases on the high-quality (left) end of the read, called the 'seed'. The sum of the quality values at all mismatched positions (not just in the seed) may not exceed 70. As reads may align to the genome one or more times, the specifying parameters '*-a-m-1-best-strata*' were used to instruct BOWTIE to choose the best alignment and to refrain from reporting any alignments for reads having more than one alignment, that is, corresponding to repeated loci (572 211 reads). A

customized Perl script (available on request) was used to assign each unique read to a specific genome location (nucleotide coordinates on scaffolds) for determination of the genomic distribution. For Supporting Information Fig. S1, expression scores for every genomic base pair position were assigned on the basis of how many sequence reads covered each 1000-bp window. The log₂ of the score was then plotted using CIRCOS (<http://mkweb.bcgsc.ca/circos/>).

Assessment of functional content

To obtain an accurate measure of gene expression, reads that were found in unique positions (16 588 455) in the genome were further mapped to coding sequences (CDSs) of *in silico*-predicted protein-coding genes and counted to generate gene coverage values. We used only those reads (*c.* 11 867 391) that mapped at a unique position and only those positions in which a read could potentially map uniquely within the framework of CDSs (Sultan *et al.*, 2008). For each CDS, the number of all possible 36-bp oligomers was determined using a sliding window of 1 bp.

To allow the transcript level to be compared both within and between samples, the number of read hits per gene was normalized according to the theoretical length and sample size, and reported as the number of reads per exon kilobase per million mapped sequence reads (RPKM) (Mortazavi *et al.*, 2008).

AS analysis by RNA-Seq data

The current gene repertoire contains 21 525 annotated junction sites. To identify all potential splice sites, we searched for the putative donor site ('GT' on the forward strand or 'AC' on the reverse strand) and the acceptor site ('AG' on the forward strand or 'CT' on the reverse strand) inside the intronic regions of gene models, and set that a novel splice site was required to be supported by at least two uniquely mapped reads. We then enumerated all possible pairs of donor sites and acceptor sites in annotated introns and detected 244 260 potential splice junctions. To

Table 1 Distribution of Illumina reads mapped to genomic features for the different cDNA libraries

Sample	FLM (%)	FB (%)	ECM (%)	Total (%)
Sequenced reads	7 820 255	8 653 984	8 034 965	24 509 204
Reads with unique matches on the genome	6 241 339 (79.81)	7 122 621 (82.31)	3 224 495 (40.13)	16 588 455 (67.68)
Reads with multiple matches on the genome ¹	260 143 (3.33)	227 776 (2.63)	84 292 (1.05)	572 211 (2.33)
Reads with no match on the genome ²	1 318 773 (16.86)	1 303 587 (15.06)	4 726 178 (58.82)	7 348 538 (29.99)
Reads with matches on CDSs	4 277 388 (54.6)	5 293 676 (61.1)	2 296 327 (28.5)	11 867 391 (48.4)

CDS, coding sequence; ECM, ectomycorrhizal root tips; FB, fruiting body; FLM, free-living mycelium.

¹Multiple matches correspond to expressed retrotransposons and nearly identical paralogs.

²Reads corresponding to plant transcripts (in ECM), mitochondrial transcripts, junction sites, nonassembled genes in assembly gaps or of poor quality. Owing to the percentage of reads with no match on the *Tuber melanosporum* genome for FLM and FB, the plant reads were estimated to be *c.* 1 285 000.

determine the number of reads supporting each splice junction, all reads that could not match to the genome were aligned against the junction sequence with a tolerance of 2 bp mismatches. A junction site was required to be supported by at least two unambiguous mapped reads within the splice junction region and also with a minimum match of five bases on each side of the junction. In total, 15 223 junction sites were identified. To identify AS events in *T. melanosporum*, the junction sequences identified were used to analyze seven types of AS, including skipped exons (SE), retained introns (RI), alternative 5'-splice sites (A5SS), alternative 3'-splice sites (A3SS), mutually exclusive exons (MXE), alternative first exons (AFE) and alternative last exons (ALE) (Wang *et al.*, 2008).

UTR analysis by RNA-Seq data

To define the UTRs of annotated protein-coding genes, we searched for a break in the transcribed region around *in silico* gene models, denoted by positions with read counts of zero, starting from either end of every gene. Genes whose ends overlapped with other genes or transposons were excluded from the analysis. If no break was found, the UTR was discarded. The rare 5'- or 3'-UTRs with length > 1000 nucleotides were also discarded as the presence of transposable elements was suspected. The μ ORFs were screened in 5'-UTR. Any μ ORF of length > 150 bp, or with the distance between the μ ORF and the gene start codon > 500 bp, was excluded from the result.

Detection of novel and antisense transcripts

Novel transcriptional regions were determined in the intergenic regions using G-Mo.R-Se (Denoeud *et al.*, 2008), a method aimed at using RNA-Seq short reads to build *de novo* gene models. First, candidate exons were built directly from the positions of the reads mapped on the genome, and all the possible splice junctions between these exons were tested against unmapped reads: the testing of junctions was directed by the information available in the RNA-Seq dataset rather than *a priori* knowledge about the genome. Exons were thus chained into stranded gene models. Novel transcripts were identified in G-Mo.R-Se models on the basis of the occurrence of ≥ 4 reads per base pair, length ≥ 70 bp, together with a separation by ≥ 500 bp from any currently annotated gene models and transposons to avoid counting missing 5'- and 3'-exons. Predicted novel transcripts were then visually validated by manual curation. All G-Mo.R-Se gene models were converted to gff3 format, visualized in the context of the existing annotation and incorporated in the Genoscope and INRA *Tuber* genome databases and browsers (see 'G-Mo.R-Se models' track).

For the detection of antisense transcripts, *ab initio* gene models and G-Mo.R-Se models were clustered, and gene

models (≥ 4 reads per base pair, length ≥ 70 bp) with opposite orientations were identified.

Gene ontology (GO)

We obtained the GO terms of each *T. melanosporum* gene by the software Blast2GO (Conesa *et al.*, 2005) using the default parameters. Blast2GO was also used for GO functional enrichment analysis of certain genes by performing Fisher's exact test with robust false discovery rate (FDR) correction to obtain an adjusted *P* value between specific tested gene groups and the *in silico* gene model repertoire.

Results and Discussion

RNA-Seq sequencing and mapping

To illuminate the transcriptional landscape of the symbiotic fungus *T. melanosporum*, RNA samples of free-living (saprobic) mycelium (FLM), FB (truffle) and ectomycorrhizal root tips of *C. avellana*/*T. melanosporum* (ECM) were obtained. We generated cDNA libraries per tissue and pooled them (per tissue) before sequencing DNA fragments using the Illumina GA analyzer. Each individual pooled cDNA library yielded *c.* eight million sequence reads of 36 bp in length per sample. The sequencing reactions generated a total of 24 509 204 sequence reads (Table 1) representing > 882 Mb of sequence data (*c.* 56 $\times$ coverage of the coding space).

To construct a coverage signal profile across the genome, all reads were then aligned to the reference genome assembly (v1.0) of *T. melanosporum* Mel28 (Fig. S1) (Martin *et al.*, 2010) using BOWTIE software (Langmead *et al.*, 2009). Of the total sequence reads, *c.* 80% matched to unique genomic locations for the FLM and FB libraries (Table 1), whereas, for the ectomycorrhizal sample, this percentage was lower (40.1%) because symbiotic tissues contain a mixture of plant and fungal transcripts. A total of 16 588 455 reads mapped to a unique location. About 2.3% matched to multiple genomic locations (e.g. active retrotransposons, ribosomal RNAs) and 30% did not match the genome. The latter corresponded mainly to plant transcripts from the ectomycorrhizal sample, reads covering exon-intron junction sites, mitochondrial transcripts (*c.* 40 000 reads) and reads of poor quality. As expected, most uniquely mapped reads (84%) were confined to predicted *ab initio* gene models, that is, 65%, 3% and 19% within exons, introns and UTRs, respectively.

Reads mapping outside the predicted UTRs and gene models (13%) could be ascribed to putative novel transcripts, as tRNAs and snoRNAs comprised < 200 reads. As shown in Fig. S1, the predicted protein-coding sequences are consistent with the number of mapped Illumina reads. In general, the gene annotation obtained is in agreement

Table 2 Number and distribution of each type of alternative splicing event in different *Tuber melanosporum* tissues

Occurrence	Alternative splicing events in all transcripts (%)	Alternative splicing events in transcripts with ≥ 100 reads (%)
All tissues	960 (54.6)	755 (58.2)
FLM and FB	327 (18.6)	256 (19.7)
FLM and ECM	160 (9.1)	100 (7.7)
FB and ECM	113 (6.4)	82 (6.3)
FLM uniquely	61 (3.5)	27 (2.1)
FB uniquely	121 (6.9)	71 (5.5)
ECM uniquely	16 (0.9)	4 (0.3)
Total	1758	1295

AS, alternative splicing; ECM, ectomycorrhizal root tips; FB, fruiting body; FLM, free-living mycelium.

with the current *Tuber* Genome Database annotation (Martin *et al.*, 2010) as shown on the genome browser of TuberDB (<http://mycor.nancy.inra.fr/cgi-bin/gbrowse/tuberDB/>). However, additional features were identified, and some of the most relevant are highlighted in the following sections as they are altered by the various developmental stages.

Based on the last TuberDB update, the current gene catalog of *T. melanosporum* consists of a total of 7496 protein-coding sequences, 1315 of which were manually curated and 564 of which, of 75 amino acids or less in size, were considered as dubious. Reads mapping to unique locations on the genome were aligned to the known CDSs of the predicted 7496 protein-coding genes of *T. melanosporum* (Table 1,

Fig. S1) to provide further experimental support to *in silico*-based annotation. Gene transcription was measured in RPKM, a normalized measure of exonic read density that allows transcript levels to be compared both within and between samples (Mortazavi *et al.*, 2008). Using a stringent coverage threshold for expression (≥ 4 reads per bp), we detected 7048 protein-coding genes expressed during different developmental stages, indicating an extensive transcription (94%) of the whole *T. melanosporum* gene repertoire.

Identification of splice variants

Post-transcriptional processing events, such as AS, play a major role in generating protein diversity and subsequent functional complexity (Modrek & Lee, 2002; McGuire *et al.*, 2008). They may have an impact on gene expression at the different developmental stages of *T. melanosporum*. To explore AS events in *T. melanosporum*, we generated an exon junction database for predicted *ab initio* gene models and then identified sequence reads mapping to these regions (see Materials and Methods section). Amongst the 15 223 splice junctions mapped by sequence reads, 98.7% corresponded to already annotated gene model junctions. AS is not a rare event in *T. melanosporum* as 15.4% (1160) of the predicted genes produced two or more AS isoforms (Table S1). This percentage is higher ($> 27\%$) in *Magnaporthe grisea* (Gowda *et al.*, 2006), but lower in *Aspergillus oryzae* genes (8.5%) (Wang *et al.*, 2010). This suggests that AS probably plays a major role in generating protein diversity in *T. melanosporum*.

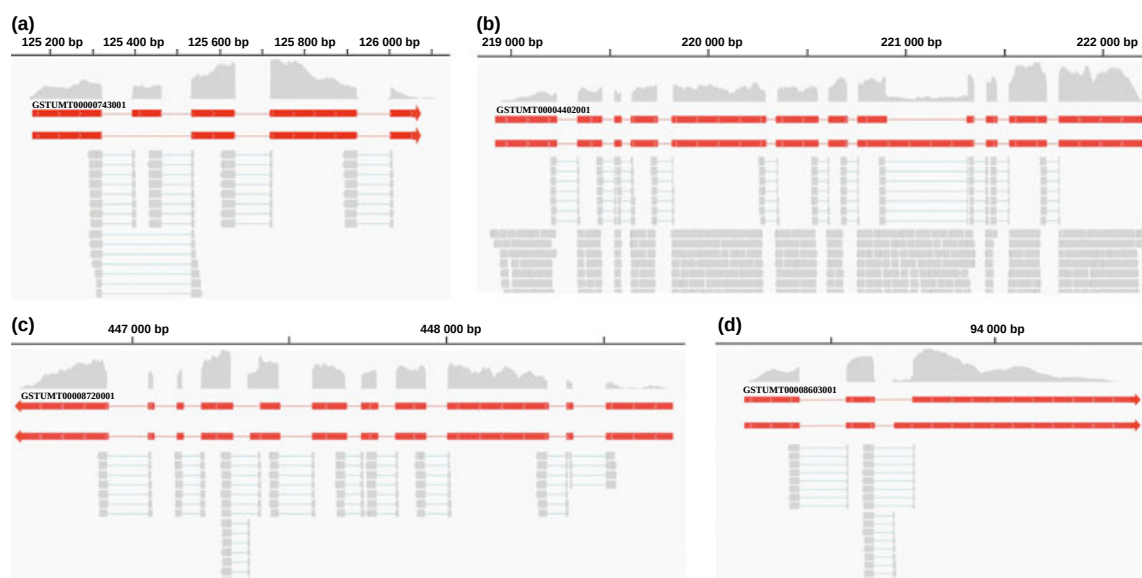


Fig. 1 Alternative splicing (AS) events in the transcriptome of *Tuber melanosporum*. Examples of the four major types of AS event. The light grey curve indicates the read coverage. The dark grey bar denotes the predicted structures of annotated genes. The mapped reads are depicted by the short light grey box and the junction reads are depicted by the broken light grey boxes below the gene structure models. (a) Exon skipping; (b) intron retention; (c) alternative 5'-splice site; (d) alternative 3'-splice site.

We found 1758 AS events (Table 2, Fig. 1) including RI (1653 events), A3SS (45 events), A5SS (35 events), SE (22 events), AFE (two events) and ALE (one event). The number of genes affected by AS varied in saprobic mycelium, ectomycorrhizas and truffles, and was higher in truffles (Table 2), indicating that AS is developmentally regulated. For example, several genes encoding homocysteine synthase and reductases showed AS events specific of symbiosis (Table 3). RI, that is, a single intron is alternatively included or spliced, was the primary AS type. RI occurred in 94% of all AS events in *T. melanosporum*, whereas cassette exons (including SE, AFE and ALE) only constituted 1.42%. Several genes, such as those coding for zinc-dependent oxidoreductase and lysophospholipase, presented several types of AS event (Table S1).

Characterization of UTRs

Because UTRs contain important regulatory motifs, the identification of untranslated boundaries is crucial for an

understanding of post-transcriptional gene regulation in *T. melanosporum*. UTRs are known to play key roles in the post-transcriptional regulation of gene expression, influencing mRNA transport, mRNA subcellular localization and mRNA turnover (Mignone *et al.*, 2002). Therefore, annotation of *T. melanosporum* has the potential to provide important insights into the gene regulatory mechanisms underlying ectomycorrhiza and fruit body development. We thus mapped the 5'- and 3'-end boundaries (5'-, 3'-UTRs) of *T. melanosporum* genes by searching for a sharp reduction in RNA-Seq read signals at both ends of predicted gene models. A few genes whose 5'- or 3'-boundaries overlapped with other genes were excluded from this UTR analysis. This RNA-based mapping identified or extended 5'- and 3'-boundary regions for 3936 and 4213 transcribed genes, respectively. The median lengths of both 5'- and 3'-UTRs were 105 and 210 bp, respectively. The median length of 5'-UTRs of *T. melanosporum* is very similar to that detected in *A. oryzae* (107 bp), whereas the median length of 3'-UTRs is substantially larger than those detected

Table 3 Selected transcripts showing an alternative isoform uniquely in ectomycorrhizal root tips or fruiting body

Gene	AS type	Tissue	Percent alternative isoform FLM	Percent alternative isoform FB	Percent alternative isoform ECM	Definition
GSTUMT00000404001	A5SS	ECM	0	0	70.83	Histone acetyltransferase
GSTUMT00009485001	RI	ECM	0	0	30.30	Homocysteine synthase
GSTUMT00000686001	RI	ECM	0	0	8.86	Carbonyl reductase
GSTUMT00007076001	RI	ECM	0	0	25.22	3-Oxoacyl reductase
GSTUMT00008321001	SE	FB	0	7.14	0	Ketol-acid reductoisomerase
GSTUMT00000585001	RI	FB	0	3.28	0	Sulfate/bicarbonate/oxalate exchanger
GSTUMT00001066001	RI	FB	0	1.21	0	Hydroxymethylglutaryl-CoA lyase
GSTUMT00001479001	A5SS	FB	0	2.67	0	Ubiquitin regulatory protein
GSTUMT00001845001	RI	FB	0	3.79	0	α -Trehalose-phosphate synthase
GSTUMT00001994001	RI	FB	0	2.23	0	Ribose-phosphate pyrophosphokinase
GSTUMT00002747001	A5SS	FB	0	0.75	0	Sulfate adenyllyltransferase
GSTUMT00003745001	RI	FB	0	1.16	0	Adenyllyl-sulfate kinase
GSTUMT00004366001	RI	FB	0	0.59	0	Glucosylase
GSTUMT00006263001	RI	FB	0	3.64	0	Translation initiation factor eIF-3
GSTUMT00006278001	RI	FB	0	2.17	0	Aconitase
GSTUMT00006718001	RI	FB	0	2.99	0	Glutathione S-transferase
GSTUMT00006745001	RI	FB	0	2.99	0	Glutamate dehydrogenase
GSTUMT00007052001	RI	FB	0	0.43	0	Alanine aminotransferase
GSTUMT00007395001	RI	FB	0	3.38	0	Fructose-6-phosphate 2-kinase
GSTUMT00008267001	RI	FB	0	12.98	0	26S proteasome protein
GSTUMT00008924001	A5SS	FB	0	33.33	0	GTPase activity
GSTUMT00009183001	RI	FB	0	14.20	0	Large subunit ribosomal protein L15
GSTUMT00009297001	A3SS	FB	0	1.81	0	Lysophospholipase
GSTUMT00009363001	RI	FB	0	9.18	0	Ribose 5-phosphate isomerase
GSTUMT00009396001	RI	FB	0	3.91	0	Acetolactate synthase
GSTUMT00011873001	RI	FB	0	43.08	0	Actin-like protein
GSTUMT00009592001	A3SS	FB	0	3.13	0	Extracellular lipase
GSTUMT00010358001	A3SS	FB	0	2.84	0	Monooxygenase cytochrome P450
GSTUMT00010528001	RI	FB	0	0.38	0	Cystathionine γ -lyase

Only genes supported by at least 100 reads and showing a sequence similarity with a protein of known function were considered. Alternative splicing (AS) type, tissue, percentage of alternative isoform and best BLAST hit are shown.

A3SS, alternative 3'-splice site; A5SS, alternative 5'-splice site; ECM, ectomycorrhizal root tips; FB, fruiting body; FLM, free-living mycelium; RI, retained intron; SE, skipped exon.

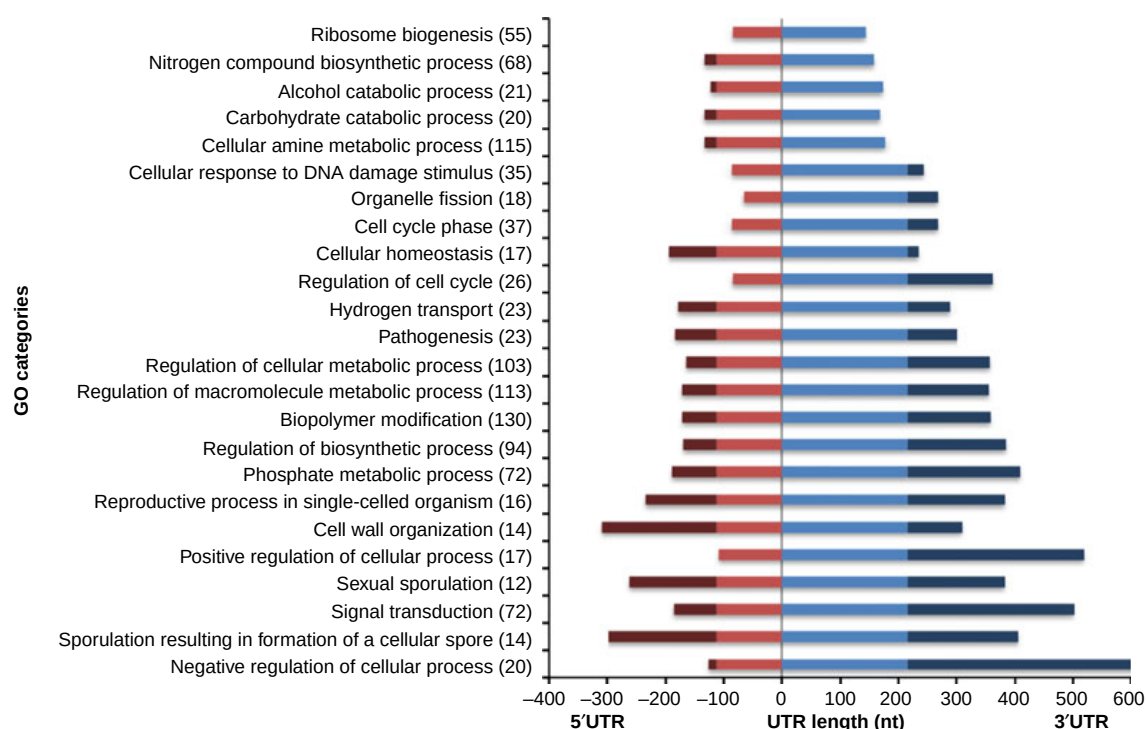


Fig. 2 Histograms showing the length distribution of 5'- and 3'-untranslated regions (UTRs) for transcripts for selected gene ontology (GO) categories. 5'-UTRs are depicted on the left and 3'-UTRs on the right. The part of the UTR exceeding the overall median length is represented in a darker color. Numbers in parentheses correspond to the number of genes in each GO category.

in other fungi (*A. oryzae*, 156 bp; *Schizosaccharomyces pombe*, 169 bp; *Saccharomyces cerevisiae*, 104 bp) (Nagalakshmi *et al.*, 2008; Wilhelm *et al.*, 2008; Wang *et al.*, 2010). We also compared the UTR length distribution for different GO functional categories (Fig. 2). Transcripts related to cell wall organization (GO:0071555), signal transduction (GO:0007165) and sporulation (GO:0034293, GO:0030435) had remarkably longer 5'- and 3'-UTRs, suggesting that these transcripts have a rapid turnover rate (Wilhelm *et al.*, 2010). By contrast, transcripts involved in ribosome biogenesis (GO:0042254) had significantly shorter UTRs and are thus probably less prone to mRNA degradation. Interestingly, this analysis of GO categories uncovered similarities to *S. pombe* and *A. oryzae* (Wilhelm *et al.*, 2008; Wang *et al.*, 2010). For example, transcripts related to signal transduction and sporulation had remarkably longer 5'- and 3'-UTRs, which, being the least stable, accelerate RNA turnover. By contrast, genes with shorter UTRs, and probably more stable, were significantly enriched in functions related to ribosome biogenesis. Intriguingly, a large number of transcripts with very long 5'-UTRs are key transduction pathway regulators, such as histidine kinase. Therefore, fine-tuning RNA turnover rates for these transcripts is potentially a major regulatory process involved in the control of environmental sensing.

In several filamentous fungi, short μ ORFs located in 5'-UTRs upstream of the start codon play a role in the translational regulation of gene expression (Mignone *et al.*, 2002; Vilela & McCarthy, 2003). μ ORFs were identified for 508 genes (12.9% of 5'-UTRs) (Fig. S2), suggesting that this transcriptional regulation mechanism might also be important in *T. melanosporum*. Our RNA-Seq analysis identified μ ORFs for *c.* 8% of *T. melanosporum* genes, whereas 11% and 6% were found in *A. oryzae* (Wang *et al.*, 2010) and *S. cerevisiae* (Nagalakshmi *et al.*, 2008), indicating that this post-transcriptional regulation mechanism might also be important in *T. melanosporum*.

GO functional enrichment analysis revealed that genes associated with membrane, transport, carbohydrate metabolic processes and nitrogen compound biosynthetic processes were significantly enriched in these μ ORFs (Table S2). Fifty-four and 106 of these genes showed differential expression on symbiosis or fructification, respectively, and μ ORFs may be involved in their regulation (Table S3). These transcripts code for proteins playing key metabolic roles. For example, transcripts coding for a serine-phosphatidyl decarboxylase, a zinc-finger transcription factor and a zinc-regulated transporter were upregulated in ECM, whereas a transcript encoding a malate synthase was upregulated in truffle.

Discovery of novel genes, antisense transcripts and improvement of genome annotation

RNA-Seq can complement automated genome annotation efforts for gene discovery (Trapnell *et al.*, 2010; Wang *et al.*, 2010). We thus questioned whether our sequence reads identified novel transcriptionally active regions (TARs) within intergenic regions of the *T. melanosporum* genome employing G-Mo.R-Se (Denoeud *et al.*, 2008), a method aimed at using RNA-Seq short reads to build *de novo* gene models. Thirteen percent of the uniquely mapped reads showed no sequence similarity to *ab initio* annotated gene model sequences, but matched the assembly. These TARs included 91 putative novel genes (average gene size of 766 bp) with plausible CDSs and a recognizable exon–intron structure (Fig. S3, Table S4). Of these, 77 are orphan genes with no sequence similarity to known genes in public databases, whereas 12 encode conserved hypothetical proteins and eight have a predicted protein sequence with a known PFAM domain. In addition, 17 and 12 of these novel transcripts were upregulated in ectomycorrhizas and truffles, respectively, and 11 and 24 were downregulated (Table S4). Their functional analysis may provide novel insights into the mechanisms regulating symbiosis and FB formation.

Next, G-Mo.R-Se-reconstructed 7837 transcript models were compared with the predicted *ab initio* gene models. There were discrepancies between *ab initio* genome annotation and RNA-Seq-predicted genes in c. 1500 of the 7496 annotated gene models (genome assembly annotation v1.0), suggesting incorrect annotations. Amongst these misannotated genes, we found 65 novel exons in introns for 56 genes, 129 novel exons in the 5' upstream region for 88 genes, and 141 novel exons in the 3' downstream region for 94 genes. Finally, 118 *ab initio* gene models with a size of 75 amino acids or less and with no RNA-Seq reads were also considered as misannotations. These revised annotations are available on the Genoscope Tuber Genome Browser (<http://www.genoscope.cns.fr/tuber>).

Thirty-three antisense transcripts were identified (Table S5). They encode for various enzymes, such as oxidoreductase, carbonyl reductase, pyrophosphatase, glyoxalase, methyltransferase and glucanase. Three of these transcripts, coding for hypothetical proteins or glyoxylase, were upregulated during symbiosis, whereas a gene encoding for a monooxygenase was upregulated in truffle. AS and antisense transcription add additional levels of complexity in the regulation of gene networks.

In conclusion, using RNA-Seq, we curated the current *in silico* annotation of the *T. melanosporum* protein-coding gene repertoire, identified 91 previously unidentified transcripts and characterized novel exons, UTRs, μ ORFs and AS events. RNA-Seq also showed that major developmental transitions in *T. melanosporum* not only involved quantitative changes in gene expression levels, but also a series of

transcriptional and translational regulation mechanisms. By providing an unbiased view of the transcriptome, RNA-Seq transcript profiling will pave the way for a better understanding of the molecular processes underlying symbiosis and truffle FB formation (Ceccaroli *et al.*, 2010; Montanini *et al.*, 2010; Rubini *et al.*, 2010).

Acknowledgements

This work was supported by the European Commission within the Network of Excellence EVOLTREE (FP6-016322) and the Project Energypoplar (FP7-211917), INRA and Région Lorraine grants (to F.M.). This work was funded, in part, by Institut de Génétique, CEA and Agence Nationale de la Recherche to the Génoscope. E.T. is supported by a PhD scholarship from the Région Lorraine. We thank Professor Simone Ottonello and Paola Bonfante for numerous exciting discussions. We would like to thank Dr Gérard Chevalier (INRA, Clermont-Ferrand, France) and Dr Elisa Zampieri (University of Torino, Italy) for providing *T. melanosporum* mycelial and fruit body samples.

References

- Ceccaroli P, Buffalini M, Saltarelli R, Barbieri E, Polidori E, Ottonello S, Kohler A, Tisserant E, Martin F, Stocchi V. 2011. Genomic profiling of carbohydrate metabolism in the ectomycorrhizal fungus *Tuber melanosporum*. *New Phytologist* **189**: 751–764.
- Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674–3676.
- Denoeud F, Aury JM, Da Silva C, Noel B, Rogier O, Delledonne M, Morgante M, Valle G, Wincker P, Scarpelli C *et al.* 2008. Annotating genomes with massive-scale RNA sequencing. *Genome Biology* **9**: R175.
- Eaton CJ, Cox MP, Scott B. 2010. What triggers grass endophytes to switch from mutualism to pathogenesis? *Plant Science*, doi: 10.1016/j.plantsci.2010.10.002.
- Gowda M, Venu RC, Raghupathy MB, Nobuta K, Li H, Wing R, Stahlberg E, Coughlan S, Haudenschild CD, Dean R *et al.* 2006. Deep and comparative analysis of the mycelium and appressorium transcriptomes of *Magnaporthe grisea* using MPSS, RL-SAGE, and oligoarray methods. *BMC Genomics* **7**: 310.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**: R25.
- Larsen PE, Trivedi G, Sreedasyam A, Lu V, Podila GK, Collart FR. 2010. Using deep RNA sequencing for the structural annotation of the *Laccaria bicolor* mycorrhizal transcriptome. *PLoS ONE* **5**: e9780.
- Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. 2008. RNA-Seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research* **18**: 1509–1517.
- Martin F, Kohler A, Murat C, Balestrini R, Coutinho PM, Jaillon O, Montanini B, Morin E, Noel B, Percudani R *et al.* 2010. Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* **464**: 1033–1038.
- McGuire AM, Pearson MD, Neafsey DE, Galagan JE. 2008. Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biology* **9**: R50.

- Mello A, Murat C, Bonfante P. 2006. Truffles: much more than a prized and local fungal delicacy. *FEMS Microbiology Letters* **260**: 1–8.
- Mignone F, Gissi C, Liuni S, Pesole G. 2002. Untranslated regions of mRNAs. *Genome Biology* **3**: reviews0004–reviews0004.10.
- Modrek B, Lee C. 2002. A genomic view of alternative splicing. *Nature Genetics* **30**: 13–19.
- Montanini B, Levati E, Bolchi A, Kohler A, Morin E, Tisserant E, Martin F, Ottonello S. 2011. Genome-wide search and functional identification of transcription factors in the mycorrhizal fungus *Tuber melanosporum*. *New Phytologist* **189**: 736–750.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**: 621–628.
- Murat C, Riccioni C, Belfiori B, Cichocki N, Labbé J, Morin E, Tisserant E, Rubini A, Paolocci F, Martin F. 2011. Distribution and localization of microsatellites in the Périgord black truffle genome and identification of new molecular markers. *Fungal Genetics and Biology*, doi: 10.1016/j.fgb.2010.10.007.
- Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**: 1344–1349.
- Rubini A, Belfiori B, Riccioni C, Tisserant E, Arcioni S, Martin F, Paolocci P. 2011. Isolation and characterization of *MAT* genes in the symbiotic ascomycete *Tuber melanosporum*. *New Phytologist* **189**: 710–722.
- Splivallo R, Ottonello S, Mello A, Karlovsky P. 2011. Truffle volatiles: from chemical ecology to aroma biosynthesis. *New Phytologist* **189**: 688–699.
- Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D *et al.* 2008. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**: 956–960.
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* **28**: 511–515.
- Vilela C, McCarthy JE. 2003. Regulation of fungal gene expression via short open reading frames in the mRNA 5′ untranslated region. *Molecular Microbiology* **49**: 859–867.
- Wang B, Guo G, Wang C, Lin Y, Wang X, Zhao M, Guo Y, He M, Zhang Y, Pan L. 2010. Survey of the transcriptome of *Aspergillus oryzae* via massively parallel mRNA sequencing. *Nucleic Acids Research* **38**: 5075–5087.
- Wang ET, Sandberg R, Luo S, Khrebukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**: 470–476.
- Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bähler J. 2010. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **453**: 1239–1243.

Supporting Information

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

Fig. S1 The distribution of transcribed regions in the genome of *Tuber melanosporum* as detected by RNA-Seq read mapping

Fig. S2 An example of upstream open reading frames (μORFs) identified in the gene model GSTUMT000081 35001 coding for a zinc-regulated transporter.

Fig. S3 An unannotated novel transcript constructed by G-Mo.R-Se software based on RNA-Seq sequences in the *Tuber melanosporum* genome.

Table S1 List of transcripts showing alternative splicing events

Table S2 List of gene ontology (GO)-enriched terms in transcripts having upstream open reading frames (μORFs)

Table S3 List of transcripts containing upstream open reading frames (μORFs) and showing a regulation in ectomycorrhizal root tips or fruiting body

Table S4 List of novel transcripts identified by RNA-Seq

Table S5 Genes with identified antisense transcripts in the *Tuber melanosporum* transcriptome

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.

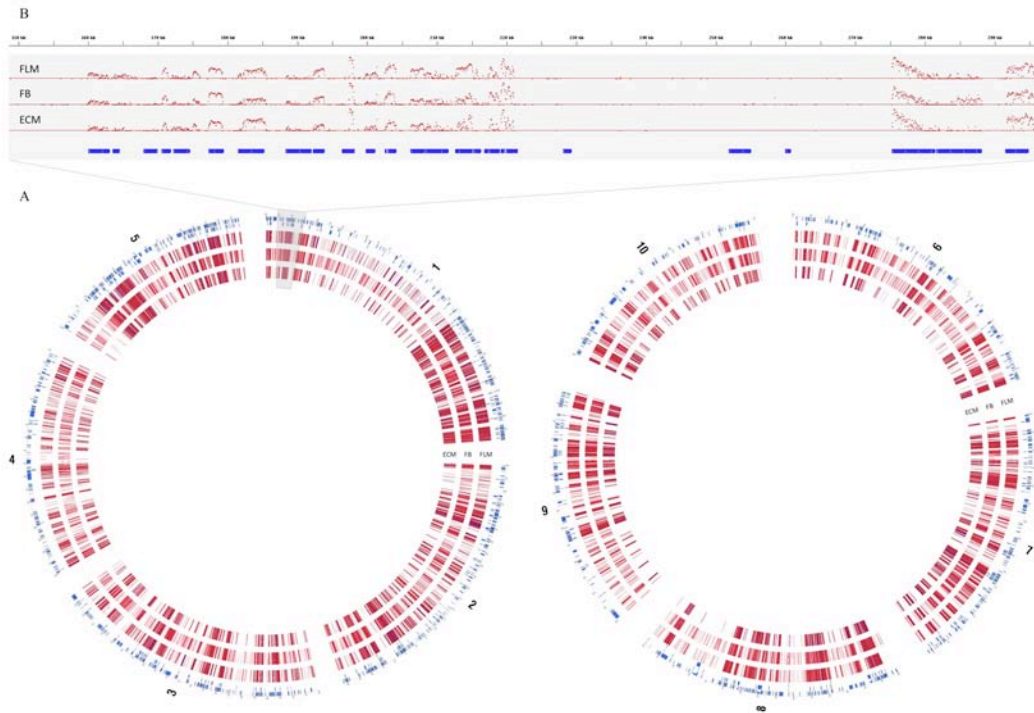


Figure S1 : The distribution of transcribed regions in the genome of *T. melanosporum* as detected by RNA-Seq read mapping. (A) Red colors in the innermost circles represent different transcription levels for each genes in ectomycorrhiza (ECM), fruiting body (FB) or free-living mycelium (FLM) for the 10 largest supercontigs of the genome assembly v. 1.0. Blue colors in the outermost circles indicated gene model tracks. (B) Plot depicting numbers of sequence reads (log-scale) across a genomic region of supercontig 1 (coordinates in kb). Predicted coding sequences are indicated in blue.

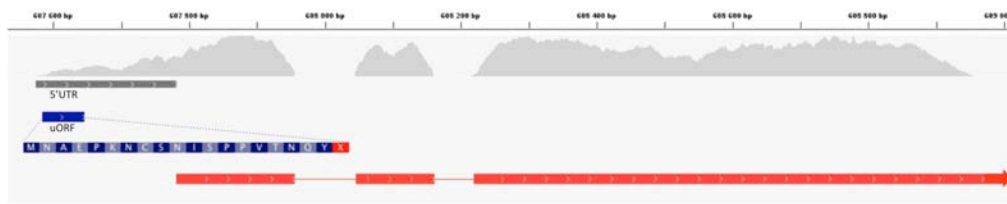


Figure S2 : An example of μ ORF identified in the gene model GSTUMT00008135001 coding for a zinc-regulated transporter. There are the following information tracks from top to bottom: supercontig sequence strand with coordinates, the total aligned RNA-Seq reads in grey, exons and 5' and 3' UTRs of the G-Mo.R-Se model in red and dark grey bars, respectively and the μ ORF in the blue box with its amino acid sequence.

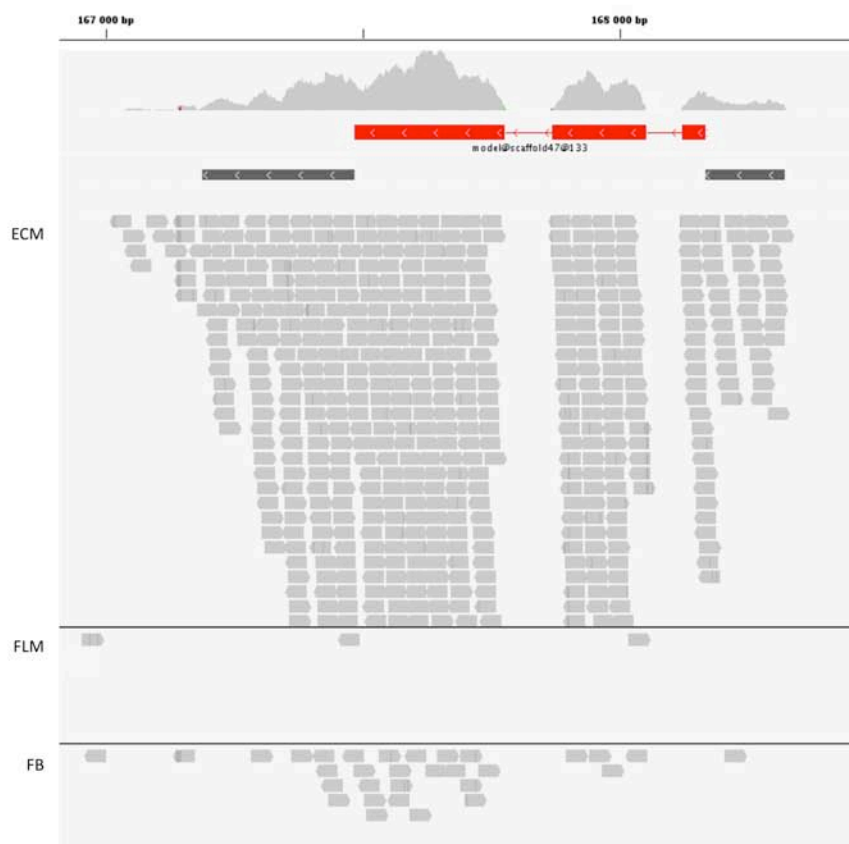


Figure S3 : An unannotated novel transcript constructed by G-Mo.R-Se software based on RNA-Seq sequences in *T. melanosporum* genome. There are the following information tracks from top to bottom: supercontig sequence strand with coordinates, the total aligned RNA-Seq reads in grey, exons and 5' and 3' UTRs of the G-Mo.R-Se model in red and dark grey bars, respectively and the mapped RNA-Seq reads for ECM (top), FLM (middle) and FB (down) in grey boxes.

**Comprehensive transcriptome analysis of the Perigord black truffle
unravels gene expression alterations during symbiosis and fruiting
body formation**

E. Tisserant, C. Da Silva, A. Kohler, E. Morin, P. Wincker et F. Martin.

Article en préparation

Comprehensive transcriptome analysis of the Perigord black truffle unravels gene expression alterations during symbiosis and fruiting body formation

E. Tisserant¹, C. Da Silva², A. Kohler¹, E. Morin¹, P. Wincker², and F. Martin^{1*}

¹ INRA, UMR INRA/Nancy Université 'Interactions Arbres/Micro-Organismes', INRA-Nancy, 54280 Champenoux, France; ² CEA, IG, Genoscope, 2 rue Gaston Crémieux CP5702, F-91057 Evry, France

Summary

- The genome of the Perigord Black truffle (*Tuber melanosporum* Vittad.) has prominent unique features. However, the functional complexity of the *T. melanosporum* transcriptome has not yet been fully elucidated.
- Here, we applied high-throughput Illumina RNA-sequencing to the transcriptome of *T. melanosporum* under different major developmental stages, free-living mycelium, fruiting body and ectomycorrhiza. Sequencing of cDNA libraries generated a total of ~24 millions sequence reads representing more than 882 Mb of sequence data.
- Many genes and pathways that might be crucial developmental factors in symbiosis and fruiting body formation were identified by comparing gene expression levels in hyphae at different developmental stages. These results may provide a blueprint for further study of the gene networks controlling the different life cycle stages.

Key words: Tuber, transcriptome, RNA-Seq, mycorrhiza, fruiting body

INTRODUCTION

The Perigord black truffle (*Tuber melanosporum* Vittad.) is endemic to calcareous soils in southern Europe and found in mutualistic symbiosis with roots of deciduous trees, mostly oaks (*Quercus* spp.) and hazelnut trees (*Corylus avellana*). This ectomycorrhizal ascomycete requires a host tree to complete its life cycle and produce hypogeous fruiting bodies (ascomata), so-called truffles (Mello *et al.*, 2006). Truffles are a highly appreciated delicacy for their complex organoleptic properties (i.e. taste and aroma). This ectomycorrhizal symbiosis is of major ecological importance in woodland and forest ecosystems (Mello *et al.*, 2006). Inoculation of *Cistus* and oak seedlings in greenhouse and nursery with *T. melanosporum* mycelium lead to an increased plant growth showing that the fungus is a mutualistic symbiont (Mello *et al.*, 2006, Martin *et al.*, 2010). It has however been suggested that *T. melanosporum* may act as a biotroph with an antagonistic nature in the later stages of the interaction (G Chevalier, pers. com.). Identification of developmental processes that condition and trigger symbiosis and fruiting body formation, ultimately leading to sustainable truffle crop management will be facilitated by a thorough analysis of genome traits, physiology and ecology of the mycobiont (Martin *et al.*, 2010, Montanini *et al.*, 2010; Rubini *et al.*, 2010; Splivallo *et al.*, 2010). The host plant, the environment and the community of soil organisms with which *T. melanosporum* interacts affect the function of its genome (Martin *et al.*, 2010), but our current knowledge on the gene, transcript and protein networks involved in this complex web of interactions is scarce.

The available draft genome sequence provides access to the ~7,500 gene repertoire of *T. melanosporum* allowing to monitoring genome-wide changes in gene expression during the successive stages of the truffle life cycle (Martin *et al.*, 2010). The transcriptome of *T. melanosporum* has been interrogated by using custom genome-wide exon oligoarrays (Martin *et al.*, 2010). Both ectomycorrhiza and fruiting body development induce alterations in gene expression; 7.6% of the total gene repertoire showed an upregulated expression in these lifecycle stages. In the symbiosis, the regulated genes identified so far encode proteins playing a role in detoxication of plant defence metabolites, nutrient exchange, and colonization of root apoplast through the deconstruction of cell walls. It appears that *T. melanosporum* is much more aggressive in its colonization of root tissue, expressing high levels of transcripts coding for laccase, tyrosinase and lipase, as compared to *L. bicolor* (Martin *et al.*, 2008). Microarray expression profiling also showed that a developmental reprogramming takes place during *T. melanosporum* fruiting body formation with an induction of transcripts coding for lineage-specific unknown proteins and enzymes involved in redox metabolism (Martin *et al.*, 2010).

Interrogation of the *T. melanosporum* transcriptome in the latter study has been greatly facilitated by the use of custom exon oligoarrays, which quantify transcript abundance by hybridization. All expression arrays are however based on *a priori* assumptions regarding the expressed parts of the genome. This limitation, together with the restricted dynamic range of signal intensity measurements (1 to 1,000), may hamper the potential of microarrays for unbiased studies of transcription and the discovery of novel transcripts. Recently, the development of next generation Illumina sequencing of steady state RNA populations (RNA-Seq) has provided unprecedented details about the RNA landscape within a cell (Mortazavi *et al.*, 2008; Nagalakshmi *et al.*, 2008; Sultan *et al.*, 2008; Wilhelm *et al.*, 2008). This cost-effective sequencing-by-synthesis produces millions of short cDNA reads which are mapped to a reference genome to obtain a quantitative measure of transcript expression levels at single-nucleotide resolution as well as transcript structure informations. RNA-Seq provides access to rare transcripts (thanks to a wide dynamic range from 1 to >1,000,000), more accurate quantification of abundant transcripts (above the signal saturation point of arrays).

Here we report the results from deep sequencing of *T. melanosporum* transcriptome in saprobic mycelium, ectomycorrhiza and truffle fruiting body using the Illumina platform. The sequence data were used to quantify expression of known genes. In these analyses we assessed differences between the tissue types. We also compared RNA-Seq profiles with previously published transcript profiling generated by custom genome-wide exon oligoarrays (Martin *et al.*, 2010).

MATERIALS & METHODS

Biological materials

Free-living mycelium (FLM) of *T. melanosporum* Mel28 was grown on 1% malt agar (Cristomalt-D, Difal, Villefranche-sur-Saône, France) for five weeks before harvesting (seven biological replicates). The Mel28 strain used in this study is the same as the reference genome strain. Ectomycorrhizal root tips (ECM) were sampled (four biological replicates of 100 ECM each) from five-month-old Common Hazel (*Corylus avellana* L.) plantlets inoculated by a mycelium slurry produced from a fruiting body of *T. melanosporum* harvested in Meuse (France) by Dr. Gérard Chevalier. Inoculated plants were grown in the greenhouse facilities of the AGRI-TRUFFE company (Saint-Maixant, France). Five different fruiting bodies (FB) of *T. melanosporum* were collected at different maturation stages below hazel trees or oak trees at different locations [(Auvergne, Meuse, and Dordogne (France), and Piceno (Italy)].

Construction of cDNA library for RNA-Seq and Illumina sequencing

FLM, FB and ECM samples (100 mg fresh weight) were snap frozen in liquid nitrogen and RNA extraction was carried out using the RNeasy Plant Mini Kit including a DNase treatment (Qiagen, Cat No. 74904). RNA quality and integrity were checked prior to cDNA synthesis using the Bio-Rad Experion analyzer and Experion RNA StdSens analysis kit (Bio-Rad, Hercules, USA). Total RNA preparations (four biological replicates for ECM, five for FB and seven for FLM) were primed by Oligo-dT primer and amplified using the SMART PCR cDNA Synthesis Kit (Takara Bio Europe / Clontech, Saint-Germain-en-Laye, France) according to the manufacturer's instructions and used either for hybridizations to NimbleGen oligoarrays (for details, see Martin *et al.*, 2010) or pooled for Illumina sequencing. Cluster generation was performed after applying each DNA sample to the individual lanes of the Illumina Genome Analyser II flowcell according to the manufacturer's recommendations (Illumina, San Diego, CA). Three lanes of massively parallel sequencing by synthesis were performed and processed into SCARF files containing millions of 36-bp sequencing reads and raw quality scores using the Illumina Genome Analyser II and the Illumina Pipeline Software according to the manufacturer's instructions.

Mapping reads to reference genome and annotated genes

The alignments of Illumina reads to the genome were used to map reads to genomic features. The genomic location of each alignment may correspond to features such as genes, tRNA, rRNA or transposable elements. *T. melanosporum* Mel 28 genomic sequence and gene models were taken from the Genoscope Tuber Genome Browser (www.genoscope.cns.fr/tuber) using publicly available data files: [Tuber melanosporum v1.0 unmasked.fa.gz](http://www.genoscope.cns.fr/tuber/tuber_melanosporum_v1.0_unmasked.fa.gz) and [Tuber melanosporum v1.0 annot.gff.gz](http://www.genoscope.cns.fr/tuber/tuber_melanosporum_v1.0_annot.gff.gz) (Martin *et al.*, 2010). Genome assemblies together with predicted gene models and annotations for *T. melanosporum* are also available at DNA Data Bank of Japan/European Molecular Biology Laboratory/GenBank under the project accession numbers CABJ01000001–CABJ01004455 (whole genome shotgun sequencing data) and FN429986–FN430383 (scaffolds and annotations). In version 1.0 of the annotation, *T. melanosporum* gene repertoire contains 7,496 best gene models. *Populus trichocarpa* gene models were downloaded from the Joint Genome Institute Phytozome 5.0 web site (<http://www.phytozome.net/poplar.php>).

After removing Illumina reads containing sequencing adapters, the remaining 24,509,204 read FASTA files were aligned to the *T. melanosporum* Mel28 genome sequence and gene models using the BOWTIE software (Langmead *et al.*, 2009). This software is capable of producing spliced alignments of the input cDNA fragments that were aligned to the genome. The default mapping criteria used by BOWTIE were the followings: alignments may have no more than 2 mismatches in the first 28 bases on the high-quality (left) end of the read, called the "seed". The sum of the quality values at all mismatched positions (not just in the seed) may not exceed 70. As reads may align to the genome one or more times, the specifying parameters '-a -m 1 -best -strata' were used to instruct BOWTIE to choose the best alignment and to refrain from reporting any alignments for reads having more than one alignments, i.e. corresponding to repeated loci (572,211 reads). A customized Perl script (available on request) was used to assign each unique read to a specific genome location (nucleotide coordinates on scaffolds) for determining the genomic distribution.

Assessment of functional content

To obtain an accurate measure of gene expression, reads that were found in unique positions (16,588,455) in the genome were further mapped to coding sequences (CDS) of *in-silico* predicted protein-coding genes and counted to generate gene coverage values (Figure S1). We used only reads (*ca.*, 11,867,391) that mapped at unique position and only position where a read could potentially map uniquely within the frame of CDS (Sultan *et al.*, 2008). For each CDS, the number of all possible 36 bp oligomers was determined, using a sliding window of one base pair. To allow transcript level to be compared both within and between samples, the number of read hits per gene was normalized according to the theoretical length and the sample size and reported in number of reads per exon kilobase per million mapped sequence (RPKM) (Mortazavi *et al.*, 2008). We compared RNA-Seq transcript coverage values against the microarray expression values of the exact samples. The complete array expression dataset is available as series (accession number GSE17529) at the Gene Expression Omnibus at NCBI (Martin *et al.*, 2010). Both microarray expression and RNA-Seq transcript coverage values were normalized by base-2 logarithm transformations. The relationship between the two sets of values were quantified using Pearson's correlation.

Gene ontology and KOG analyses

We obtained the gene ontology (GO) terms of each *T.melanosporum* gene by the software Blast2GO (Conesa *et al.*, 2005) using the default parameters. Blast2GO was also used for GO functional enrichment analysis of certain genes, by performing Fisher's exact test with robust false discovery rate (FDR) correction to obtain an adjusted *p*-value between specific tested gene groups and *in silico* gene model repertoire. We obtained the KOG terms of each *T.melanosporum* gene by the software RPSBlast against the KOG database.

RESULTS

RNA-Seq sequencing and mapping

To illuminate the transcriptional landscape of the symbiotic fungus *T.melanosporum* and characterize changes in gene expression during its major developmental stages, RNA samples of free-living (saprobic) mycelium (FLM), fruiting body (truffle, FB), and ectomycorrhizal roots tips of *C. avellana*/*T. melanosporum* (ECM) were obtained. We generated several cDNA libraries per tissue and pooled them (per tissue) before sequencing DNA fragments using the Illumina GA analyzer. Each individual pooled cDNA library yielded ~8 millions sequence reads of 36 bp in length per sample. The sequencing reactions generated a total of 24,509,204 sequence reads (Table 1) representing more than 882 Mb of sequence data (*ca.*, 56X coverage of the coding space). To construct a coverage signal profile across the genome, all reads were then aligned to the reference genome assembly (v1.0) of *T. melanosporum* Mel28 (Martin *et al.*, 2010) using the BOWTIE software (Langmead *et al.*, 2009). Of the total sequence reads, 81.1% matched to unique genomic locations for the FLM and FB libraries, whereas for the ectomycorrhiza sample, this percentage was lower (40.1%) due to fact that symbiotic tissues contain a mixture of plant and fungal transcripts. As expected, most uniquely mapped reads (84%) were confined to predicted *ab initio* gene models, i.e., 65%, 3% and 19% within exons, introns or untranslated regions (UTRs), respectively (Figure 1). Reads mapping outside the predicted UTRs and gene models (13%) could be ascribed to putative novel transcripts as tRNAs and snoRNAs comprised less than 200 reads.

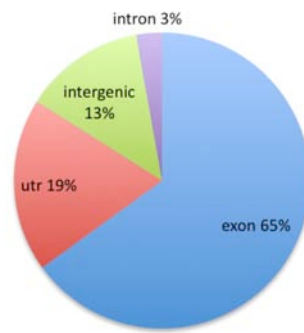


Figure 1 : Pie chart summary of Illumina read mapping statistics. Sectors indicate successful matches of reads to the specified *T. melanosporum* genome regions according to the annotated gene set (v.1.0).

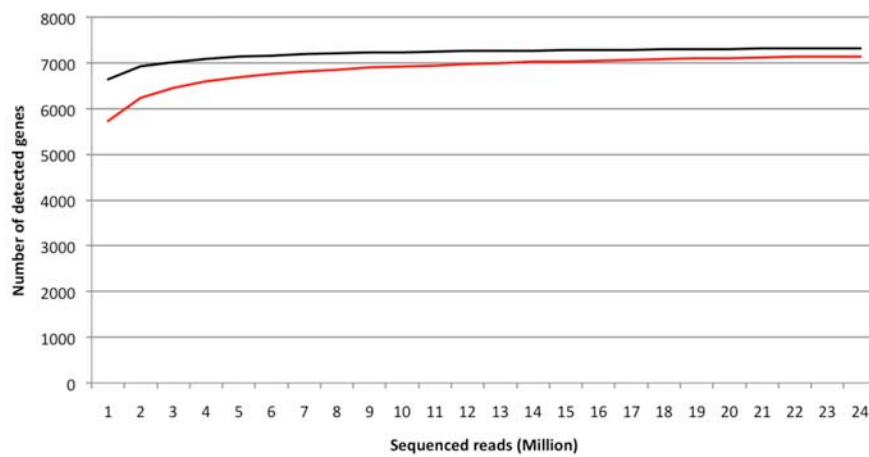


Figure 2 : Rarefaction curves depicting the effect of sequence number (in millions) on the number of *T. melanosporum* annotated genes identified by using a one- (black) or four-read threshold (red).

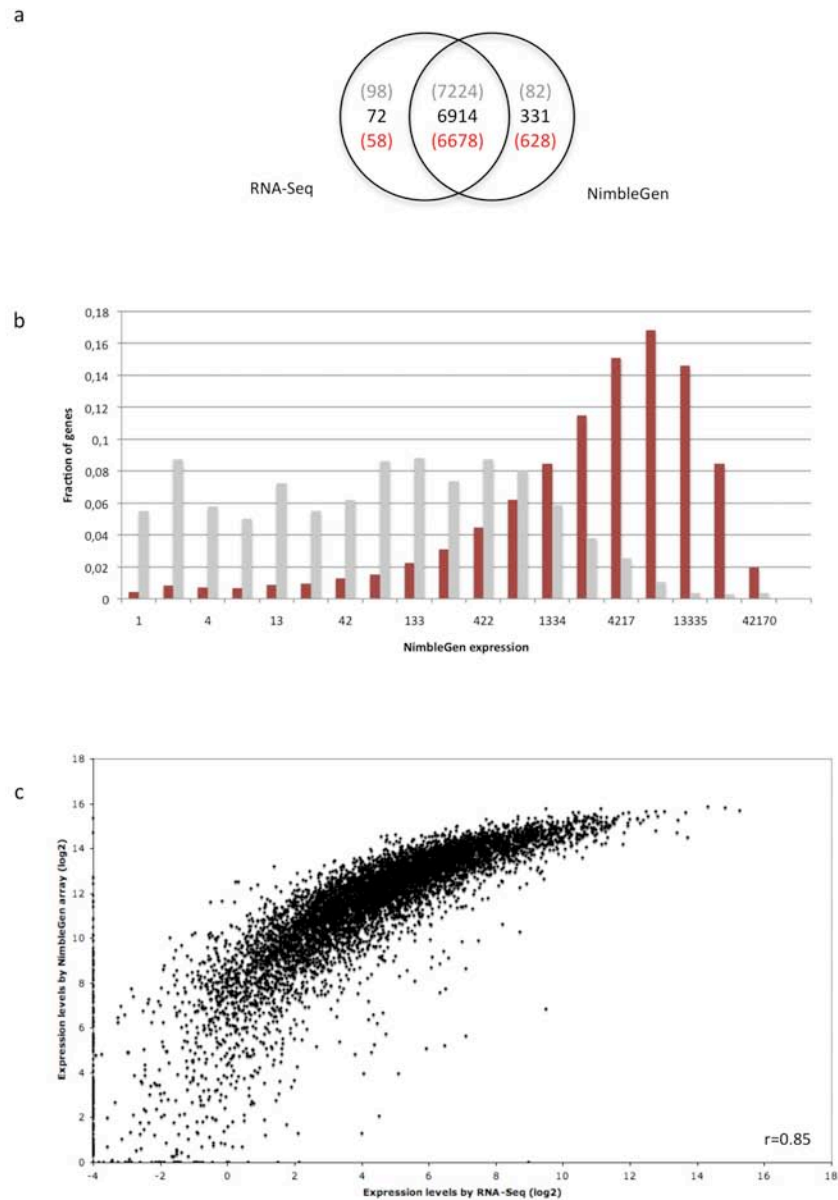


Figure 3 : Correlation between mRNA expression levels measured by NimbleGen exon oligoarray and RNA-Seq. **(A)** Venn diagram showing the number of transcripts detected by RNA-Seq and NimbleGen microarray in the same samples. RNA-Seq coverage values ≥ 1 , 5 and 10 reads are written in grey, black, and red, respectively. **(B)** Distribution of transcript expression values detected by the NimbleGen microarray. Transcripts missed by RNA-Seq are shown in grey bars, whereas transcripts also detected by RNA-Seq are shown in red bars. **(C)** Scatter plot of the relationship between microarray transcript expression (Y axis, $\log_2$) and RNA-Seq coverage in RPKM (X axis, $\log_2$). Note the strong association between the two variables (Pearson's $r = 0.85$).

We then tested the impact of sequencing depth on our ability to detect expressed genes (Figure 2). The number of detected genes (based on a stringent 4 reads per bp threshold) increased with the number of reads, and a plot of predicted genes vs the number of sequence reads resulted in rarefaction curves that did plateau at eight millions reads (Figure 2). Doubling the number of sequences only increased the number of detected genes by 2.5%, suggesting that the entire transcriptome of saprobic mycelium, truffles and ectomycorrhizal root tips was nearly covered. The remaining 6% of the annotated genomic features not detected in this study could represent transcripts that are not expressed in the analyzed developmental conditions or mis-annotations; 84 of these transcripts have no similarity in the GenBank database and a size ≤ 75 AA, suggesting that they are mis-annotations.

Comparison with gene expression arrays

One of the goals of this study was the assessment of Illumina sequencing as a tool for expression analysis. However, biological variation between samples was not addressed as pooled cDNA libraries were used for sequencing. The comparison of RNA-Seq profiles with profiles generated by microarrays using the same material was thus important to support our data. To test whether there was a positive correlation between transcript expression levels measured by microarrays and those measured by RNA-Seq, we thus compared the expression scores of saprobic mycelial, ectomycorrhizal and truffle RNA samples from our recent *T. melanosporum* oligoarray study (Martin *et al.*, 2010) against the coverage values of their corresponding RNA-Seq values. Using a stringent false discovery rate (FDR) (p -value of 0.05) for arrays and a detection cut off value ≥ 4 reads per base for RNA-Seq, 94.4% of *T. melanosporum* genes were identified to be expressed by both methods (Figure 3a). At variance with previous studies (Sultan *et al.*, 2008), the number of transcripts detected exclusively by oligoarray (331) was greater than those only detected by RNA-Seq (72). However, oligoarray-unique transcripts showed very low signal intensities values, in the range of the background intensity values, suggesting that they may be false positive signals (Figure 3b). We then plotted the $\log_2$ normalized expression levels obtained by the two methods (Figure 3c). We found a strong positive correlation between microarray expression and RNA-Seq coverage (Pearson's $r = 0.85$), confirming that RNA-Seq provides a quantitative measure of transcript levels.

Comparing expression levels in ectomycorrhiza, fruiting body and saprobic mycelium

We assessed the functional information content of RNA-Seq and microarray data through several different analyses. As expected, GO analyses showed that highly expressed genes were primarily involved in general biological processes such as primary metabolism, translation and transcription, and transport although several transcripts coding for unknown proteins were also present at high levels (Figure S2a). The GO analysis of molecular functions revealed that several enzyme transcripts were more abundant in fruiting body, e.g. oxidoreductases acting on sulfur group, carbon-sulfur lyases, transferases acting on nitrogenous compounds and superoxide radicals, and steroid binding (Figure S2b). These transcripts were not detected by microarrays because their probe signal was saturated (data not shown). An analysis of KOG categories uncovered similar transcript patterns in saprobic mycelium, ectomycorrhiza and truffle, with translational, posttranslational, signal transduction and energy mechanisms being the most prevalent housekeeping functions (Figure S2c).

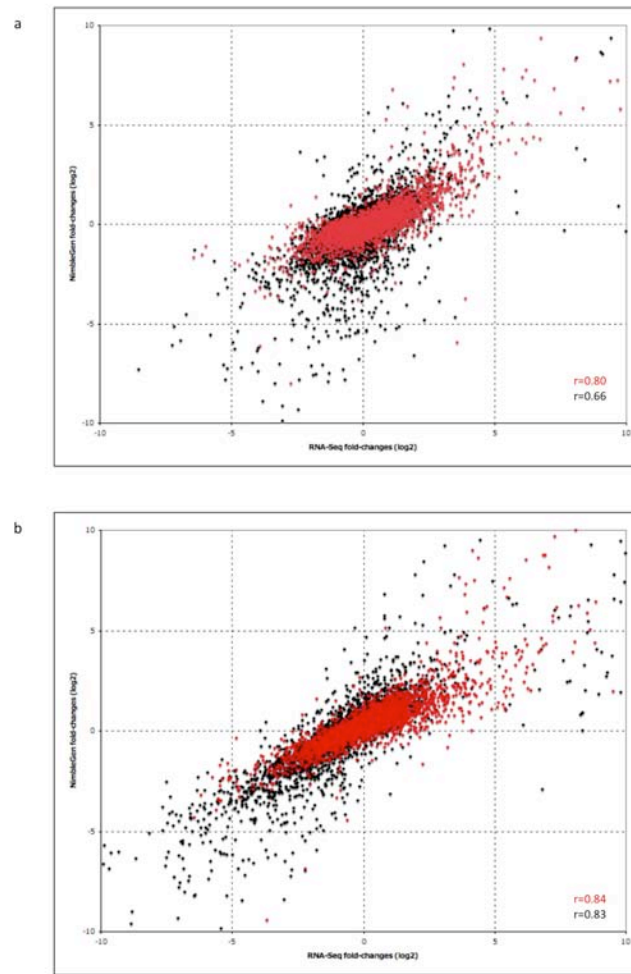


Figure 4 : Correlation between mRNA expression fold-changes measured by NimbleGen exon oligoarray and RNA-Seq for differentially-expressed transcripts in fruiting body and ectomycorrhiza of *T. melanosporum*/*C. avellana*. Only transcripts with ≥ 4 reads (RPKM) for ECM or FB were taken in account. Ratio for transcripts ≥ 100 reads (RPKM) are showed as red dots, whereas transcripts < 100 reads are depicted in black dots. Note the strong association between the two variables (Pearson's $r > 0.80$). A: ECM/FLM expression ratio. B: FB/FLM expression ratio.

The most abundant transcript constitutively expressed in saprobic mycelium, truffles and ectomycorrhizas coded for a small *Tuber*-specific protein with unknown function (GSTUMT00004096001). In the free-living mycelium, transcripts coding for endo-1,3-beta-glucanase, glyoxalase, elongation factor 1-alpha, glyceraldehyde-3-phosphate dehydrogenase and NADP-dependent mannitol dehydrogenase were amongst the most abundant. In truffles, the most highly expressed transcripts encoded 30 kDa heat shock protein, Cu-Zn superoxide dismutase, pyruvate decarboxylase and *Tuber*-specific proteins of unknown function. In ectomycorrhizal root tips, transcripts coding for polyubiquitin, protein translation factor SU11, chorismate mutase, histones and unknown proteins were the most abundant sequences.

We next compared differentially-expressed genes called from RNA-Seq with those identified by microarray analysis. Estimates of the $\log_2$ fold-change of gene expression levels across the two platforms were highly correlated (Pearson's $r = 0.66$ and 0.83) (Figure 4). As expected, the positive association was higher for the most highly expressed genes that were mapped by a large number of sequence reads, i.e. ≥ 100 reads (Pearson's $r = 0.80$ and 0.84), although the

strength of correlation did not depend much on the sequence coverage. While Illumina sequencing of pooled cDNA libraries was the only affordable option when this study was carried out and it did not affect the identification of differentially expressed transcripts. However, it is now both advisable and affordable to sequence individual cDNA samples.

By using the microarray, we detected 449 differentially-expressed transcripts (5.9% of the total gene repertoire) in ectomycorrhiza, including 219 and 230 up- and down-regulated transcripts (≥ 5 -fold, p -value <0.05), respectively. Using RNA-Seq, we found 739 differentially-expressed transcripts (9.8% of the total gene repertoire) in ectomycorrhiza, including 392 and 347 up- and down-regulated transcripts, respectively. In truffles, we found 280 and 453 up- and down-regulated transcripts (9.8% of the total gene repertoire) by microarrays, whereas 399 and 935 up- and down-regulated transcripts (17.8% of the total gene repertoire) were identified by RNA-Seq. The upregulated transcripts identified by both methods were 162 and 183 genes in ECM vs. FLM and FB vs. FLM, respectively. Most of the transcripts identified as upregulated by microarrays and not found by RNA-Seq showed expression scores near the background intensity value.

GO term enrichment analysis was used to examine upregulated genes (Tables 1 & 2). Genes upregulated in ectomycorrhizal root tips were mainly associated with amino sugar and cell wall polysaccharide metabolism (Table 1). During the symbiotic stage, *T. melanosporum* also activates genes associated with membrane transport, including ion and polyol translocation, and ion homeostasis. In addition, GO terms related to ribosome machinery and glutamate/oxoglutarate metabolism were also enriched in ectomycorrhizas. The GO term enrichment analysis confirmed that upregulated genes in truffles were different from those induced in ectomycorrhizas. They are mainly involved in sulfate assimilation and biosynthesis of sulfur-containing amino acids, isoprenoid biosynthesis and related glyoxylate cycle (Table 2). The most highly induced transcripts coded for *Tuber*-specific proteins.

The most highly upregulated gene encoded a phosphatidylserine decarboxylase, a phospholipid metabolism enzyme, followed by genes coding for proteins involved in metabolite translocation (major facilitator superfamily, amino acid and oligopeptide transporters), redox metabolism (cytochrome and oxygenase redox enzymes) and polysaccharide/protein degradation (cellulase, laccase and peptidase) (Table 3). In truffles, a majority of the secreted proteins are lineage-specific proteins of unknown function, whereas other proteins are coding for hydrolytic enzymes. However, genes coding for sugar and amino acid transporters, redox enzymes, together with methyltransferase, acyl-CoA synthetase, and malate synthase (Table 4) were also activated.

Table 1. GO terms significantly enriched in sets of *T. melanosporum* genes upregulated (≥ 5 fold) in ectomycorrhiza compared to free-living mycelium. Fisher's exact tests (FDR<0.05) were performed for each GO term.

GO id	GO term (Biological Process)	FDR
GO:0005975	carbohydrate metabolic process	1.12682E-4
GO:0006040	amino sugar metabolic process	0.00251444
GO:0046348	amino sugar catabolic process	8.46378E-4
GO:0006041	glucosamine metabolic process	0.00798669
GO:0006043	glucosamine catabolic process	8.46378E-4
GO:0006044	N-acetylglucosamine metabolic process	0.00798669
GO:0006046	N-acetylglucosamine catabolic process	8.46378E-4
GO:0005976	polysaccharide metabolic process	0.0122671
GO:0000272	polysaccharide catabolic process	0.0291912
GO:0006030	chitin metabolic process	0.00798669
GO:0006032	chitin catabolic process	8.46378E-4
GO:0044036	cell wall metabolic process	0.00798669
GO:0016998	cell wall catabolic process	8.46378E-4
GO:0009065	glutamine family amino acid catabolic process	0.0291912
GO:0006538	glutamate catabolic process	0.0291912
GO:0019551	glutamate catabolic process to 2-oxoglutarate	0.0291912
GO:0042180	cellular ketone metabolic process	0.0291912
GO:0043648	dicarboxylic acid metabolic process	0.0122671
GO:0006103	2-oxoglutarate metabolic process	0.0291912
GO:0006810	transport	7.4499E-5
GO:0006811	ion transport	0.0407213
GO:0015791	polyol transport	0.0291912
GO:0015798	myo-inositol transport	0.0291912
GO:0048878	chemical homeostasis	0.0266286
GO:0050801	ion homeostasis	0.0266286
GO:0055080	cation homeostasis	0.0266286
GO:0055069	zinc ion homeostasis	0.0291912
GO:0006882	cellular zinc ion homeostasis	0.0291912
GO:0051179	localization	1.34281E-4
GO:0051234	establishment of localization	7.80558E-5
GO:0033750	ribosome localization	0.0291912
GO:0033753	establishment of ribosome localization	0.0291912
GO:0000054	ribosome export from nucleus	0.0291912
GO:0000055	ribosomal large subunit export from nucleus	0.0291912
GO:0000746	conjugation	0.0117516
GO:0000747	conjugation with cellular fusion	0.0117516
GO:0009847	spore germination	0.0211108
GO:0007109	cytokinesis, completion of separation	0.0291912
GO:0006467	protein thiol-disulfide exchange	0.0291912
GO:0010033	response to organic substance	0.0326572

Table 2. GO terms significantly enriched in sets of *T. melanosporum* genes upregulated (≥ 5 fold) in fruiting body compared to free-living mycelium. Fisher's exact tests (FDR<0.05) were performed for each GO term.

GO id	GO term (Biological Process)	FDR
GO:0008152	metabolic process	1.87839E-6
GO:0055114	oxidation reduction	0.0390231
GO:0006519	cellular amino acid and derivative metabolic process	2.63502E-5
GO:0006520	amino acid metabolic process	1.7985E-5
GO:0009063	amino acid catabolic process	0.0482495
GO:0009069	serine family amino acid metabolic process	4.55829E-4
GO:0009070	serine family amino acid biosynthetic process	0.0117584
GO:0006544	glycine metabolic process	0.0482495
GO:0000096	sulfur amino acid metabolic process	9.57802E-4
GO:0000097	sulfur amino acid biosynthetic process	0.00131799
GO:0006534	cysteine metabolic process	0.0210605
GO:0019344	cysteine biosynthetic process	0.0117584
GO:0006555	methionine metabolic process	0.0210605
GO:0006807	nitrogen compound metabolic process	1.16443E-5
GO:0034641	cellular nitrogen compound metabolic process	7.33556E-6
GO:0044270	nitrogen compound catabolic process	0.0482495
GO:0009308	cellular amine metabolic process	2.63502E-5
GO:0009310	amine catabolic process	0.0482495
GO:0006790	sulfur metabolic process	6.60002E-4
GO:0044272	sulfur compound biosynthetic process	0.00358956
GO:0006791	sulfur utilization	6.60002E-4
GO:0000103	sulfate assimilation	9.57802E-4
GO:0005975	carbohydrate metabolic process	0.0262538
GO:0044262	cellular carbohydrate metabolic process	0.0260927
GO:0006000	fructose metabolic process	0.034373
GO:0006003	fructose 2,6-bisphosphate metabolic process	0.034373
GO:0006629	lipid metabolic process	0.0117584
GO:0044255	cellular lipid metabolic process	0.00172077
GO:0008610	lipid biosynthetic process	0.00999431
GO:0006720	isoprenoid metabolic process	5.13901E-5
GO:0008299	isoprenoid biosynthetic process	1.87839E-6
GO:0006082	organic acid metabolic process	1.7985E-5
GO:0019752	carboxylic acid metabolic process	5.13901E-5
GO:0046487	glyoxylate metabolic process	0.0184958
GO:0006097	glyoxylate cycle	0.0184958
GO:0006725	cellular aromatic compound metabolic process	0.0250295

Table 3. The most highly upregulated transcripts in *T. melanosporum*/*Corylus avellana* ECM root tips. Expression values in ECM were compared to FLM and FB expression values based on microarray and RNA-Seq profilings.

SEQ_ID	RNA-Seq			Array			Definition	Location
	ECM	FB	FLM	ECM	FB	FLM		
GSTUMT00000499001	1193	0.92	0.01	49524	634	77	Phosphatidylserine decarboxylase	-
GSTUMT000008973001	192	0.28	0.01	10866	2	2	Endoglucanase GH5	-
GSTUMT000006579001	1082	346	0.29	14241	8701	22	DUF1479-domain protein	M
GSTUMT000009894001	463	0.44	0.16	18205	2	2	Cytochrome P450	M
GSTUMT000006890001	140	51	0.11	5436	3787	2	Sporulation-induced protein	-
GSTUMT000005583001	1191	29	1	42051	1866	764	Major Facilitator Superfamily	-
GSTUMT00000122001	243	3	0.29	28848	1221	190	Dipeptidyl peptidase	S
GSTUMT000006891001	1467	363	2	36291	17771	252	Peroxidase/oxygenase	-
GSTUMT000008992001	166	0.33	0.37	19305	9	4	Laccase	S
GSTUMT000008954001	165	0.09	0.44	19493	35	18	MFS permease	-
GSTUMT00010075001	322	1	1	12626	223	223	hypothetical protein	-
GSTUMT000005760001	268	4	0.81	12966	989	9	Major Facilitator Superfamily (MFS) permease	S
GSTUMT000008628001	118	11	0.43	16903	3740	55	oligopeptide transporter protein	-
GSTUMT00010297001	723	2	4	35250	805	724	hypothetical protein	-

GSTUMT00004342001	204	1	1	20705	234	186	E1-E2 ATPase	-
GSTUMT00004342001	326	45	3	7848	1926	12	hypothetical protein	-
GSTUMT00001488001	624	393	6	30628	24511	1592	Amino acid permease	-
GSTUMT00003580001	286	0.01	3	5014	1	33	phospholipase A2	S
GSTUMT00003976001	533	0.3	6	32628	170	1568	Major intrinsic protein	-
GSTUMT00010098001	189	10	2	15146	1975	466	Amino acid permease	-
GSTUMT00000632001	421	14.8	6	28122	5975	1665	ATPase family (AAA)	-
GSTUMT00006630001	232	0.01	3	18170	3	110	Class II Aldolase	-
GSTUMT00008512001	155	0.18	2	8505	110	442	beta-carotene 15,15'-dioxygenase	-
GSTUMT00011897001	673	0.61	12	45752	424	3856	Domain of unknown function DUF1996	S
GSTUMT00004451001	102	0.39	2	9814	164	291	Homeobox domain	-

Transcript profiling was performed on free-living mycelium (FLM), fruiting body (FB) and ectomycorrhiza (ECM) using RNA-Seq or custom genome-wide exon oligoarrays. For RNA-Seq, a value of 0.01 was assigned to genes without read. For microarrays, signals below the cut-off values were assigned a signal intensity value of 1. Abbreviations: S, secreted; M, mitochondrial

Table 4. The most highly upregulated transcripts in *T. melanosporum* fruiting body. Expression values in FB were compared to FLM and ECM expression values based on microarray and RNA-Seq profilings.

SEQ_ID	RNA-Seq			Array			Definition	Location
	ECM	FB	FLM	ECM	FB	FLM		
GSTUMT00009588001	0.01	6426	0.01	Not on array	Not on array	Not on array	<i>Tuber</i> -specific protein	-
GSTUMT00007268001	1.9	3053	0.01	74	27526	448	<i>Tuber</i> -specific protein	-
GSTUMT00011091001	0.01	773	0.01	1	15694	106	<i>Tuber</i> -specific protein	M
GSTUMT00001784001	0.01	974	0.16	4	41047	3	<i>Tuber</i> -specific protein	-
GSTUMT00007269001	2.8	6234	2.0	169	34625	219	<i>Tuber</i> -specific protein	S
GSTUMT00006579001	1082	346	0.29	14241	8701	22	DUF1479-domain protein	M
GSTUMT00009589001	56	33138	44	11177	39128	10029	Methyltransferase	S
GSTUMT00010375001	34	491	1	7054	20089	234	Acyl-CoA synthetase	-
GSTUMT00002556001	0.9	218	0.48	11	10845	522	DHN1 [<i>Tuber borchii</i>]	-
GSTUMT00009161001	9	1116	3	5	2902	90	Amidohydrolase	-
GSTUMT00012200001	0.84	495	1	15	25314	440	Malate synthase	-
GSTUMT00009497001	40	450	1	8047	24157	322	Amino acid permease	-
GSTUMT00001878001	0.01	116	0.42	1	14686	14	hypothetical protein	S
GSTUMT00004061001	7	445	2	1658	17281	817	<i>Tuber</i> -specific protein	-

GSTUMT00004391001	19	2030	8	2950	248010	1650	Flavin-containing monooxygenase	M
GSTUMT00006891001	1477	363	2	36291	17771	252	Peroxidase/oxygenase	-
GSTUMT00006097001	5	282	2	105	13736	17	<i>Tuber</i> -specific small secreted protein	S
GSTUMT00006937001	48	402	3	10519	27863	645	Sugar (and other) transporter	S
GSTUMT00006936001	39	355	2	11275	27318	518	Sugar (and other) transporter	-
GSTUMT00012074001	8	1182	9	2510	27383	1368	Glycine cleavage system P-protein	-
GSTUMT00004141001	34	209	2	2669	13898	33	GMC oxidoreductase	-
GSTUMT00005006001	67	549	5	11706	18751	44	<i>Tuber</i> -specific secreted protein	S
GSTUMT00002382001	4	478	4	1586	23601	1491	hypothetical protein	-
GSTUMT00011282001	17	890	9	1327	22058	1482	Fumarylacetoacetate (FAA) hydrolase family	S
GSTUMT00001310001	27	1012	11	8464	33854	3487	Sugar (and other) transporter	S

Transcript profiling was performed on free-living mycelium (FLM), fruiting body (FB) and ectomycorrhiza (ECM) using RNA-Seq or custom genome-wide exon oligoarrays. For RNA-Seq, a value of 0.01 was assigned to genes without read. For microarrays, signals below the cut-off values were assigned a signal intensity value of 1. Abbreviations: S, secreted; M, mitochondrial

A large part of the ectomycorrhizal reads belongs to the host plant (~3,250,000), but the lack of a draft genome for *C. avellana* precluded their mapping. We therefore performed a BLASTN search against *P. trichocarpa* gene models; 625,000 reads showed a similarity to these protein-coding sequences. This corresponded to 14,800 plant genes using a ≥ 4 reads per bp coverage threshold. Metallothionein-like protein, cytosolic glutamine synthetase, ubiquitin, elongation factor 1- α , and calmodulin were the most prevalent transcripts in the plant mycorrhizal tissues.

DISCUSSION

In this study, we investigated the potential utility of Illumina RNA-Seq for transcriptome studies of ectomycorrhizal symbionts. With more than 17 million genome-mapped short reads from three major developmental stages of *T. melanosporum*, this is the most comprehensive dataset for a mycorrhizal symbiont to date. We evaluated the quality of cDNA sequences by comparing them with the reference genome sequence of this ascomycete (Martin *et al.*, 2010). The genomic distribution of the mapped reads according to current gene annotation shows that most reads are confined to exons of protein coding genes (65%), although we detected a significant number of reads in UTRs (19%) and intronic regions (3%) (Figure 1). Using a stringent, conservative threshold for expression, we detected a core set of ~7,050 expressed genes, indicating that ~94% of the gene repertoire was expressed; 30% being very abundant transcripts covered by >1,000 reads. Massive redundancy was observed for a small number of genes, i.e., 26% of the reads were derived from only 50 genes. A core set of 6,815 transcripts was expressed in all tissues analyzed and only ~10% of the transcripts were differentially expressed in a tissue-specific manner. We also detected a sizable fraction of reads in intergenic regions (13%), which correspond to potentially new coding transcripts.

RNA-Seq data, due to its digital characteristics, provides accurate information about gene expression levels as confirmed by a strong correlation (Pearson's $r = 0.85$) of RNA expression levels with those determined from our previous *T. melanosporum* oligoarray transcript profiling (Martin *et al.*, 2010) (Figure 2). Correlation of the differential expression ratio between developmental stages assessed by the two platforms using the same RNA samples was also good (Pearson's $r > 0.80$) (Figure 3), in spite of RNA-Seq appears to be more sensitive in detecting differentially-expressed genes than oligoarrays, as a result of its much wider dynamic range achievable by RNA-Seq (from 0.04 to 54,474 RPKM). RNA-Seq was thus used to confirm and extend our previous survey of changes in gene expression upon fruiting body and symbiosis developments in *T. melanosporum* (Martin *et al.*, 2010).

Although most genes are not affected by ECM establishment, this transcript profiling study confirmed that the development of the symbiosis between *T. melanosporum* and *C. avellana* encompasses a series of complex and overlapping developmental genetic processes in the colonizing mycelium. Genome-wide microarray analysis of upregulated genes in *T. melanosporum* and *L. bicolor* symbiotic tissues using microarray method has showed that several different families of membrane transporters are consistently upregulated during symbiosis (Martin *et al.*, 2008, 2010). The end result of their activity – provision of nutrients to the plant host – is a key factor in driving and maintaining symbiosis. The present RNA-Seq analysis confirmed that several genes encoding membrane transporters involved in phosphorus, amino acid and carbohydrate exchange were induced upon symbiosis, reflecting the intense metabolite fluxes occurring between the symbiotic partners (Tables 1 & 3).

The GO term enrichment analysis also identified alterations in amino-sugar and glutamate metabolism consistent with the nutritional shift resulting from the plant colonization (Table 1). Interestingly, Montanini *et al.* (2010) have identified and confirmed the nuclear activity of a glucose-responsive transcription factor (TF) (TmelAzf1), together with several nitrogen regulatory proteins (TmelAreA, TmelTamA, TmelAreB, TmelNirA-3, TmelArcA), that are induced during symbiosis. These findings suggest that these TFs may be part of the transcriptional control networks for metabolic pathway regulation in ectomycorrhiza. In addition, we have observed the up-regulation of genes related to plant cell wall polysaccharides and chitin catabolism likely to participate in the cell wall remodelling of the novel symbiotic interface. Cell wall deconstruction may be driven by differential expression of TFs, such as TmelXLNR, an activator of xylan degradation (Montanini *et al.*, 2010). RNA-Seq transcript profiling also revealed the induction of several redox enzymes in ectomycorrhizal root tips which may be used by the mycobiont to overcome the host defence response. The present transcriptome profile is consistent with findings in other ectomycorrhizal systems (Voiblet *et al.*, 2001; Peter *et al.*, 2003; Johansson *et al.*, 2004; Duplessis *et al.*, 2005; Le Quéré *et al.*, 2005; Heller *et al.*, 2008; Martin *et al.*, 2008).

Changes in gene expression are taking place during fruiting body development in *T. melanosporum*. Genes involved in lipid and carbohydrate metabolism, glyoxylate cycle and nitrogen compound catabolism are upregulated in fruiting bodies (Tables 2 & 4). This confirms the results obtained on *T. borchii* (Lacourt *et al.*, 2002; Gabella *et al.*, 2005) showing that the glyoxylate cycle and gluconeogenesis sustains the carbon flux needed for fruiting body development and spore maturation (*e.g.*, lipid and glycogen store formation and chitin synthesis). The positive regulator of gluconeogenesis, TmelAcuK, and the activators of fatty acid metabolism, TmelFarA and TmelFarB, which are induced in FB (Montanini *et al.*, 2010) could be among the key regulators controlling these processes. Finally, a striking induction of genes involved in sulfur, aromatic compound and amino acid metabolism, required for the synthesis of various flavor-related volatiles released by *T. melanosporum* (Martin *et al.*, 2010; Splivallo *et al.*, 2010), was observed. Alterations in sulfur metabolism may be regulated by TmelCbf1, an activator of sulfur metabolism which is overexpressed in FB (Montanini *et al.*, 2010).

We obtained comprehensive expression maps that cover three of the most critical developmental stages of *T. melanosporum*, saprobic mycelium, fruiting body and ectomycorrhiza. By providing an unbiased view of the transcriptome, RNA-Seq transcript profilings is paving the way for a better understanding of the molecular processes underlying the symbiosis and truffle fruiting body formation.

Acknowledgements

This work was supported by the European Commission within the Network of Excellence EVOLTREE (FP6-016322) and the Project Energypoplar (FP7-211917), INRA and Région Lorraine grants (to FM). This work was funded, in part, by Institut de Génomique, CEA, and Agence Nationale de la Recherche to the Génomoscope. ET is supported by a Ph.D. scholarship from the Région Lorraine. We thank Prof. Simone Ottonello and Paola Bonfante for numerous exciting discussions. We would like to thank Dr. Gérard Chevalier (INRA- Clermont-Ferrand, France) and Dr. Elisa Zampieri (University of Torino, Italy) for providing *T. melanosporum* mycelial and fruit body samples.

REFERENCE

- Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005.** Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674–3676.
- Denoeud F, Aury JM, Da Silva C, Noel B, Rogier O, Delledonne M, Morgante M, Valle G, Wincker P, Scarpelli C, Jaillon O, Artiguenave F. 2008.** Annotating genomes with massive-scale RNA sequencing. *Genome Biology* **9**: R175.
- 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.
- Gabella S, Abbà S, Duplessis S, Montanini B, Martin F, Bonfante P. 2005.** Transcript profiling reveals novel marker genes involved in fruiting body formation in *Tuber borchii*. *Eukaryotic Cell* **4**: 1599-602.
- Gowda M, Venu RC, Raghupathy MB, Nobuta K, Li H, Wing R, Stahlberg E, Coughlan S, Haudenschild CD, Dean R, Nahm BH, Meyers BC, Wang GL. 2006.** Deep and comparative analysis of the mycelium and appressorium transcriptomes of *Magnaporthe grisea* using MPSS, RL-SAGE, and oligoarray methods. *BMC Genomics* **7**: 310
- Heller G, Adomas A, Li G, Osborne J, van Zyl L, Sederoff R, Finlay RD, Stenlid J, Asiegbu FO. 2008.** Transcriptional analysis of *Pinus sylvestris* roots challenged with the ectomycorrhizal fungus *Laccaria bicolor*. *BMC Plant Biology* **8** :19
- Johansson T, Le Quéré A, Ahren D, Söderström B, Erlandsson R, Lundeborg J, Uhlen M, Tunlid A. 2004.** Transcriptional responses of *Paxillus involutus* and *Betula pendula* during formation of ectomycorrhizal root tissue. *Molecular Plant Microbe Interaction* **17**: 202-215.
- Lacourt I, Duplessis S, Abbà S, Bonfante P, Martin F. 2002.** Isolation and characterization of differentially expressed genes in the mycelium and fruit body of *Tuber borchii*. *Applied Environmental Microbiology* **68**: 4574-82.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009.** Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**: R25.
- Larsen PE, Trivedi G, Sreedasyam A, Lu V, Podila GK, Collart FR. 2010.** Using Deep RNA Sequencing for the Structural Annotation of the *Laccaria Bicolor* Mycorrhizal Transcriptome. *PloS ONE* **5** (7) e9780
- 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. *Molecular Plant Microbe Interaction* **18**: 659-673.
- Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. 2008.** RNA-seq: an assessment of

- technical reproducibility and comparison with gene expression arrays. *Genome Research* **18**: 1509–1517.
- Martin F, Aerts A, Ahren D, Brun A, Danchin EGJ, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V et al. 2008.** The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **452**: 88–92.
- Martin F, Kohler A, Murat C, Balestrini R, Coutinho PM, Jaillon O, Montanini B, Morin E, Noel B, Percudani R et al. 2010.** Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* **464**: 1033–1038.
- McGuire AM, Pearson MD, Neafsey DE, Galagan JE. 2008.** Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biology* **9**: R50.
- Mello A, Murat C, Bonfante P. 2006.** Truffles: much more than a prized and local fungal delicacy. *FEMS Microbiology Letters* **260**: 1–8.
- Mignone F, Gissi C, Liuni S, Pesole G. 2002.** Untranslated regions of mRNAs. *Genome Biology* **3**: 0004.
- Modrek B, Lee C. 2002.** A genomic view of alternative splicing. *Nature Genetics* **30**: 13–19.
- Montanini B, Levati E, Bolchi A, Kohler A, Morin E, Tisserant E, Martin F, Ottonello S. 2010.** Genome-wide search and functional identification of transcription factors in the mycorrhizal fungus *Tuber melanosporum*. *New Phytol*, in press.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008.** Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**: 621–628.
- Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. 2008.** The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**: 1344–1349.
- Peter M, Courty PE, Kohler A, Delaruelle C, Martin D, Tagu D, Frey-Klett P, Duplessis S, Chalot M, Podila G et al. 2003.** Analysis of expressed sequence tags from the ectomycorrhizal basidiomycetes *Laccaria bicolor* and *Pisolithus microcarpus*. *New Phytologist* **159**: 117–129.
- Rubini A, Belfiori B, Riccioni C, Tisserant E, Arcioni S, Martin F, Paolocci P. 2010.** Isolation and characterization of *MAT* genes in the symbiotic ascomycete *Tuber melanosporum*. *New Phytol*, in press.
- Splivallo R, Ottonello S, Mello A, Karlovsky P. 2010.** Truffle volatiles: from chemical ecology to aroma biosynthesis. *New Phytol*, in press.
- Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D, et al. 2008.** A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**: 956–960.
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010.** Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* **28**: 511–515.
- Vilela C, McCarthy JE. 2003.** Regulation of fungal gene expression via short open reading frames in the mRNA 5' untranslated region. *Molecular Microbiology* **49**: 859–67.
- 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 cDNAs. *Plant Journal* **25**: 181–191.
- Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008.** Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**: 470–476.
- Wang B, Guo G, Wang C, Lin Y, Wang X, Zhao M, Guo Y, He M, Zhang Y, Pan L. 2009.** Survey of the transcriptome of *Aspergillus oryzae* via massively parallel mRNA sequencing. *Nucleic Acids Research*, in press.
- Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bahler J. 2008.** Dynamic repertoire of a eukaryotic transcriptome surveyed at single-

nucleotide resolution. *Nature* **453**: 1239–1243.

Chapitre III : Analyse globale du transcriptome de *G. intraradices*

Chapitre III : Analyse globale du transcriptome de *Glomus intraradices*

La symbiose mycorhizienne à arbuscules (AM) entre les Gloméromycètes et les plantes implique plus des deux tiers de toutes les espèces végétales connues, y compris des espèces cultivées importantes. Cette symbiose mutualiste est sans doute la symbiose la plus importante d'un point de vue écologique et agricole dans les écosystèmes terrestres. Le séquençage des génomes des champignons ECM, et l'analyse à grande échelle de leurs profils d'expression symbiotiques a permis d'identifier plusieurs facteurs régulant le développement et la fonctionnement des ectomycorhizes. A l'heure actuelle, aucune séquence génomique n'est disponible pour le phylum des Gloméromycètes, et peu de choses sont connues sur l'évolution, l'organisation des génomes, la variation génétique et le mode de reproduction de ces champignons ancestraux. Alors que les gènes spécifiquement régulés lors du fonctionnement de la symbiose AM ont été identifiés chez le partenaire végétal par des profils d'expression à l'échelle du génome entier pour plusieurs plantes modèles cultivées, la régulation de l'expression génique chez les symbiotes fongiques a été jusqu'ici peu explorée. Ces études ont principalement été basées sur des approches ciblées offrant seulement un aperçu étroit de l'adaptation fongique au mode symbiotique. Afin d'obtenir une compréhension plus large de cette symbiose importante, le « Glomus Genome Consortium » a entrepris la caractérisation du transcriptome de *G. intraradices* par séquençage Sanger et 454. L'objectif de cette étude a été d'accéder à l'ensemble du contenu en gène du génome de *G. intraradices* DAOM197198 afin d'ébaucher son paysage transcriptomique. Nous avons également utilisées ces données de séquences transcrites pour construire une puce à ADN, qui nous a permis d'explorer les profils d'expression pour différents stades de développement de *G. intraradices*, afin d'identifier les bases moléculaires impliquées dans l'établissement de la symbiose mycorhizienne arbusculaire.

Le manuscrit intitulé « **The transcriptome of the arbuscular mycorrhizal fungus *Glomus intraradices* (DAOM 197198) reveals functional tradeoffs in an obligate symbiont** » décrit cette analyse du transcriptome de *G. intraradices*.

The transcriptome of the arbuscular mycorrhizal fungus *Glomus intraradices* (DAOM 197198) reveals functional tradeoffs in an obligate symbiont

E Tisserant, A Kohler, P Dozolme-Seddas, R Balestrini, K Benabdellah, A Colard, D Croll, C Da Silva, SK Gomez, R Koul, N Ferrol, V Fiorilli, D Formey, Ph Franken, N Helber, M Hijri, L Lanfranco, E Lindquist, Y Liu, M Malbreil, E Morin, J Poulain, H Shapiro, D van Tuinen, A Waschke, C Azcón-Aguilar, G Bécard, P Bonfante, MJ Harrison, H Küster, P Lammers, U Paszkowski, N Requena, SA Rensing, C Roux, IR Sanders, Y Shachar-Hill, G Tuskan, JPW Young, V Gianinazzi-Pearson et F Martin.

Article accepté dans la revue New Phytologist (Septembre 2011)

The transcriptome of the arbuscular mycorrhizal fungus *Glomus intraradices* (DAOM 197198) reveals functional tradeoffs in an obligate symbiont

E Tisserant^{1#}, A Kohler^{1#}, P Dozolme-Seddas²⁺, R Balestrini³, K Benabdellah⁴, A Colard⁵, D Croll⁵, C Da Silva⁶, SK Gomez⁷, R Koul⁸, N Ferrol⁴, V Fiorilli³, D Formey⁹, Ph Franken¹⁰, N Helber¹¹, M Hijri¹², L Lanfranco³, E Lindquist¹³, Y Liu², M Malbreil⁹, E Morin¹, J Poulain⁶, H Shapiro¹³, D van Tuinen², A Waschke¹⁰, C Azcón-Aguilar⁴, G Bécard⁹, Bonfante P³, MJ Harrison⁷, H Küster¹⁴, P Lammers⁸, U Paszkowski¹⁵, N Requena¹¹, SA Rensing¹⁶, C Roux⁹, IR Sanders⁵, Y Shachar-Hill¹⁷, G Tuskan¹⁸, JPW Young¹⁹, V Gianinazzi-Pearson² and F Martin^{1*}

¹ INRA, UMR1136 INRA/UHP, Interactions Arbres/ Micro-organismes, Centre de Nancy, 54280 Champenoux, France;

² UMR 1088 INRA/5184 CNRS/ Burgundy University Plante-Microbe-Environnement, INRA-CMSE, BP 86510, 21065 Dijon, France;

³ Istituto per la Protezione delle Piante del CNR, sez. di Torino and Dipartimento di Biologia Vegetale, Università degli Studi di Torino, Viale Mattioli, 25, 10125 Torino, Italy;

⁴ Departamento de Microbiología del Suelo y Sistemas Simbióticos, Estación Experimental del Zaidín, CSIC, C. Profesor Albareda, 1, 18008 Granada, Spain;

⁵ University of Lausanne, Department of Ecology and Evolution, Biophore Building, 1015 Lausanne, Switzerland & ETH Zürich, Plant Pathology, Universitätsstrasse 3, CH-8092 Zürich, Switzerland;

⁶ CEA, IG, Genoscope, 2 rue Gaston Crémieux CP5702, F-91057 Evry, France;

⁷ Boyce Thompson Institute for Plant Research, Tower Road Ithaca, New York 14853-1801 USA;

⁸ Department of Chemistry and Biochemistry, New Mexico State University, Department 3MLS, PO Box 3001, Las Cruces, NM, 88003–8001, USA;

⁹ Université de Toulouse & CNRS ; UPS ; UMR 5546, Laboratoire de Recherche en Sciences Végétales; BP 42617, F-31326, Castanet-Tolosan, France;

¹⁰ Leibniz-Institute of Vegetable and Ornamental Crops, Department of Plant Nutrition, Theodor-Echtermeyer-Weg 1, D-14979 Grossbeeren, Germany;

¹¹ University of Karlsruhe, Institute for Applied Biosciences, Heisenberg-Group 'Plant-Fungal-Interactions', Hertzstrasse 16, D-76187 Karlsruhe, Germany;

¹² Institut de la Recherche en Biologie Végétale, Département de sciences biologiques, Université de Montréal, 4101 Rue Sherbrooke est, Montréal, Que., Canada H1X 2B2;

¹³ Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA;

¹⁴ Institut für Pflanzengenetik, Naturwissenschaftliche Fakultät, Leibniz Universität Hannover, D-30419 Hannover, Germany;

¹⁵ Université de Lausanne, Department de Biologie Moléculaire Végétale, Biophore, 4419, CH-1015 Lausanne, Suisse;

¹⁶ BIOS Centre for Biological Signalling Studies, Freiburg Initiative for Systems Biology and Faculty of Biology, University of Freiburg, Hauptstr. 1, D-79104 Freiburg;

¹⁷ Michigan State University, Department of Plant Biology, East Lansing, MI 48824-1312, USA;

¹⁸ Oak Ridge National Laboratory, BioSciences, PO Box 2008, Oak Ridge, TN 37831, USA;

¹⁹ University of York, Department of Biology, York, YO10 5DD, UK.

+ Present address: UR1264 Mycologie et Sécurité des Aliments (MycSA) – QUALIS, INRA, Centre Bordeaux – Aquitaine, B.P. 81, 33883 Villenave d'Ornon Cedex

* Author for correspondence: Francis Martin. E-mail: fmartin@nancy.inra.fr, Tel: +33 383 39 40 80

Joint first author with an equal contribution to the work.

Summary

- The arbuscular mycorrhizal (AM) symbiosis is arguably the most ecologically important eukaryotic symbiosis, yet it is poorly understood at the molecular level. To provide novel insights into the molecular basis of the symbiosis-associated traits, we report the first genome-wide analysis of the transcriptome from *Glomus intraradices* DAOM 197198.
- We generated a set of 25,906 nonredundant virtual transcripts (NRVTs) transcribed in germinated spores, extraradical mycelium (ERM) and symbiotic roots using Sanger- and 454-sequencing. NRVTs were used to construct an oligoarray for investigating gene expression.
- We identified transcripts coding for the meiotic recombination machinery, as well as meiosis-specific proteins, suggesting that the lack of a known sexual cycle in *G. intraradices* is not due to major deletions of genes essential for sexual reproduction and meiosis. Induced expression of genes encoding membrane transporters and small secreted proteins (SSPs) in intraradical mycelium, together with the lack of expression of hydrolytic enzymes acting on plant cell wall polysaccharides, are all features of *G. intraradices* that are shared with ectomycorrhizal symbionts and obligate biotrophic pathogens.
- Our results illuminate the genetic basis of symbiosis-related traits of the most ancient lineage of plant biotrophs, advancing future research on these agriculturally and ecologically important symbionts.

Key words: Glomeromycota, *Glomus*, mycorrhiza, meiosis and recombination genes, small secreted proteins, symbiosis, transcript profiling

Short title: *Glomus intraradices* transcriptome

INTRODUCTION

The arbuscular mycorrhizal (AM) symbiosis between fungi in the Glomeromycota (Schüssler *et al.*, 2001) and plants involves over two thirds of all known plant species, including important crop species, such as wheat, rice, maize, soybean and poplar. This mutualistic symbiosis, involving one of the oldest fungal lineages, is arguably the most ecologically and agriculturally important symbiosis in terrestrial ecosystems (Fitter *et al.*, 2011). The extraradical mycelium (ERM) of the symbiont acts as an extension of the root system and increases the uptake of key nutrients, particularly phosphorus (P) and zinc (Zn) and possibly also nitrogen (N) (Smith & Smith, 2011). As such, they are crucial to plant growth (Smith & Read, 2008) and also define the diversity of plants in ecosystems (Van der Heijden *et al.*, 1998). Furthermore, since the colonization of plants by AM fungi can also result in a 20% net increase in photosynthesis (Smith & Read, 2008), these universal mycosymbionts make a very large, poorly understood contribution to the global carbon cycling budget of ecosystems.

The Glomeromycota are unique in that their spores and coenocytic hyphae contain multiple nuclei in a common cytoplasm. No sexual cycle is known, although anastomosis and nuclear movement between hyphae of the same species have been described (Giovannetti *et al.*, 2001), as well as genetic exchange between AM fungal individuals (Croll *et al.*, 2008, 2009). Furthermore, the concept of an individual is unclear as nuclei within a single AM fungus appear to be genetically different in some species, which raises substantial questions about natural selection, population genetics and gene expression of these highly unusual organisms (Jany and Pawloska, 2010; Sanders and Croll, 2010). AM fungi are thought to grow clonally, but controversy clouds their ploidy, their genome size, and whether or not they can reproduce sexually (for review, see Sanders and Croll, 2010). Evidence for recombination has been provided, but whether mating and meiosis are involved is unknown (Croll & Sanders, 2009). There is therefore a need to identify genes whose products are required for proper completion of meiotic recombination.

Root colonization by AM fungi follows a series of distinct steps (for review, see Bonfante & Genre, 2010) starting with a presymbiotic molecular dialogue that involves root-released strigolactones (Akiyama *et al.*, 2005) and AM fungal signaling molecules, such as lipochitooligosaccharides (Maillet *et al.*, 2011). AM fungi are obligate mutualistic symbionts that can only grow for a limited time without colonizing a susceptible host root (Bécard *et al.*, 2004). This has been suggested to be a consequence of some nutritional deficiencies and loss of metabolic pathways, such as an absence of *de novo* fatty acid synthesis (Bago *et al.*, 2000), during the asymbiotic and presymbiotic phases related to putative genome erosion (Ercolin & Reinhardt, 2011). After spore germination, the first contact at the root surface is marked by the differentiation of the fungal hyphae into an appressorium from which a penetration hypha invades the root epidermis (Genre *et al.*, 2005). After the fungus has traversed the outer cell layers it spreads in the inner cortex, i.e. intraradical mycelium (IRM), and forms highly branched structures inside cortical cells, so-called arbuscules. Arbuscules are the main site of nutrient transfer from the fungus to the plant (Javot *et al.*, 2007; Bonfante & Genre, 2010; Smith & Smith, 2011). Arising from the colonized roots, the extraradical mycelium (ERM) proliferates in the growth medium where it absorbs and assimilates the available nutrients before their transfer to the host plant. Major transcriptome shifts are thus expected during these different key developmental stages, but little is known of the repertoire of effector-like proteins, membrane transporters and assimilative enzymes involved in these different steps of symbiosis development and functioning.

The ability to establish a sophisticated zone of interaction, like the haustorium in pathogenic oomycetes and fungi interacting with plants, requires sophisticated host defence suppression (Dodds and Rathjen, 2010), which is predominantly achieved via secreted proteins delivered into the host cell (Kamoun, 2007). The protein SP7 of *G. intraradices* is secreted and transferred to the plant cell nucleus in colonized roots of *Medicago truncatula* where it binds to the host ethylene-responsive transcriptional factor regulating the expression of several defence-related genes (Kloppholz *et al.*,

2011). It remains to be determined whether *G. intraradices* expresses genes coding for additional secreted effector-like proteins during its interaction with the host plant.

The mycorrhizal colonization leads to quantitative and qualitative changes in the host transcriptome. Plant genes that are specifically regulated during the establishment and/or functioning of the AM symbiosis have been identified by both targeted and high-throughput expression profiling in several sequenced model and crop plants (Güimil *et al.* 2005; Balestrini & Lanfranco, 2006; Guether *et al.*, 2009a; Gomez *et al.*, 2009; Grunwald *et al.*, 2009). In contrast, regulation of gene expression in the AM fungal symbionts caused by interactions with roots has so far been poorly explored, and studies have mainly been based on targeted approaches providing only narrow insights into fungal adaptation to the symbiotic mode (Seddas *et al.*, 2009; Gomez *et al.*, 2009; Tani *et al.*, 2009; Kuznetsova *et al.*, 2010). Most studies have focused on P transport and metabolism (Viereck *et al.*, 2004; Balestrini *et al.*, 2007; Javot *et al.*, 2007; Benedetto *et al.*, 2005; Gómez-Ariza *et al.*, 2009; Grunwald *et al.*, 2009) and primary carbon metabolism (Bago *et al.*, 2000, 2003; Schaarschmidt *et al.*, 2006). Further studies adopting a microarray or high-throughput sequencing approach are highly warranted, because they could give a wider understanding of how genetic changes in the plant affect overall patterns of fungal gene expression, as well as of the implication of fungal genetics for the symbiotic effects on plants (Angelard *et al.*, 2010).

Biotrophs are widely accepted to intimately interact and co-evolve with their hosts. The genetic changes that brought about the evolution of obligate biotrophy in the Glomeromycota are, however, unknown. Recent research on obligate biotrophic pathogens, such as the powdery mildew fungus *Blumeria graminis* (Spanu *et al.*, 2010), the downy mildew oomycete *Hyaloperonospora arabidopsidis* (Baxter *et al.*, 2010), the oomycetous white rust *Albugo laibachii* (Kemen *et al.*, 2011), and the poplar and stem rusts, *Melampsora larici-populina* and *Puccinia graminis* (Duplessis *et al.*, 2011), reveals a close correlation between the biotrophic life style and gene losses in primary and secondary metabolism. In the oomycetes, all haustorium-forming species have lost the thiamine biosynthetic pathway, suggesting that haustorial oomycetes obtain thiamine from the host (Kemen *et al.*, 2011). It appears that biotrophy in these obligate plant pathogens is also associated with a convergent loss of secondary metabolic enzymes and an extremely reduced set of carbohydrate active enzymes devoted to plant cell wall depolymerization. Genes for nitrate and nitrite reductases, a nitrate transporter, and sulfite reductase are also often missing. This streamlining of metabolism presumably reflects adaptation to life within host cells, as it is not been observed in non-obligate biotrophic pathogens. Whether similar events have also characterized AM fungal evolution remains to be determined.

As mentioned above, transcriptome studies have been pursued on AM roots in multiple plant species under a variety of experimental conditions, but they have mainly focused on the host plant transcriptome. No single study has yet brought together genome-wide transcriptomic data for the fungal component of this important symbiosis. The aim of this study was to establish a comprehensive, genome-wide inventory of gene expression in *G. intraradices* DAOM197198 by sequencing cDNA libraries from different fungal structures (germinated spores, extra- and intraradical mycelium, arbuscules). Through transcriptomic analyses, we also wished to identify which of these genes were induced upon symbiosis and, further, to assign the differentially-expressed transcripts to key symbiotic mechanisms, such as nutrient transport and assimilation, host colonization, and signaling pathways. Several reasons have led to the choice of *G. intraradices* for the first large-scale transcriptome sequencing of an AM fungus. As a symbiont, *G. intraradices* is highly effective in mobilizing, taking up and transferring mineral nutrients, such as inorganic orthophosphate ions (Pi), N and sulfur (S), from soils to plants (Govindarajulu *et al.*, 2005; Allen & Shachar-Hill, 2009; Smith & Smith, 2011; Tian *et al.*, 2010) and it readily colonizes many plants including agriculturally important crop species (e.g., alfalfa, poplar, rice, wheat) as well as model plants such as *Medicago truncatula* and *Lotus japonicus*. *G. intraradices* is one of the most studied AM fungi as it rapidly colonizes its host plants and it is a model species for dissecting the molecular

and cellular biology of the Glomeromycota (Seddas *et al.*, 2009; Sanders & Croll, 2010). It is easily amenable to *in vitro* culture on transformed host roots (Chabot *et al.*, 1992) and is the only species for which spores are available commercially in pure form in large quantities.

The comparative analyses of gene repertoires in *G. intraradices*, pathogenic obligate biotrophs and ectomycorrhizal fungi offer insights into genes that may be involved in obligate biotrophy and mycorrhizal symbioses. In the absence of whole-genome sequence of any member of the phylum Glomeromycota (Martin *et al.*, 2008a), the availability of large-scale expressed sequenced tags (EST) collections therefore represents the core foundation for understanding genome functionality in the Glomeromycota.

MATERIALS & METHODS

Biological material

Glomus intraradices Schenck & Smith DAOM 197198 (recently reassigned to *Rhizophagus irregularis* (Błaszk., Wubet, Renker & Buscot) C. Walker & A. Schüßler comb. nov., see Stockinger *et al.*, 2009) was produced in monoxenic cultures maintained on *Agrobacterium rhizogenes*-transformed carrot (clone DC2) roots (Bécard & Fortin, 1988). Germinated spores were produced as described by Chabot *et al.* (1992). Biological materials used for cDNA library construction and microarray transcript profiling are summarized in Table 1. Protocols for producing the biological materials, RNA purification and cDNAs are described in the online Supporting Information S1. For array profiling, *G. intraradices* ERM was grown on liquid M medium without sucrose for three weeks before harvesting. *G. intraradices*- and mock-inoculated roots of *M. truncatula* were prepared as described for cDNA library AKNA (Supporting Information S1). Germinated spores were sampled as described for the cDNA library AKND (Supporting Information S1) (Balestrini *et al.*, 2007). Samples were snap frozen in liquid N₂.

EST sequencing, filtering and assembly

Sanger sequencing was performed on ABI3730xl analyzers (Applied Biosystems, Nutley, NJ, USA) and pyrosequencing on the Genome Sequencer FLX System (454 Life Sciences/Roche, Nutley, NJ, USA). Reads (797,394) were filtered and trimmed for low quality, low complexity, and adaptor sequences using SeqClean (TIGR, http://137.111.92.51/resources_programs/SeqClean.html) (Fig. S1). Sequences ≤ 100 bp were discarded. The resulting high-quality sequences were screened to detect potential bacterial and fungal contaminant sequences, and plant sequences present in cDNAs from symbiotic tissues. Sequences having a GC% higher than *G. intraradices* (i.e., $\geq 45\%$) were removed (Fig. S2). Sequences with a high nucleotide sequence similarity (BLASTN e-value $\leq 1e-5$, score ≥ 150 , % identity ≥ 95) with identified contaminants, *Aspergillus* spp. and *Chromobacterium violaceum*, as well as *M. truncatula* genome and transcripts, and other plant DNA (e.g., Carrot) were also removed (Fig. S1). Finally, to avoid the loss of *G. intraradices* outlier sequences, removed sequences were aligned to the *G. intraradices* draft genome using BLASTN (e-value $\leq 1e-5$, score ≥ 150) at the INRA GlomusDB web site (<http://mycor.nancy.inra.fr/IMGC/GlomusGenome/index3.html>) and retained when sequence similarity was $\geq 95\%$. Filtered reads were then assembled using the MIRA assembler (Chevreux *et al.* 2004). To choose the best assembly, several assemblies were constructed using different parameters and several indexes were measured including number of nonredundant NRVs, percentage of assembled reads and percentage of NRVs with hits against gene models from the taxonomically-related, although distant, Mucoromycotina *Rhizopus oryzae* (<http://genome.jgi-psf.org/Rhior3/Rhior3.home.html>) (Table S1). The assembly generated by MiraSearchESTSNPs was selected. The number of detected NRVs increased with the number of reads and the rarefaction curve did not reach a plateau (Fig. S3A). However, many of these sequences may be the result of

DNA pyrosequencing errors creating false transcripts. This is acute for the AT-rich sequences of *G. intraradices*. In contrast, the number of protein sequences documented in Swiss-Prot DNA database (<http://www.ebi.ac.uk/uniprot/>) and *R. oryzae* gene repertoire showing a similarity to *G. intraradices* NRVTs plateaued around 5,500 homologs (Fig. S3B), suggesting that most of the transcriptome of *G. intraradices* was covered. This was confirmed by a TBLASTN search (cutoff e-value of $1e-5$) using the protein sequences of the Core Eukaryotic Genes (CEG) (Parra *et al.*, 2009); 245 (98.7%) of the 248 CEG proteins were found in the current *G. intraradices* transcriptome. The missing protein sequences were a metal-binding protein, a spindle assembly checkpoint protein and a monooxygenase involved in coenzyme Q (ubiquinone) biosynthesis. Eighty-six % of NRVTs matched the current *G. intraradices* 52.5 Mb-genome assembly (Martin *et al.*, 2008a) confirming that most filtered, high-quality ESTs were from *G. intraradices*.

Single-nucleotide polymorphisms (SNPs) in NRVTs were identified with MiraSearchESTSNPs using the following parameters: minimum reads, 2 (Sanger) or 4 (454); minimum quality value, 30 (Sanger) or 25 (454); minimum neighbour quality value, 20.

For the digital Northern gene expression, the number of reads for each NRVT in each library was counted and the relative frequencies (reads of a given NRVT divided by the total number of reads) was obtained. Significant differences in gene expression between libraries for each NRVT were calculated (Audic & Claverie, 1997).

EST sequences are available at the National Center for Biotechnology Information (NCBI) (accession numbers GW091323-GW125581, GW086621-GW090678). A MySQL/PHP database compiling NRVT sequences and annotations is available at the INRA GlomusDB web site. This server can also be used to query the ESTs and the draft genome assembly using BLAST programs.

Functional annotation

NRVTs were compared against Swiss-Prot using BLASTX with a significance threshold (e-value $<1e-5$). Gene annotations were assigned to each NRVT based on the best BLAST hits. Sequences were searched against the *G. intraradices* draft genome contigs (version 'test14') at the INRA GlomusDB using BLASTN (e-value $\leq 1e-5$). NRVTs were also compared to proteins of Basidiomycota (*Cryptococcus neoformans*, *L. bicolor*, *M. larici-populina*, *Phanerochaete chrysosporium*, *Ustilago maydis*), Ascomycota (*A. nidulans*, *B. graminis*, *Botrytis cinerea*, *Magnaporthe grisea*, *Neurospora crassa*, *T. melanosporum*) and Mucoromycotina (*R. oryzae*, *M. circinelloides*) using BLASTX (e-value $\leq 1e-5$). To identify conserved protein domains in NRVTs, predicted protein sequences were compared to the KOG (Eukaryotic Orthologous Groups) database (Tatusov *et al.*, 2003). The counts of each KOG domain by species were transformed into a z-scores matrix to center the data, and visualization was performed using MeV (Saeed *et al.*, 2006). The Gene Ontology (GO) terms (Ashburner *et al.*, 2000) were assigned to each sequence using BLAST2GO (Conesa *et al.*, 2005). Enrichment analysis of GO annotations was carried out with BINGO using the hypergeometric test and Benjamini & Hochberg's false discovery rate (FDR) multiple testing correction ($p \leq 0.05$) (Maere *et al.*, 2005).

To predict *G. intraradices* metabolic pathways, sequences were queried for KEGG (Kyoto Encyclopedia of Genes and Genomes) Orthology (KO) assignments (Kanehisa *et al.*, 2000) using KAAS (KEGG Automatic Annotation Server) (Moriya *et al.*, 2007) and results were then plotted into the KEGG Global metabolic atlas (map01100) using iPATH (Letunic *et al.*, 2008). A high resolution version of this metabolic pathway map can be downloaded at: http://mycor.nancy.inra.fr/IMGC/GlomusGenome/download/Gi_metabolic_map.pdf.

To identify sequences encoding secreted proteins, NRVTs were translated into protein using FrameFinder (<http://www.ebi.ac.uk/~guy/estate/>). *In silico* predictions of secreted proteins were then carried out using SignalP 3.0 (Bendtsen *et al.*, 2004), TargetP 1.1 (Emanuelsson *et al.*, 2000) and TMHMM 2.0 (prediction of transmembrane helices in proteins) (Krogh *et al.*, 2001) as previously reported in Duplessis *et al.* (2011). Small secreted proteins (SSPs) were selected based on an arbitrary cut-off of 300 amino acids (Duplessis *et al.*, 2011).

Oligoarray design

For the design of custom oligoarrays, a MIRA assembly (v. 1.0) comprising 22,254 NRVTs was constructed using cDNA libraries CACE, CCHU, EXTA and EXTB. At the time, sequences from the AKNA, AKNB, AKNC and AKND libraries were not available. To ensure that all sequences were sufficiently long to allow the design of high-quality probes, we selected the subset of sequences ≥ 200 bp in length ($n=19,465$). To remove redundancy among sequences that would produce cross-hybridization, the assembled NRVTs were clustered based on nucleotide sequence similarity using BLASTClust (<http://www.ncbi.nlm.nih.gov/Web/Newsltr/Spring04/blastlab.html>). Within each cluster, sequences were ranked by length, and the longer member of each cluster was selected ($n=15,932$). We also selected singleton sequences with matches in NCBI databases and with length ≥ 200 bp ($n=2,046$). We have then added 2,327 NRVTs generated by the Paracel Transcript Assembler (Striking Development, <http://www.paracel.com/>) (S Rensing, unpublished results) which are not found in the MIRA NRVt set. We also included 12 *G. intraradices* reference sequences downloaded from the NCBI, 23 *G. intraradices* mitochondrial sequences (C Roux, unpublished results) and 70 *G. intraradices* sequences generated from the manual annotation of the *G. intraradices* genome and not found in the NRVTs. Finally, 4,471 remaining singleton Sanger sequences without matches in public databases were added. For 2,477 sequences of this set, no probe could be designed probably due to their relative high A+T content, which resulted in a higher-than-usual prevalence of homopolymer runs and other forms of low-complexity sequences. This procedure produced a total set of 22,404 sequences containing 14,828 MIRA NRVTs, 1,976 PARACEL NRVTs, 5,496 singletons and 104 additional sequences. This set did not include NRVTs from symbiotic tissues not available when the custom array was constructed. However, ~80% of the NRVTs used to construct the oligoarray were found by BLASTN in the 25,906 MiraSearchESTSNPs NRVTs.

The *G. intraradices* expression array (4 x 72K) manufactured by Roche NimbleGen Systems Limited (Madison, WI) (<http://www.nimblegen.com/products/exp/index.html>) contained three independent, non identical, 60-mer probes per sequence. Included in the oligoarray were *G. intraradices* sequences, and 5,785 random 60-mer control probes.

Transcript profiling

Whereas biological samples used for cDNA sequencing were produced by several different collaborators, biological materials used for microarray transcript profiling were generated in triplicate in a small number of laboratories: spores at the Laboratoire de Recherche en Sciences Végétales (Université de Toulouse, France), *M. truncatula* ERM at the Departamento de Microbiología del Suelo y Sistemas Simbióticos (Estación Experimental del Zaidín, CSIC, Granada, Spain) and *M. truncatula* IRM at the Boyce Thompson Institute for Plant Research (Tower Road Ithaca, New York, USA). cDNA synthesis was carried out at INRA-Nancy.

RNA was extracted from mycorrhizal roots with Trizol reagent (Invitrogen Corporation, Carlsbad, CA) with additional phenol:chloroform (1:1, v/v) purification steps, from ERM and spores using the RNeasy Plant Mini Kit (Qiagen) with RLT lysis buffer. RNA from mycorrhizal roots was treated with Turbo DNase I (Ambion Inc., Austin, TX) and purified with the RNeasy MinElute CleanUp kit (Qiagen). DNase treatment of RNA from ERM was performed using RNase-free

DNase (Qiagen). RNA quality was checked prior to cDNA synthesis using the Bio-Rad Experion analyzer. RNA preparations (three biological replicates) were amplified using the SMART PCR cDNA Synthesis Kit (Clontech) according to the manufacturer's instructions. Single dye labeling of cDNA samples, hybridization procedures and data acquisition, were performed at the NimbleGen facilities (NimbleGen Systems, Reykjavik, Iceland) following their standard protocol. As ERM and IRM materials were not produced in strictly similar conditions, transcript profiles of ERM and IRM should be compared with caution.

Average expression levels were calculated for each gene from the independent probes on the array and were used for further analysis. Raw array data were normalized by the RMA (Robust Multiarray Average) routine using the ARRAYSTAR software (DNASTAR Inc. Madison, WI, USA). A transcript was deemed expressed when its signal intensity was three-fold higher than the mean signal-to-noise threshold (cut-off value) of the random oligonucleotide probes present on the array (50 to 100 arbitrary units). The maximum signal intensity values for the most abundant transcripts were ~65,000 arbitrary units. A control hybridization was performed with cDNA from uninfected *M. truncatula* roots to evaluate any plant RNA cross-hybridization with *G. intraradices* sequences and to detect any plant-derived probes on the array. The later probe signals were discarded. A Student *t*-test with FDR correction (Benjamini-Hochberg) was applied to the data using ARRAYSTAR. Transcripts with a significant *p*-value (<0.05) and $\geq$ five-fold change in transcript level were considered to be differentially expressed. The complete expression dataset is available as series (accession number GSE29866) at the Gene Expression Omnibus at NCBI (<http://www.ncbi.nlm.nih.gov/geo/>).

Validation of oligoarray data

To obtain the biological material for the PCR validation of oligoarray data, 6,000 arbuscule-containing cells from *M. truncatula* roots (2,000 cells for each of three biological replicates) were microdissected using a laser microdissection system (LMD) according to Balestrini *et al.* (2007). ERM from monoxenic *M. truncatula* cultures and about 6,000 germinated spores were sampled, snap frozen in liquid N₂ and stored at -80 °C until used. Total RNA was extracted using the RNeasy Microarray Tissue Mini Kit (Qiagen, Courtaboeuf, France) following the manufacturer's instructions. All reverse transcriptase (RT)-PCR assays were carried out using the One Step RT-PCR kit (Qiagen). DNA contamination in RNA samples was evaluated using *GintEF1 α* specific primers for *G. intraradices*. To test the amount of extracted RNA, a semi-quantitative RT-PCR using *GintEF1 α* specific primers was performed. Reactions were carried out in a final volume of 25 μ l. Amplification reactions were run for 40 cycles of 94°C for 30 s, 58°C for 30 s, and 72°C for 30 s, and an aliquot of the PCR reaction was taken after 35, 37 and 40 cycles.

RESULTS & DISCUSSION

An obligate biotroph with a large gene repertoire

The cDNA libraries of *G. intraradices* listed in Table 1 were sequenced to assess the size and diversity of the fungus gene repertoire and explore the responses of the AM fungus at key developmental stages, germinated spores, ERM and IRM. After sequencing, the *de-novo* hybrid clustering of the 437,411 filtered reads produced 25,906 NRVTs (Table S1); 87% of the reads were in the assembly corresponding to 20.3 Mbp of nonredundant sequences. Only 10,823 NRVTs (41.8%) matched database sequences (cutoff *e*-value of 1e-5) in BLAST searches against genes from fungi other than Glomeromycota (Fig. S4). Homologous NRVTs corresponded to 8,773 unique known genes in the DNA databases. Among the predicted peptide sequences, 9,035 and 9,265 showed significant sequence similarity to proteins from *R. oryzae* and *Mucor circinelloides* respectively, distantly related fungi in the order Mucorales (Fig. S4). NRVTs that did not match any

of the known genes still gave highly relevant hits against the *G. intraradices* draft genome, suggesting that these represent novel orphan genes. As many *G. intraradices* transcripts are represented by two NRVTs (i.e., aligning to the 5'- or the 3'-end of the transcript) with a NRVt/gene ratio of 1.4, the number of nonredundant expressed transcripts in *G. intraradices* was ~18,500 (25,906/1.4).

Table 1: Tissues used in the present study of *Glomus intraradices* transcriptome. GR24: strigolactone GR24; API: apigenin; ERM: extraradical mycelium; IRM: intraradical mycelium; Arbuscules: microdissected arbuscule-containing cells. cDNA library identifications: AKNA, *M. truncatula* IRM; AKNB, germinated spores incubated in apigenin for 24 h; AKNC, germinated spores incubated in the strigolactone GR24 for 24 h; AKND, laser microdissected *M. truncatula* arbusculed cells; CACE, *M. truncatula* ERM; CCHU, germinated spores; EXTA, germinated spore mix; EXTB, *M. truncatula* ERM.

Samples	Sanger sequencing	454 sequencing	Array profiling
Germinated spores	CCHU		3 replicates
Germinated spores mix = Spores + spores in GR24 + spores in API		EXTA	
Germinated spores in GR24		AKNC	
Germinated spores in API		AKNB	
<i>Daucus carota</i> ERM	CACE	EXTB	
<i>Medicago truncatula</i> ERM			3 replicates
<i>M. truncatula</i> IRM		AKNA*	3 replicates
Laser microdissected <i>Medicago</i> arbusculed cells		AKND	

* cDNA library AKNA has been constructed using mRNA from a mix of the three replicates also used for microarray transcript profiling.

Based on BLASTClust analysis, the number of NRVTs found in multigene families was low with 120 (90% sequence identity, 80% sequence coverage) to 500 (90% sequence identity, 50% sequence coverage) NRVTs, most of gene families having only two members.

Although the vast majority of genes encoding enzymes of primary metabolism are retained in obligate biotrophic pathogens (downy and powdery mildews, white rust, poplar and stem rusts), notable exceptions include anaerobic fermentation, biosynthesis of glycerol from glycolytic intermediates, biosynthesis of thiamine, and nitrate and sulfate assimilation (Baxter *et al.*, 2010; Spanu *et al.*, 2010; Duplessis *et al.*, 2011; Kemen *et al.*, 2011). Out of the 25,906 *G. intraradices* NRVTs, 5,296 (20.4%) had significant matches in the KEGG pathway database and were assigned to 175 KEGG pathways. Most KEGG enzymes were mapped to NRVTs (see the metabolic pathway map online at http://mycor.nancy.inra.fr/IMG/GlomusGenome/download/Gi_metabolic_map.pdf), indicating the occurrence of these active metabolic processes in *G. intraradices* mycelium. In contrast to oomycete and fungal biotrophic pathogens, transcripts coding for nitrate and nitrite reductases, nitrate transporter, and sulfite reductase, were highly expressed in *G. intraradices*. On the other hand, *G. intraradices* has lost the thiamine biosynthetic pathway (Table S2) as have most haustorium-forming species (Kemen *et al.*, 2011). We infer that *G. intraradices* obtains thiamine from the host. There is no invertase transcript in the *G. intraradices* transcriptome (Table S2), implying that this fungus depends on the plant to both provide and hydrolyse sucrose, with the glucose moiety then preferentially transferred to the mycobiont. This is consistent with earlier observations (Schaarschmidt *et al.*, 2006). Additional genes missing from biotrophic pathogens (Spanu *et al.*, 2010), such as those coding for the allantoin permease DAL4p, the uracil permease FUR4p, and several enzymes involved in the aromatic amino acid metabolism and detoxification, were also missing from *G. intraradices* (Table S2). Although some metabolic pathways are missing in *G. intraradices*, we hypothesize that evolution to biotrophy in *G. intraradices* was not initiated by massive loss of metabolic complexity as observed in obligate biotrophic pathogens (Spanu *et al.*,

2010). The mycobiont retains the ability to take up and assimilate nutrients from its soil environment.

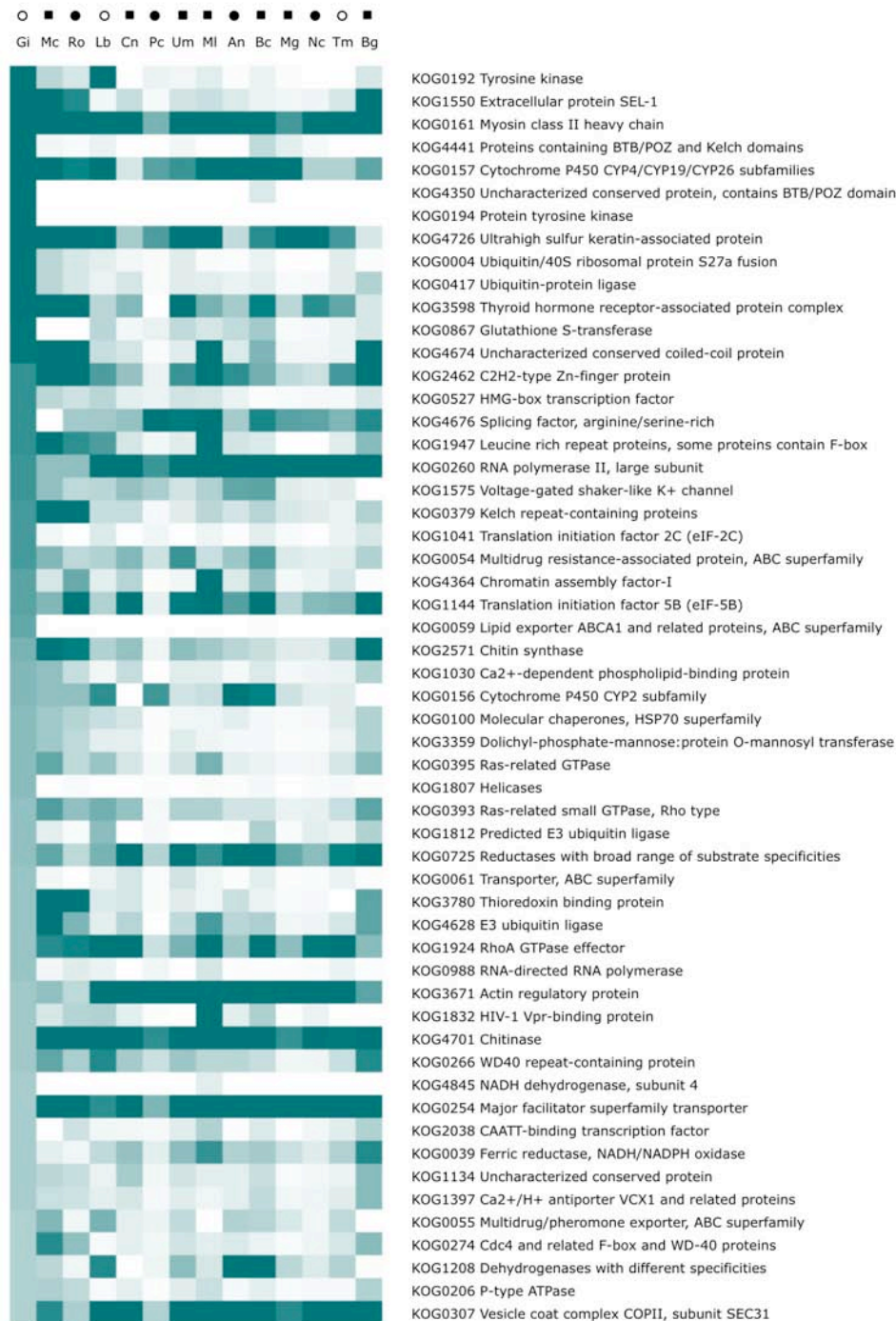


Figure 1: Most abundant KOG domains in *Glomus intraradices* compared to representative saprotrophic (black circle), pathogenic (black square) and mutualistic fungi (open circle). Heatmap based on the relative z-score of KOG conserved domains. The top 55 KOG domains found in *G. intraradices* were selected. From dark green: highly abundant to white: weakly abundant. Abbreviations: An, *Aspergillus nidulans*; Bc, *Botrytis cinerea*; Bg, *Blumeria graminis*; Cn, *Cryptococcus neoformans*; Lb, *Laccaria bicolor*; Ml, *Melampsora larici-populina*; Mg, *Magnaporthe grisea*; Mc, *Mucor circinellus*; Nc, *Neurospora crassa*; Pc, *Phanerochaete chrysosporium*; Ro, *Rhizopus oryzae*; Tm, *Tuber melanosporum*; and Um, *Ustilago maydis*.

Protein domain distributions

About half (5,056) of the 10,823 peptide sequences from *G. intraradices* with sequence similarity to documented proteins in databases were shared with all other fungal species and can thus be considered ubiquitous among fungi. The comparison of protein sets of *G. intraradices* with *R. oryzae* and *M. circinelloides* showed that 1,076 proteins were unique to these basal lineages of aseptate, coenocytic fungi. Interestingly, a small set of *G. intraradices* protein sequences were only shared with one or other of the sequenced mycorrhizal symbionts, the ascomycete *T. melanosporum* (49 NRVTs) or the basidiomycete *L. bicolor* (107 NRVTs). They mainly code for hypothetical proteins with no known function, although a few predicted proteins belong to signaling pathways. A total of 9,321 NRVTs contained at least a part of a conserved protein domain in the KOG database. Compared to other fungal gene repertoires, *G. intraradices* showed an over-representation of proteins involved in signaling pathways and ubiquitin-related metabolism (Fig. 1). The expansion of the tyrosine kinase encoding gene family involved in signaling pathways is also a feature of the ectomycorrhizal *L. bicolor* genome (Fig. 1; Martin *et al.*, 2008b).

Sequence polymorphism of transcripts

There were a total of 43,872 SNPs in 3,963 NRVTs (15.3% of the total number of NRVTs; 2.1 SNP/kb); 1,102 NRVTs contained less than five SNPs (Fig. 2A). When this SNP analysis was conducted on NRVT regions having a similarity with known protein coding sequences, 846 polymorphic NRVTs (8%; 1.3 SNP/kb) were identified among 10,823 homologs. Most polymorphic NRVTs contained less than five SNPs (Fig. 2B). The NRVTs with known function having the highest SNPs were a 60S ribosomal protein L17 (15 SNPs) and a Ras-related Rab (11 SNPs). The presence of multiple SNPs in hundreds of NRVTs confirmed the within-isolate DNA sequence polymorphism that has repeatedly been reported for a number of genes in *G. intraradices* (Croll *et al.*, 2008; Sanders & Croll, 2010). The transcriptome data show that this polymorphism is widespread in the genome, and that more than one variant of each polymorphic gene is transcriptionally active. This implies that the sequence variation may be of functional importance.

Symbiosis induces alterations in *G. intraradices* transcriptome

Little is known about the impact of symbiosis on the mycobiont transcriptome and molecular factors driving developmental pathways (Güimil *et al.*, 2005; Gomez *et al.*, 2009; Seddas *et al.*, 2009; Kuznetsova *et al.*, 2010; Sanders & Croll, 2010). Our transcript profiling thus provides the first large scale discovery of fungal symbiosis-related genes. Of the 18,751 coding sequences detected by oligoarrays, 395 (2.1%) and 569 (3.0%) were up- or downregulated (≥ 5 fold, p -value ≤ 0.05) in IRM in comparison to germinated spores, whereas 202 (1.1%) and 74 (0.4%) were up- or downregulated respectively in ERM in comparison to germinated spores. Highly upregulated genes in IRM and ERM are shown in Tables 2, 3, S3 and S4. In this study, mRNA concentrations were taken as a proxy of protein concentrations. Gene expression is however a multistep process that involves the transcription, translation and turnover of messenger RNAs and proteins, and transcript levels are not always related to protein levels and enzyme activities (Schwanhäusser *et al.*, 2011). It remains to be determined whether the observed changes in transcripts lead to alteration of the proteome, enzyme activities and/or metabolic fluxes. In addition, regulatory enzymes exhibiting allosteric properties are likely effective agents for fine tuning of metabolic fluxes in symbiotic tissues.

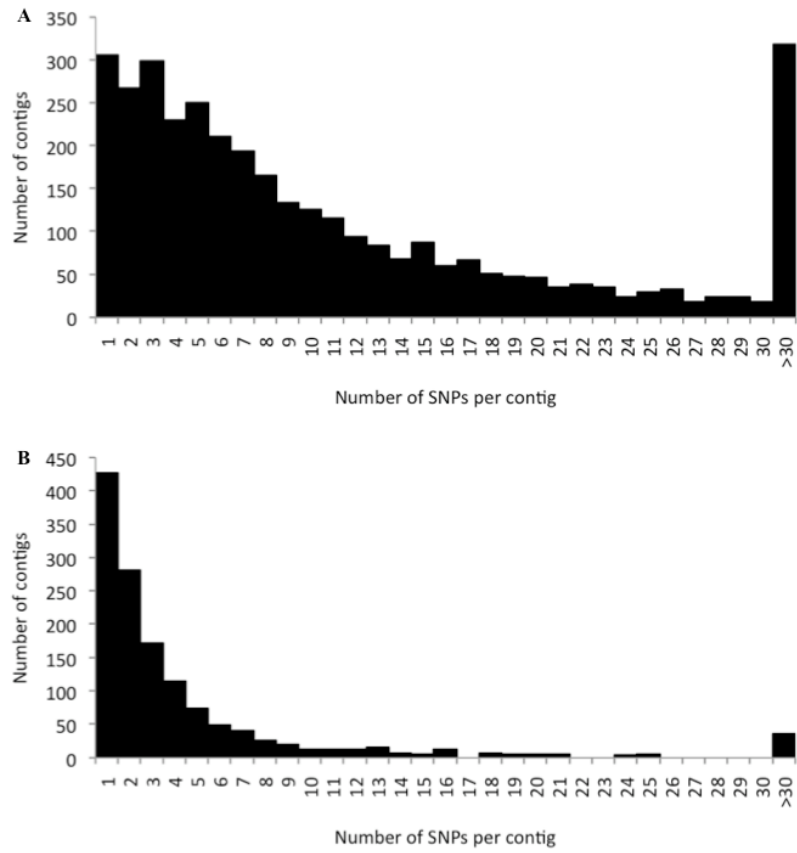


Figure 2: Single nucleotide polymorphisms in NRVt sequences from *Glomus intraradices*. (A) SNPs were identified by the MiraSearchESTSNPs program by aligning the complete nucleotide sequences from NRVts, including 5'- and 3'-UTRs. (B) SNPs were calculated in the coding regions of NRVts having a homolog documented in the Swiss-Prot database.

Table 2. Most highly upregulated *Glomus intraradices* transcripts (NRVTs) in the IRM compared with germinated spores identified using expression oligoarrays and ranked by decreasing fold changes. A value of 1 was given to transcripts not detected in germinated spores.

NRVT ID	Ratio IRM/Spore	IRM expression level	Spore expression level	Putative function
EXTB106676.b0	36516	38488	1	No Hit
Glomus_c7119	10427	10427	1	Cytochrome P450
Glomus_c13017	4057	7579	2	No hit
Gi.12932_C1	3919	3919	1	No hit
EXTB106330.b0	3428	3428	1	No hit
Glomus_lrc986	2759	6911	3	No hit
EXTB28921.b0	2581	2581	1	No hit
Gi.304_C1	1996	30567	15	No hit
CCHU1751.g1	1481	8199	6	Centromere-related protein
Glomus_c14959	1395	9891	7	Hypothetical protein
Gi.167_C1	1339	44618	33	No hit
CEXTB101194.b0	1206	1307	1	No hit
Glomus_c19831	878	6131	7	Hypothetical protein
Glomus_c5958	761	46426	61	No hit
Glomus_c3804	667	667	1	No hit
Glomus_c8796	638	638	1	No hit
EXTB112691.b0	576	576	1	No hit
Glomus_c8747	561	561	1	No hit
Glomus_c17585	426	426	1	Hypothetical protein
Gi.7237_C1	425	13107	31	No hit
Glomus_c10269	378	3150	8	No hit
EXTB121165.b0	339	3765	11	No hit
EXTA118370.b0	333	1353	4	No hit
Glomus_c8579	329	16088	49	No hit
Glomus_c2327	309	2838	9	No hit
Glomus_c12964	246	25666	105	Hypothetical protein
Gi.8222_C1	236	18627	79	Zinc-regulated transporter

Table 3. All transcripts (NRVTs) with functional annotation that were upregulated by at least ten-fold in *Glomus intraradices* IRM compared with germinated spores based on oligoarray profiling (ratio ≥ 10 ; p -value <0.05).

NRVT ID	Ratio IRM/Spores	IRM expression level	Putative function	Functional categories
EXTB25580.b0	17.6	8653	Major Facilitator Family (MFS) monocarboxylic acid transporter	Transport
Glomus_c20826	13.7	25099	Amino acid transporter	Transport
Glomus_c3620	27.4	21270	MFS nitrate transporter	Transport
Glomus_c10810	34.5	22106	MFS nitrate transporter	Transport
EXTA118370.b0	333.4	1353	Iron permease	Transport
Gi.8222_C1	236.2	18627	ZIP (Zrt- and Irt-like) Zn transporter	Transport
Glomus_c13887	15.8	8222	P-type ATPase	Transport
Gi.5720_C1	12.5	3337	Cation transport-related protein	Transport
EXTA115701.b0	24.7	7920	(ATP-binding Cassette) ABC transporter	Transport
Glomus_c14346	16.1	1921	P-type ATPase	Transport
Glomus_c9272	77.1	3766	Fatty acyl-CoA elongase	Lipid metabolism
EXTB126208.b0	10.6	9156	Peroxisomal 3-ketoacyl-CoA-thiolase	Lipid metabolism
Glomus_c8371	40.2	9435	Phosphatidylglycerol/phosphatidylinositol transfer protein	Lipid metabolism
Gi.6806_C1	48.2	23206	Lipase	Lipid metabolism
Glomus_c11625	15.4	26531	Acetyl-CoA acetyltransferase	Lipid metabolism
Glomus_c15563	10.0	10357	Adenine phosphoribosyl transferase	Nucleotide metabolism
Gi.1390_C2	32.1	15107	Ribonucleoside-diphosphate reductase	Nucleotide metabolism
Glomus_c2075	23.9	12720	Thymidylate synthase	Nucleotide metabolism
Glomus_c12729	26.4	10113	Endopolyphosphatase	Phosphate metabolism
EXTB107807.b0	18.8	17134	Endopolyphosphatase	Phosphate metabolism
Glomus_c4491	27.0	26571	Carbonic anhydrase	Carbon metabolism
Gi.8293_C1	43.4	7083	Cytochrome P450	Secondary metabolism
Glomus_c7119	10426.8	10427	Cytochrome P450	Secondary metabolism
EXTB18295.b0	27.7	20682	Multicopper oxidase	Secondary metabolism
Glomus_c1906	10.0	11236	Glutathione S-transferase	Secondary metabolism
Glomus_c1140	10.9	24021	Iron sulfur cluster assembly protein	Protein modification
Glomus_c14063	21.0	7074	Ubiquitin-conjugating enzyme	Protein modification
Glomus_c22386	58.6	13312	Protease, Ulp1 family	Protein modification
Glomus_c3253	13.6	3141	Ubiquitin/ribosomal protein S27a	Protein modification
Glomus_c8960	10.0	8892	Menaquinone biosynthesis methyltransferase	Coenzyme metabolism
Glomus_c10425	13.6	22927	Steroidogenic acute regulatory protein	Signal transduction
CACE6804.g1	11.3	2292	Mitogen-activated protein kinase kinase kinase	Signal transduction
EXTB27526.b0	13.5	2738	Tyrosine-like kinase	Signal transduction
Glomus_lrc4409	50.5	37996	Calcium binding protein	Signal transduction
Glomus_lrc7412	33.1	50044	Calcium binding protein	Signal transduction
Glomus_c9394	14.4	5290	Chitin synthase	Cell wall biogenesis
Glomus_c1601	40.8	50339	Cell wall anchor domain protein	Cell wall biogenesis
CCHU1751.g1	1480.7	8199	Centromere protein-related protein	Cell cycle control
Glomus_c10811	27.2	9906	Kinetochore protein	Cell cycle control
Glomus_c3780	17.9	20774	Targeting protein for Xklp2	Cell cycle control
Glomus_c5080	50.3	16455	Kinesin-like protein	Cytoskeleton

Glomus_c7425	35.1	4759	DNA topoisomerase type II	Chromatin structure
Glomus_lrc174	1214.8	22989	Chromatin assembly factor-I	Chromatin structure
Glomus_c505	26.3	37147	Histone H3.1	Chromatin structure
CCHU9624.1b	12.5	450	DNA repair protein RAD51	Replication, repair
Glomus_c9197	12.7	7425	DNA replication licensing factor	Replication, repair
Glomus_c16484	73.5	2903	Uracil-DNA glycosylase	Replication, repair
Glomus_c13628	31.5	3433	BTB/POZ domain-containing protein	Transcription
Glomus_c18585	13.2	4751	Zn finger-containing protein	Transcription
Glomus_c10966	80.6	9500	Zn finger-containing protein	Transcription
Glomus_c1198	11.2	25711	Yabby-like transcription factor	Transcription
Glomus_lrc739	11.3	11898	Translation initiation factor 3	Translation

The majority of IRM- and ERM-induced genes were lineage-specific genes as 80% of the most highly upregulated transcripts coded for orphan proteins (Tables 2 & S3). Differentially expressed genes coding for proteins of known function were categorized into functional classes (Tables 3 & S4). Genes overrepresented in IRM vs. germinated spores showed a GO enrichment (FDR correction, $p \leq 0.05$) for several categories of biological processes associated with ion transport, lipid/steroid metabolism and DNA replication (data not shown). Out of 25 transcripts that appeared to be highly upregulated in IRM, 17 were validated by RT-PCR (data not shown). While 13 out of 17 transcripts were detected in at least one other stage of the life cycle besides arbuscules, four were arbuscule-specific transcripts (Table S5). Three were orphan genes, while the fourth showed similarity to a multidrug resistance protein (ABC transporter).

As the oligoarray was constructed before the 454 ESTs from IRM and LMD arbuscule-containing cells were obtained, we also carried out a digital Northern analysis based on differential frequency of 454 reads (Audic & Claverie, 1997). This transcript profiling confirmed the above oligoarray expression patterns (data not shown). However, we detected a significant set of IRM transcripts coding for small secreted proteins (SSPs) not detected in ERM and spores. These SSPs showed a striking upregulation in IRM and arbuscule-containing cells (Table 4). The most highly upregulated IRM orphan transcript ‘*step3_c3163*’ coded for a SSP of 280 amino acids of unknown function with similarity to proteins of several biotrophic basidiomycetes (*L. bicolor*, *M. larici-populina*). This transcript was only detected in mycorrhizal roots and arbuscules.

Recent studies of plant-microbe interactions involving pathogenic and ectomycorrhizal fungi have shown that many lineage-specific orphan genes code for effector proteins playing key roles in the host colonisation and *in planta* accommodation by controlling the plant immune system (de Wit *et al.*, 2009; Plett *et al.*, 2011). The set of differentially-expressed *G. intraradices* orphan genes included the effector protein SP7, which interacts with the pathogenesis-related transcription factor ERF19 in the host nucleus (Kloppholz *et al.*, 2011) (Table S4). Given the key results obtained for SP7, the role played by other mycorrhiza-upregulated small secreted proteins of *G. intraradices* should be elucidated, as well as the identity of plant-based signals that may control their expression within the root space.

Table 4. The transcripts coding for mycorrhizal-induced small secreted proteins (MiSSPs) showing upregulation in *Glomus intraradices* IRM compared with germinated spores based on 454 read frequency. These sequences have no probe on oligoarrays. Abbreviation: AA, amino acids; Cys, cysteine residue; SP, secreted proteins.

NRVT ID	IRM (reads)	Arbuscules (reads)	Spores (reads)	Putative function	Size (AA)	# Cys
step3_c3163	62	4	0	Hypothetical protein	280	12
remain_c2375	24	16	0	No hit	156	3
step3_c3587	15	2	0	No hit	149	0
remain_c4548	12	4	0	MD-2-related lipid recognition domain	173	4
step3_c2834	11	19	0	No hit	61	9
remain_c2379	11	7	0	No hit	120	0
step3_c3282	10	1	0	No hit	86	2
remain_c8131	9	0	0	Hypothetical protein	150	3
remain_c2779	7	20	0	No hit	67	0
remain_c8897	7	1	0	No hit	71	7
remain_c12045	5	3	0	No hit	122	4

Stealth colonisation to evade host defenses

Several obligate biotrophic pathogens and ectomycorrhizal symbionts have a decreased repertoire of carbohydrate-acting enzymes involved in the degradation of plant cell wall (PCW) polysaccharides (Baxter *et al.*, 2010; Martin *et al.*, 2008b, 2010; Spanu *et al.*, 2010; Duplessis *et al.*, 2011). We identified 139 NRVTs encoding carbohydrate-active enzymes in the *G. intraradices* transcriptome (Table S6). They included enzymes that are involved in acquiring and converting carbohydrates and in maintaining fungal cell wall plasticity. Genes encoding glycosyl hydrolases (GHs) involved in degrading PCW lignocellulosic polymers were not found in the IRM predicted proteome. No genes encoding GH from families GH6 and GH7, which include cellobiohydrolases and are involved in attack of crystalline cellulose, polysaccharide lyases or proteins with a CBM1 cellulose-binding motif were identified (Table S6). Enzymes that attack amorphous cellulose, hemicellulose, and pectin (e.g., β -glucosidases and cellobiose dehydrogenases of family GH3, GH10 xylanases, and GH28 pectinases) were absent. Genes encoding carbohydrate esterases and secreted feruloyl esterase that could hydrolyze cross-links in PCW were also not expressed. However, genes encoding GH5 endoglucanases (that act on mannan and/or cellulose) and a GH9 cellobiohydrolase were identified (Table S6). As observed in ectomycorrhizal symbionts (Martin *et al.*, 2008b, 2010; Nagendran *et al.*, 2009), the minimal set of PCW-degrading enzymes in the transcriptome of *G. intraradices* might be an evolutionary adaptation to avoid the release of polysaccharide fragments and their detection by the host immune system during the obligate biotrophic phase of the fungus. In contrast, a large set of genes encoding enzymes involved in chitin metabolism were expressed, including chitin synthases, chitin deacetylases, and chitinases (Table S6). They probably play a role in remodeling the fungal cell wall during growth and symbiosis. The abundance of transcripts coding for dolichyl-related mannosylation enzymes suggests an intense biosynthesis of cell wall mannans and mannoproteins.

Nutrient assimilation

We identified numerous transcripts coding for different predicted proteins associated with the uptake and assimilation of major soil nutrients, such as nitrate and Pi, and conversion of metabolites shuttled between partners (amino acids, carbohydrates) (Table S7). Although a few genes involved in nutrient transport and assimilation have previously been identified in *G. intraradices* and *G. mosseae* (Benedetto *et al.*, 2005; Govindarajulu *et al.*, 2005; Javot *et al.*, 2007; Tian *et al.*, 2010), there is no comprehensive molecular description of N and P assimilation pathways despite the crucial importance of metabolite exchange in this symbiosis (Kiers *et al.*, 2011; Smith & Smith, 2011). The striking induction of several genes coding for metal transporters suggests that metals, such as Zn, are important for plant colonization and the expression pattern of genes coding for assimilative enzymes fully supports a rapid uptake, translocation and transfer of metabolites as discussed in the following sections.

Nitrogen metabolism. In the *G. intraradices*-colonized *M. truncatula* roots, transcripts for transporters and enzymes involved in N uptake and assimilation showed a high constitutive expression, supporting the contention that a high N turnover is taking place in the symbiosis (Gomez *et al.*, 2009; Guether *et al.*, 2009b). The upregulation of transcripts for nitrate permease and nitrate reductase (Table S8), together with the higher expression of the gene coding for ammonium transporter gene *GintAMT2* (Table S8; Pérez-Tienda *et al.*, 2011), might favor the accumulation of NH_4^+ for incorporation into amino acids and further translocation of N to the host cells. The IRM transcript profile is consistent with assimilation of NH_4^+ via the glutamine synthetase/glutamate synthase (GS/GOGAT) cycle. Transcripts encoding enzymes of the arginine biosynthesis and degradation were expressed highly in spores, ERM and IRM (Table S8), indicating an intense cycling of N through the arginine pathway (Tian *et al.*, 2010). The higher expression of genes coding for arginase, urease and urease accessory protein (*UreG*) in IRM might lead to a higher release of NH_4^+ able to feed the GS/GOGAT cycle and/or the passive efflux of ammonia to the interfacial apoplast and the uptake of NH_4^+ by the host ammonium transporter(s) (Govindarajulu *et al.*, 2005; Gomez *et al.*, 2009; Tian *et al.*, 2010). High expression of amino acid transporter genes in ERM and IRM (Table S7) suggests that the mycobiont can taken up amino acids present in both soil and in the interfacial apoplast.

Phosphate metabolism. *G. intraradices* expressed a wide spectrum of secreted phosphatase transcripts, including those coding for the *p*-nitrophenylphosphatase Pho13p, acid phosphatase Pho3p, repressible alkaline phosphatase Pho8p, and magnesium-dependent phosphatase, able to act on various phosphate esters (Table S9). No phytase transcript was detected. *G. intraradices* has a combination of low-affinity and high-affinity transporter genes, including the high-affinity Pi transporter *GintPT* (equivalent to Pho84p from *Saccharomyces cerevisiae*), the low-affinity Pi transporter Pho91p, and the Na⁺/Pi symporter Pho89p (Table S9). The transcripts coding for the interactor protein Pho88p that promotes maturation and trafficking of Pho84p in yeast have also been identified. Pi transporter genes are expressed in spores, ERM and IRM. The expression pattern of *GintPT* is in agreement with the results obtained on *GmosPT* in *G. mosseae* (Benedetto *et al.*, 2005) and the observed expression of *GmosPT* in arbuscules (Balestrini *et al.*, 2007; Gomez-Ariza *et al.*, 2009). Although high net transfer of P from soil to the plant is taking place via the mycobiont (Kiers *et al.*, 2011; Smith & Smith, 2011), our findings suggest that the IRM may re-absorb some of the Pi released in the symbiotic apoplastic space. As shown by Kiers *et al.* (2011), the fungus might exert some control on the delivery of nutrients to its host plant and reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. Following Pi uptake, PolyP accumulates in *G. intraradices* hyphae and is stored in the vacuolar compartments where it buffers cytoplasmic Pi concentrations. PolyP is translocated along hyphae towards the interfacial apoplast (Viereck *et al.*, 2004; Javot *et al.*, 2007; Hijikata *et al.*, 2010). The rapid synthesis of PolyP is thus crucial for the maintenance of effective hyphal Pi uptake (Tani *et al.*, 2009). Transcripts coding for the PolyP polymerase Vtc4p complex

involved in synthesis and transfer of PolyP to the vacuole (Hothorn *et al.*, 2009), and for its accessory proteins, such as Vtc1p, were expressed in *G. intraradices* (Table S9). Transcripts for vacuolar endopolyphosphatase Ppn1p and the exopolyphosphatase Ppx1p were also identified. The endopolyphosphatase transcript was strikingly upregulated in IRM. Several components of the complex phosphate regulatory system (PHO regulon) found in *S. cerevisiae* and *Neurospora crassa* were expressed in *G. intraradices*, including the ankyrin-repeat containing NUC2-related protein (Pho81p), the cyclin-dependent protein kinase regulator Pho80p and the mitogen-activated serine/threonine Cdc2 cyclin-dependent protein kinase Pho85p (Table S9), suggesting putative regulatory similarities.

Lipid metabolism. The profuse development of fungal membranes associated with arbuscule development in host cells dramatically increases the need for plasma membrane fatty acids and sterols. The oligoarray analysis identified transcripts in IRM coding for the subunits a and b of the fatty acid synthase complex, together with the acetyl-CoA carboxylase transcript (Table S10), confirming that the mycobiont does possess fatty acid synthetic capacities (Table S10). Similarly, mycorrhiza-induced genes predicted to be involved in fatty acid metabolism are upregulated in *M. truncatula* (Gomez *et al.*, 2009), pointing to a complementary regulation of fungal and plant lipid metabolism. In this host plant, the gene *MtMSBP1*, encoding a membrane steroid (progesterone)-binding protein, is also induced early by a diffusible AM fungal signal produced by *G. intraradices* branched hyphae (Kuhn *et al.*, 2010). The high activity of enzymes involved in sterol and steroid metabolism in IRM, together with the AM fungal induction of *MtMSBP1*, might be related to the need to alter sterol metabolism (e.g., lanosterol) to allow plasma membrane invagination and intracellular accommodation of the fungal symbiont in the cortical cells.

An expanded inventory of conserved meiotic genes provides evidence for cryptic sex

The lack of an observed sexual stage in any member of the Glomeromycota led to the suggestion, which has been debated, that AM fungi were ancient asexuals. Results from Croll and Sanders (2009) strongly suggest that recombination occurred among some *G. intraradices* genotypes in the field, although the majority of populations examined so far have been clonal (Rosendahl, 2008). Since Glomeromycota could represent one of the earliest diverging fungal lineages, their meiotic processes could represent an ancestral state. We surveyed the transcriptome of *G. intraradices* for a set of sex and meiotic genes conserved among eukaryotes (Malik *et al.*, 2008) (Table S11) and identified several ‘meiosis-specific’ genes (HOP2, MND1, and MER3) that are only known to function in meiosis in other eukaryotes (Malik *et al.*, 2008). These genes are hypothesized to be present in organisms with sexual ancestry (Table S11). The homolog of a transcript coding for the key meiotic recombinase SPO11 (REC12) was not found, but has been identified in the genomic sequences of *G. diaphanum* (MUCL 43196), *G. intraradices* (DAOM 197198), *G. clarum* (DAOM 234281) and *G. cerebriforme* (DAOM 227022) (Halary *et al.*, 2011). Transcripts coding for high-mobility group (HMG) domain containing transcriptional factors with a significant similarity (55%) to the *sexP* and *sexM* genes from *Phycomyces blakesleeanus* (Idnurm *et al.*, 2008) were also retrieved. These genes are master switches controlling mating type in fungi. Mechanisms controlling the *G. intraradices* sexual cycle need to be further examined within AM fungal populations as they are likely to allow the mixing of nuclei and subsequent recombination among different individuals of this fungus interacting in soil.

In summary, induced expression of genes coding for membrane transporters and SSPs during the symbiotic interaction, together with the lack of expression of hydrolytic enzymes acting on PCW polysaccharides, are hallmarks of *G. intraradices*. These results extend conserved patterns of gene expression profiles observed in obligate biotrophic pathogens (Spanu *et al.*, 2010; Kemen *et al.*, 2011) and ectomycorrhizal symbionts (Plett & Martin, 2011) to the Glomeromycota lineage. In contrast, obligate biotrophy in *G. intraradices* is not associated with a striking reduction of metabolic complexity (e.g. lack of N and S assimilation pathways), as observed in many obligate biotrophic pathogens, so that the ability to interact with the soil environment with respect to nutrient uptake is maintained in the symbiotic fungus. Finally, we can hypothesize that biotrophy in AM fungi has evolved through a series of steps requiring effectors, such as the secreted SP7 (Kloppholz *et al.*, 2011), coupled with a reduced inventory of PCW hydrolyzing enzymes to suppress or attenuate host defence reactions, and weak selection forces to maintain certain biosynthetic pathways if products (e.g. thiamine, sucrose) can be directly obtained from the host. The present comprehensive repertoire of *G. intraradices* genes, the first for Glomeromycota, provides a basis for future research in environmental genomics and for accessing symbiosis-related functional features in other members of this unique phylum.

Acknowledgements

ET is supported by a scholarship from the Région Lorraine and the European Commission ENERGYPOPLAR project (FP7-211917), and YL by a scholarship from the Education Ministry of China. This work was supported by the Network of Excellence EVOLTREE (FP6-016322) and the U.S. Department of Energy (DOE) – Oak Ridge National Laboratory Scientific Focus Area for Genomics Foundational Sciences (to FM). Ph F was supported by the German Research Foundation and the Ministries of Food, Agriculture, and Consumer Protection of the FRG and the Brandenburg and Thüringen States. PB was funded by the Regional Project Converging technologies-BIOBIT, and NF and CAA by the project AGL2009-08868 from the Spanish Ministry of Science and Education. Support for development of mycorrhizal *M. truncatula* root materials was provided by the U.S National Science Foundation IOS-0842720 to MJH. A portion of the writing of this manuscript was sponsored by the DOE Genomic Science Program (DE-AC05-00OR22725). EST sequencing was funded by INRA, Genoscope, New Mexico University, and Michigan State University. The genomic and Sanger EST sequencing was conducted by JGI (DE-AC02-05CH11231). We would like to thank Nicolas Corradi for helpful discussions on the meiotic genes, and three anonymous referees for their insightful comments and suggestions.

REFERENCES

- Akiyama K, Matsuzaki KI, Hayashi H. 2005. Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature* **435**: 824-827.
- Allen JW, Shachar-Hill Y. 2009. Sulfur transfer through an arbuscular mycorrhiza. *Plant Physiology* **149**: 549-560.
- Angelard C, Colard A, Niculita-Hirzel H, Croll D, Sanders IR. 2010. Segregation in a mycorrhizal fungus alters rice growth and symbiosis-specific gene transcription. *Current Biology* **20**: 1216-1221.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al.* 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* **25**: 25-29.

- Audic S, Claverie JM. 1997.** The significance of digital gene expression profiles. *Genome Research* 7: 986-995.
- Bago B, Pfeffer PE, Shachar-Hill Y. 2000.** Carbon metabolism and transport in arbuscular mycorrhizas. *Plant Physiology* 124: 949-957.
- Bago B, Pfeffer PE, Abubaker J, Jun J, Allen JW, Brouillette J, Douds DD, Lammers PJ, Shachar-Hill Y. 2003.** Carbon export from arbuscular mycorrhizal roots involves the translocation of carbohydrate as well as lipid. *Plant Physiology* 131: 1496-1507.
- Balestrini R, Lanfranco L. 2006.** Fungal and plant gene expression in arbuscular mycorrhizal symbiosis. *Mycorrhiza* 16: 509-524.
- Balestrini R, Gómez-Ariza J, Lanfranco L, Bonfante P. 2007.** Laser microdissection reveals that transcripts for five plant and one fungal phosphate transporter genes are contemporaneously present in arbusculated cells. *Molecular Plant-Microbe Interactions* 20: 1055-1062.
- Baxter L, Tripathy S, Ishaque N, Boot N, Cabral A, Kemen E, Thines M, Ah-Fong A, Anderson R, Badejoko W, et al. 2010.** Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. *Science* 330:1549-1551.
- Bécard G, Fortin JA. 1988.** Early events of vesicular arbuscular mycorrhiza formation on Ri T-DNA transformed roots. *New Phytologist* 108: 211-218.
- Bécard G, Kosuta S, Tamasloukht M, Séjalon-Delmas N, Roux C. 2004.** Partner communication in the arbuscular mycorrhizal interaction. *Canadian Journal of Botany* 8 : 1186-1197.
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004.** Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* 340: 783-795.
- Benedetto A, Magurno F, Bonfante P, Lanfranco L. 2005.** Expression profiles of a phosphate transporter gene (*GmosPT*) from the endomycorrhizal fungus *Glomus mosseae*. *Mycorrhiza* 15: 620-627.
- Bonfante P, Genre A. 2010.** Mechanisms underlying beneficial plant-fungus interactions in mycorrhizal symbiosis. *Nature Communications* 1: 48.
- Cantarel BL, Continho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009.** The Carbohydrate-Active enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Research* 37: D233-D238.
- Chabot S, Bécard G, Piché Y. 1992.** The life cycle of *Glomus intraradices* in root organ culture. *Mycologia* 84: 315-321.
- Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, Suhai S. 2004.** Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Research* 14: 1147-1159.
- Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005.** Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674-3676.
- Croll D, Wille L, Gamper HA, Mathimaran N, Lammers PJ, Corradi N, Sanders IR. 2008.** Genetic diversity and host plant preferences revealed by simple sequence repeat and mitochondrial markers in a population of the arbuscular mycorrhizal fungus *Glomus intraradices*. *New Phytologist* 178: 672-687.
- Croll D, Giovannetti M, Koch AM, Sbrana C, Ehinger M, Lammers PJ, Sanders IR. 2009.** Nonself vegetative fusion and genetic exchange in the arbuscular mycorrhizal fungus *Glomus intraradices*. *New Phytologist* 181: 924-937.
- Croll D, Sanders IR. 2009.** Recombination in *Glomus intraradices*, a supposed ancient asexual arbuscular mycorrhizal fungus. *BMC Evolutionary Biology*. 9: 13.
- De Wit PJGM, Mehrabi R, Van Den Burg HA, Stergiopoulos I 2009.** Fungal effector proteins: past, present and future. *Journal Molecular Plant Pathology* 10: 735-747.
- Dodds PN, Rathjen JP. 2010.** Plant immunity: towards an integrated view of plant-pathogen interactions. *Nature Reviews Genetics*. 8: 539-548.
- Duplessis S, Cuomo CA, Lin YC, Aerts A, Tisserant E, Veneault-Fourrey C, Joly DL, Hacquard S, Amselem J, Cantarel BL et al. 2011.** Obligate biotrophy features unraveled by the

- genomic analysis of the rust fungi, *Melampsora larici-populina* and *Puccinia graminis* f. sp. *tritici*. *Proceedings of the National Academy of Sciences of the USA* **108**: 9166-9171.
- Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000.** Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* **300**: 1005-1016.
- Ercolin F, Reinhardt D. 2011.** Successful joint ventures of plants: arbuscular mycorrhiza and beyond. *Trends in Plant Science* **16**: 356-362.
- Fitter AH, Helgason T, Hodge A. 2011.** Nutritional exchanges in the arbuscular mycorrhizal symbiosis: Implications for sustainable agriculture. *Fungal Biology Reviews* **25**: 68-72.
- Genre A, Chabaud M, Timmers T, Bonfante P, Barker DG. 2005.** Arbuscular mycorrhizal fungi elicit a novel intracellular apparatus in *Medicago truncatula* root epidermal cells before infection. *Plant Cell* **17**: 3489-3499.
- Giovannetti M, Fortuna P, Citernes AS, Morini S, Nuti MP. 2001.** The occurrence of anastomosis formation and nuclear exchange in intact arbuscular mycorrhizal networks. *New Phytologist* **151**: 717-724.
- Gomez SK, Javot H, Deewatthanawong P, Torres-Jerez I, Tang Y, Blancaflor EB, Udvardi MK, Harrison MJ. 2009.** *Medicago truncatula* and *Glomus intraradices* gene expression in cortical cells harboring arbuscules in the arbuscular mycorrhizal symbiosis. *BMC Plant Biology* **22**: 10.
- Gómez-Ariza J, Balestrini R, Novero M, Bonfante P. 2009.** Cell-specific gene expression of phosphate transporters in mycorrhizal tomato roots. *Biology and Fertility of Soils* **45**: 845-853.
- Govindarajulu M, Pfeffer PE, Jin HR, Abubaker J, Douds DD, Allen JW, Bucking H, Lammers PJ, Shachar-Hill Y. 2005.** Nitrogen transfer in the arbuscular mycorrhizal symbiosis. *Nature* **435**: 819-823.
- Grunwald U, Guo W, Fischer K, Isayenkov S, Ludwig-Müller J, Hause B, Yan X, Küster H, Franken P. 2009.** Overlapping expression patterns and differential transcript levels of phosphate transporter genes in arbuscular mycorrhizal, Pi-fertilised and phytohormone-treated *Medicago truncatula* roots. *Planta* **229**: 1023-1034.
- Guether M, Balestrini R, Hannah MA, Udvardi MK, Bonfante P. 2009a.** Genome-wide reprogramming of regulatory networks, transport, cell wall and membrane biogenesis during arbuscular mycorrhizal symbiosis in *Lotus japonicus*. *New Phytologist* **182**: 200-212.
- Guether M, Neuhäuser B, Balestrini R, Dynowski M, Ludewig U, Bonfante P. 2009b.** A mycorrhizal-specific ammonium transporter from *Lotus japonicus* acquires nitrogen released by arbuscular mycorrhizal fungi. *Plant Physiology* **150**: 73-83.
- Güimil S, Chang H-S, Zhu T, Sesma A, Osbourn A, Roux C, Ioannidis V, Oakeley EJ, Docquier M, Descombes P, et al. 2005.** Comparative transcriptomics of rice reveals an ancient pattern of response to microbial colonization. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 8066-8070.
- Halary S, Malik SB, Lildhar L, Slamovits CH, Hijri M, Corradi N. 2011.** Conserved meiotic machinery in *Glomus* spp., a putatively ancient asexual fungal lineage. *Genome Biology and Evolution*, in press.
- Hijikata N, Murase M, Tani C, Ohtomo R, Osaki M, Ezawa T. 2010.** Polyphosphate has a central role in the rapid and massive accumulation of phosphorus in extraradical mycelium of an arbuscular mycorrhizal fungus. *New Phytologist* **186**: 285-289.
- Hothorn M, Neumann H, Lenherr ED, Wehner M, Rybin V, Hassa PO, Uttenweiler A, Reinhardt M, Schmidt A, Seiler J, et al. 2009.** Catalytic core of a membrane-associated eukaryotic polyphosphate polymerase. *Science* **324**: 513-516.
- Idnurm A, Walton FJ, Floyd A, Heitman J. 2008.** Identification of the sex genes in an early diverged fungus. *Nature* **451**: 193-197.
- Jany JL, Pawlowska TE. 2010.** Multinucleate spores contribute to evolutionary longevity of asexual Glomeromycota. *The American Naturalist* **175**: 424-435.

- Javot H, Pumplin N, Harrison MJ. 2007.** Phosphate in the arbuscular mycorrhizal symbiosis: transport properties and regulatory roles. *Plant, Cell and Environment* **30**: 310-322.
- Kamoun S. 2007.** Groovy times: filamentous pathogen effectors revealed. *Current Opinion in Plant Biology*. **10**: 358-65.
- Kanehisa M, Goto S. 2000.** KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**: 27-30.
- Kemen E, Gardiner A, Schultz-Larsen T, Kemen AC, Balmuth AL, Robert-Seilanianantz A, Bailey K, Holub E, Studholme DJ, Maclean D, et al. 2011.** Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis thaliana*. *PLoS Biology* **9**: e1001094.
- Kiers ET, Duhamel M, Beesetty Y, Mensah JA, Franken O, Verbruggen E, Fellbaum CR, Kowalchuk GA, Hart MM, Bago A, et al. 2011.** Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science* **333**: 880-882.
- Kloppholz S, Kuhn H, Requena N. 2011.** A secreted fungal effector of *Glomus intraradices* promotes symbiotic biotrophy. *Current Biology* **21**,1204-1209.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001.** Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* **305**: 567-580.
- Kuhn H, Küster H, Requena N. 2010.** Membrane steroid-binding protein 1 induced by a diffusible fungal signal is critical for mycorrhization in *Medicago truncatula*. *New Phytologist* **185**: 716-733.
- Kuznetsova E, Seddas-Dozolme PMA, Arnould C, Tollot M, van Tuinen D, Borisov A, Gianinazzi S, Gianinazzi-Pearson V. 2010.** Symbiosis-related pea genes modulate fungal and plant gene expression during the arbuscule stage of mycorrhiza with *Glomus intraradices*. *Mycorrhiza* **20**: 427-443.
- Letunic I, Yamada T, Kanehisa M, Bork P. 2008.** iPath: interactive exploration of biochemical pathways and networks. *Trends in Biochemical Sciences* **33**: 101-103.
- Maere S, Heymans K, Kuiper M. 2005.** BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*. **21**: 3448-3449.
- Maillet F, Poinot V, André O, Puech-Pagès V, Haouy A, Gueunier M, Cromer L, Giraudet D, Formey D, Niebel A, et al. 2011.** Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature*. **469**: 58-63.
- Malik SB, Pightling AW, Stefaniak LM, Schurko AM, Logsdon JM. 2008.** An expanded inventory of conserved meiotic genes provides evidence for sex in *Trichomonas vaginalis*. *PLoS One* **3**: e2879.
- Martin F, Gianinazzi-Pearson V, Hijri M, Lammers P, Requena N, Sanders IR, Shachar-Hill Y, Shapiro H, Tuskan GA, Young JPW. 2008a.** The long hard road to a completed *Glomus intraradices* genome. *New Phytologist* **180**: 747-750.
- Martin F, Aerts A, Ahren D, Brun A, Danchin EGJ, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V, et al. 2008b.** The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **452**: 88-92.
- Martin F, Kohler A, Murat C, Balestrini R, Coutinho PM, Jaillon O, Montanini B, Morin E, Noel B, Percudani R et al. 2010.** Périgord Black Truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* **464**: 1033-1038.
- Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007.** KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* **35**: W182-185.
- Nagendran S, Hallen-Adams HE, Paper JM, Aslam N, Walton JD. 2009.** Reduced genomic potential for secreted plant cell-wall-degrading enzymes in the ectomycorrhizal fungus *Amanita bisporigera*, based on the secretome of *Trichoderma reesei*. *Fungal Genetics and Biology* **46**: 427-35.
- Parra G, Bradnam K, Ning Z, Keane T, Korf I. 2009.** Assessing the gene space in draft genomes. *Nucleic Acids Research* **37**: 289-297.

- Pérez-Tienda J, Testillano PS, Balestrini R, Fiorilli V, Azcón-Aguilar C, Ferrol N. 2011.** GintAMT2, a new member of the ammonium transporter family in the arbuscular mycorrhizal fungus *Glomus intraradices*. *Fungal Genetics and Biology*, in press.
- Plett JM, Martin F. 2011.** Blurred boundaries: lifestyle lessons from ectomycorrhizal fungal genomes. *Trends in Genetics* **27**: 14-22.
- Plett JM, Kemppainen M, Kale SD, Kohler A, Legué V, Brun A, Tyler B, Pardo A, Martin F. 2011.** A secreted effector protein of *Laccaria bicolor* is required for symbiosis development. *Current Biology* **21**: 1197-1203.
- Rosendahl S. 2008.** Communities, populations and individuals of arbuscular mycorrhizal fungi. *New Phytologist* **178**: 253-266.
- Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, Quackenbush J. 2006.** TM4 microarray software suite. *Methods in Enzymology* **411**: 134-193.
- Sanders IR, Croll D. 2010.** Arbuscular mycorrhiza: the challenge to understand the genetics of the fungal partner. *Annual Review of Genetics* **44**: 271-92.
- Schaarschmidt S, Roitsch T, Hause B. 2006.** Arbuscular mycorrhiza induces gene expression of the apoplastic invertase LIN6 in tomato (*Lycopersicon esculentum*) roots. *Journal of Experimental Botany* **57**: 4015-4023.
- Schüssler A, Schwarzott D, Walker C. 2001.** A new fungal phylum, the Glomeromycota: phylogeny and evolution. *Mycological Research* **105**: 1413-1421.
- Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. 2011.** Global quantification of mammalian gene expression control. *Nature* **473**: 337-342.
- Seddas PM, Arias C, Arnould C, van Tuinen D, Godfroy O, Benhassou H, Gouzy J, Morandi D, Dessaint F, Gianinazzi-Pearson V. 2009.** Symbiosis-related plant genes modulate molecular responses in an arbuscular mycorrhizal fungus during early root interactions. *Molecular Plant-Microbe Interactions* **22**: 341-351.
- Smith SE, Read DJ. 2008.** *Mycorrhizal symbiosis*. Academic press.
- Smith SE, Smith FA (2011)** Roles of arbuscular mycorrhizas in plant nutrition and growth: New paradigms from cellular to ecosystem scales. *Annual Review of Plant Biology* **62**: 227-250.
- Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, van Themaat EVL, Brown JKM et al. 2010.** Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* **330**: 1543-1546.
- Stockinger H, Walker C, Schüssler A. 2009.** ‘*Glomus intraradices* DAOM197198’, a model fungus in arbuscular mycorrhiza research, is not *Glomus intraradices*. *New Phytologist* **183**: 1176-1187.
- Tani C, Ohtomo R, Osaki M, Kuga Y, Ezawa T. 2009.** ATP-dependent but proton gradient-independent polyphosphate-synthesizing activity in extraradical hyphae of an arbuscular mycorrhizal fungus. *Applied and Environmental Microbiology* **75**: 7044-7050.
- Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, et al. 2003.** The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**: 41.
- Tian C, Kasiborski B, Koul R, Lammers PJ, Bucking H, Shachar-Hill Y. 2010.** Regulation of the nitrogen transfer pathway in the arbuscular mycorrhizal symbiosis: gene characterization and the coordination of expression with nitrogen flux. *Plant Physiology* **153**: 1175-1187.
- van der Heijden MGA, Klironomos JN, Ursic M, Moutoglis P, Streitwolf-Engel R, Boller T, Wiemken A, Sanders IR. 1998.** Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* **396**: 69-72.
- Viereck N, Hansen PE, Jakobsen I. 2004.** Phosphate pool dynamics in the arbuscular mycorrhizal fungus *Glomus intraradices* studied by *in vivo* ³¹P NMR spectroscopy. *New Phytologist* **162**: 783-794.

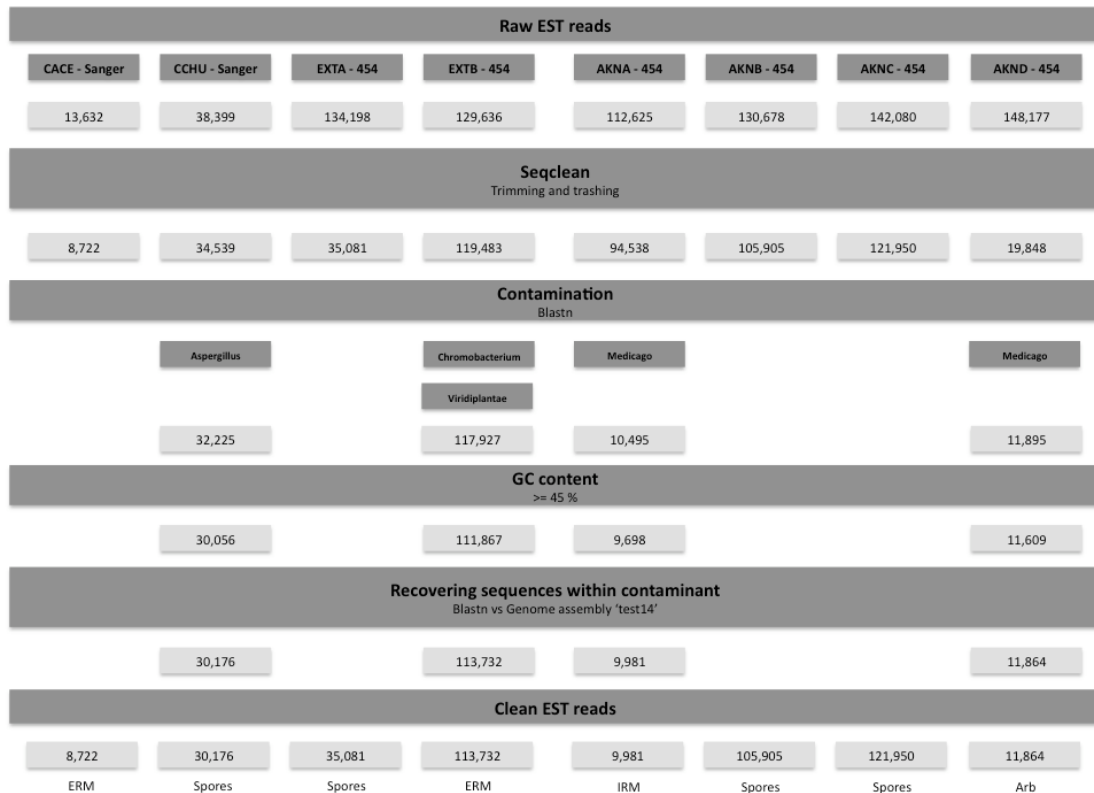


Figure S1. Flow chart describing the *in silico* filtering process of *Glomus intraradices* ESTs. Reads were filtered and trimmed based on their sequence length, low quality, low complexity, occurrence of polyA tails, vector/sequencing adaptors, ribosomal RNA genes, and transposable elements. Contaminant sequences were filtered and removed based on their GC content ($\geq 45\%$) and BLASTN (e-value $\leq 1e-5$, score ≥ 150 , % identity ≥ 95) against *Aspergillus fumigatus*, *Chromobacteria violaceum*, *Medicago truncatula* genome and transcriptome, as well as plant nucleotide databases. Finally, filtered sequences were aligned to the *G. intraradices* draft genome using BLASTN (e-value $\leq 1e-5$, score ≥ 150 , % identity ≥ 95) and kept when sequence identity was $\geq 95\%$. Abbreviations: ERM, extraradical mycelium; IRM, intraradical mycelium; Arb, microdissected arbuscular cells. cDNA library identifications: AKNA, *M. truncatula* IRM; AKNB, germinated spores incubated in apigenin; AKNC, germinated spores incubated in strigolactone GR24; AKND, *M. truncatula* arbuscules; CACE, *Daucus carota* ERM; CCHU, germinated spores; EXTA, germinated spore mix; EXTB, *Daucus carota* ERM. (see Table 1).

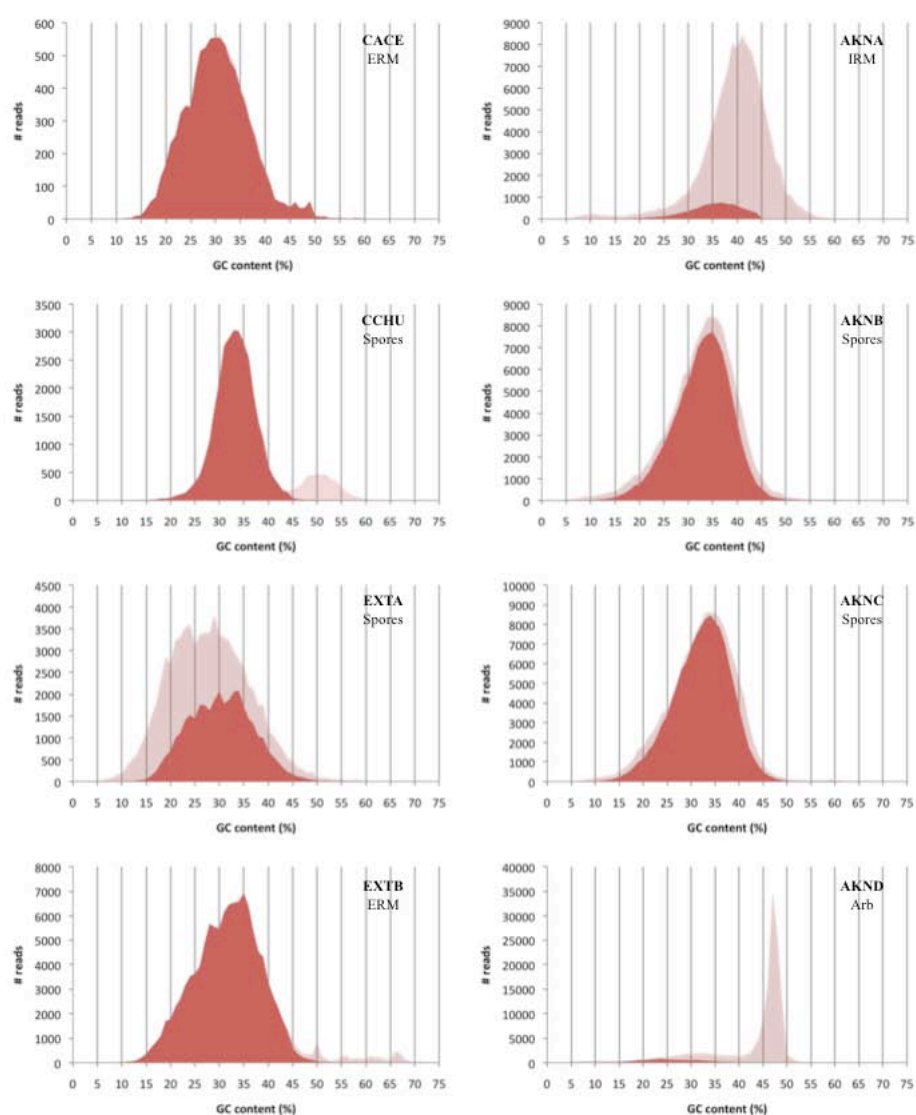


Figure S2. GC% distribution of *Glomus intraradices* raw (pink) and filtered (red) EST sequences for each cDNA library. Abbreviations: ERM, extraradical mycelium; IRM, intraradical mycelium; Arb, microdissected arbuscular cells; for types of spores, see Table 1.

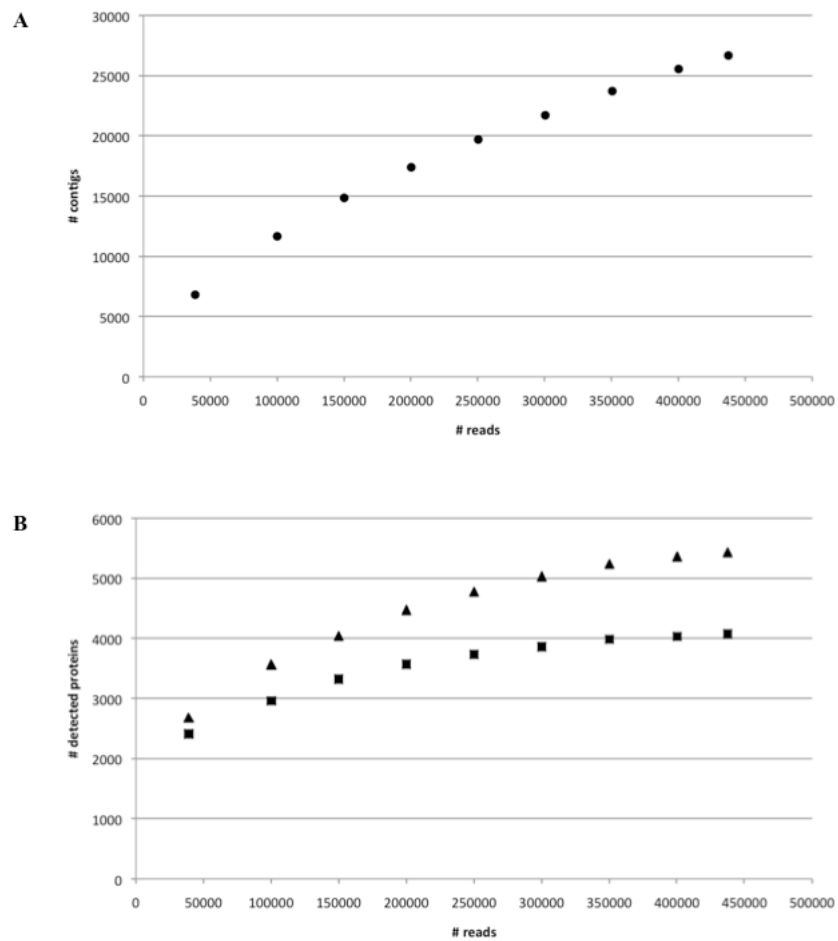


Figure S3. Rarefaction curves depicting the effect of *Glomus intraradices* read number (A) on the number of assembled NRVs and (B) on the number of detected proteins in Swiss-Prot database (triangle) and *R. oryzae* gene repertoire (square). Eight serial assemblies were made, first using the 38,898 Sanger reads, and then by randomly choosing 50,000 454 reads for each assembly. Number of NRVs and number of proteins matched in the *R. oryzae* and Swiss-Prot databases (BLASTX, e-value <1e-5) were then counted for each assembly, resulting in the rarefaction curves.

A

	ASCOMYCOTA						BASIDIOMYCOTA		MUCOROMYCOTINA		
	An	Mg	Nc	Bc	Tm	Lb	Pc	Cn	Um	Ro	Mc
# hits	7656	7619	7708	7745	7957	8058	7766	7354	7444	9035	9265
% hits	28.46	28.32	28.65	28.79	29.57	29.95	28.86	27.33	27.67	33.58	34.44
% identity	45.11	44.76	44.99	44.94	45.94	46.03	46.66	45.39	45.76	49.42	50.15

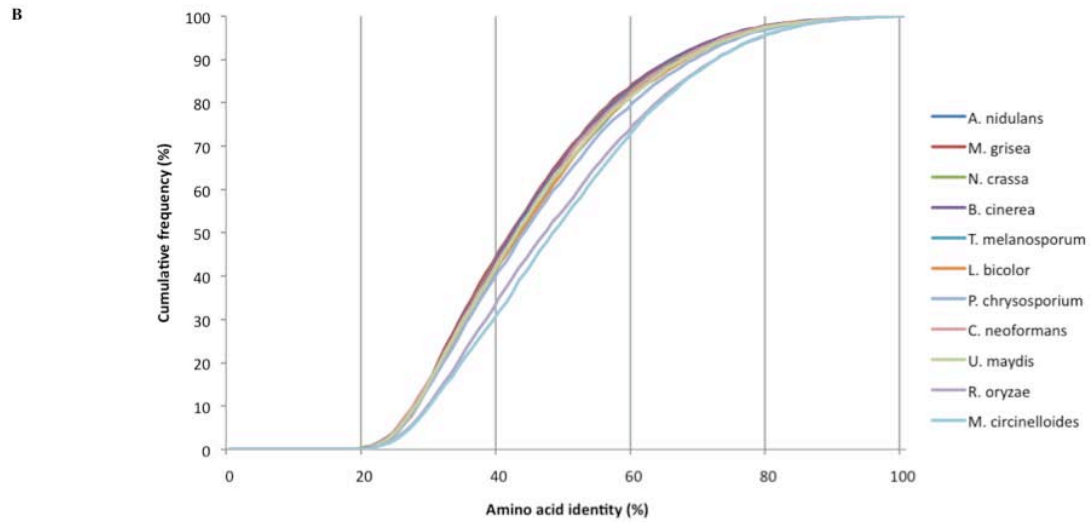


Figure S4. Similarity of *Glomus intraradices* NRVs to gene coding sequences from other selected fungi. (A) % of NRVs that had a good hit in the target genome using BLASTX (e-value $\leq 1e-5$). (B) Cumulative frequencies of amino acid identity across each set of homologous pairs. (An: *Aspergillus nidulans*; Mg: *Magnaporthe grisea*; Nc: *Neurospora crassa*; Bc: *Botrytis cinerea*; Tm: *Tuber melanosporum*; Lb: *Laccaria bicolor*; Pc: *Phanerochaete chrysosporium*; Cn: *Cryptococcus neoformans*; Um: *Ustilago maydis*; Ro: *Rhizopus oryzae*; Mc: *Mucor circinelloides*).

Chapitre IV : Discussion générale

Chapitre IV : Discussion générale

I - Le transcriptome de *Tuber melanosporum* par RNA-Seq

Le séquençage du génome nucléaire de *T. melanosporum* et l'étude de son transcriptome ont permis de dresser un catalogue de 7496 gènes codant des protéines, d'identifier des catégories de gènes dont le rôle pourrait être essentiel pour l'établissement de la symbiose et d'étudier le niveau d'expression de l'ensemble de ces gènes durant différents stades développementaux par l'utilisation de puce à ADN (publication n°1). Ce séquençage a donc permis une avancée considérable dans la compréhension des bases moléculaires de la symbiose. Récemment, le développement du séquençage à haut débit des populations d'ARN (RNA-Seq) a permis d'obtenir une vision très précise du paysage transcriptomique (Mortazavi et al, 2008 ; Nagalakshmi et al, 2008 ; Sultan et al, 2008 ; Wilhelm et al, 2008). Nous avons appliqué cette approche au transcriptome de *T. melanosporum* pour différents stades développementaux majeurs : le mycélium, les fructifications et les ectomycorhizes. Ces séquences ont été utilisées pour quantifier l'expression des gènes (publication n°3) et explorer la complexité fonctionnelle du transcriptome de *T. melanosporum* (publication n°2).

L'alignement des séquences issues du RNA-Seq au génome de référence a permis de générer un profil de couverture des transcrits le long de la séquence génomique. Celui-ci montre que la plupart des régions génomiques ne sont pas transcrites, les transcrits étant confinés pour la plupart (87%) aux gènes codant des protéines. Cependant, nous avons également détecté une fraction importante de lectures de transcrits dans les régions intergéniques (13%), qui pourraient correspondre à des protéines non annotées ou des ARN non-codant. Cette approche de séquençage d'ARN peut, en outre, permettre d'améliorer le processus d'annotation automatisé du génome et d'identifier de nouveaux transcrits (Trapnell et al, 2010 ; Wang et al, 2010b). Plus de 7800 modèles de gènes ont été reconstruits *de novo* à partir de ces données en utilisant G-Mo.R-Se (Denoeud et al, 2008). La comparaison de ces modèles avec les modèles de gènes prédits *ab initio* a montré l'existence de divergence entre ces deux annotations, suggérant des annotations incorrectes. Par exemple, nous avons détecté de nouveaux exons pour plus de 200 gènes. Ces données nous ont également permis d'identifier environ une centaine de nouveaux modèles de gènes avec une structure exon-intron. La plupart de ces nouveaux gènes potentiels sont de petite taille et ne montrent pas de similitudes avec des séquences connues, ce qui expliquerait qu'ils n'aient pas été détectés par l'annotation automatique réalisée par GAZE, le pipeline d'annotation du Génoscope. Certains d'entre eux semblent spécifiques de la phase symbiotique et pourraient donc apporter de nouvelles connaissances sur les mécanismes régulant la symbiose.

Cette approche RNA-Seq nous a également permis d'explorer les événements transcriptionnels et post-transcriptionnels affectant la structure des ARNm. Nous avons détecté des événements d'épissage alternatif pour environ 15% des gènes. Cette proportion importante suggère que l'épissage alternatif joue probablement un rôle majeur dans la diversité protéique et la complexité fonctionnelle de *T. melanosporum*. De plus, certains de ces événements d'épissage alternatif semblent spécifiques des stades développementaux et pourraient donc jouer un rôle important dans l'établissement et le fonctionnement de la symbiose. Toutefois, il faut nuancer ces résultats, par le fait que le type d'événement le plus commun, à savoir la rétention d'intron (94%), est un événement difficile à distinguer des transcrits n'ayant pas encore subi l'épissage présents dans la population d'ARNm. Cette analyse des profils de transcription nous a également permis d'étendre les limites des régions transcrites non traduites (UTRs) qui peuvent contenir d'importants motifs de régulation. Ces régions sont en effet connues pour

jouer un rôle clé dans la régulation post-transcriptionnelle de l'expression génique, influençant le transport, la localisation subcellulaire et le renouvellement des ARNm (Mignone et al, 2002). Nous avons ainsi observé que des transcrits liés à certains processus cellulaires ont des UTRs remarquablement plus longues, ou plus courtes, ce qui influe sur leurs stabilité et donc sur leurs taux de renouvellement. Nous confirmons ainsi les observations faites chez d'autres champignons, tels que *S. pombe* et *A. oryzae* (Wilhelm et al, 2008 ; Wang et al, 2010b). Chez ces deux champignons et chez *T. melanosporum*, les transcrits impliqués dans la transduction du signal et la sporulation ont des UTRs plus longues et donc un taux de renouvellement plus rapide, alors que les transcrits liés à la biogenèse des ribosomes ont des UTRs plus courtes et sont donc probablement plus stables. Cette analyse des UTRs nous a également permis d'identifier des μ ORF courtes situées en amont du codon d'initiation et qui jouent un rôle dans la régulation traductionnelle de l'expression génique chez de nombreux champignons (Mignone et al, 2002; Vilela & McCarthy, 2003). Notre analyse a identifié des μ ORF dans 8% des gènes, indiquant que ce mécanisme de régulation post-transcriptionnelle pourrait également être important chez *T. melanosporum*. Les gènes contenant ces μ ORF codent des protéines associées aux membranes, au transport et au métabolisme des sucres et de l'azote. De plus, certains de ces gènes montrent une expression différentielle lors de la symbiose. Ces μ ORF pourraient donc être impliqués dans leur régulation. Enfin, bien que l'approche de séquençage utilisée ici ne soit pas parfaitement appropriée pour ce type d'analyse, nous avons identifié des transcrits antisens dont certains présentent une expression différentielle lors de la symbiose, rajoutant ainsi un niveau de régulation supplémentaire. Toutefois, notre approche d'identification est basée sur l'utilisation des modèles de gènes G-Mo.R-Se, limitant la détection des transcrits antisens aux gènes ayant une structure exon-intron et une couverture de séquençage suffisante. Une approche de séquençage conservant l'information de l'orientation du brin d'ARN permettrait une analyse plus exhaustive de ce type de transcrit (Martin et Wang, 2011).

Les séquences issues du ARN-Seq, en raison de leur forte profondeur de couverture, permettent d'obtenir une mesure numérique du niveau d'expression des gènes. Lors de notre analyse, nous avons observé que les niveaux d'expression génique déterminés par RNA-seq et par puce à ADN étaient fortement corrélés ($r=0,85$). De plus, la corrélation des ratios d'expression différentielle évalués par les deux plateformes est également correct ($r=0,80$). Toutefois, il existe certaines divergences principalement en raison de la gamme dynamique de mesure beaucoup plus large pour l'approche basée sur le séquençage. Cette approche semble être plus sensible pour la détection des gènes différentiellement exprimés. Toutefois, la comparaison directe du nombre de gènes différentiellement exprimés identifié par les deux plateformes est délicate en raison du manque de répétitions biologiques dans notre analyse basée sur le RNA-Seq. Par ailleurs, contrairement à ce qui a été observé dans d'autres études, le nombre de gènes exprimés détectés par RNA-seq (6986) est plus faible que le nombre de gènes exprimés détectés par puce à ADN (7245). Ceci suggère que la couverture de séquençage est encore trop faible et/ou que nous avons sous-estimé le bruit de fond des signaux de puce à ADN. Une mesure de l'expression de ces gènes par PCR quantitative pourrait déterminer la précision des deux méthodes.

Toutefois, les deux approches utilisées pour l'étude des profils d'expression génique durant les stades développementaux majeurs de *T. melanosporum* donnent la même vision globale. La plupart (95%) des gènes prédits du répertoire de *T. melanosporum* sont exprimés dans au moins un des trois stades de développement. D'autre part, une faible proportion de transcrits (~10%) sont exprimés différentiellement lors de la symbiose et très peu de gènes spécifiques sont détectés. Ceci confirme les patrons d'expression obtenus par le profil transcriptomique de la symbiose chez *L. bicolor*, suggérant un rôle important des mécanismes post-transcriptionnels et traductionnels dans la régulation des gènes au cours du développement symbiotique.

Parmi les transcrits fortement enrichis dans la symbiose, nous avons identifié des transcrits codant des lectines, des protéines contenant un motif LysM, des laccases, des tyrosinases et des lipases. Ces gènes pourraient avoir un rôle dans l'adhésion aux cellules de l'hôte, la détoxification des métabolites de défense de la plante et la colonisation de l'apoplaste racinaire à travers la déconstruction des parois végétales. Plusieurs dizaines de gènes codant des transporteurs transmembranaires impliqués dans le transport de phosphore, d'acides aminés et de sucres sont induits par la symbiose, suggérant une augmentation intense des flux de métabolites à l'interface symbiotique. L'analyse RNA-Seq a également identifié des altérations dans les réseaux d'ARNm impliqués dans le métabolisme des sucres et du glutamate, ce qui est cohérent avec la modification nutritionnelle résultant de la colonisation de la plante. En outre, nous avons observé l'induction de quelques gènes liés au catabolisme des polysaccharides des et de la chitine susceptibles de participer au remodelage des parois cellulaires de la nouvelle interface symbiotique. Le profil de transcription RNA-Seq révèle également l'induction de plusieurs enzymes d'oxydoréduction dans les ECM qui peuvent être utilisées par le mycobionte pour surmonter la réponse de défense de l'hôte.

Les profils d'expression symbiotiques chez les deux champignons ectomycorhiziens *L. bicolor* et *T. melanosporum* se sont révélés être divergents, suggérant que ces deux symbiotes utilisent des mécanismes moléculaires différents pour établir et maintenir la symbiose. Un des rares points communs entre ces champignons ECM est l'induction d'un ensemble de transcrits codant pour des transporteurs de métabolites lors de la symbiose, indiquant que l'apport des nutriments à la plante hôte est un des facteurs clé guidant la symbiose ECM. Ce trait transcriptomique semble pour être l'une des caractéristiques des symbioses. Cependant, alors que *L. bicolor* exprime plutôt des protéines sécrétées de type expansine, lors de la colonisation des tissus racinaires, pour remodeler les parois des cellules végétales (Martin et al, 2008a) *T. melanosporum* est beaucoup plus agressif, exprimant à des niveaux élevés des transcrits codant pour une laccase, une tyrosinase et une lipase. Ces enzymes pourraient aider à briser les liens entre les cellules de la plante, facilitant ainsi la pénétration des hyphes fongiques et la mise en place du réseau de Hartig. Une autre différence frappante entre ces deux symbiotes est l'absence dans le jeu de transcrits fortement induits dans les ectomycorhizes de *T. melanosporum* de petites protéines sécrétées effectrices (MiSSPs). Celles-ci sont très fortement exprimées au sein des mycorhizes de *L. bicolor*. Ces protéines effectrices, qui agissent dans l'espace apoplastique ou dans la cellule végétale pour reprogrammer les voies de signalisation des plantes, jouent un rôle très important dans l'établissement de la symbiose chez *L. bicolor* (Plett et al, 2010). D'autre part, contrairement aux autres champignons ectomycorhiziens *T. melanosporum* possède un gène d'invertase exprimé lors de la symbiose, le rendant moins dépendant du contrôle de la plante pour la conversion du saccharose en glucose (Ceccaroli et al, 2010).

Montanini et al (2010) ont identifié et confirmé par analyse fonctionnelle (« transcriptional activator trap ») l'activité nucléaire de plusieurs facteurs de transcription chez *T. melanosporum*. Les gènes codant un facteur de transcription impliqué dans la réponse au glucose (*TmelAzf1*), un activateur de la dégradation du xylane (*TmelXLNR*), ainsi que plusieurs protéines régulatrices du métabolisme de l'azote (*TmelAreA*, *TmelTamA*, *TmelAreB*, *TmelNirA-3*, *TmelArcA*) sont induits lors de la symbiose. Ces facteurs de transcription pourraient faire partie du réseau de contrôle transcriptionnel de la symbiose.

Bien que la plupart des gènes ne soit pas affectés par l'établissement des ECM, cette analyse des profils de transcription confirme que le développement de la symbiose chez l'ascomycète ectomycorhizien *T. melanosporum* englobe une série de processus génétiques complexes. L'image globale du transcriptome symbiotique de *T. melanosporum* est conforme aux conclusions tirées par l'analyse d'autres systèmes ectomycorhiziens (Voiblet et al, 2001 ; Peter et al, 2003 ; Johansson et al, 2004 ; Duplessis et al, 2005 ; Le Quéré et al, 2005 ; Heller et

al, 2008 ; Martin et al, 2008a). De plus, le séquençage à haut débit de transcrits a également permis de montrer que les grandes transitions développementales chez *T. melanosporum* sont également accompagnées par une série de mécanismes de régulation post-transcriptionnelle et traductionnelle (épissage alternatif, μ ORF), ajoutant ainsi un niveau supplémentaire de complexité dans la régulation des réseaux de gènes. Ces résultats, en fournissant une vision globale du transcriptome, ouvrent la voie à une meilleure compréhension, *via* l'analyse fonctionnelle des gènes candidats, des processus moléculaires qui sous-tendent la formation de la symbiose et de l'ascocarpe.

II - Caractérisation du transcriptome de *Glomus intraradices*

La symbiose mycorhizienne arbusculaire est certainement la symbiose eucaryotique la plus importante au sein des écosystèmes naturels et agricoles. Plusieurs études ont été réalisées sur le transcriptome de la plante hôte lors du développement de la symbiose. Toutefois, en l'absence de séquence génomique, aucune étude transcriptomique à grande échelle n'a encore été réalisée sur le partenaire fongique. Afin d'avoir un aperçu des bases moléculaires associées à la symbiose AM pour le partenaire fongique, nous avons entrepris la première analyse du transcriptome de *G. intraradices* à l'échelle du génome entier. Notre objectif était d'inventorier le répertoire de gènes de ce champignon et de mettre en évidence d'éventuels changements de profil d'expression génique lors des principaux stades de développement. Pour ce faire, des banques d'ADNc provenant de spores, d'ERM, d'IRM et d'arbuscules de *G. intraradices* ont été séquencées par séquençage Sanger et 454 et plus de 400000 lectures ont été assemblées en un jeu de 25000 séquences de transcrits non redondantes. Ces séquences ont ensuite été utilisées pour construire une puce à ADN et étudier la régulation du transcriptome à différents stades développementaux (publication n°4).

Bien qu'il n'existe pas de méthode établie pour apprécier la qualité d'un assemblage de transcrits, en particulier pour les organismes non modèle pour lesquels il n'existe pas de génome séquencé, nous avons tenté d'estimer l'exhaustivité de notre jeu de séquences. Pour cela, nous nous sommes basé sur un jeu de gènes conservés chez les eucaryotes (« Core Eucaryotic Genes ») utilisé pour mesurer le niveau de qualité des projets de séquençage de génomes (Parra et al, 2009). Nous avons également réalisé différents assemblages, que nous avons comparé à des séquences protéiques connues répertoriées dans Uniprot, afin d'obtenir des courbes de raréfaction. Ces deux approches montrent que notre inventaire de transcrits est relativement complet. Toutefois, ces deux tests sont basés sur la comparaison à des séquences connues. Les gènes conservés ayant souvent des fonctions importantes, ils sont exprimés à un niveau élevé. Il est donc possible que des transcrits faiblement exprimés aient échappé à notre analyse.

Seulement 40% des transcrits de *G. intraradices* présentent une similarité de séquence avec des séquences connues, ce qui reflète le caractère ancestral de ce champignon et la carence des bases de données en séquences provenant des Gloméromycètes. La plupart des transcrits orphelins de *G. intraradices* ont toutefois une homologie avec les fragments de la séquence génomique (80%) suggérant qu'il s'agit bien de nouveaux gènes spécifiques de ce clade. Environ la moitié des transcrits ayant une similarité de séquence dans les bases de données la partagent avec toutes les autres espèces fongiques analysées et sont donc des gènes conservés au sein des Mycota. Fait intéressant, *G. intraradices* partage quelques séquences uniquement avec l'un ou l'autre des symbiotes mycorhiziens séquencés, l'ascomycète *T. melanosporum* ou le basidiomycète *L. bicolor*. Ces gènes codent principalement pour des protéines hypothétiques n'ayant pas de fonction connue. De plus, lorsque l'on compare le transcriptome de *G. intraradices* avec d'autres répertoires de gènes fongiques, il montre une surreprésentation des domaines impliqués dans les voies de signalisation, ce qui est également une caractéristique de *L. bicolor*. Ces gènes pourraient donc être impliqués dans les voies de signalisation liées à l'établissement de la symbiose.

En se basant sur les transcrits ayant une similarité de séquence avec des séquences connues, nous constatons que de nombreux transcrits sont représentés par deux ESTs correspondant aux extrémités 5' et 3', ce qui indique que la couverture de séquençage actuelle est encore trop faible pour obtenir des transcrits pleine longueur. Le polymorphisme des séquences d'ADN chez *G. intraradices* est connu pour un certain nombre de gènes (Croll et al, 2008 ; Sanders et Croll, 2010). Nous avons donc estimé celui-ci au sein du transcriptome de

G. intraradices en utilisant la dernière étape de l'algorithme d'assemblage MIRA qui rassemble les transcrits présentant des variations nucléotidiques (SNP) issus d'un même gène. La présence de multiples SNP dans plusieurs centaines de transcrits suggère que ce polymorphisme nucléotidique est très répandu dans le génome (Figure 2, publication n°4). Plusieurs variants des gènes polymorphes sont exprimés, suggérant que ce polymorphisme pourrait avoir une importance fonctionnelle.

Afin de caractériser les voies métaboliques de *G. intraradices*, nous avons comparé les transcrits à la base de données des voies métaboliques de KEGG (Moriya et al, 2007). La plupart des enzymes du métabolisme primaire est codé au sein du transcriptome de *G. intraradices*, ainsi que les gènes codant pour l'assimilation du nitrate et du sulfate, habituellement absent chez les pathogènes biotrophes obligatoires (Baxter et al, 2010 ; Spanu et al, 2010 ; Duplessis et al, 2011 ; Kemen et al, 2011). Nous avons également identifié de nombreux transcrits codant différentes protéines prédites associées à l'absorption et l'assimilation des éléments minéraux du sol. Il semble donc que *G. intraradices* ait conservé la capacité d'assimiler les éléments nutritifs de son environnement. Toutefois, nous avons constatés l'absence de certains gènes tels que ceux impliqués dans la voie de biosynthèse de la thiamine, ce qui est un trait de caractère partagé avec les pathogènes obligatoires et les ectomycorhiziens. Aucun transcrit codant l'invertase fongique n'a été détecté au sein du transcriptome de *G. intraradices*, suggérant que le champignon dépendrait de la plante pour l'hydrolyse du saccharose, tout comme de nombreux champignons ectomycorhiziens (Parrent et al, 2009). Nous avons également constaté l'absence de transcrits codant les enzymes dégradant les polysaccharides des parois cellulaires végétales, notamment les cellulases des familles GH6 et GH7, comme observé chez les symbiotes ectomycorhiziens (Martin et al, 2008a, 2010 ; Nagendran et al, 2009). Ce jeu réduit d'enzymes hydrolytiques pourrait donc constituer une adaptation évolutive pour éviter la libération de fragments polysaccharidiques et leur détection par le système de défense de l'hôte lors de la phase biotrophique du champignon.

Les transcrits assemblés ont été utilisées pour construire une puce à ADN portant des oligonucléotides complémentaires des séquences d'ARNm. Celle-ci a permis de suivre les profils d'expression génique des structures symbiotiques (ERM et IRM). Une proportion très faible de gènes (~1 à 2%) montre une expression différentielle dans les tissus symbiotiques par rapport aux spores germés. La majorité des gènes induits dans l'IRM et l'ERM sont des gènes orphelins. Les gènes les plus fortement induits lors de la symbiose sont impliqués notamment dans le métabolisme des lipides et du phosphate, le métabolisme secondaire, la transduction du signal, la biosynthèse des parois cellulaires et le mécanisme de réplication de l'ADN. De nombreux gènes codant pour des transporteurs sont également induits lors de la symbiose, comme c'est le cas pour les champignons ectomycorhiziens. De plus, le profil d'expression de gènes codant les enzymes d'assimilation de l'azote et du phosphore va dans le sens d'une absorption intense d'éléments minéraux, de leur assimilation et leur transfert vers la plante-hôte.

L'induction des transcrits codant pour le transporteur de nitrate et la nitrate réductase, ainsi que le transporteur d'ammonium GintAMT2 (Pérez-Tienda et al, 2011), pourrait favoriser l'accumulation de l'ammonium dans les arbuscules et ainsi, l'incorporation dans les acides aminés exportés vers les cellules de l'hôte. Toutefois, la forte expression de gènes codant pour l'arginase, l'uréase et les protéines accessoires de l'uréase (UreG) dans l'IRM devrait conduire à une forte libération d'ammonium permettant d'alimenter l'efflux passif de l'ammonium à l'interface apoplastique racinaire (Govindarajulu et al, 2005 ; Gomez et al, 2009 ; Tian et al, 2010). L'expression élevée des transcrits codant pour des transporteurs d'acides aminés dans l'ERM et l'IRM suggère que le champignon peut absorber les acides aminés présents dans le sol et dans l'interface apoplastique.

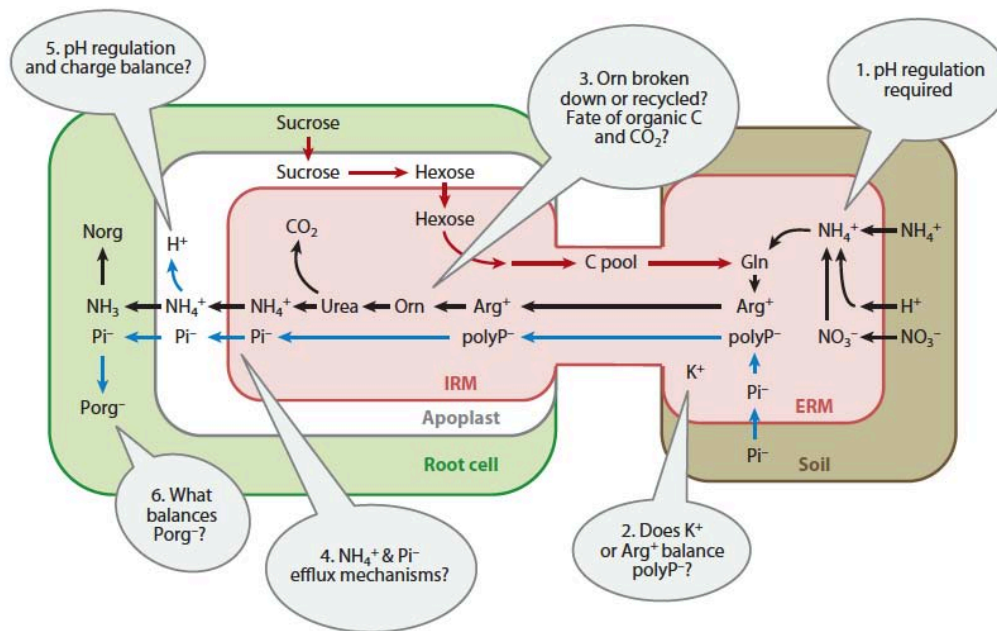


Figure 22 : Représentation hypothétique du transfert de l'azote (flèches noires), du carbone (flèches rouge) et du phosphate (flèches noires) chez les champignons mycorhiziens à arbuscules. L'ammonium (NH_4^+) et/ou le nitrate (NO_3^-) sont absorbés par le mycélium extraradical puis assimilés en glutamine (Gln), puis en arginine (Arg^+). L'assimilation va générer un excès de H^+ et/ou de OH^- de sorte que la régulation du pH est nécessaire. L'équilibre de charge ionique au cours la translocation de l'arginine vers le mycélium intraradical pourrait être maintenue par la charge négative du polyphosphate (polyP^-), mais l'ion K^+ est une alternative possible. L'arginine serait ensuite dégradée, générant du CO_2 , du nitrate et de l'ammonium. Les mécanismes d'efflux du phosphate et de l'ammonium vers l'apoplaste interfaciales sont inconnus. Le transfert de l'ammonium dans les cellules végétales et son assimilation génère des ions H^+ et de sorte que la régulation du pH est nécessaire. Les ions équilibrant les charges négatives du phosphate organique (Porg^-) dans les cellules végétales sont inconnus. (Smith et Smith, 2011).

Les transcrits codant plusieurs transporteurs de Pi à forte et faible affinité sont exprimés dans les spores, l'ERM et l'IRM. La forte activité de *GintPT* dans l'IRM suggère que le champignon peut dans une certaine mesure réabsorber le Pi libéré dans l'espace apoplastique. Comme l'a montré Kiers et al. (2011), le champignon exerce un contrôle sur l'allocation du Pi à sa plante-hôte ; apport qui est récompensé par la fourniture de sucres, stabilisant ainsi la coopération au sein de la symbiose mycorhizienne. La synthèse rapide de polyP qui accompagne l'absorption intense de Pi est cruciale pour le maintien de l'absorption du Pi par les hyphes (Tani et al, 2009). Le transcrit codant pour l'endopolyphosphatase Ppn1p est fortement surexprimés dans les IRM confirmant la dégradation de ce polymère au niveau de l'arbuscule (Figure 22).

Nous avons également réalisé une analyse d'expression numérique basée sur le nombre de lectures 454, qui nous a permis de détecter plusieurs transcrits codant des petites protéines sécrétées (SSPs) putatives exprimées spécifiquement dans les tissus symbiotiques. Des études récentes sur les interactions plantes-pathogènes et ectomycorhiziennes montrent que beaucoup de gènes orphelins codant des protéines effectrices jouent des rôles clés dans la colonisation de l'hôte en contrôlant le système immunitaire des plantes (de Wit et al, 2009 ; Plett et al, 2011). La protéine effectrice SP7 de *G. intraradices* interagit avec le facteur de transcription nucléaire ERF19 contrôlant les réactions de défense de la plante hôte (Kloppholz et al, 2011). Ce type d'effecteur sécrété joue donc un rôle clé dans l'établissement de la symbiose. Cet ensemble de SSPs exprimées spécifiquement lors de la symbiose pourraient donc être des gènes clés pour

l'établissement de la symbiose AM et sont de bons candidats pour de futures analyses fonctionnelles.

Cette caractérisation du transcriptome de *G. intraradices* a permis d'établir le premier répertoire de gènes à l'échelle du génome pour un champignon AM. L'analyse des profils d'expression nous a permis d'identifier les gènes qui pourraient être impliqués dans la symbiose mycorhizienne arbusculaire. Parmi les résultats marquants de cette étude, il faut souligner l'existence de caractéristiques communes aux profils d'expression de *G. intraradices* et des champignons ectomycorhiziens *L. bicolor* et *T. melanosporum*. En particulier, l'induction des gènes codant pour des transporteurs membranaires et des SSPs durant l'interaction symbiotique, et l'absence d'expression des enzymes hydrolytiques agissant sur les polysaccharides de la paroi végétale. En l'absence de séquence génomique complète de Glomeromycètes, ces données de transcriptomique à grande échelle représentent une ressource solide pour la compréhension de la symbiose mycorhizienne arbusculaire et pour de futures recherches fonctionnelles.

Chapitre V : Conclusions et perspectives

Chapitre V : Conclusions et perspectives

- Conclusions biologiques :

Le travail entrepris dans le cadre de ce projet de thèse visait à caractériser *in silico* le transcriptome symbiotique des champignons ectomycorhizien *T. melanosporum* et endomycorhizien *G. intraradices*. Cette étude a permis de caractériser des modifications dans l'expression génique lors l'établissement de la symbiose chez ces organismes et de mettre en évidence plusieurs points communs entre les patrons d'expression des différents symbiotes. Ainsi, bien que ces différents champignons mettent en place des stratégies plus ou moins agressives pour coloniser leurs hôtes, certaines fonctions cellulaires semblent être partagées par ces symbiotes. Celles-ci pourraient refléter différents niveaux de contrôle clés s'exerçant entre les partenaires végétaux et fongiques. La principale caractéristique commune à ces interactions mutualistes est l'induction de transporteurs membranaires et des enzymes assimilatrices nécessaires à l'échange réciproque de nutriments. Par exemple, l'activité invertase de la plante pourrait lui permettre de réguler la quantité de carbone allouée au champignon (Nehls et al, 2007). Les petites protéines sécrétées effectrices (SSPs), telles que MiSSP7 et SP7, permettant au champignon de manipuler les défenses des cellules végétales, sont une autre voie de contrôle impliquée dans la symbiose chez *L. bicolor* et *G. intraradices* (Plett et al, 2011 ; Kloppeholz et al, 2011). Toutefois, les SSPs ne semblent pas jouer un rôle crucial lors du développement de la symbiose chez *T. melanosporum*. Des SSPs régulées par la symbiose sont induites chez le Peuplier lors de l'interaction avec *L. bicolor* (Ting Li et al, comm. pers.). Leur caractérisation devrait apporter de nouvelles informations sur les voies de signalisation jouant un rôle important dans la régulation de la symbiose. En effet, *G. intraradices* et *L. bicolor* montrent tous les deux une surreprésentation des domaines impliqués les voies de traduction du signal (ex. tyrosine kinases). Il est urgent d'étudier la contribution de ces voies de signalisation dans le fonctionnement de la symbiose. De plus, étant donné l'importance de ces voies de signalisation pour le partenaire végétal de la symbiose mycorhizienne arbusculaire, celles-ci devraient être également explorées chez les plantes formant des ectomycorhizes.

La régulation transcriptionnelle n'est pas le seul mécanisme impliqué dans la régulation génique du développement symbiotique. En effet, l'analyse du transcriptome de *T. melanosporum* suggère que des mécanismes de régulation post-transcriptionnels et traductionnels pourraient jouer un rôle clef dans la régulation de l'expression génique et donc dans le fonctionnement de la symbiose. Afin d'obtenir une meilleure vision des mécanismes qui sous-tendent la symbiose mycorhizienne, il est donc nécessaire d'étudier la régulation traductionnelle à l'échelle génomique.

- Conclusions méthodologiques :

En parallèle, le principal objectif méthodologique de ce travail était l'analyse *in silico* des ressources transcriptomiques issues des nouvelles technologies de séquençage (454, Illumina). En effet, ces nouvelles approches de séquençage ont créé de nouveaux challenges bio-informatiques, déplaçant les limitations de l'analyse génomique de la production de données expérimentales à leur analyse statistique et informatique. En pratique, différents outils et procédures bio-informatiques sont mis en place pour permettre le traitement et l'exploitation de ces données. Ceux-ci sont actuellement utilisés pour analyser de nouveaux jeux de données transcriptomiques. Au cours de ce travail, différentes bases de données interactives compilant les données acquises sur ces transcriptomes ont également été mises en place.

Ces approches de séquençage à haut débit du transcriptome offrent de réels avantages pour explorer les interactions plantes-champignons. Tout d'abord, la forte couverture de séquençage autorisée par le RNA-Seq permet la détection d'une large gamme de transcrits, indépendamment du niveau d'expression. La nature numérique de ces approches permet également une mesure plus précise de l'expression. De plus, ces approches offrent un accès direct à la séquence, fournissant ainsi des informations importantes, notamment sur les événements post-transcriptionnels. Enfin, les transcrits sont générés et quantifiés à partir du même ensemble de données, sans la nécessité d'un génome de référence. Dans le cas de la symbiose, ces approches permettent également l'analyse simultanée des transcrits des deux partenaires au sein du même échantillon.

Cependant, des défis importants restent à relever en ce qui concerne ces approches. Tout d'abord, le volume de données produit par ces approches nécessite une infrastructure informatique appropriée ayant une capacité de stockage, de calcul et de RAM très élevée. Alternativement, l'utilisation du « cloud computing » qui permet de délocaliser le traitement informatique des données sur des serveurs distants pourrait être une solution efficace si le transfert de données brutes de séquençage n'était pas aussi long. Par ailleurs, la dépense engagée par échantillon est encore un facteur limitant pour ces approches, ce qui restreint parfois l'usage de répétitions biologiques nécessaires à l'analyse statistique. Toutefois, avec l'amélioration en continue des technologies de séquençage et le développement des techniques de multiplexage, le coût par échantillon pourrait être considérablement réduit (<200€/échantillon). En outre, la profondeur de séquençage requise pour une couverture suffisante du transcriptome, qui détermine la précision de l'analyse, est assez difficile à estimer *a priori*. Pour des transcriptomes complexes, un minimum de 100 à 200 millions de séquences (10 Gbp) est cependant recommandé (RNA-Seq Minimal Standards and Guidelines). Enfin, ces approches étant à un stade de développement précoce, les outils bio-informatiques sont continuellement en développement, de sorte qu'il est difficile de normaliser la procédure d'analyse. Le développement de méthodes statistiques spécifiques et standardisées pour la normalisation et la détection de l'expression différentielle est également essentielle pour traiter efficacement les données issues de ces approches.

Une autre considération importante pour l'analyse des transcriptomes est l'annotation fonctionnelle. L'approche couramment utilisée est de déduire la fonction des gènes par la recherche de similarité de séquence avec un ensemble d'organismes modèles. Toutefois, le transfert des fonctions des gènes d'un organisme modèle à un autre organisme conduit souvent à des erreurs d'interprétation et n'apportent pas d'information sur les gènes n'ayant pas de similarité de séquence. La seule stratégie pour comprendre la fonction cellulaire et biochimique des gènes est leur caractérisation fonctionnelle expérimentale. La plupart des gènes régulés par la symbiose chez les champignons mycorhiziens n'ont pas d'homologues au sein des champignons modèles couramment étudiés, tels que la Levure ou *Neurospora crassa*. Il est donc essentiel de caractériser leurs fonctions. Ceci est rendu possible par notre capacité à manipuler génétiquement différents champignons mycorhiziens (*L. bicolor*, *Oidiodendron maius*, *Hebeloma cylindrosporum*). À l'avenir, des progrès importants dans la compréhension de la symbiose pourraient découler de la combinaison des approches d'inactivation de gènes avec la puissance des approches de séquençage à haut débit de transcrits et la microdissection laser de tissus symbiotiques.

- Perspectives :

Avec l'arrivée des nouvelles technologies de séquençage, plusieurs perspectives s'ouvrent pour améliorer la compréhension des mécanismes moléculaires de la symbiose mycorhizienne.

- Bien que *G. intraradices* ait été choisi, en 2004, comme candidat pour le premier projet de séquençage de champignon mycorhizien arbusculaire, l'assemblage de son génome s'est révélé problématique. Ces difficultés sont dûes au niveau de polymorphisme élevé chez ce champignon. Toutefois, d'après une récente estimation, la taille de son génome (150 Mb) s'est révélée être 10 fois supérieure à celle précédemment estimée (Sędziewska et al, 2011). De nouvelles données de séquences génomiques issues des plateformes 454 et Illumina ont récemment été obtenues. Ces données, avec celles précédemment acquises, feront l'objet d'un nouvel assemblage qui permettra, en premier lieu, d'estimer la quantité de données nécessaires pour un assemblage correct du génome. Des filtres seront tout d'abord appliqués aux données brutes afin d'éliminer les contaminants potentiels. Différents assembleurs et paramètres d'assemblage seront ensuite utilisés pour obtenir un assemblage optimum. La réussite de ce projet de séquençage conduira une meilleure compréhension de la symbiose mycorhizienne arbusculaire.
- Afin d'explorer la diversité génétique des champignons, le programme FGP (Fungal Genomic Program) a pour objectif le séquençage de centaines de génomes d'espèces fongiques provenant de clades et de niches écologiques variés. Dans le cadre de ce programme, notre équipe coordonne l'analyse des génomes de 30 champignons saprotrophes et 25 champignons symbiotiques mycorhiziens. L'analyse comparative des répertoires de gènes, de la structure des génomes, des régions régulatrices et des profils expression génique de ces différents organismes permettra d'expliquer les traits de vie de ces champignons et d'identifier des mécanismes et des gènes impliqués dans le fonctionnement et l'évolution de la symbiose mycorhizienne.
- Enfin, à plus long terme, le nombre grandissant de jeux de données d'expression acquis pour de nombreuses conditions différentes, nous permet d'entrevoir de nouvelles méthodes d'analyse pour l'étude de la symbiose mycorhizienne, telles que les analyses de réseaux de co-expression génique. Ces méthodes permettent de réduire la complexité des données, de mettre à jour des réseaux de gènes co-exprimés et d'identifier des gènes centraux susceptibles de réguler ces réseaux, tels que des facteurs de transcription. De plus, les différents types de données expérimentales générés les approches à haut débit permettent d'envisager des méthodes de fouilles de données avancées à grande échelle. Ainsi, ces méthodes exploratoires permettent d'intégrer différents types de données, telles que des données de séquences (annotation structurale des gènes, sites de régulation), des données d'expression (gènes co-régulés, expression différentielle), des données d'interaction (par exemple, issues d'expériences de ChIP-Seq), des données d'annotation fonctionnelle (description des fonctions et des voies métaboliques). Elles peuvent ainsi générer des hypothèses biologiques, par exemple, prédire les réseaux de régulation, les interactions protéine-protéine ou les fonctions des gènes. De telles approches pourraient être des outils puissants pour la découverte de nouveaux mécanismes moléculaires contrôlant la symbiose mycorhizienne.

Liste des publications

Liste des publications

Perigord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. 2010. Martin F, Kohler A, Murat C, Balestrini R, Coutinho PM, Jaillon O, Montanini B, Morin E, Noel B, Percudani R, Porcel B, Rubini A, Amicucci A, Amselem J, Anthouard V, Arcioni S, Artiguenave F, Aury JM, Ballario P, Bolchi A, Brenna A, Brun A, Buée M, Cantarel B, Chevalier G, Couloux A, Da Silva C, Denoeud F, Duplessis S, Ghignone S, Hilselberger B, Iotti M, Marçais B, Mello A, Miranda M, Pacioni G, Quesneville H, Riccioni C, Ruotolo R, Splivallo R, Stocchi V, **Tisserant E**, Viscomi AR, Zambonelli A, Zampieri E, Henrissat B, Lebrun MH, Paolocci F, Bonfante P, Ottonello S, Wincker P. *Nature*. 464:1033-8.

Laser capture microdissection of uredinia formed by *Melampsora larici-populina* revealed a transcriptional switch between biotrophy and sporulation. 2010. Hacquard S, Delaruelle C, Legué V, **Tisserant E**, Kohler A, Frey P, Martin F, Duplessis S. *Mol Plant Microbe Interact*. 23:1275-86.

Comparative transcriptomics of drought responses in *Populus*: a meta-analysis of genome-wide expression profiling in mature leaves and root apices across two genotypes. 2010. Cohen D, Bogeat-Triboulot MB, **Tisserant E**, Balzergue S, Martin-Magniette ML, Lelandais G, Ningre N, Renou JP, Tamby JP, Le Thiec D, Hummel I. *BMC Genomics*.

Distribution and localization of microsatellites in the Perigord black truffle genome and identification of new molecular markers. 2010. Murat C, Riccioni C, Belfiori B, Cichocki N, Labbé J, Morin E, **Tisserant E**, Paolocci F, Rubini A, Martin F. *Fungal Genet Biol*. 48:592-601.

Genome-wide inventory of metal homeostasis-related gene products including a functional phytochelatin synthase in the hypogeous mycorrhizal fungus *Tuber melanosporum*. 2010. Bolchi A, Ruotolo R, Marchini G, Vurro E, di Toppi LS, Kohler A, **Tisserant E**, Martin F, Ottonello S. *Fungal Genet Biol*. 48:573-84.

Deep RNA sequencing improved the structural annotation of the *Tuber melanosporum* transcriptome. 2011. **Tisserant E**, Da Silva C, Kohler A, Morin E, Wincker P, Martin F. *New Phytol*. 189:883-91.

Genome wide search and functional identification of transcription factors in the mycorrhizal fungus *Tuber melanosporum*. 2011. Montanini B, Levati E, Bolchi A, Kohler A, Morin E, **Tisserant E**, Martin F, Ottonello S. *New Phytol*. 189:736-50.

Genomic profiling of carbohydrate metabolism in the ectomycorrhizal fungus *Tuber melanosporum*. 2011. Ceccaroli P, Buffalini M, Saltarelli R, Barbieri E, Polidori E, Ottonello S, Kohler A, **Tisserant E**, Martin F, Stocchi V. *New Phytol*. 189:751-64.

Isolation and characterization of MAT genes in the symbiotic ascomycete *Tuber melanosporum*. 2011. Rubini A, Belfiori B, Riccioni C, **Tisserant E**, Arcioni S, Martin F, Paolocci F. *New Phytol*. 189:710-22.

454-pyrosequencing of *Coffea arabica* leaves infected by the rust fungus *Hemileia vastatrix* reveals in planta-expressed pathogen-secreted proteins and plant functions in a late compatible plant-rust interaction. 2011. Fernandez D, **Tisserant E**, Talhinhos P, Azinheira H, Vieira A, Petitot AS, Loureiro A, Poulain J, DA Silva C, Silva MD, Duplessis S. *Mol Plant Pathol*. doi: 10.1111/j.1364-3703.2011.00723.x.

***Melampsora larici-populina* transcript profiling during germination and timecourse infection of poplar leaves reveals dynamic expression patterns associated with virulence and biotrophy.** 2011. Duplessis S, Hacquard S, Delaruelle C, Tisserant E, Frey P, Martin F, Kohler A. Mol Plant Microbe Interact. 24:808-18.

Obligate biotrophy features unraveled by the genomic analysis of rust fungi. 2011. Duplessis S, Cuomo CA, Lin YC, Aerts A, Tisserant E, Veneault-Fourrey C, Joly DL, Hacquard S, Amselem J, Cantarel BL, Chiu R, Coutinho PM, Feau N, Field M, Frey P, Gelhaye E, Goldberg J, Grabherr MG, Kodira CD, Kohler A, Kües U, Lindquist EA, Lucas SM, Mago R, Mauceli E, Morin E, Murat C, Pangilinan JL, Park R, Pearson M, Quesneville H, Rouhier N, Sakthikumar S, Salamov AA, Schmutz J, Selles B, Shapiro H, Tanguay P, Tuskan GA, Henrissat B, Van de Peer Y, Rouzé P, Ellis JG, Dodds PN, Schein JE, Zhong S, Hamelin RC, Grigoriev IV, Szabo LJ, Martin F. Proc Natl Acad Sci U S A. 108:9166-71.

The transcriptome of the arbuscular mycorrhizal fungus *Glomus intraradices* (DAOM 197198) reveals functional tradeoffs in an obligate symbiont. 2011. E Tisserant, A Kohler, P Dozolme-Seddas, R Balestrini, K Benabdellah, A Colard, D Croll, C Da Silva, SK Gomez, R Koul, N Ferrol, V Fiorilli, D Formey, Ph Franken, N Helber, M Hijri, L Lanfranco, E Lindquist, Y Liu, M Malbreil, E Morin, J Poulain, H Shapiro, D van Tuinen, A Waschke, C Azcón-Aguilar, G Bécard, Bonfante P, MJ Harrison, P Lammers, U Paszkowski, N Requena, SA Rensing, C Roux, IR Sanders, Y Shachar-Hill, G Tuskan, JPW Young, V Gianinazzi-Pearson, F Martin. New Phytol. *In press*.

Références bibliographiques

Références bibliographiques

- 1000 Genomes Project Consortium. 2010. A map of human genome variation from population-scale sequencing. *Nature*. 467:1061–73.
- Akiyama K, Matsuzaki KI, Hayashi H. 2005. Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature* 435: 824-827.
- Allen JW, Shachar-Hill Y. 2009. Sulfur transfer through an arbuscular mycorrhiza. *Plant Physiology* 149: 549-560.
- 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 Res.* 25:3389-3402.
- Alvarez LA, Exton DA, Timmis KN, Suggett DJ, McGenity TJ. 2009. Characterization of marine isoprene-degrading communities. *Environ Microbiol.* 11: 3280–3291.
- Angelard C, Colard A, Niculita-Hirzel H, Croll D, Sanders IR. 2010. Segregation in a mycorrhizal fungus alters rice growth and symbiosis-specific gene transcription. *Current Biology* 20: 1216-1221.
- Aono T, Maldonado-Mendoza IE, Dewbre GR, Harrison MJ, Saito M. 2004. Expression of alkaline phosphatase genes in arbuscular mycorrhizas. *New Phytol.* 162:525–553.
- Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*. 408:796–815.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* 25: 25-29.
- Audic S, Claverie JM. 1997. The significance of digital gene expression profiles. *Genome Research* 7: 986-995.
- Bago B, Pfeffer PE, Shachar-Hill Y. 2000. Carbon metabolism and transport in arbuscular mycorrhizas. *Plant Physiology* 124: 949-957.
- Bago B, Zipfel W, Williams R, Jun J, Arreola R, Lammers P, Pfeffer PE, Shachar-Hill Y. 2002. Translocation and utilization of fungal lipid in the arbuscular mycorrhizal symbiosis. *Plant Physiol.* 128:108–124.
- Bago B, Pfeffer PE, Abubaker J, Jun J, Allen JW, Brouillette J, Douds DD, Lammers PJ, Shachar-Hill Y. 2003. Carbon export from arbuscular mycorrhizal roots involves the translocation of carbohydrate as well as lipid. *Plant Physiology* 131: 1496-1507.
- Balestrini R, Lanfranco L. 2006. Fungal and plant gene expression in arbuscular mycorrhizal symbiosis. *Mycorrhiza* 16: 509-524.
- Balestrini R, Gómez-Ariza J, Lanfranco L, Bonfante P. 2007. Laser microdissection reveals that transcripts for five plant and one fungal phosphate transporter genes are contemporaneously present in arbusculated cells. *Molecular Plant-Microbe Interactions* 20: 1055-1062.
- Baxter L, Tripathy S, Ishaque N, Boot N, Cabral A, Kemen E, Thines M, Ah-Fong A, Anderson R, Badejoko W, et al. 2010. Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. *Science* 330:1549-1551.
- Bécard G, Fortin JA. 1988. Early events of vesicular arbuscular mycorrhiza formation on Ri T-DNA transformed roots. *New Phytologist* 108: 211-218.

- Bécard G, Piché Y. 1989. Fungal growth stimulation by CO₂ and root exudates in the vesicular-arbuscular mycorrhizal symbiosis. *Appl Environ Microbiol.* 55:2320–2325.
- Bécard G, Kosuta S, Tamasloukht M, Séjalon-Delmas N, Roux C. 2004. Partner communication in the arbuscular mycorrhizal interaction. *Canadian Journal of Botany* 8 : 1186-1197.
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* 340: 783-795.
- Benedetto A, Magurno F, Bonfante P, Lanfranco L. 2005. Expression profiles of a phosphate transporter gene (GmosPT) from the endomycorrhizal fungus *Glomus mosseae*. *Mycorrhiza* 15: 620-627.
- Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature.* 456:53-9.
- Bertone P, Gerstein M, Snyder M. 2005. Applications of DNA tiling arrays to experimental genome annotation and regulatory pathway discovery. *Chromosome Res.* 13:259–274.
- Birol I, Jackman SD, Nielsen CB, Qian JQ, Varhol R, Stazyk G, Morin RD, Zhao Y, Hirst M, Schein JE, et al. 2009. De novo transcriptome assembly with ABySS. *Bioinformatics* 25:2872–2877.
- Bloom J, Khan Z, Kruglyak L, Singh M, Caudy A. 2009. Measuring differential gene expression by short read sequencing: quantitative comparison to 2-channel gene expression microarrays. *BMC Genomics.* 10:221.
- Boguski MS, Tolstoshev CM, Bassett DE Jr. 1994. Gene discovery in dbEST. *Science* 265:1993-1994.
- Bonfante P. 2001. At the interface between mycorrhizal fungi and plants: the structural organization of cell wall, plasma membrane and cytoskeleton. *The Mycota IX, Fungal Associations*, Hock ed., Springer- Verlag Berlin, Heidelberg, 45–61.
- Bonfante P, Genre A. 2010. Mechanisms underlying beneficial plant-fungus interactions in mycorrhizal symbiosis. *Nature Communications* 1: 48.
- Breuninger M, Requena N. 2004. Recognition events in AM symbiosis: analysis of fungal gene expression at the early appressorium stage. *Fungal Genet Biol.* 41:794–804.
- Bücking H, Heyser W. 2001. Microautoradiographic localization of phosphate and carbohydrates in mycorrhizal roots of *Populus tremula* x *Populus alba* and the implications for transfer processes in ectomycorrhizal associations. *Tree Physiology.* 21:101-107.
- Buée M, Reich M, Murat C, Morin E, Nilsson RH, Uroz S, Martin F. 2009. 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytologist.* 184:449–456.
- Bullard JH, Purdom E, Hansen KD, Dudoit S. 2010. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics.* 18:94.
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Research* 37: D233-D238.
- Casneuf T, Van de Peer Y, Huber W. 2007. In situ analysis of crosshybridisation on microarrays and inference of expression correlation. *BMC Bioinformatics.* 8:461.
- Ceccaroli P, Buffalini M, Saltarelli R, Barbieri E, Polidori E, Ottonello S, Kohler A, Tisserant E, Martin F, Stocchi V. 2011. Genomic profiling of carbohydrate metabolism in the ectomycorrhizal fungus *Tuber melanosporum*. *New Phytologist* 189: 751–764.

- Chabot S, Bécard G, Piché Y. 1992. The life cycle of *Glomus intraradices* in root organ culture. *Mycologia* 84: 315-321.
- Chalot M, Blaudez D, Brun A. 2006. Ammonia: a candidate for nitrogen transfer at the mycorrhizal interface. *Trends Plant Sci.* 11:263-6.
- Chen LQ, Hou BH, Lalonde S, Takanaga H, Hartung ML, Qu XQ, Guo WJ, Kim JG, Underwood W, Chaudhuri B et al. 2010. Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature*. 468:527-532.
- Chevreur B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, Suhai S. 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Research* 14: 1147-1159.
- Choi J, Park J, Kim D, Jung K, Kang S, Lee YH. 2010. Fungal secretome database: integrated platform for annotation of fungal secretomes. *BMC Genomics*. 11:105-120.
- Clapp JP, Rodriguez A, Dodd JC. 2001. Inter- and intra-isolate rRNA large subunit variation in *Glomus coronatum* spores. *New Phytol.* 149:539–54
- Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G, et al. 2008. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods*. 5:613–619.
- Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674-3676.
- Corradi N, Kuhn G, Sanders IR. 2004. Monophyly of beta-tubulin and H⁺-ATPase gene variants in *Glomus intraradices*: consequences for molecular evolutionary studies of AM fungal genes. *Fungal Genet. Biol.* 41:262–73.
- Corradi N, Ruffner B, Croll D, Colard A, Horak A, Sanders IR. 2009. High-level molecular diversity of copper-zinc superoxide dismutase genes among and within species of arbuscular mycorrhizal fungi. *Appl. Environ. Microbiol.* 75:1970–78.
- Couturier J, Montanini B, Martin F, Brun A, Blaudez D, Chalot M. 2007. The expanded family of ammonium transporters in the perennial poplar plant. *New Phytologist*. 174:137–150.
- Croll D, Wille L, Gamper HA, Mathimaran N, Lammers PJ, Corradi N, Sanders IR. 2008. Genetic diversity and host plant preferences revealed by simple sequence repeat and mitochondrial markers in a population of the arbuscular mycorrhizal fungus *Glomus intraradices*. *New Phytologist* 178: 672-687.
- Croll D, Giovannetti M, Koch AM, Sbrana C, Ehinger M, Lammers PJ, Sanders IR. 2009. Nonspecific vegetative fusion and genetic exchange in the arbuscular mycorrhizal fungus *Glomus intraradices*. *New Phytologist* 181: 924-937.
- Croll D, Sanders IR. 2009. Recombination in *Glomus intraradices*, a supposed ancient asexual arbuscular mycorrhizal fungus. *BMC Evolutionary Biology*. 9: 13.
- De Wit PJGM, Mehrabi R, Van Den Burg HA, Stergiopoulos I 2009. Fungal effector proteins: past, present and future. *Journal Molecular Plant Pathology* 10: 735-747.
- Denoeud F, Aury JM, Da Silva C, Noel B, Rogier O, Delledonne M, Morgante M, Valle G, Wincker P, Scarpelli C, Jaillon O, Artiguenave F. 2008. Annotating genomes with massive-scale RNA sequencing. *Genome Biology* 9: R175.
- Deveau A, Kohler A, Frey-Klett P, Martin F. 2008. The major pathways of carbohydrate metabolism in the ectomycorrhizal basidiomycete *Laccaria bicolor* S238N. *New Phytol.* 175:743-755.

- Dexheimer J, Pargney JC. 1991. Comparative anatomy of the host-fungus interface in mycorrhizas. *Experientia*. 47:312–320.
- Dodds PN, Rathjen JP. 2010. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nature Reviews Genetics*. 8: 539-548.
- 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.
- Duplessis S, Cuomo CA, Lin YC, Aerts A, Tisserant E, Veneault-Fourrey C, Joly DL, Hacquard S, Amselem J, Cantarel BL et al. 2011. Obligate biotrophy features unraveled by the genomic analysis of the rust fungi, *Melampsora larici-populina* and *Puccinia graminis* f. sp. *tritici*. *Proceedings of the National Academy of Sciences of the USA* 108: 9166-9171.
- Eastwood DC, Floudas D, Binder M, Majcherczyk A, Schneider P, Aerts A, Asiegbu FO, Baker SE, Barry K, Bendiksby M, et al. 2011. The plant cell wall-decomposing machinery underlies the functional diversity of forest fungi. *Science*. 333:762-5.
- Eaton CJ, Cox MP, Scott B. 2010. What triggers grass endophytes to switch from mutualism to pathogenism? *Plant Science* 180:190-195.
- Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, et al. 2009. Real-time DNA sequencing from single polymerase molecules. *Science*. 323:133-8.
- Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* 300: 1005-1016.
- Ercolin F, Reinhardt D. 2011. Successful joint ventures of plants: arbuscular mycorrhiza and beyond. *Trends in Plant Science* 16: 356-362.
- Fajardo López M, Männer P, Willmann A, Hampp R, Nehls U. 2007. Increased trehalose biosynthesis in Hartig net hyphae of ectomycorrhizas. *New Phytol*. 174:389-98.
- Fajardo López M, Dietz S, Grunze N, Bloeschies J, Weiss M, Nehls U. 2008. The sugar porter gene family of *Laccaria bicolor*: function in ectomycorrhizal symbiosis and soil-growing hyphae. *New Phytol*. 180:365-78
- Falgueras J, Lara AJ, Fernández-Pozo N, Cantón FR, Pérez-Trabado G, Claros MG. 2010. SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read. *BMC Bioinformatics* 11:38.
- Felten J, Kohler A, Morin E, Bhalerao RP, Palme K, Martin F, Ditengou FA, Legué V. 2009. The ectomycorrhizal fungus *Laccaria bicolor* stimulates lateral root formation in poplar and *Arabidopsis* through auxin transport and signaling. *Plant Physiol*. 151:1991-2005.
- Fitter AH, Helgason T, Hodge A. 2011. Nutritional exchanges in the arbuscular mycorrhizal symbiosis: Implications for sustainable agriculture. *Fungal Biology Reviews* 25: 68-72.
- Gabella S, Abbà S, Duplessis S, Montanini B, Martin F, Bonfante P. 2005. Transcript profiling reveals novel marker genes involved in fruiting body formation in *Tuber borchii*. *Eukaryotic Cell* 4: 1599-602.
- Garcia-Reyero N, Griffitt RJ, Liu L, Kroll KJ, Farmerie WG, Barber DS, Denslow ND. 2008. Construction of a robust microarray from a non-model species largemouth bass, *Micropterus salmoides* (Lacepede), using pyrosequencing technology. *J Fish Biol*. 72: 2354–2376.
- Genre A, Chabaud M, Timmers T, Bonfante P, Barker DG. 2005. Arbuscular mycorrhizal fungi elicit a novel intracellular apparatus in *Medicago truncatula* root epidermal cells before infection. *Plant Cell* 17: 3489-3499.

- Gerhard DS, Wagner L, Feingold EA, Shenmen CM, Grouse LH, Schuler G, Klein SL, Old S, Rasooly R, Good P, et al. 2004. The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Res.* 14:2121-2127.
- Gibbons JG, Janson EM, Hittinger CT, Johnston M, Abbot P, Rokas A. 2009. Benchmarking next-generation transcriptome sequencing for functional and evolutionary genomics. *Mol Biol Evol.* 26: 2731–2744.
- Giovannetti M, Fortuna P, Citernesi AS, Morini S, Nuti MP. 2001. The occurrence of anastomosis formation and nuclear exchange in intact arbuscular mycorrhizal networks. *New Phytologist* 151: 717-724.
- Gomez SK, Javot H, Deewatthanawong P, Torres-Jerez I, Tang Y, Blancaflor EB, Udvardi MK, Harrison MJ. 2009. *Medicago truncatula* and *Glomus intraradices* gene expression in cortical cells harboring arbuscules in the arbuscular mycorrhizal symbiosis. *BMC Plant Biology* 22: 10.
- Gómez-Ariza J, Balestrini R, Novero M, Bonfante P. 2009. Cell-specific gene expression of phosphate transporters in mycorrhizal tomato roots. *Biology and Fertility of Soils* 45: 845-853.
- Govindarajulu M, Pfeffer PE, Jin HR, Abubaker J, Douds DD, Allen JW, Bucking H, Lammers PJ, Shachar-Hill Y. 2005. Nitrogen transfer in the arbuscular mycorrhizal symbiosis. *Nature* 435: 819-823.
- Gowda M, Venu RC, Raghupathy MB, Nobuta K, Li H, Wing R, Stahlberg E, Coughlan S, Haudenschild CD, Dean R, Nahm BH, Meyers BC, Wang GL. 2006. Deep and comparative analysis of the mycelium and appressorium transcriptomes of *Magnaporthe grisea* using MPSS, RL-SAGE, and oligoarray methods. *BMC Genomics* 7: 310
- Grigoriev IV, Cullen D, Hibbett D, Goodwin SB, Jeffries TW, Kubicek CP, Kuske C, Magnuson J, Martin F, Spatafora J et al. 2011. Fueling the future with fungal genomics. *Mycology*, in press.
- Grunwald U, Nyamsuren O, Tamasloukht M, Lapopin L, Becker A, Mann P, Gianinazzi-Pearson V, Krajinski F, Franken P. 2004. Identification of mycorrhiza-regulated genes with arbuscule development-related expression profile. *Plant Mol Biol.* 55:553–566.
- Grunwald U, Guo W, Fischer K, Isayenkov S, Ludwig-Müller J, Hause B, Yan X, Küster H, Franken P. 2009. Overlapping expression patterns and differential transcript levels of phosphate transporter genes in arbuscular mycorrhizal, Pi-fertilised and phytohormone-treated *Medicago truncatula* roots. *Planta* 229: 1023-1034.
- Guether M, Balestrini R, Hannah MA, Udvardi MK, Bonfante P. 2009a. Genome-wide reprogramming of regulatory networks, transport, cell wall and membrane biogenesis during arbuscular mycorrhizal symbiosis in *Lotus japonicus*. *New Phytologist* 182: 200-212.
- Guether M, Neuhäuser B, Balestrini R, Dynowski M, Ludewig U, Bonfante P. 2009b. A mycorrhizal-specific ammonium transporter from *Lotus japonicus* acquires nitrogen released by arbuscular mycorrhizal fungi. *Plant Physiology* 150: 73-83.
- Güimil S, Chang H-S, Zhu T, Sesma A, Osbourn A, Roux C, Ioannidis V, Oakeley EJ, Docquier M, Descombes P, et al. 2005. Comparative transcriptomics of rice reveals an ancient pattern of response to microbial colonization. *Proceedings of the National Academy of Sciences of the United States of America* 102: 8066-8070.
- Halary S, Malik SB, Lildhar L, Slamovits CH, Hijri M, Corradi N. 2011. Conserved meiotic machinery in *Glomus* spp., a putatively ancient asexual fungal lineage. *Genome Biology and Evolution*, in press.
- Harbers M, Carninci P. 2005. Tag-based approaches for transcriptome research and genome annotation. *Nat. Methods.* 2:495–502.

- Heller G, Adomas A, Li G, Osborne J, van Zyl L, Sederoff R, Finlay RD, Stenlid J, Asiegbu FO. 2008. Transcriptional analysis of *Pinus sylvestris* roots challenged with the ectomycorrhizal fungus *Laccaria bicolor*. *BMC Plant Biology* 8 :19
- Hibbett DS, Matheny PB. 2009. The relative ages of ectomycorrhizal mushrooms and their plant hosts estimated using Bayesian relaxed molecular clock analyses. *BMC Biol.* 7:13.
- Hijikata N, Murase M, Tani C, Ohtomo R, Osaki M, Ezawa T. 2010. Polyphosphate has a central role in the rapid and massive accumulation of phosphorus in extraradical mycelium of an arbuscular mycorrhizal fungus. *New Phytologist* 186: 285–289
- Hijri M, Sanders IR. 2004. The arbuscular mycorrhizal fungus *Glomus intraradices* is haploid and has a small genome size in the lower limit of eukaryotes. *Fungal Genet Biol.* 41:253–61.
- Hijri M, Sanders IR. 2005. Low gene copy number shows that arbuscular mycorrhizal fungi inherit genetically different nuclei. *Nature.* 433:160–63
- Hinton JC, Hautefort I, Eriksson S, Thompson A, Rhen M. 2004. Benefits and pitfalls of using microarrays to monitor bacterial gene expression during infection. *Curr Opin Microbiol.* 7: 277–282.
- Hohnjec N, Vieweg MF, Pühler A, Becker A, Küster H. 2005. Overlaps in the transcriptional profiles of *Medicago truncatula* roots inoculated with two different *Glomus* fungi provide insights into the genetic program activated during arbuscular mycorrhiza. *Plant Physiol.* 137:1283–1301.
- Homer N, Merriman B, Nelson SF. 2009. BFAST: an alignment tool for large scale genome resequencing. *PLoS One.* 4:e7767.
- Horan, DP, Chilvers GA, Lapeyrie FF. 1988. Time sequence of the infection process in eucalypt ectomycorrhizas. *New Phytol.* 109:451-458.
- Hothorn M, Neumann H, Lenherr ED, Wehner M, Rybin V, Hassa PO, Uttenweiler A, Reinhardt M, Schmidt A, Seiler J, et al. 2009. Catalytic core of a membrane-associated eukaryotic polyphosphate polymerase. *Science* 324: 513-516.
- Hudson ME. 2008. Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol Ecol Notes.* 8:3–17.
- Hutchison CA. 2007. DNA sequencing: bench to bedside and beyond. *Nucleic Acids Res.* 35:6227–6237.
- Idnurm A, Walton FJ, Floyd A, Heitman J. 2008. Identification of the sex genes in an early diverged fungus. *Nature* 451: 193-197.
- International Rice Genome Sequencing Project. 2005. A map-based sequence of the rice genome. *Nature* 437:693-698.
- Jany JL, Pawlowska TE. 2010. Multinucleate spores contribute to evolutionary longevity of asexual Glomeromycota. *The American Naturalist* 175: 424-435.
- Javelle A, Morel M, Rodríguez-Pastrana BR, Botton B, Andre B, Marini AM, Brun A, Chalot M, 2003. Molecular characterization, function and regulation of ammonium transporters (Amt) and ammonium-metabolizing enzymes (GS, NADPGDH) in the ectomycorrhizal fungus *Hebeloma cylindrosporum*. *Mol. Microbiol.* 47:411-430.
- Javot H, Pumplin N, Harrison MJ. 2007. Phosphate in the arbuscular mycorrhizal symbiosis: transport properties and regulatory roles. *Plant, Cell and Environment* 30: 310-322.
- 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. *Mol Plant Microbe Interact.* 17:202-215.

- Johnson DS, Mortazavi A, Myers RM, Wold B. 2007. Genome wide mapping of in vivo protein-DNA interactions. *Science*. 316:1497–1502.
- Kammenga JE, Herman MA, Ouborg NJ, Johnson L, Breitling R. 2007. Microarray challenges in ecology. *Trends Ecol Evol*. 22: 273–279.
- Kamoun S. 2007. Groovy times: filamentous pathogen effectors revealed. *Current Opinion in Plant Biology*. 10: 358-65.
- Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28: 27-30.
- Kapranov P, Willingham AT, Gingeras TR. 2007. Genome wide transcription and the implications for genomic organization. *Nat Rev Genet*. 8:413–423.
- Kemen E, Gardiner A, Schultz-Larsen T, Kemen AC, Balmuth AL, Robert-Seilanianantz A, Bailey K, Holub E, Studholme DJ, Maclean D, et al. 2011. Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis thaliana*. *PLoS Biology* 9: e1001094.
- Kemppainen M, Duplessis S, Martin F, Pardo AG. 2009. RNA silencing in the model mycorrhizal fungus *Laccaria bicolor*: gene knock-down of nitrate reductase results in inhibition of symbiosis with *Populus*. *Environ Microbiol*. 11:1878-1896.
- Kent WJ. 2002. BLAT--the BLAST-like alignment tool. *Genome Res*. 12:656-664.
- Kiers ET, Duhamel M, Beesetty Y, Mensah JA, Franken O, Verbruggen E, Fellbaum CR, Kowalchuk GA, Hart MM, Bago A, et al. 2011. Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science* 333: 880-882.
- Kistner C, Parniske M. 2002. Evolution of signal transduction in intracellular symbiosis. *Trends Plant Sci*. 7:511–518.
- Kistner C, Winzer T, Pitzschke A, Mulder L, Sato S, Kaneko T, Tabata S, Sandal N, Stougaard J, Webb KJ, et al. 2005. Seven *Lotus japonicus* genes required for transcriptional reprogramming of the root during fungal and bacterial symbiosis. *Plant Cell*. 17:2217–2229.
- Kloppholz S, Kuhn H, Requena N. 2011. A secreted fungal effector of *Glomus intraradices* promotes symbiotic biotrophy. *Current Biology* 21,1204-1209.
- Kottke I, Oberwinkler F. 1987. The cellular structure of the Hartig net: coenocytic and transfer cell-like organization. *Nord J Bot*. 7:85–95.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* 305: 567-580.
- Kuhn G, Hijri M, Sanders IR. 2001. Evidence for the evolution of multiple genomes in arbuscular mycorrhizal fungi. *Nature*. 414:745–48.
- Kuhn H, Küster H, Requena N. 2010. Membrane steroid-binding protein 1 induced by a diffusible fungal signal is critical for mycorrhization in *Medicago truncatula*. *New Phytologist* 185: 716-733.
- Kuznetsova E, Seddas-Dozolme PMA, Arnould C, Tollot M, van Tuinen D, Borisov A, Gianinazzi S, Gianinazzi-Pearson V. 2010. Symbiosis-related pea genes modulate fungal and plant gene expression during the arbuscule stage of mycorrhiza with *Glomus intraradices*. *Mycorrhiza* 20: 427-443.
- Lacourt I, Duplessis S, Abbà S, Bonfante P, Martin F. 2002. Isolation and characterization of differentially expressed genes in the mycelium and fruit body of *Tuber borchii*. *Applied Environmental Microbiology* 68: 4574-82.

- Lagrange H, Jay-Allemand C, Lapeyrie F. 2001. Rutin, the phenolglycoside from eucalyptus root exudates, stimulates *Pisolithus* hyphal growth at picomolar concentrations. *New Phytol.* 149:349–355
- Lammers PJ, Jun J, Abubaker J, Arreola R, Gopalan A, Bago B, Hernandez-Sebastia C, Allen JW, Douds DD, Pfeffer PE et al. 2001. The glyoxylate cycle in an arbuscular mycorrhizal fungus: gene expression and carbon flow. *Plant Physiol.* 127:1287–1298.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. *Nature.* 409:860–921.
- Lanfranco L, Novero M, Bonfante P. 2005. The mycorrhizal fungus *Gigaspora margarita* possesses a CuZn superoxide dismutase which is up-regulated during the symbiosis with legume hosts. *Plant Physiol.* 137:1319–1330.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25.
- Larsen PE, Trivedi G, Sreedasyam A, Lu V, Podila GK, Collart FR. 2010. Using Deep RNA Sequencing for the Structural Annotation of the *Laccaria Bicolor* Mycorrhizal Transcriptome. *PloS ONE.* 5 :e9780.
- Larsen PE, Sreedasyam A, Trivedi G, Podila GK, Cseke LJ, Collart FR. 2011. Using next generation transcriptome sequencing to predict an ectomycorrhizal metabolome. *BMC Syst Biol.* 2011. 13:70.
- Lassmann T, Hayashizaki Y, Daub CO. 2009. TagDust—a program to eliminate artifacts from next generation sequencing data. *Bioinformatics.* 25:2839–2840.
- 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.
- LePage BA, Currah RS, Stockey RA, Rothwell GW. 1997. Fossil ectomycorrhizae from the middle Eocene. *Am J Bot.* 84:410–412
- Letunic I, Yamada T, Kanehisa M, Bork P. 2008. iPath: interactive exploration of biochemical pathways and networks. *Trends in Biochemical Sciences* 33: 101–103.
- Li R, Li Y, Kristiansen K, Wang J. 2008a. SOAP: short oligonucleotide alignment program. *Bioinformatics.* 24:713–4.
- Li H, Ruan J, Durbin R. 2008b. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* 18:1851–8.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 25:1754–60.
- Li R, Fan W, Tian G, Zhu H, He L, Cai J, Huang Q, Cai Q, Li B, Bai Y, et al. 2010. The sequence and de novo assembly of the giant panda genome. *Nature.* 463:311–7.
- Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. 2008. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell.* 133:523–536.
- Liu J, Blaylock LA, Harrison MJ. 2004. cDNA arrays as a tool to identify mycorrhiza-regulated genes: identification of mycorrhiza induced genes that encode or generate signaling molecules implicated in the control of root growth. *Can J Bot.* 82:1177–1185.
- Lucic E, Fourrey C, Kohler A, Martin F, Chalot M, Brun-Jacob A. 2008. A gene repertoire for nitrogen transporters in *Laccaria bicolor*. *New Phytol.* 180:343–364.

- Maere S, Heymans K, Kuiper M. 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*. 21: 3448-3449.
- Maillet F, Poinso V, André O, Puech-Pagès V, Haouy A, Gueunier M, Cromer L, Giraudet D, Formey D, Niebel A, et al. 2011. Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature*. 469: 58-63.
- Malik SB, Pightling AW, Stefaniak LM, Schurko AM, Logsdon JM. 2008. An expanded inventory of conserved meiotic genes provides evidence for sex in *Trichomonas vaginalis*. *PLoS One* 3: e2879.
- Marguerat S, Bähler J. 2010. RNA-seq: from technology to biology. *Cell Mol Life Sci*. 67:569–579.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 437:376-80.
- Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research* 18: 1509–1517.
- Martin F Marchal JP, Timinska A, Canet D. 1985. The metabolism and physical state of polyphosphates in ectomycorrhizal fungi. A ³¹P nuclear magnetic resonance study. *New Phytologist* 101:257-290.
- Martin F, Rubini P, Cote R, Kottke I. 1994. Aluminum polyphosphate complexes in the mycorrhizal basidiomycete *Laccaria bicolor*: a ²⁷Al NMR study. *Planta* 19:241-246.
- Martin F, Tuskan GA, DiFazio SP, Lammers P, Newcombe G, Podila GK. 2004. Symbiotic Sequencing for the *Populus* Mesocosm. *New Phytologist*. 161:330-335.
- Martin F, Kohler A, Duplessis S. 2007. Living in harmony in the wood underground: ectomycorrhizal genomics. *Curr Opin Plant Biol*. 10:204-10.
- Martin F, Aerts A, Ahren D, Brun A, Danchin EGJ, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V, et al. 2008a. The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* 452: 88-92.
- Martin F, Gianinazzi-Pearson V, Hijri M, Lammers P, Requena N, Sanders IR, Shachar-Hill Y, Shapiro H, Tuskan GA, Young JPW. 2008b. The long hard road to a completed *Glomus intraradices* genome. *New Phytologist* 180: 747-750.
- Martin F, Tunlid A. 2008. The ectomycorrhizal symbiosis: a marriage of convenience. *The Mycota V, Plant Relationships*, edn 2. Edited by Deising HB, 237-257.
- Martin F, Nehls U. 2009. Harnessing ectomycorrhizal genomics for ecological insights. *Curr Opin Plant Biol*. 12:508-15.
- Martin F, Kohler A, Murat C, Balestrini R, Coutinho PM, Jaillon O, Montanini B, Morin E, Noel B, Percudani R et al. 2010a. Périgord Black Truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* 464: 1033-1038.
- Martin J, Bruno VM, Fang Z, Meng X, Blow M, Zhang T, Sherlock G, Snyder M, Wang Z. 2010b. Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-seq reads. *BMC Genomics*. 11:663.
- Martin F, Cullen D, Hibbett D, Pisabarro A, Spatafora JW, Baker SE, Grigoriev IV. 2011. Sequencing the fungal tree of life. *New Phytol*. 190:818-21.
- Martin JA, Wang Z. 2011. Next-generation transcriptome assembly. *Nat Rev Genet*. 12:671-82.

- Martinez D, Challacombe J, Morgenstern I, Hibbett D, Schmoll M, Kubicek CP, Ferreira P, Ruiz-Duenas FJ, Martinez AT, Kersten P et al. 2009. Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion. *Proc Natl Acad Sci USA*. 106:1954-1959.
- Martinez DA, Nelson MA. 2010. The next generation becomes the now generation. *PLoS Genet*. 6:e1000906.
- Maxam AM, Gilbert W. 1977. A new method for sequencing DNA. *Proc Natl Acad Sci U S A*. 74:560-4.
- McGuire AM, Pearson MD, Neafsey DE, Galagan JE. 2008. Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biology* 9: R50.
- Mello A, Murat C, Bonfante P. 2006. Truffles: much more than a prized and local fungal delicacy. *FEMS Microbiology Letters* 260: 1-8.
- Menotta M, Amicucci A, Sisti D, Gioacchini AM, Stocchi V. 2004. Differential gene expression during pre-symbiotic interaction between *Tuber borchii* Vitad. and *Tilia americana* L. *Curr Genet*. 46:158-165.
- Metzker ML. 2010. Sequencing technologies - the next generation. *Nat Rev Genet*. 11:31-46.
- Meyer E, Aglyamova GV, Wang S, Buchanan-Carter J, Abrego D, Colbourne JK, Willis BL, Matz MV. 2009. Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx. *BMC Genomics*. 10: 219.
- Mignone F, Gissi C, Liuni S, Pesole G. 2002. Untranslated regions of mRNAs. *Genome Biology* 3: 0004.
- Miller JR, Koren S, Sutton G. 2010. Assembly algorithms for next-generation sequencing data. *Genomics*. 95:315-27.
- Modrek B, Lee C. 2002. A genomic view of alternative splicing. *Nature Genetics* 30: 13-19.
- Montanini B, Levati E, Bolchi A, Kohler A, Morin E, Tisserant E, Martin F, Ottonello S. 2010. Genome-wide search and functional identification of transcription factors in the mycorrhizal fungus *Tuber melanosporum*. *New Phytol*, in press.
- Morel M, Jacob C, Kohler A, Johansson T, Martin F, Chalot M, Brun A. 2005. Identification of genes differentially expressed in extraradical mycelium and ectomycorrhizal roots during *Paxillus involutus*-*Betula pendula* ectomycorrhizal symbiosis. *Appl Environ Microbiol*. 71:382-391.
- Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* 35: W182-185.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5: 621-628.
- Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320: 1344-1349.
- Nagendran S, Hallen-Adams HE, Paper JM, Aslam N, Walton JD. 2009. Reduced genomic potential for secreted plant cell-wall-degrading enzymes in the ectomycorrhizal fungus *Amanita bisporigera*, based on the secretome of *Trichoderma reesei*. *Fungal Genetics and Biology* 46: 427-35.
- Nehls U, Wiese J, Guttenberger M, Hampp R. 1998. Carbon allocation in ectomycorrhizas: identification and expression analysis of an *Amanita muscaria* monosaccharide transporter. *Molecular Plant-Microbe Interactions*. 11:167-176.

- Nehls U, Mokolajewski S, Magel E, Hampp R. 2001. Carbohydrate metabolism in ectomycorrhizas: gene expression, monosaccharide transport and metabolic control. *New Phytologist* 150: 533–541.
- Nehls U. 2004. Carbohydrates and nitrogen: nutrients and signals in ectomycorrhizas. In: Varma A, Abbott L, Werner D, Hampp R, eds. *Plant surface microbiology*. Berlin, Germany: Springer Verlag, 373–392.
- Nehls U, Grunze N, Willmann M, Reich M, Küster H. 2007. Sugar for my honey: carbohydrate partitioning in ectomycorrhizal symbiosis. *Phytochemistry*. 68: 82–91.
- Nehls U. 2008. Mastering ectomycorrhizal symbiosis: the impact of carbohydrates. *Journal of Experimental Botany*. 59: 1097–1108.
- Novaes E, Drost DR, Farmerie WG, Pappas GJ Jr, Grattapaglia D, Sederoff RR, Kirst M. 2009. High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genomics*. 9: 312.
- Nowrousian M, Stajich JE, Chu M, Engh I, Espagne E, Halliday K, Kamerewerd J, Kempken F, Knab B, Kuo HC, et al. 2010. De novo assembly of a 40 Mb eukaryotic genome from short sequence reads: *Sordaria macrospora*, a model organism for fungal morphogenesis. *PLoS Genet*. 6:e1000891.
- Okoniewski MJ, Miller CJ. 2006. Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. *BMC Bioinformatics*. 7:276.
- Orlovich DA, Ashford AE. 1993. Polyphosphate granules are an artefact of specimen preparation in the ectomycorrhizal fungus *Pisolithus tinctorius*. *Protoplasma*. 173:91-102.
- Papanicolaou A, Stierli R, French-Constant RH, Heckel DG. 2009. Next generation transcriptomes for next generation genomes using est2assembly. *BMC Bioinformatics*. 24:447.
- Parchman TL, Geist KS, Grahnen JA, Benkman CW, Buerkle CA. 2010. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics*. 11:180.
- Parniske M. 2008. Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nat. Rev. Microbiol*. 6:763–75.
- Parra G, Bradnam K, Ning Z, Keane T, Korf I. 2009. Assessing the gene space in draft genomes. *Nucleic Acids Research* 37: 289-297.
- Parrent JL, James TY, Vasaitis R, Taylor AF. 2009. Friend or foe? Evolutionary history of glycoside hydrolase family 32 genes encoding for sucrolytic activity in fungi and its implications for plant-fungal symbioses. *BMC Evol Biol*. 9:148.
- Paszkowski U. 2006. A journey through signaling in arbuscular mycorrhizal symbioses. *New Phytol*. 172:35-46.
- Pawlowska TE, Taylor JW. 2004. Organization of genetic variation in individuals of arbuscular mycorrhizal fungi. *Nature*. 427:733–37.
- Pepke S, Wold B, Mortazavi A. 2009. Computation for ChIPseq and RNA-seq studies. *Nat Meth*. 6:S22–S32.
- Pérez-Tienda J, Testillano PS, Balestrini R, Fiorilli V, Azcón-Aguilar C, Ferrol N. 2011. GintAMT2, a new member of the ammonium transporter family in the arbuscular mycorrhizal fungus *Glomus intraradices*. *Fungal Genetics and Biology*, in press.
- Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J. 2003. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*. 19:651-652.

- Peter M, Courty PE, Kohler A, Delaruelle C, Martin D, Tagu D, Frey-Klett P, Duplessis S, Chalot M, Podila G et al. 2003. Analysis of expressed sequence tags from the ectomycorrhizal basidiomycetes *Laccaria bicolor* and *Pisolithus microcarpus*. *New Phytologist* 159: 117-129.
- Plett JM, Kemppainen M, Kale SD, Kohler A, Legué V, Brun A, Tyler B, Pardo A, Martin F. 2011. A secreted effector protein of *Laccaria bicolor* is required for symbiosis development. *Current Biology* 21: 1197-1203.
- Plett JM, Martin F. 2011. Blurred boundaries: lifestyle lessons from ectomycorrhizal fungal genomes. *Trends in Genetics* 27: 14-22.
- Read DJ, Perez-Moreno J. 2003. Mycorrhizas and nutrient cycling in ecosystems – a journey towards relevance? *New Phytologist*. 57:475–492.
- Requena N, Breuninger M, Franken P, Ocón A. 2003. Symbiotic status, phosphate, and sucrose regulate the expression of two plasma membrane H⁺-ATPase genes from the mycorrhizal fungus *Glomus mosseae*. *Plant Physiol.* 132:1540-1549.
- Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, Mungall K, Lee S, Okada HM, Qian J, et al. 2010. De novo assembly and analysis of RNA-seq data. *Nature Methods*. 7:909–912.
- Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*. 11:R25.
- Rodriguez A, Clapp JP, Dodd JC. 2004. Ribosomal RNA gene sequence diversity in arbuscular mycorrhizal fungi (glomeromycota). *J. Ecol.* 92:986–89
- Rosendahl S. 2008. Communities, populations and individuals of arbuscular mycorrhizal fungi. *New Phytologist* 178: 253-266.
- Royce TE, Rozowsky JS, Bertone P, Samanta M, Stolc V, Weissman S, Snyder M, Gerstein M. 2005. Issues in the analysis of oligonucleotide tiling microarrays for transcript mapping. *Trends Genet.* 21:466–475.
- Rubini A, Belfiori B, Riccioni C, Tisserant E, Arcioni S, Martin F, Paolocci P. 2010. Isolation and characterization of MAT genes in the symbiotic ascomycete *Tuber melanosporum*. *New Phytol*, in press.
- Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, Quackenbush J. 2006. TM4 microarray software suite. *Methods in Enzymology* 411: 134-193.
- Sanders IR, Alt M, Groppe K, Boller T, Wiemken A. 1995. Identification of ribosomal DNA polymorphisms among and within spores of the glomales—application to studies on the genetic diversity of arbuscular mycorrhizal fungal communities. *New Phytol.* 130:419–27.
- Sanders IR, Croll D. 2010. Arbuscular mycorrhiza: the challenge to understand the genetics of the fungal partner. *Annual Review of Genetics* 44: 271-92.
- Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* 74:5463–7.
- Schaarschmidt S, Roitsch T, Hause B. 2006. Arbuscular mycorrhiza induces gene expression of the apoplastic invertase LIN6 in tomato (*Lycopersicon esculentum*) roots. *Journal of Experimental Botany* 57: 4015–4023.
- Schena M, Heller RA, Thériault TP, Konrad K, Lachenmeier E, Davis RW. 1998. Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol.* 16:301–306.
- Schüssler A, Schwarzott D, Walker C. 2001. A new fungal phylum, the Glomeromycota: phylogeny and evolution. *Mycological Research* 105: 1413-1421.

- Schüssler A, Martin H, Cohen D, Fitz M, Wipf D. 2006. Characterization of a carbohydrate transporter from symbiotic glomeromycotan fungi. *Nature*. 444:933-6.
- Schuster SC. 2008. Next-generation sequencing transforms today's biology. *Nat Meth*. 5:16–18.
- Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. 2011. Global quantification of mammalian gene expression control. *Nature* 473: 337–342.
- Seddas PM, Arias C, Arnould C, van Tuinen D, Godfroy O, Benhassou H, Gouzy J, Morandi D, Dessaint F, Gianinazzi-Pearson V. 2009. Symbiosis-related plant genes modulate molecular responses in an arbuscular mycorrhizal fungus during early root interactions. *Molecular Plant-Microbe Interactions* 22: 341-351.
- Sędziewska KA, Fuchs J, Temsch EM, Baronian K, Watzke R, Kunze G. 2011. Estimation of the *Glomus intraradices* nuclear DNA content. *New Phytol*. doi: 10.1111/j.1469-8137.2011.03937.x.
- Selle A, Willmann M, Grunze N, Geßler A, Weiß M, Nehls U. 2005. The high-affinity poplar ammonium importer PttAMT1.2 and its role in ectomycorrhizal symbiosis. *New Phytologist*. 168:697–706.
- Selosse MA and Le Tacon F. 1998. The land flora: a phototroph - fungus partnership. *Trends in Ecology and Evolution*. 13:15-20.
- Shalon D, Smith SJ, Brown PO. 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res*. 6:639–645.
- Shendure J, Ji H. 2008. Next-generation DNA sequencing. *Nat Biotech*. 26:1135–1145.
- Smith SE, Read DJ. 2008. *Mycorrhizal symbiosis*. Academic press.
- Smith SE, Smith FA. 2011. Roles of arbuscular mycorrhizas in plant nutrition and growth: New paradigms from cellular to ecosystem scales. *Annual Review of Plant Biology* 62: 227–250.
- Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, van Themaat EVL, Brown JKM et al. 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* 330: 1543-1546.
- Splivallo R, Ottonello S, Mello A, Karlovsky P. 2010. Truffle volatiles: from chemical ecology to aroma biosynthesis. *New Phytol*, in press.
- Stergiopoulos I, de Wit PJ. 2009. Fungal effector proteins. *Annu Rev Phytopathol*. 47:233-263.
- Stockinger H, Walker C, Schüssler A. 2009. ‘*Glomus intraradices* DAOM197198’, a model fungus in arbuscular mycorrhiza research, is not *Glomus intraradices*. *New Phytologist* 183: 1176-1187.
- Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D, et al. 2008. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 321: 956–960.
- Surget-Groba Y, Montoya-Burgos JI. 2010. Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Res*. 20:1432–1440.
- Swindell SR, Plasterer TN. 1997. *SEQMAN. Sequence Data Analysis Guidebook* Humana Press 70:75-89.
- Tani C, Ohtomo R, Osaki M, Kuga Y, Ezawa T. 2009. ATP-dependent but proton gradient-independent polyphosphate-synthesizing activity in extraradical hyphae of an arbuscular mycorrhizal fungus. *Applied and Environmental Microbiology* 75: 7044-7050.
- Tatry MV, Kassiss EE, Lambilliotte R, Corratgé C, van Aarle I, Amenc LK, Alary R, Zimmermann S, Sentenac H, Plassard C. 2008. Two differentially regulated phosphate

transporters from the symbiotic fungus *Hebeloma cylindrosporum* and phosphorus acquisition by ectomycorrhizal *Pinus pinaster*. *Plant J.* 57:1092-1102.

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, et al. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4: 41.

Tian C, Kasiborski B, Koul R, Lammers PJ, Bucking H, Shachar-Hill Y. 2010. Regulation of the nitrogen transfer pathway in the arbuscular mycorrhizal symbiosis: gene characterization and the coordination of expression with nitrogen flux. *Plant Physiology* 153: 1175-1187.

Toth AL, Varala K, Newman TC, Miguez FE, Hutchison SK, Willoughby DA, Simons JF, Egholm M, Hunt JH, Hudson ME et al. 2007. Wasp gene expression supports an evolutionary link between maternal behavior and eusociality. *Science*. 318 :441–444.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 25:1105-11.

Trapnell C, Salzberg SL. 2009. How to map billions of short reads onto genomes. *Nat Biotechnol*. 27:455–457.

Trapnell C, Williams BA , Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010 Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* 28 511-515

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*. 313: 1596-1604.

van der Heijden MGA, Klironomos JN, Ursic M, Moutoglis P, Streitwolf-Engel R, Boller T, Wiemken A, Sanders IR. 1998. Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* 396: 69-72.

Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, Fitzgerald LM, Vezzulli S, Reid J, et al. 2007. A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS One*. 2:e1326.

Velculescu VE, Zhang L, Vogelstein B, Kinzler K. 1995. Serial analysis of gene expression. *Science*. 270:484–487.

Veneault-Fourrey C, Martin F. 2011. Mutualistic interactions on a knife-edge between saprotrophy and pathogenesis. *Curr Opin Plant Biol*. 14:444-50.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. 2001. The sequence of the human genome. *Science*. 291:1304–51.

Vera JC et al. 2008. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Mol Ecol*. 17:1636–1647

Viereck N, Hansen PE, Jakobsen I. 2004. Phosphate pool dynamics in the arbuscular mycorrhizal fungus *Glomus intraradices* studied by in vivo ³¹P NMR spectroscopy. *New Phytologist* 162: 783-794.

Vilela C, McCarthy JE. 2003. Regulation of fungal gene expression via short open reading frames in the mRNA 5' untranslated region. *Molecular Microbiology* 49: 859-67.

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 cDNAs. *Plant J.* 25:181-191.

Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* 456: 470–476.

- Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 10:57–63.
- Wang D, Griffitts J, Starker C, Fedorova E, Limpens E, Ivanov S, Bisseling T, Long S. 2010a. A nodule-specific protein secretory pathway required for nitrogen-fixing symbiosis. *Science.* 327:1126-9.
- Wang B, Guo G, Wang C, Lin Y, Wang X, Zhao M, Guo Y, He M, Zhang Y, Pan L. 2010b. Survey of the transcriptome of *Aspergillus oryzae* via massively parallel mRNA sequencing. *Nucleic Acids Res.* 38:5075-87.
- Weigel D, Mott R. 2009. The 1001 genomes project for *Arabidopsis thaliana*. *Genome Biol.* 10:107.
- Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, et al. 2008. The complete genome of an individual by massively parallel DNA sequencing. *Nature.* 452:872–6.
- Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bahler J. 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453: 1239–1243.
- Willmann A, Weifl M, Nehls U. 2007. Ectomycorrhiza-mediated repression of the high-affinity ammonium importer gene AmAMT2 in *Amanita muscaria*. *Current Genetics.* 51: 71–78.
- Wright DP, Johansson T, Le Quéré A, Söderström B, Tunlid A. 2005. Spatial patterns of gene expression in the extramatrical mycelium and mycorrhizal root tips formed by the ectomycorrhizal fungus *Paxillus involutus* in association with birch (*Betula pendula*) seedlings in soil microcosms. *New Phytol.* 167:579-596.
- Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science.* 328:916-9.
- Zerbino DR and Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829.
- Zhang G, Guo G, Hu X, Zhang Y, Li Q, Li R, Zhuang R, Lu Z, He Z, Fang X, Chen L, Tian W, Tao Y, Kristiansen K, Zhang X, Li S, Yang H, Wang J, Wang J. 2010. Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res.* 20:646-54

Résumé

La symbiose mycorhizienne est une interaction mutualiste formée entre les racines des plantes terrestres et des champignons du sol. Cette association symbiotique améliore essentiellement la nutrition de la plante-hôte. Les champignons mycorhiziens jouent ainsi un rôle essentiel dans les cycles biogéochimiques et la santé des plantes. Ils sont d'une importance écologique majeure au sein de la plupart des écosystèmes terrestres. Les changements morphoanatomiques associés au développement de la symbiose sont accompagnés de modifications dans la régulation de l'expression génique. L'étude des profils transcriptomiques est donc fondamentale afin de caractériser les mécanismes moléculaires gouvernant la symbiose mycorhizienne et en particulier, identifier les gènes de symbiose, contrôlant la différenciation des structures symbiotiques et le métabolisme symbiotique. Le développement récent des approches de transcriptomique à haut débit offre de nouvelles perspectives pour la compréhension de ces mécanismes. Ces nouvelles technologies (pyroséquençage 454, Illumina RNA-Seq) rendent possible le séquençage massif de transcriptome d'organismes non modèles, ainsi que le développement de nouvelles méthodes de quantification numérique de l'expression des gènes. Le travail entrepris dans le cadre de ce projet de thèse visait à caractériser *in silico* le transcriptome symbiotique du champignon ectomycorhizien *Tuber melanosporum* et du champignon endomycorhizien *Glomus intraradices*. Il s'agissait de mettre en place les outils et les protocoles bioinformatiques permettant le stockage, le traitement et l'exploitation des données transcriptomiques issues des ces nouvelles technologies de séquençage, afin de caractériser les transcrits exprimés par les symbiotes et d'identifier les gènes régulés au cours de la symbiose. Ce travail original a permis (1) la mise en place de plusieurs bases de données interactives et (2) de souligner l'existence de traits communs aux profils d'expression des champignons mycorhiziens. Nous avons notamment mise en évidence l'induction de gènes codant pour des transporteurs membranaires et des petites protéines sécrétées durant l'interaction symbiotique, ainsi que l'absence d'expression des enzymes hydrolytiques agissant sur les polysaccharides de la paroi végétale. De plus, la caractérisation du transcriptome de *G. intraradices* a permis d'établir le premier répertoire de gènes à l'échelle du génome pour un champignon endomycorhizien. Enfin, l'analyse du transcriptome *T. melanosporum* suggère que le développement symbiotique est accompagné par une série de mécanismes de régulation transcriptionnelle et traductionnelle (épissage alternatif, μ ORF). Cette étude de génomique contribue à l'amélioration des connaissances sur les processus moléculaires qui sous-tendent la symbiose mycorhizienne et constitue une ressource unique pour de futures recherches sur les réseaux de gènes contrôlant la symbiose.

Mots-clés: Symbiose, Champignons, Mycorhize, Transcriptomique, Bio-informatique, 454, RNA-Seq.