

Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par:
Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par:

Damien FORMEY DE SAINT LOUVENT

Le Mercredi 19 septembre 2012

Titre:
**La symbiose mycorhizienne à arbuscules
Rhizophagus irregularis / *Medicago truncatula* : données génomiques,
transcriptomiques et microtranscriptomiques**

Ecole doctorale et spécialité:
ED SEVAB : Interactions plantes-microorganismes

Unité de recherche:
Laboratoire de Recherche en Sciences Végétales - UMR5546 CNRS-UPS

Directeur de Thèse:
Pr. Christophe ROUX

Rapporteurs:
Dr. Martin CRESPI / Pr. Dirk REDECKER

Autres membres du jury :
**Dr. Christine LELANDAIS-BRIERE / Dr. Faouzi ATTIA
/ Pr. Sylvain JEANDROZ / Pr. Christophe ROUX**

REMERCIEMENTS

J'aimerais tout d'abord adresser mes plus sincères remerciements à Christophe Roux qui, au-delà de son rôle de directeur de cette thèse assuré avec vigilance et professionnalisme, a su être présent et réactif pour les aspects extra-thèse durant ces trois ans et demi écoulés. Ces remerciements ne suffiront pas à exprimer ma gratitude, il mérite donc son titre, unanimement décerné, de « Papa ».

Ma profonde reconnaissance va également à l'ensemble de l'entreprise Agronutrition. Particulièrement, j'aimerais remercier Cédric Cabanes de m'avoir fait confiance pour mener ce projet et de m'avoir accueilli au sein de « la grande famille ». Merci également à Faouzi Attia qui a suivi avec assiduité l'avancement de mon projet tout en prodiguant de bons conseils et orientant mes travaux avec brio. Merci aussi à Marie-Dominique Da Silva pour sa disponibilité, son organisation et ses explications claires !

Je remercie Andreas Niebel, pour son implication dans ce projet, ses conseils judicieux lors du comité de thèse et tout au long du doctorat, ainsi que pour sa bonne humeur. Il mérite également son titre de « Tonton ».

Je tiens aussi à exprimer ma gratitude à Jean-Philippe Combier pour son savoir-faire, son aide et ses très bons conseils et astuces « Jean-phiesques ». Titre de « Cousin » décerné.

Un grand merci à toute l'équipe « symbiose mycorhizienne et signalisation cellulaire » avec qui on a partagé beaucoup de bons moments, débattu avec ferveur durant les pauses déjeuner, ou bien brainstormé en réunion.

Merci à tous les membres du LRSV pour leur implication de près ou de loin dans mon projet de thèse, pour leurs discussions de couloir ou tout simplement pour leur sympathie. Remerciements particuliers pour Bruno Savelli et Hélène San Clemente pour leur aide précieuse en tant que guides dans ce monde si sauvage de la bioinformatique.

Je remercie également Christine Lelandais et Diederik Van Tuinen d'avoir participer activement et efficacement à mes comités de thèse.

Merci aussi à vous les « potos », toujours présents ou partis vers de nouveaux horizons. Chacun, vous avez su prendre soin de moi à votre manière (et me supporter dans tous les sens du terme...). Ce n'est pas la tribune idéale pour détailler vos qualités, mais vous en regorgez tous et cette thèse n'aurait pas été aussi agréable, vivante et humaine sans vous tous à mes côtés. Vous faites partie intégrante de ma réussite. MERCI, Alex, Buchette, Coco, Dédé, Domi, Jérôme, Jojo, Kéké le coco, Laëtitia, Maël, « Maman » Marion, « sœurlette » Matosh, bro PM et tous les autres.

Enfin, plus que tout, j'aimerais remercier mes proches qui m'ont soutenu tout au long de ce « marathon ». Mes parents à qui je dois tout simplement tout, de mon éducation à ma joie de vivre, en passant par ma réussite. Ma sœur, Henri et les p'tits qui ont su me donner des moments de rigolade, de détente et de bonheur. Ju, pour ces weekends à nuits rallongées qui m'ont fait le plus grand bien (TWSS). Ben, Eric, Fred, Raph, même si vous êtes loin. Merci à toi, sous-marin jaune pour « avoir tenté le coup » ce soir des Solidays et un an après faire semblant d'être intéressée par les moisissures, les gènes, les souches et les génomes, et de supporter tout de même ma bipolarité de fin de thèse qui, grâce à toi, a été plus ferroviaire, rock'n'roll, reposante, riche en sommeil, gastronomique, fêtarde, boueuse, douce. C'est maintenant à moi d'être là pour toi. Merci pour ton amour.

Merci au O'Shannon, au Killarney, au Mulligan's, au Vasco le Gamma, au Delicatessen, à La Tireuse, au Connexion, au P'tit Bouchon, au Bikini, le Zénith, Bercy, La Cigale, le Stade de France, le Top14, la NFL, FMR, Néo, Nova, Ouï FM, Le Mouv', Virgin, au Moe's, à la Maison drôle, Yoshi Sushi, le Gaucho, Le Ponte Vecchio, Els Petits Catalans, l'Asian Kanteen, Serj, Daron, Shavo et John D., Anthony, Flea, Chad et John F., Paul, Ringo, George et John L., Jack et Meg, Dan et Patrick, Tom, Mark et Travis, Charles, Sharon, Marvin, La Rotule, Soko, RATM, Cocorosie, C2C, Beirut, Stupéflip, Girl in a coma, Buckethead, Jack Johnson, Cage the Elephant, Donavon Frankenreiter, The Moldy Peaches, Motörhead, The Ramones, Didier Wampas, Black Sabbath, John Butler Trio, Birdy Nam Nam, Chinese Man, Spotify, Deezer, YouTube, MySpace, Facebook, Twitter, Madmoizelle, 9gag, Cha, Koreus, Photobomb, Google, Wikipédia, PubMed, La SNCF, Tisséo, la RATP, EasyJet, Covoiturage.fr, Voyage-privé, Pope of Dope, Pilule et André, Les Musclés, R.kelly, Le Buffalo Grill de Biarritz, Ikéa, Carrefour Labège, L'amour est dans le pré, Pékin Express, FNL, HIMYM, TBBT, DH, Grey's Anatomy, Lost, Heroes, Californication, J'irai loler sur vos tombes, Bref, Taratata, les Solidays, Ida y Vuelta, Rio Loco, 1664, Ratz, Leffe, Desperados, Heineken, Kronembourg, Cap Dona, Granola, Kango, Prince, Cookies Hello, Petits écoliers, le café décaféiné, Pénélope B., Boulet, Terri P., Boris V., Harper L., Bobby F., UGC Wilson, Gaumont Wilson, Utopia, L'ABC, Le Cranium, le Loup-Garou, le Saboteur, le Jungle Speed, Hanabi, le Time's Up, Cort, Roland, Aria, Behringer, Dell, Asus, Acer, Apple, l'ADAS, les Ours de Toulouse, La ligue Réunionnaise des Sopalins Vidéos, mes baskets, GT, Madeuh, Vinci Autoroutes, Invader et Banksy.

PS : Merci à toi, personne oubliée dans ces remerciements. Et merci à toi aussi stagiaire, doctorant, ou jeune chercheur de les avoir lu: tu peux ~~peux~~ t'inspirer allègrement de ce manuscrit.

ABREVIATIONS

ADNc: Acide DésoxyriboNucléique complémentaire

ADP: Adénosine DiPhosphate

ARNm: Acide RiboNucléique messenger

ARNr: Acide RiboNucléique ribosomique

ARNt: Acide RiboNucléique de transfert

ATP: Adénosine TriPhosphate

BLAST: Basic Local Alignment Search Tool

DMI: Doesn't Make Infection

DPDCO: DNA-Polymerase Domain Containing ORF

ERO: Espèce Réactive de l'Oxygène

HEO: Homing Endonuclease ORF

ITS: Internal Transcribed Spacer

JGI: Joint Genome Institute

LCO : Lipo-Chito-Oligosaccharide

MA : Mycorhizien(ne) à Arbuscules

miR* : microARN star (brin complémentaire du miR)

NGS : Next-Generation Sequencing

ORF : Open Reading Frame

PCR : Polymerase Chain Reaction

PPA : Pre-Penetration Apparatus

RACE : Rapid Amplification of CDNA Ends

RIN : RNA Integrity Number

RL : Racine Latérale

RNAi : RNA interference

RT : Reverse Transcriptase(ion)

SIR : Small Inverted Repeats

SQO : Sulphide :Quinone Oxidoreductase

UTR : UnTranslated Region

VGE : Variability Generating Elements

TABLE DES MATIERES

TABLE DES MATIERES	1
PREAMBULE	7
INTRODUCTION	9
<i>Bénéfices de la symbiose endomycorhizienne à arbuscules</i>	11
<i>Intérêt agronomique</i>	13
<i>La plante hôte</i>	13
<i>Le symbiote fongique</i>	15
<i>Taxonomie de DAOM197198: De Rhizophagus populinus à Rhizophagus irregularis, en passant par Glomus intraradices</i>	17
<i>Etablissement de la symbiose endomycorhizienne à arbuscules</i>	19
a. Etapes présymbiotiques	19
b. La voie de signalisation commune	21
c. Etapes symbiotiques	23
<i>Présentation des travaux de thèse</i>	27
CHAPITRE I: LE MITOCHONDRIOME DE RHIZOPHAGUS IRREGULARIS	29
INTRODUCTION	31
<i>La production d'énergie mitochondriale</i>	31
<i>Importance de la régulation énergétique lors de l'établissement de la symbiose endomycorhizienne à arbuscules</i>	33
<i>Les génomes mitochondriaux fongiques</i>	35
<i>Génomique de Rhizophagus irregularis : présentation du projet international Glomus Genome</i>	39
A- LE METABOLISME ENERGETIQUE MITOCHONDRIAL CHEZ RHIZOPHAGUS IRREGULARIS DEDUIT DES ETUDES GENOMIQUES	43
<i>Matériels et méthodes</i>	43
<i>Résultats et discussion</i>	45
La chaîne respiratoire : voie principale et voie alternative	45
La respiration anoxique chez <i>Rhizophagus irregularis</i>	47
B- ORGANISATION ET COMPARAISON DES GENOMES MITOCHONDRIAUX DE 5 SOUCHES DE RHIZOPHAGUS IRREGULARIS	51
<i>Résumé de l'article</i>	53
<i>Comparative analysis of mitochondrial genomes of Rhizophagus irregularis - syn. Glomus irregulare - reveals a polymorphism induced by Variability Generating Elements.</i>	55
Abstract	57
Introduction	59
Materials and methods	63
DNA preparation from fungal strains	63
Sequencing and assembling procedures	63

Sequences annotation and comparison	65
Single-nucleotide polymorphism (SNP) analysis	67
Microsatellite analysis	69
Phylogenetic analyses	69
Primer design and PCR	71
Results	73
Direct sequencing of mitochondrial genomes	73
SNP analysis	73
Phylogenetic analyses	75
Genome composition and comparison	77
PCR-based strain identification strategies	83
Discussion	85
Multiplexed direct sequencing of mt genomes	85
Mt genomes showed no variation within strains.	87
Interspecific phylogenetic relationships	87
Intraspecific phylogenetic relationships and cosegregation of mitochondrial and nuclear genomes	89
Mt genome plasticity is generated by three types of VGE	91
Homing Endonuclease ORFs (HEOs)	91
DNA-Polymerase-domain Containing ORF (DPDCO)	91
Small Inverted Repeats (SIRs)	95
Mitochondrial markers are suitable for strain differentiation and monitoring.	97
Acknowledgements	99
Supplementary data	101
<i>Compléments aux travaux publiés</i>	121
Séquençage additionnel d'une souche de <i>Rhizophagus irregularis</i>	121
Expression des gènes mitochondriaux	123
<i>Conclusions et perspectives</i>	123

CHAPITRE II: ANALYSE DU MICROTRANSCRIPTOME DE *MEDICAGO TRUNCATULA* LORS DE LA SYMBIOSE ENDOMYCORHIZIENNE A

ARBUSCULES -----127

INTRODUCTION	127
La symbiose MA: une co-évolution racine-champignon MA	127
Incidence des champignons MA sur le développement racinaire	129
Les microARN: de petits régulateurs aux grands effets	131
Les microARN: incidences sur la symbiose MA et le développement racinaire	133
Objectif et contexte de l'étude: Le projet MIRMED	137
A - ANALYSE DU MICROTRANSCRIPTOME DE <i>MEDICAGO TRUNCATULA</i> LORS DES ETAPES TARDIVES DE LA SYMBIOSE MYCORHIZIENNE A ARBUSCULES	139
Matériels et méthodes	139
Préparation du matériel végétal	139
Mycorhization	139
Nodulation	139

Préparation des banques de petits ARN-----	141
Traitement informatique des séquences-----	141
Quantification des microARN par northern blot-----	143
Validation de l'expression de microARN candidats par la technique de Stem-loop qRT-PCR-----	143
Obtention de transformants pour l'analyse fonctionnelle des microARN -----	145
Résultats et discussion-----	147
Séquençage HiSeq des banques de petits ARN-----	147
Quantification des microARN candidats dans les banques de symbiose MA tardive-----	149
Phénotypage du caractère mycorhizien de plantes sur-exprimant des microARN d'intérêt-----	153
Conclusion -----	157
B - EFFET DES MYC-LCO ET DES FACTEURS NOD SUR LE MICROTRANSCRIPTOME DE MEDICAGO TRUNCATULA -----	159
<i>1) Mise au point des conditions expérimentales pour la réalisation des analyses microtranscriptomiques -----</i>	<i>159</i>
Matériel et méthode-----	159
Préparation du matériel végétal -----	159
Préparation des échantillons et hybridations -----	161
Validation par PCR en temps réel -----	163
Résultats et discussions-----	163
<i>2) Analyse du microtranscriptome de Medicago truncatula en réponse aux Myc-LCO et facteurs Nod-----</i>	<i>165</i>
Matériels et methods-----	167
Préparation du matériel végétal -----	167
Préparation des banques de petits ARN-----	167
Résultats et discussion-----	167
Conclusion -----	173
C - LIMITATION DES APPROCHES DE SEQUENÇAGE MASSIF POUR LA DETECTION ET LA QUANTIFICATION DES MICROARN -----	175
CONCLUSION GENERALE: APPORT SCIENTIFIQUE DES TRAVAUX DE THESE, DEVELOPPEMENTS ET APPLICATIONS ENVISAGEES-----	183
<i>La mitochondrie : un organe révélateur de l'état physiologique du champignon</i>	<i>183</i>
<i>Le génome mitochondrial : une cible de choix pour réaliser des empreintes génétiques et définir un protocole de quantification du champignon -----</i>	<i>185</i>
<i>Les microARN : des éléments régulateurs de la symbiose permettant de définir des lignées végétales supermycorhiziennes -----</i>	<i>187</i>
BIBLIOGRAPHIE -----	189
ANNEXES -----	229

PREAMBULE

Cette thèse a été réalisée dans le cadre d'une convention CIFRE n°1039/2008 liant l'Université Paul Sabatier, établissement de rattachement du Laboratoire de Recherche en Sciences Végétales (équipe Symbiose Endomycorhizienne, resp. Pr G. Bécard) à l'entreprise Agronutrition, une PME toulousaine spécialisée dans le développement, la production et la vente de fertilisants (www.agronutrition.fr). Les champignons mycorhiziens à arbuscules (MA) constituant des bio-fertilisants favorisant la nutrition minérale et hydrique des plantes, l'entreprise Agronutrition s'est orientée vers la production massive d'inoculum afin de les commercialiser. Parallèlement à ce contrat CIFRE, Agronutrition et l'équipe de recherche qui m'accueille sont impliqués dans un projet (Neofertil – Fonds Unique Interministériel) dont l'objet est la collecte, le typage moléculaire et la production de microorganismes d'intérêt agronomique. L'enjeu de mon projet de thèse était donc de définir des séquences génomiques du champignon permettant le typage moléculaire des souches et leur quantification, autant lors de la phase de production que lors de la dissémination au champ. Ce travail a été rendu possible grâce à la participation de l'équipe de recherche, au sein de laquelle j'ai effectué mon doctorat, au projet de séquençage du champignon MA modèle *Rhizophagus irregularis*. La délimitation précise des objectifs de la thèse par l'entreprise m'a par ailleurs laissé un grand degré de liberté dans l'orientation de mes recherches, me permettant de réaliser des études complémentaires sur le partenaire végétal. De par ma formation initiale en génomique fonctionnelle et régulations géniques et épigénétiques, ainsi que la participation de mon équipe d'accueil à un projet visant à étudier des régulations épigénétiques chez la plante hôte *Medicago truncatula* lors de l'établissement de la symbiose MA, j'ai pu m'intéresser à un second volet, en accord avec les intérêts de l'entreprise, portant sur les petits ARN régulateurs. J'évoquerai en fin de manuscrit que ces travaux sont loin d'être dénués de retombées potentielles pour l'entreprise.

En accord avec ces objectifs expérimentaux, le manuscrit de thèse, ici présenté, est divisé en deux chapitres distincts traitant, d'une part, du mitochondriome de *Rhizophagus irregularis* et, d'autre part, de l'analyse du microtranscriptome de *Medicago truncatula* lors de la symbiose MA.

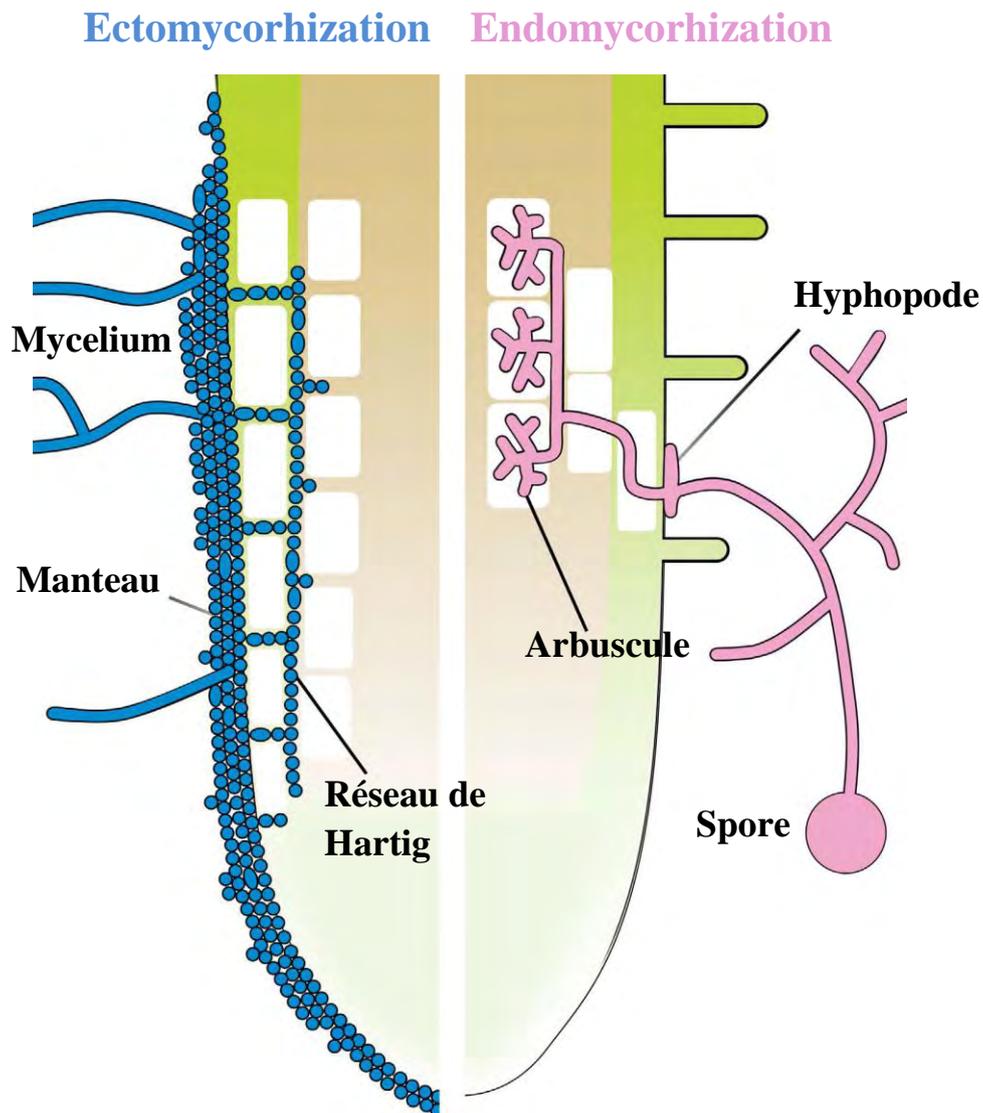


Figure 1 : Illustration de structures de colonisation de la racine lors de la symbiose ectomycorhizienne (gauche) et endomycorhizienne à arbuscules (droite). D'après Bonfante & Genre, (2010).

INTRODUCTION

Les symbioses sont des interactions entre des individus d'espèces différentes qui ont co-évolués afin de tirer parti d'une des ressources du partenaire en échange, ou non, de ses propres ressources. Au sens étymologique et dans sa conception anglo-saxonne, le terme symbiose - « vivre avec » - ne présuppose pas un bénéfice réciproque entre les deux organismes. Ainsi certaines de ces symbioses sont unilatérales comme le parasitisme, lors duquel le parasite tire profit des ressources de l'hôte sans compensation pour ce dernier, ou le commensalisme, au cours duquel le commensal profite des ressources de l'hôte sans jamais lui nuire, enfin d'autres associations sont à bénéfices réciproques : les symbioses mutualistes (Margulis & Fester, 1991). Parmi ces interactions les plus connues, le lichen fait intervenir un champignon et une algue unicellulaire. Le champignon apporte à l'algue des sels minéraux, de l'humidité ainsi qu'un support pour se développer et l'algue fournit au champignon des photosynthétats, l'algue possédant les organites nécessaires à la photosynthèse. Dans ces interactions, on distingue les ectosymbioses, où l'interaction se déroule à l'extérieur de la cellule, des endosymbioses, où les échanges se font de manière intracellulaire (Selosse & Le Tacon, 1998). Les champignons interviennent dans ces deux types de symbiose en association avec les plantes, en produisant, d'une part, des ectomycorhizes, dont les structures d'échanges les plus intimes avec les racines des plantes restent à la surface des cellules (Figure 1) et dont l'organe de fructification issu du mycélium, le sporophore, est la partie hors-sol visible qui constitue le « champignon » tel qu'il est représenté dans l'imaginaire commun. Les champignons réalisant cette symbiose appartiennent pour la quasi-totalité au groupe des dicaryomycètes (principalement basidiomycètes) et peuvent s'associer à près de 8000 espèces végétales qui sont pour la plupart des plantes ligneuses (Rinaldi *et al.*, 2008). D'autre part, certains champignons produisent des endomycorhizes, qui représentent la symbiose fongique la plus répandue chez les espèces végétales (Smith & Read, 2008), en terme de représentativité (près de 80% des espèces végétales sont hôtes) et de répartition écologique (ces champignons sont représentés dans toutes les zones terrestres). Cette interaction consiste en une association entre les racines des plantes et des champignons mycorhiziens à arbuscules (MA) (Figure 1), champignons microscopiques du sol du groupe des Gloméromycètes. Ces champignons assurent aux végétaux qui leurs sont associés un apport minéral, notamment

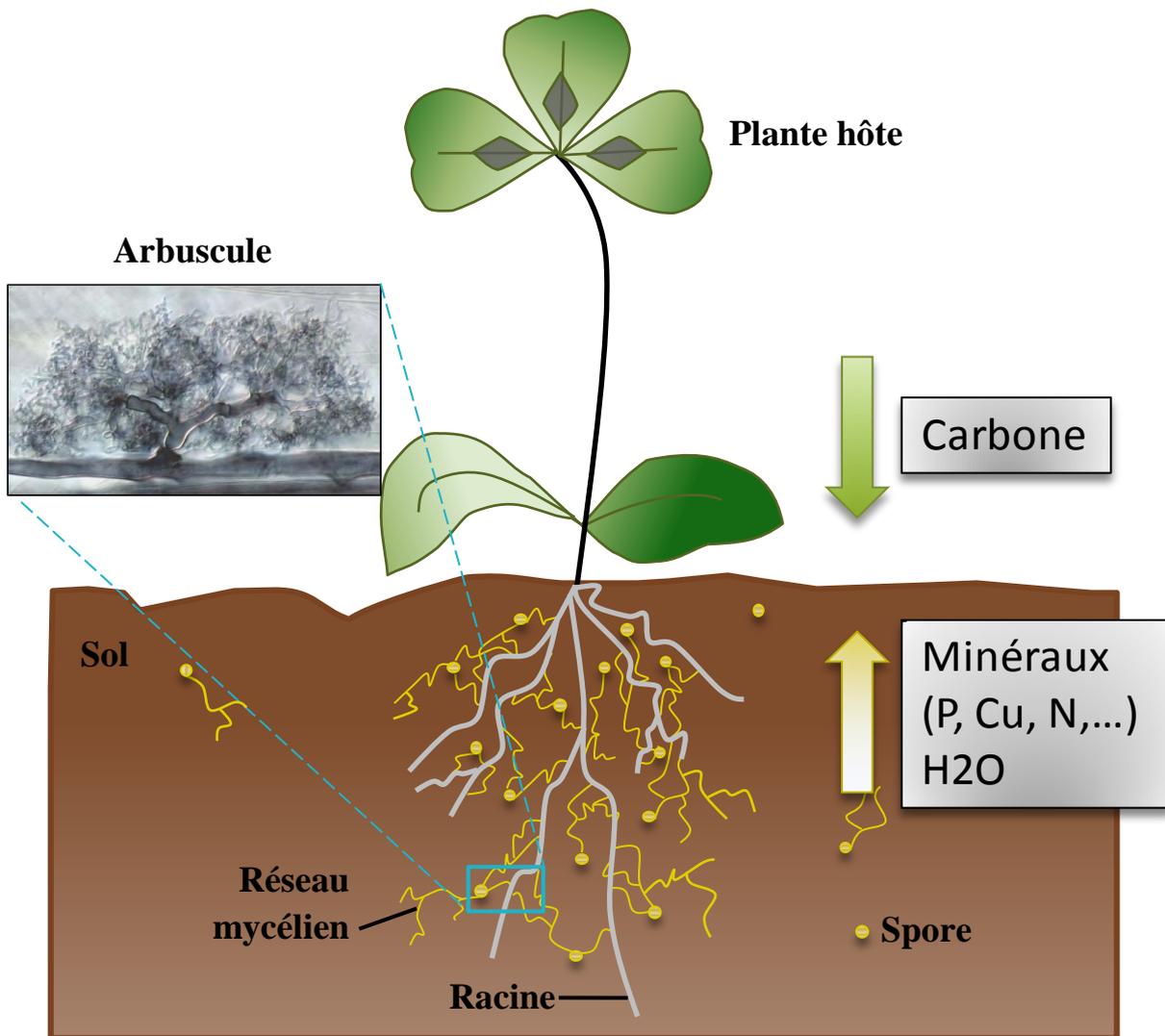


Figure 2 : Représentation schématique de l'établissement de la symbiose endomycorhizienne, du réseau mycélien et des échanges de ressources entre les partenaires.
Photo microscopique d'arbuscule par Mark Brundrett (1989).

du phosphate, et hydrique essentiel à leur développement, accroissant ainsi leur compétitivité dans le milieu naturel.

Bénéfices de la symbiose endomycorhizienne à arbuscules

En complément des systèmes d'acquisition des nutriments des plantes *via* l'épiderme et les poils absorbants racinaires, les champignons MA sont considérés comme des « organes additionnels » d'absorption des nutriments de la plupart des plantes vasculaires (Smith *et al.*, 2009). En effet les champignons une fois associés à la plante hôtes forment un réseau mycélien extraracinaire qui constitue un prolongement du réseau racinaire et permet un recrutement des nutriments et de l'eau dans un volume de sol plus important que celui exploré par les racines seules (Figure 2). Le champignon fournit principalement du phosphore qui est un nutriment peu disponible pour les plantes en raison de sa mobilité réduite dans le sol et sa solubilité partielle. Les plantes peuvent tout de même moduler leur colonisation par le champignon en fonction de différents facteurs environnementaux et notamment en fonction de leur besoin en nutriments. Par exemple, lorsque la concentration en phosphate du sol est élevée, l'établissement de la symbiose est inhibé (Graham *et al.*, 1981) et inversement. Le champignon fournit également de l'eau à la plante : une plante mycorhizée présente une meilleure capacité d'acquisition de l'eau qu'une plante non mycorhizée, que celle-ci soit dans des conditions de carence hydrique ou non, *a contrario* de la régulation par le phosphore (Bárzana *et al.*, 2012). Outre une meilleure tolérance au stress hydrique, une plante mycorhizée est, par ailleurs, moins sensible aux stress biotiques qu'une plante non mycorhizée. Par exemple, le riz *Oryza sativa* mycorhizé présente une induction systémique de gènes régulateurs et marqueurs de la défense, cette pré-stimulation immunitaire lui conférant une meilleure résistance aux futures attaques d'agents pathogènes (Campos-soriano *et al.*, 2011a). La symbiose MA étant une symbiose à bénéfices réciproques, le champignon en retire des photosynthétats carbonés que la plante lui fournit sous forme d'hexoses et qui peuvent constituer jusqu'à 20 % du carbone fixé par la plante lors de la photosynthèse (Smith & Read, 2008). Ces transports de nutriments et photosynthétats s'effectuent *via* des transporteurs spécifiques (Figure 3) : *e.g.* le transporteur de phosphate *MtPt4* chez *Medicago truncatula* qui est spécifiquement exprimé dans les cellules abritant les structures d'échange du champignon (Harrison *et al.*, 2002) ou, chez le champignon, le transporteur de monosaccharide *Mst2* qui permet le transport de carbone de la plante au champignon (Helber *et al.*, 2011). Il a

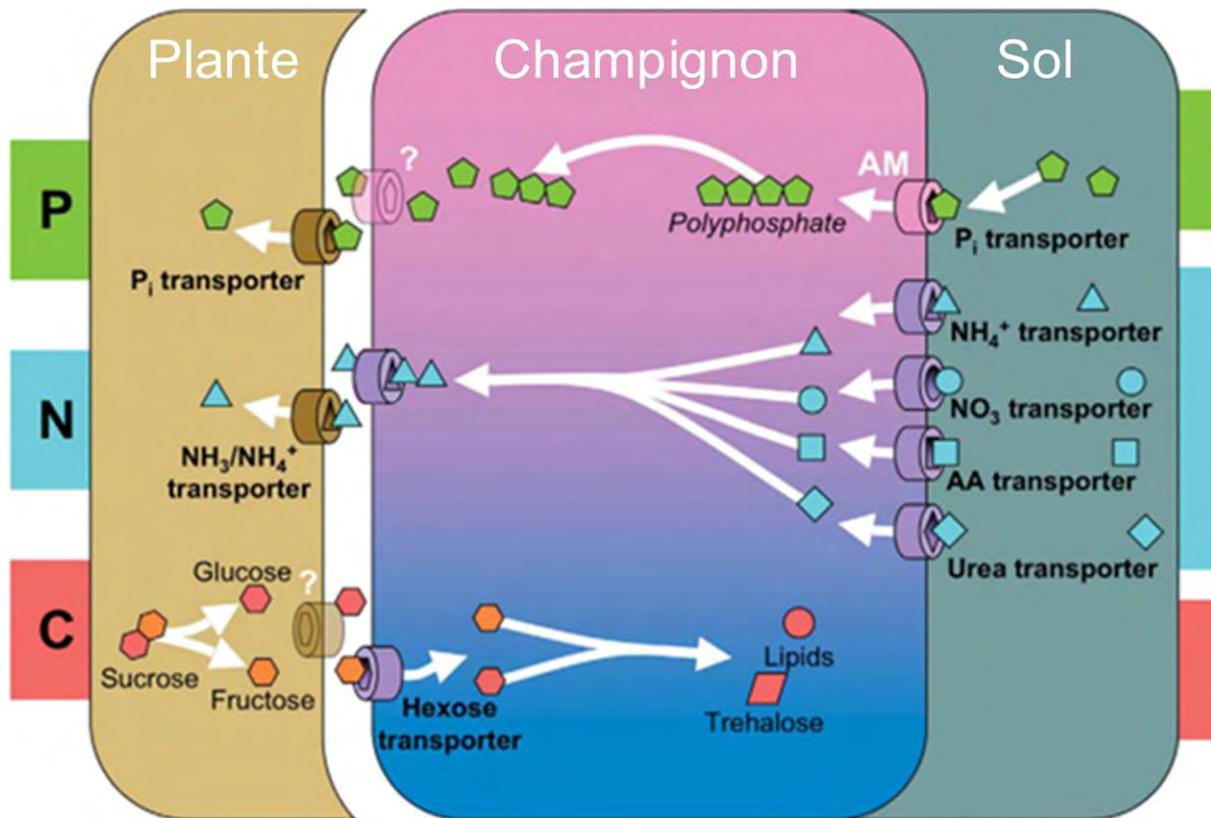


Figure 3 : Schéma résumant les principaux échanges, ainsi que les transporteurs les permettant mis en évidence à ce jour, entre la plante, le champignon et le sol, durant la symbiose endomycorhizienne à arbuscules. D'après Bonfante & Genre (2010), Helber *et al.* (2011) et Fellbaum *et al.* (2012).

été démontré que le blocage de ces transporteurs, soit de plante (Javot *et al.*, 2007) soit du champignon (Helber *et al.*, 2011), bloque également l'établissement de la symbiose.

Intérêt agronomique

En agriculture intensive, l'apport d'engrais et l'utilisation de fongicides ont limité la présence dans les sols de ces micro-organismes bénéfiques. Par ailleurs, la sélection variétale a totalement ignoré ce trait agronomique dans les protocoles de sélection et les variétés élités actuelles ont parfois un mauvais comportement mycorhizien (Zhu *et al.*, 2001). L'agriculture est confrontée à de nouveaux enjeux qui impliquent une nouvelle révolution des systèmes de production : maintenir un niveau optimal de production tout en limitant l'apport d'intrants chimiques et d'eau. L'exploitation des symbioses racinaires est une des pistes qui permettront de relever ces défis. Un projet ANR Recherche & Innovation en Biotechnologie initié en

2006, portant sur la production intensive de ces micro-organismes et particulièrement sur les champignons MA, a débouché sur la création d'une activité de production au sein de l'entreprise Agronutrition (Carbonne, France). En effet, les champignons MA étant des biotrophes obligatoires, ils ne se cultivent qu'en association avec leur hôte. Une stratégie basée sur la co-culture de champignons et de racines transformées par *Agrobacterium rhizogenes* a permis d'obtenir un niveau de production compatible avec une exploitation industrielle de ces champignons.

La plante hôte

La symbiose MA concerne de l'ordre de 80 % des espèces végétales terrestres (Trappe, 1987) représentant plus de 350 000 espèces, parmi lesquelles la grande majorité des espèces d'intérêt agronomique (tournesol, blé, maïs, riz, légumineuses, arbres fruitiers...). Parmi les plantes qui n'établissent pas de symbiose MA, certaines ont développé des associations avec d'autres champignons, notamment les champignons ectomycorhiziens, comme par exemple la plupart des Myrtaceae (Malloch & Pirozynski, 1980). La plante modèle *Arabidopsis thaliana*, comme la quasi totalité des Brassicacées, n'établit pas de mycorhize rendant les ressources et les avantages de ce modèle inaccessibles à l'étude de la symbiose MA et amenant les équipes de recherche à développer des légumineuses modèles telle que la luzerne tronquée *Medicago truncatula*.

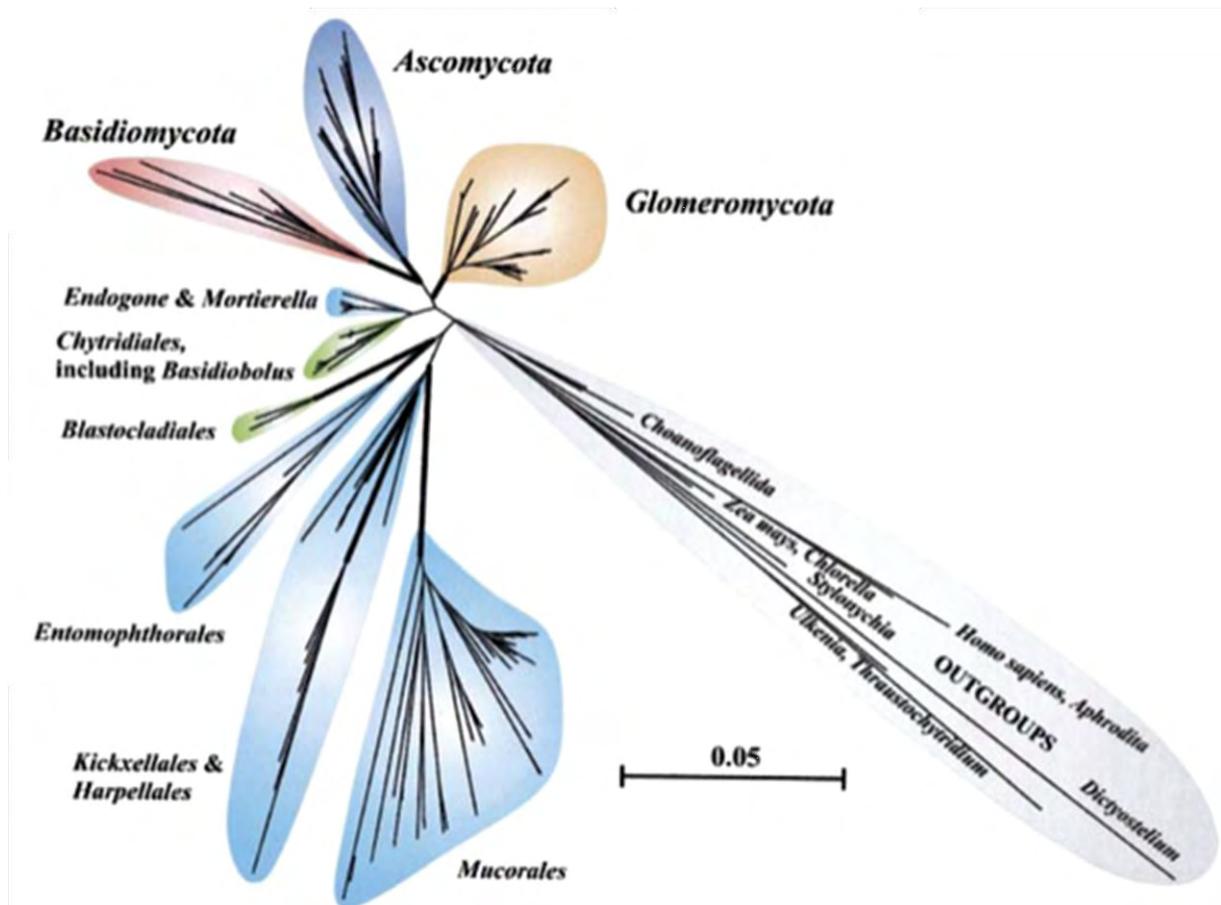


Figure 4 : Phylogénie du règne des champignons basée sur les séquences de la petite sous-unité de l'ARN ribosomique. D'après Schüßler *et al.* (2001).

Cette Fabacée est très proche d'espèces d'intérêt agronomique (luzerne cultivée, soja, pois) rendant les recherches fondamentales réalisées sur ce modèle facilement extrapolables et applicables à ces plantes cultivées. En laboratoire, cette espèce à l'avantage de croître dans un cycle court de développement d'environ 12 semaines, cultivable toute l'année en culture contrôlée, d'être autogame et de posséder un génome relativement petit (équivalent à celui du riz et 7 fois moins grand que le génome du pois) d'environ 450 millions de paires de bases, composé d'environ 48000 gènes, réparties sur 8 chromosomes (Young *et al.*, 2011). Une collection d'outils conséquente, dont la possibilité de réaliser de la transgénèse, la disposition d'une collection de mutants insertionnels et de variants naturels, ainsi que des outils d'analyses transcriptomiques, font de *M. truncatula* un modèle de génétique et génomique fonctionnelles complet. Dans le cadre des symbioses, la luzerne tronquée étant une légumineuse, elle permet l'étude des associations racinaires telles que la nodulation ou la mycorhization.

Le symbiote fongique

Les seconds partenaires de la symbiose MA sont des champignons microscopiques, retrouvés dans le sol sous forme de mycélium et de spores coenocytiques multinucléées, appartenant aux groupes, au sens large, des Gloméromycètes (Figure 4). Ces champignons sont ubiquistes, ne possèdent quasiment pas de spectre d'hôte restrictif et sont actuellement répartis en trois classes (Archaeosporomycetes, Glomeromycetes, et Paraglomeromycetes), cinq ordres (Archaeosporales, Diversisporales, Gigasporales, Glomerales et Paraglomerales), 14 familles, 29 genres et approximativement 230 espèces (Schüßler & Walker, 2010; Oehl *et al.*, 2011). Les Gloméromycètes auraient divergé il y a 450 millions d'années au cours de l'Ordovicien (Schüssler *et al.*, 2001), les premiers fossiles morphologiquement proches des champignons MA datant de cette ère géologique (Remy *et al.*, 1994; Redecker *et al.*, 2000). Ces champignons ont des morphologies variables, les diamètres des spores allant d'environ 30 µm pour certains Glomo-archaeosporoides et Glomo-intrasporoides (Oehl *et al.*, 2011; Krüger *et al.*, 2011) à 400 µm pour les spores de *Gigaspora gigantea* (INternational culture collection of Vesicular Arbuscular Mycorrhizal fungi – INVAM). En tant que biotrophes obligatoires, les champignons MA ont besoin d'un hôte végétal afin de compléter leur cycle de développement. Le champignon peut tout de même se retrouver à l'état de spores dans le sol, consommant avec parcimonie ses réserves carbonées afin de rester viable et de germer

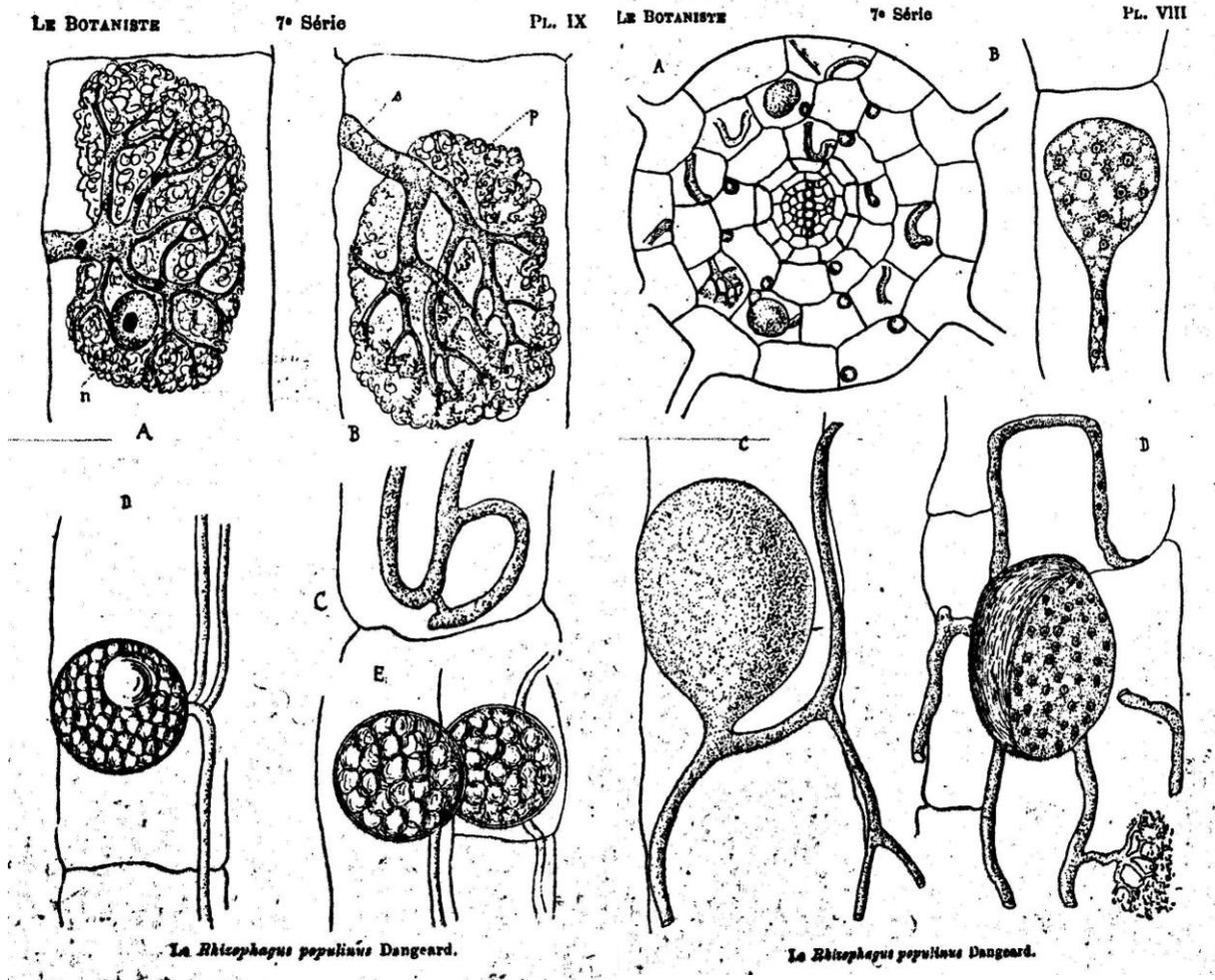


Figure 5 : Structures fongiques intraracinaires décrites et reproduites par Dangeard (1898) lors de la première identification de *Rhizophagus populinus*.

lorsque les conditions de température et d'hygrométrie sont satisfaisantes (Smith & Read, 2008). Lorsque qu'aucun partenaire végétale n'est présent à proximité du champignon, la septation des hyphes germés s'initie (Mosse, 1988), associée à une vacuolisation importante et au rapatriement du contenu du cénocyte vers la spore (Logi *et al.*, 1998). En laboratoire, la souche DAOM197198 de *Rhizophagus irregularis*, isolée à partir de *Fraxinum americana* à Pont Rouge au Québec par C. Plenchette, a été sélectionnée comme modèle des champignons MA en raison de sa capacité de croissance et de sporulation rapide dans un système de culture *in vitro* avec des racines « hairy root » transformées par *Agrobacterium rhizogenes* (Chabot *et al.*, 1992). La disponibilité d'une collection d'EST importante et un projet de séquençage du génome en cours ont désormais imposé cette souche comme modèle.

Taxonomie de DAOM197198: De Rhizophagus populinus à Rhizophagus irregularis, en passant par Glomus intraradices

Jusqu'en 2010, la souche fongique modèle DAOM197198 a été nommée *Glomus intraradices* (N.C. Schenck & G.S. Sm.). Se basant sur la traçabilité des différentes cultures disponibles de DAOM197198 et sur une collection de souches en mycothèques, l'analyse des régions ribosomiques espaceurs internes transcrits (ITS) a révélé que la souche DAOM197198 appartient à un clade distinct de la souche type de *Glomus intraradices* décrite précédemment (Stockinger *et al.*, 2009), mais se regroupe avec l'espèce récemment nommée *Glomus irregulare* (Blaszkowski *et al.*, 2008). La distinction de *Glomus irregulare* / *Glomus intraradices* en deux clades indépendants a été appuyée par des études sur d'autres régions génomiques, comme la grande sous-unité ribosomique mitochondriale (mtLSU) (Thiéry *et al.*, 2010). Plus récemment, Walker et Schüßler ont proposé pour cette espèce le nom de *Rhizophagus irregularis* (NC Schenck & GS Sm.) C. Walker & A. Schüßler (Schüßler & Walker, 2010). *Rhizophagus* a été le premier nom de genre historiquement attribué à cet organisme (Dangeard, 1898) avant d'être ultérieurement mis en synonymie avec *Glomus* par Gerdemann et Trappe (1974). Depuis les travaux taxonomiques de Walker et Schüßler (2010), le nouveau binôme *Rhizophagus irregularis* a été adopté par plusieurs auteurs (Labidi *et al.*, 2011; Sykorova *et al.*, 2011; Derelle *et al.*, 2012; Liao *et al.*, 2012; Pérez-Tienda *et al.*, 2012). Il faut souligner que la nomenclature latine proposée par Dangeard est fondée sur une diagnose erronée. Mandaté pour travailler sur une maladie du peuplier, Dangeard décrit les structures fongiques

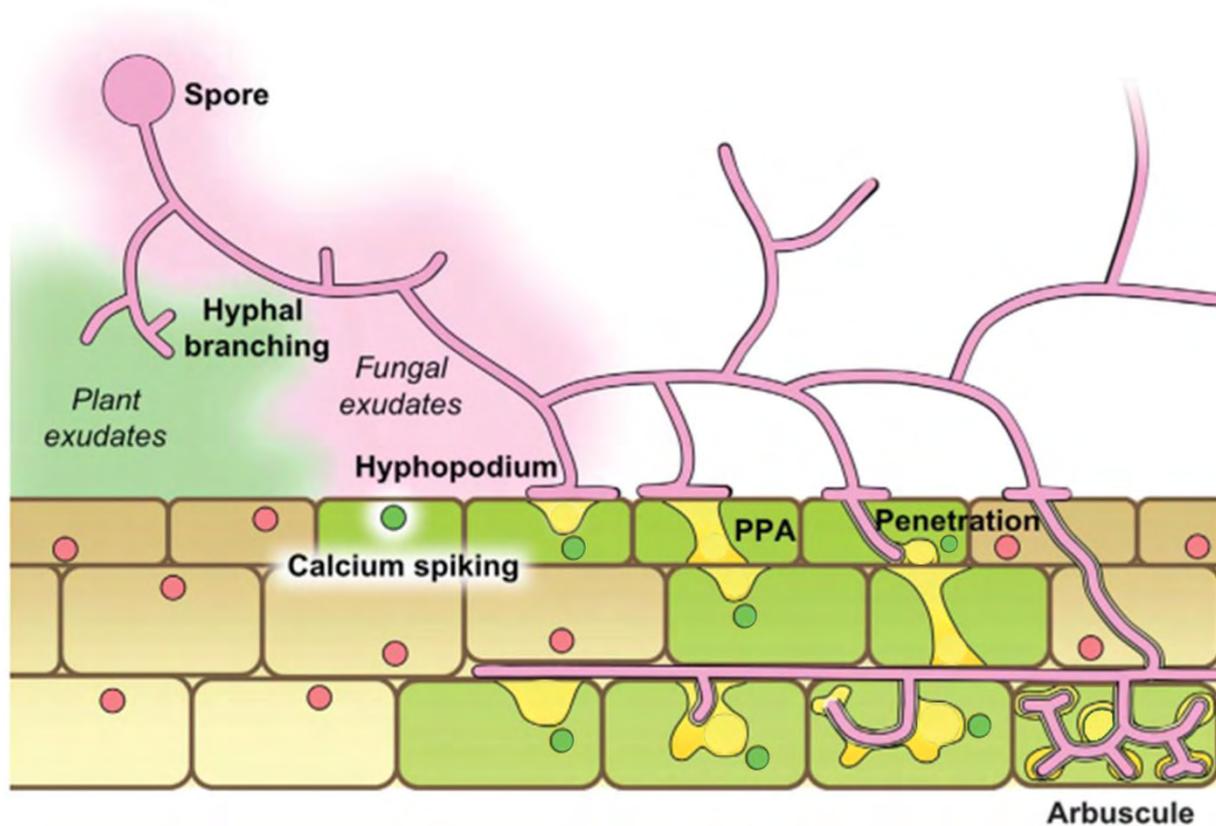


Figure 6 : Schéma représentant les différentes étapes de l'établissement de la symbiose endomycorhizienne, de la diffusion des composés des partenaires au développement des arbuscules. D'après Bonfante & Genre (2010).

observées dans les racines du peuplier, en faisant les premières analyses cytologiques et histologiques de cette interaction (Figure 5) (Dangeard, 1898). En raison de la densité des structures fongiques dans les racines, il a proposé que la maladie des peupliers fût causée par le champignon observé et a donné à ce champignon le nom de *Rhizophagus populinus* (étymologiquement «dévoreur de racines du peuplier»). Cependant, dans une publication ultérieure (Dangeard, 1901), cet auteur explique qu'il n'a pu obtenir les fonds du Ministère de l'Agriculture afin de déterminer si ce champignon est l'agent causal de la maladie. Malgré un diagnostic spéculatif antérieur non confirmé, le champignon conserva son nom de genre *Rhizophagus*. Bien que Koch ait défini dès 1890 les conditions pour établir un lien de causalité entre microbe et symptomatologie, cette erreur ne semble pas empêcher les taxonomistes d'aujourd'hui de conserver cette dénomination de genre, le nom d'espèce *populinus* ayant, lui, été révisé. De même, le fait que «*Rhizophagus*» soit également un nom de genre d'insectes décrit dès 1790 ne semble pas non plus être une limitation car ils appartiennent à deux groupes évolutifs régis par des règles taxonomiques distinctes, même s'il est établi depuis de nombreuses années que champignons et animaux appartiennent au même phylum des Opisthokonta.

Etablissement de la symbiose endomycorhizienne à arbuscules

a. Etapes présymbiotiques

Les spores de champignons MA germent spontanément dans le sol, en puisant dans leurs réserves carbonées, produisant ainsi quelques hyphes germinatifs potentiellement infectieux et leur permettant une plus grande probabilité de mise en contact avec un partenaire. La présence d'une plante hôte potentielle dans l'environnement proche du champignon est détectée grâce à des composés diffusibles sécrétés par la plante (Figure 6) parmi lesquelles des composés appelés strigolactones (Akiyama *et al.*, 2005; Besserer *et al.*, 2006), caractérisés par la suite comme faisant parti d'une nouvelle classe d'hormones végétales (Gomez-Roldan *et al.*, 2007; Umehara *et al.*, 2008). Ces molécules activent en quelques minutes le métabolisme énergétique du champignon, la multiplication de ses mitochondries et sa prolifération cellulaire (ramification) (Besserer *et al.*, 2008). Des travaux antérieurs en présence d'exsudats racinaires ont également montré que ces modifications cellulaires et morphologiques étaient accompagnées de modifications d'expression génique (Tamasloukht *et al.*, 2003). Le champignon, ayant étendu son réseau

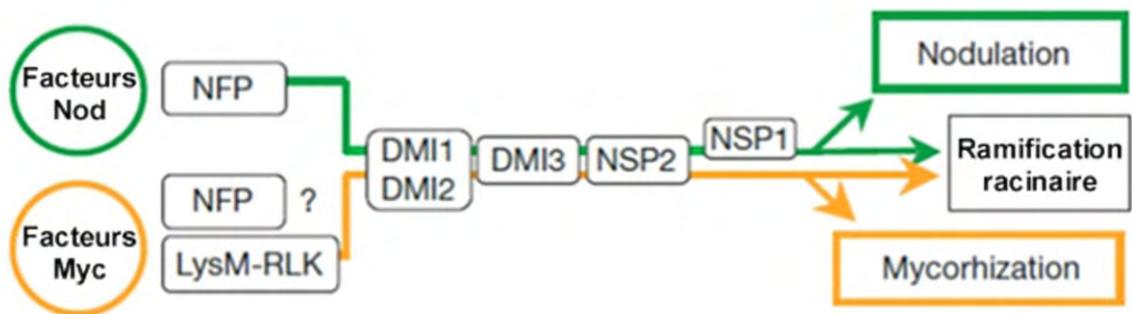


Figure 7 : Schéma de la voie de signalisation symbiotique commune aux symbioses rhizobiennes et endomycorhiziennes, stimulée par les facteurs Nod et les facteurs Myc putatifs. D'après Maillet *et al.* (2011).

mycélien, est alors près à rentrer en contact avec la plante. De son côté, l'hôte végétal perçoit également le champignon grâce à des signaux diffusibles, modifiant ainsi l'expression de ses gènes et son métabolisme (Kosuta *et al.*, 2003; Weidmann *et al.*, 2004; Gutjahr *et al.*, 2009; Kuhn *et al.*, 2010). Parmi ces molécules sécrétées par le champignon, les Myc-LCO ont été récemment mis en évidence comme jouant un rôle dans la préparation de la plante à la colonisation par le champignon (Maillet *et al.*, 2011). Chez *Medicago truncatula*, et chez d'autres espèces mycotrophes non légumineuses, la perception de ce signal favorise la mycorhization, possiblement via la formation de racines latérales (Olah *et al.*, 2005; Maillet *et al.*, 2011). Ces événements présymbiotiques de reconnaissance conduisent donc à la prolifération et ramification de chacun des partenaires, augmentant ainsi leurs chances de contact et l'établissement de la symbiose.

b. La voie de signalisation commune

Une fois ces premiers signaux échangés, les deux partenaires mettent en place une régulation génique propre à l'établissement de la symbiose. Chez la plante, une série de gènes est indispensable à la formation des symbioses rhizobienne et endomycorhizienne. Ces gènes ont été mis en évidence à partir de l'obtention et la caractérisation de mutants non nodulants (Nod-), notamment les mutants de *Medicago truncatula dmi1*, *dmi2* et *dmi3* (Doesn't Make Infection). Ces mutations impliquent des gènes qui contrôlent la transduction du signal lors de la perception des facteurs symbiotiques diffusibles chez *M. truncatula*, qu'ils soient d'origine rhizobienne ou endomycorhizienne, montrant le partage d'une voie commune de signalisation de ces deux symbioses (Wais *et al.*, 2000; Oldroyd & Long, 2003; Kosuta *et al.*, 2008). Cette voie de signalisation symbiotique commune (appelée Common Symbiotic Pathway, CSP) (Figure 7), qui contrôle la mycorhization et la nodulation, est activée par des lipo-chito-oligo-saccharides - facteurs Nod et Myc-LCO - via des récepteurs LysM-RLK (Radutoiu *et al.*, 2003). Les LCOs déclenchent des oscillations du calcium épidermique chez la plante (Oldroyd & Long, 2003; Kosuta *et al.*, 2008; Sieberer *et al.*, 2009; Chabaud *et al.*, 2011), perçues par une calmoduline calcium-dépendante (Lévy *et al.*, 2004) entraînant la modification d'expression de gènes en aval de la CSP (Kuhn *et al.*, 2010; Maillet *et al.*, 2011). Ces réponses oscillatoires du calcium présentent des patterns différents selon les molécules utilisées (Bonfante & Requena, 2011), probablement pour mettre en place des réponses cellulaires et expressions géniques spécifiques de chaque symbiose (Kosuta *et al.*, 2008), et sont transmises grâce aux

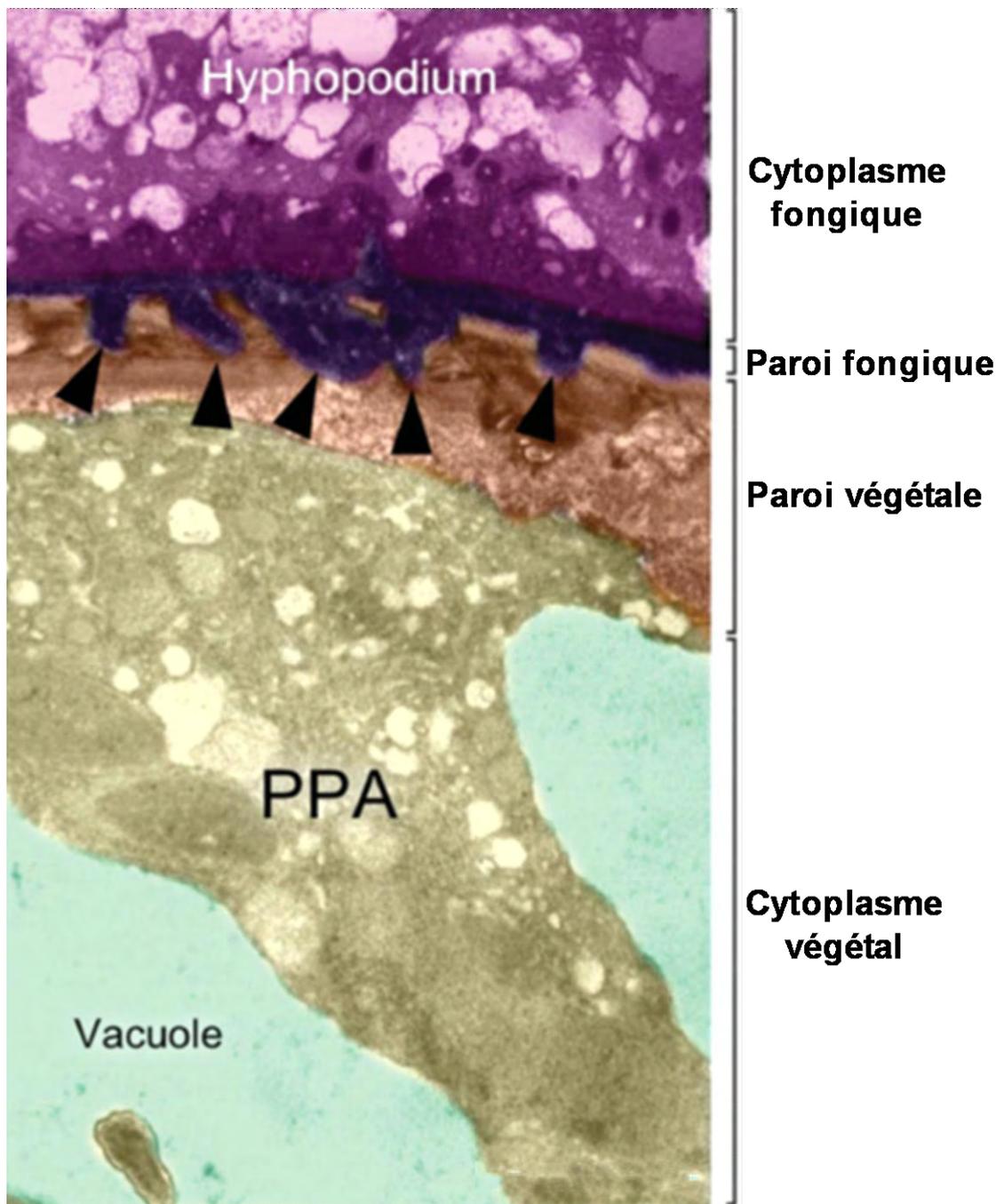


Figure 8 : Vue microscopique d'une cellule épidermique de carotte lors de la formation de l'appareil de pré-pénétration (PPA). L'hyphopode fongique (violet en haut de l'image) est ancré dans la paroi cellulaire de la plante par des protrusions (flèches). D'après Bonfante & Genre (2010).

protéines des gènes *Dmi*. Tout d'abord, la transduction du signal se fait *via* le gène *Dmi2* qui encode une protéine kinase à motif LRR (Répétitions Riches en Leucine) (Endre *et al.*, 2002; Stracke *et al.*, 2002). Suite aux oscillations calciques, le déséquilibre de charge, causé par ces dernières, est compensé grâce au canal cationique perméable au potassium présent sur la membrane nucléaire encodé par le gène *Dmi1* (Ané *et al.*, 2004; Charpentier *et al.*, 2008). Afin de décoder les oscillations calciques nucléaires, une kinase dépendante au calcium, encodée par le gène *Dmi3*, est présente à la fin de la cascade de signalisation (Gleason *et al.*, 2006). Ce décodage permet alors l'activation du facteur de transcription *Nsp2*, appartenant à la famille GRAS (Kaló *et al.*, 2005; Maillet *et al.*, 2011), synchronisant probablement l'expression de gènes impliqués dans l'établissement de ces symbioses. Récemment, Czaja *et al.* (2012) ont montré par PCR en temps réel que cinq gènes induits par les myc-LCO sur dix testés restent induits dans le mutant *nsp1-1*, apportant des preuves d'une implication partielle de ce facteur de transcription dans la réponse aux myc-LCO et, *a fortiori*, dans la CSP.

c. Etapes symbiotiques

Une fois que les signaux respectifs des partenaires ont produit leurs effets, le contact, alors facilité par une ramification importante, s'établit et les premiers hyphes du champignon atteignant la surface racinaire se différencient en une structure appelée hyphopode (Figure 6 et 8) (Genre *et al.*, 2005). Ces structures sont des extrémités aplaties d'hyphes, de forme lenticulaire, attachées à la racine, qui permettent la fixation et la pénétration du champignon à l'intérieure de la racine (Garriock *et al.*, 1989). La formation de ces structures est hautement contrôlée par la concentration en phosphate et, plus cette concentration est grande, moins d'hyphopodes sont formés (Balzergue *et al.*, 2011). Lorsque les conditions sont favorables, notamment un taux réduit de phosphate, la plante se prépare à la pénétration du champignon en développant un appareil de pré-pénétration (PrePenetration Apparatus - PPA) (Figure 6 et 8) (Genre *et al.*, 2005). Cette structure se met en place, de manière dépendante à la CSP, 4 à 5 heures après la formation de l'hyphopodium, et constitue un chemin subcellulaire prédéterminé que le champignon emprunte. La formation du PPA est guidée par la migration du noyau à travers la cellule. Le « tunnel » en résultant est traversé par un tube creux composé de microtubules associés aux citernes du réticulum endoplasmique. Une fois ce tunnel en place, les hyphes du champignon peuvent alors y pénétrer et atteindre le cortex racinaire où il va quitter les

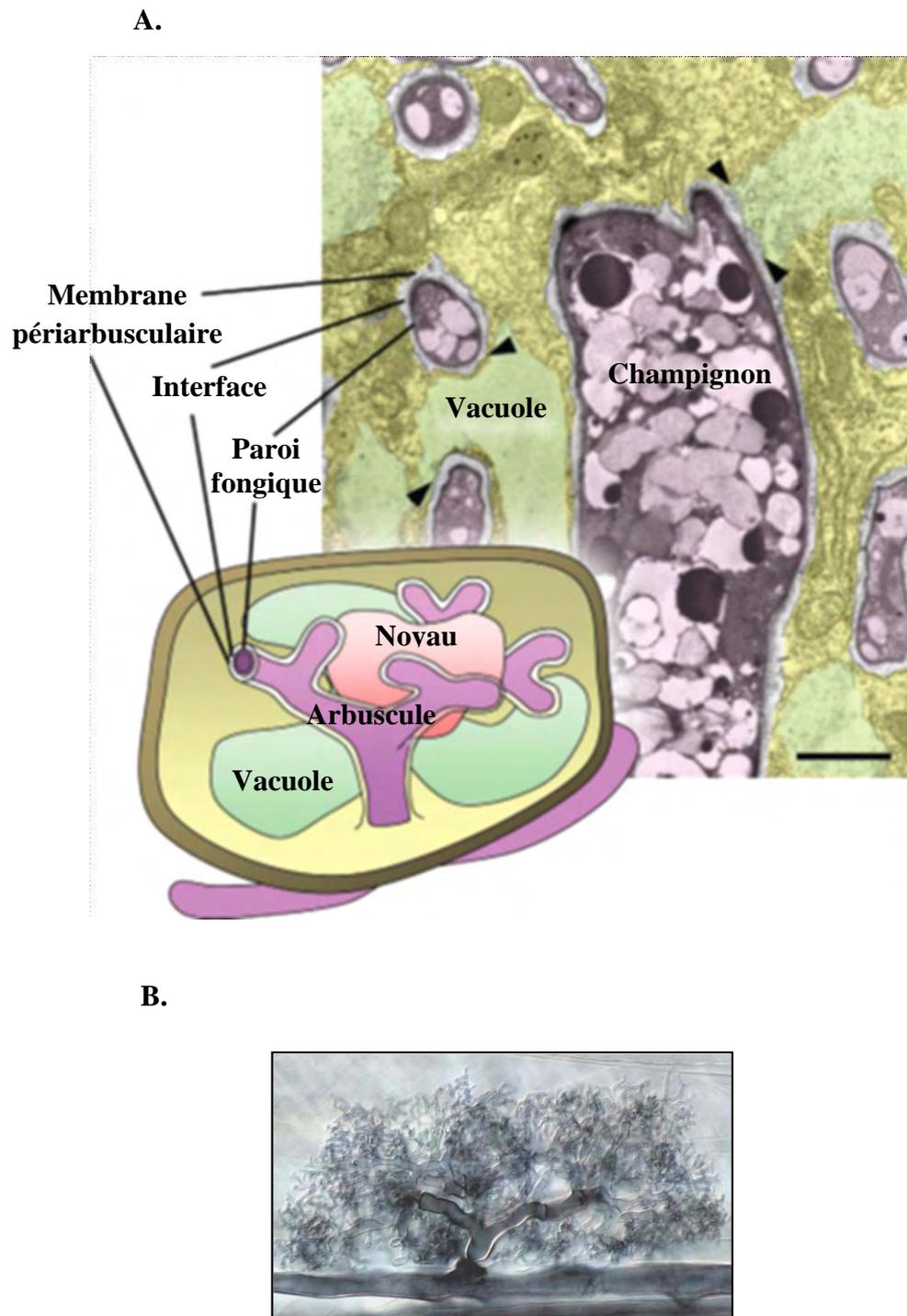


Figure 9 : A. Organisation d'une cellule corticale de plante abritant un arbuscule de champignon (schéma et prise de vue microscopique). D'après Bonfante & Genre (2010). B. Vue microscopique d'un arbuscule à l'intérieur d'une cellule de plante prise grâce à la technique de Nomarski, par Mark Brundrett (1989).

cellules et pénétrer l'apoplaste afin de croître latéralement le long de l'axe racinaire. Cette étape de pénétration est, entre autres, régulée par une protéine, la vapyrine, dont le mutant du gène la produisant présente une formation normale d'hyphopodes et du PPA mais une pénétration avortée des hyphes du champignon (Pumplin *et al.*, 2010). Une fois les cellules du cortex atteintes, le champignon induit la formation d'un PPA, dans les cellules corticales, similaire à celui retrouvé à travers les cellules de l'épiderme et du cortex externe. Les hyphes du champignon empruntent alors ce chemin et, une fois dans la cellule, se ramifient intensément pour donner une structure d'échange appelée arbuscule, en relation avec sa structure en forme d'arbuste (Figure 9). Cet organe fongique est isolé du cytoplasme de l'hôte par une membrane périarbusculaire dérivée de la plante, sur une surface accrue par la ramification importante du champignon, favorisant ainsi les échanges. Les arbuscules et les cellules les abritant sont des lieux d'aménagements cellulaires spécifiques à la symbiose (Fester *et al.*, 2001) donnant lieu à des régulations géniques présentes exclusivement dans ces cellules de l'interface champignon-plante (Gaude *et al.*, 2011) comme, par exemple, le transporteur de phosphate *MtPt4* (Harrison *et al.*, 2002) qui est essentiel au maintien de la symbiose. Le mutant *mtpt4* présente une dégénérescence prématurée des arbuscules, mais l'état de développement des arbuscules redevient égal à celui du sauvage correspondant, et le champignon peut terminer son cycle de vie, dès lors où la plante est carencée en azote (Javot *et al.*, 2011). Le développement des arbuscules est donc régulé par plusieurs paramètres et gènes dont certains ont été identifiés et étudiés. Par exemple, deux gènes codant pour des transporteurs ABC, *Str* et *Str2*, ont été retrouvés induits dans les cellules corticales abritant des arbuscules et le mutant *str*, ainsi que les transformants dont l'expression du gène *Str2* est inhibée par ARN interférant, ne produisent pas d'arbuscules, montrant leur implication dans le processus de développement de ces derniers (Zhang *et al.*, 2010). Récemment, Ivanov *et al.* (2012) ont mis en évidence que deux gènes homologues codant pour des protéines associées aux membranes de vésicules exocytiques, *Vamp721d* et *Vamp721e*, impliquées dans la formation de la membrane périarbusculaire, sont essentiels au bon développement des arbuscules.

Présentation des travaux de thèse

Mes travaux de thèses sont ici présentés en deux chapitres distincts.

Le premier traite du mitochondriome de *Rhizophagus irregularis*, d'une part, en répertoriant les gènes impliqués dans le métabolisme mitochondrial à partir des données génomiques fournies dans le cadre du « consortium Glomus », d'autre part, en étudiant le génome mitochondrial par séquençage de ce génome chez différentes souches. La description d'un répertoire de gènes mitochondriaux a pour objectif de disséquer précisément le système de respiration cellulaire de ces champignons dont la régulation semble jouer un rôle majeur dans les étapes précoces de la symbiose MA. L'étude du mitochondriome a pour autre objectif d'étudier la variabilité des génomes de cet organite au sein de l'espèce *Rhizophagus irregularis*, permettant le développement de marqueurs moléculaires afin de différencier les souches chez cette espèce et également d'affiner la phylogénie des Gloméromycètes au sein du règne fongique.

Le second chapitre porte sur le partenaire végétal de la symbiose MA, ici *Medicago truncatula*, et plus particulièrement sur la régulation de ses gènes lors des différentes étapes de l'établissement de cette symbiose. Deux volets portent sur la recherche de microARN, des petits éléments modulant l'expression génique, potentiellement impliqués dans la régulation de l'établissement des symbioses mycorhizienne MA et rhizobienne, et dont l'expression est modifiée dans les étapes tardives et précoces, respectivement. Un sous chapitre supplémentaire présente l'identification de gènes marqueurs de l'induction par les Myc-LCO lors des étapes présymbiotiques, permettant de définir les conditions expérimentales de réponses à ces signaux fongiques afin de réaliser les études du microtranscriptome de *M. truncatula* lors des étapes présymbiotiques.

CHAPITRE I: LE MITOCHONDRIOME DE *RHIZOPHAGUS IRREGULARIS*

Présentes dans tout système cellulaire eucaryote autonome, les mitochondries sont des organites présentant un rôle majeur dans le métabolisme énergétique, depuis la régulation du métabolisme lipidique à la production énergétique en tant que telle, et sont également considérées comme une des plateformes de signalisation de la cellule. De ce fait, les mitochondries jouent un rôle dans l'adaptation des cellules à leur environnement. Chez *Rhizophagus irregularis* en particulier, les mitochondries jouent un rôle prépondérant durant les phases précoces de croissance et d'interaction avec la plante hôte. Outre leur intérêt dans les études sur la physiologie fongique, les mitochondries ont également un intérêt dans les études de phylogénie et de génétique des populations des champignons, en général, et des champignons MA, en particulier. Le génome de cet organite, appelé génome mitochondrial (mt), est à la base de la définition de nombreux marqueurs moléculaires exploités dans l'étude des populations de champignons MA.

Dans ce chapitre, je présente mes travaux réalisés sur le génome mitochondrial de *Rhizophagus irregularis* apportant un regard nouveau sur A) la caractérisation des gènes impliqués dans le métabolisme énergétique mitochondrial et B) l'intérêt de l'étude du génome mitochondrial à des fins de typage des souches fongiques de *R. irregularis*. En introduction de ces deux volets, après un rappel sur la production énergétique mitochondriale, il sera précisé l'importance de la régulation de la production énergétique dans la mise en place de la symbiose MA, ainsi que l'état des connaissances sur la génomique de *Rhizophagus irregularis* à l'initiation de mes travaux de thèse.

INTRODUCTION

La production d'énergie mitochondriale

Les mitochondries sont considérées comme étant la centrale de production d'énergie des cellules eucaryotes. Un rôle essentiel est la synthèse d'ATP après l'achèvement du cycle de Krebs suivi de la phosphorylation oxydative réalisée par la chaîne respiratoire. Quelles que soient les sources de carbone, l'acétyl-CoA sera la molécule utilisée par le cycle de Krebs qui va permettre la libération d'énergie sous forme de NADH, H⁺ et FADH₂, futurs donneurs d'électrons pour la chaîne respiratoire. Grâce aux complexes enzymatiques I et II (NADH déhydrogenase et Succinate déhydrogenase) présents dans cette chaîne, les électrons sont alors transférés à la coenzyme ubiquinone qui va alors les transférer à son tour aux complexes III et IV (Cytochrome c réductase et Cytochrome oxydase), catalyseurs de la réaction de réduction de l'O₂, accepteur final des électrons, en H₂O. En parallèle, les réactions d'oxydoréductions ont entraîné la formation d'un gradient de proton entre la matrice et l'espace intermembranaire permettant alors au complexe V (ATP synthase) d'utiliser ce gradient afin de produire de l'ATP à partir d'ADP. Les mitochondries doivent donc être constamment fonctionnelles, s'adapter à l'environnement de l'organisme et au besoin énergétique de la cellule, gérer le stock de carbone et ne pas produire un excès d'espèces réactives à l'oxygène (ERO). La mitochondrie est le site majeur de production des ERO qui peuvent entraîner un vieillissement plus rapide associé à une perte de fonction de certaines protéines, des mutations ou une altération des membranes. Au cours de l'évolution, les différents organismes ont dû s'adapter aux changements de ressources disponibles pour produire de l'énergie, et certains d'entre eux possèdent des systèmes hybrides leur permettant de changer de type de respiration, selon l'environnement où ils se trouvent. La levure *Saccharomyces cerevisiae* est un modèle de cette capacité d'adaptation des mitochondries. En effet, cet organisme est un anaérobe facultatif : en présence d'oxygène, la levure produit de l'énergie par la voie principale décrite précédemment et, lorsque l'oxygène se raréfie, cette dernière change son métabolisme et produit de l'énergie par la fermentation alcoolique qui consiste à transformer le pyruvate en éthanol et en dioxyde de carbone. De la même manière, la source de carbone utilisée peut varier. Bien que privilégiant le glucose, la levure est capable de métaboliser le maltose ou le saccharose. Lorsque la levure utilise le glucose en tant que source de carbone en anaérobie, la glycolyse permet la

production de 18 fois moins d'énergie et produit alors moins d'ERO. Tous ces changements sont accompagnés de modifications transcriptomiques et il a été montré que l'expression de près de 500 gènes diffère significativement entre des levures cultivées en condition aérobie et anaérobie, dont 28 sont essentiels à la croissance anaérobie (Snoek & Steensma, 2007). Parmi ces gènes essentiels se trouve un gène codant pour une fumarate réductase (Arikawa *et al.*, 1998). Cette enzyme catalyse irréversiblement la réduction du fumarate en succinate, lors de la respiration anaérobie, qui fait alors office d'accepteur final d'électron en remplacement du dioxygène lors de la respiration aérobie. Récemment, la présence et l'activité de cette enzyme ont été mises en évidence chez des spores en germination de *Rhizophagus irregularis* suggérant ainsi la capacité de ce champignon à produire de l'énergie par une voie anaérobie (Sędziewska *et al.*, 2012). Ces résultats sont à ajouter à la description faite dans notre équipe de la stimulation du métabolisme énergétique des champignons MA par les exsudats racinaires (Besserer *et al.*, 2006). La quasi absence de connaissances sur le génome mitochondrial de ces champignons symbiotiques m'a donc conduit à identifier les différentes voies du métabolisme énergétique mitochondrial par une stratégie d'analyse génomique.

Importance de la régulation énergétique lors de l'établissement de la symbiose endomycorhizienne à arbuscules

Plusieurs travaux précisent l'implication des mitochondries lors des étapes présymbiotiques de l'association entre la racine des plantes et les champignons MA. Tamasloukht et co-auteurs (2003) ont montré que la régulation des gènes de *G. rosea* est également modifiée lors de la perception des composés végétaux. Dès 30 min après cette perception, le transcriptome de *G. rosea* est modifié atteignant un maximum à 1 h. Des transcrits provenant de dix gènes ont été retrouvés sur-accumulés comme, par exemple, des séquences ribosomiques nucléaires et mitochondriales ou bien un gène montrant des similarités avec une monooxygénase cytochrome P450 qui permet la détoxification des défenses de plante (Van Etten *et al.*, 1989). Certains de ces gènes ont également été identifiés comme impliqués dans l'activité mitochondriale comme, par exemple, une translocase mitochondriale permettant d'acheminer l'ADP aux organelles, ou une pyruvate carboxylase qui catalyse la carboxylation du pyruvate en oxaloacetate. Les composés végétaux stimulent donc une induction rapide de gènes mitochondriaux, ou de gènes impliqués dans des fonctions mitochondriales, ce qui suggère que l'activité

mitochondriale est une cible primaire de la stimulation et qu'elle joue un rôle essentiel dans le changement d'un état asymbiotique à un état présymbiotique du champignon. Après la caractérisation des strigolactones comme composés inducteurs de la ramification des champignons MA (Bécard *et al.*, 2004), Besserer et co-auteurs (2006, 2008, 2009) ont montré que dans les minutes qui suivent la perception des strigolactones par *Gigaspora rosea*, les mitochondries fongiques produisent et consomment une grande quantité de NADH reflétant ainsi une augmentation de leur activité. Après une heure, la production d'ATP cellulaire augmente à son tour, les mitochondries prennent une conformation bacilliforme, modifient leurs mouvements et la prolifération mitochondriale s'intensifie fortement (23% d'augmentation de densité). L'inhibition de la voie principale de phosphorylation oxydative par différents inhibiteurs (Rotenone, Antimycine A et KCN) entraîne une réduction de la réponse de ramification des hyphes de 36,7% à 50%. La suppression n'étant pas complète, les auteurs ont suggéré la présence d'une voie alternative qui permet le maintien de l'activité du cycle de l'acide citrique alors que la chaîne respiratoire principale est inhibée. Ces voies alternatives sont déjà connues chez d'autres champignons (Joseph-horne *et al.*, 2001) et caractérisées par des NADH dehydrogenases alternatives et des oxydases alternatives (AOX). Chez *Rhizophagus irregularis*, l'organisme modèle des champignons MA, très peu de connaissances sont établies sur sa respiration (Tamasloukht *et al.*, 2003). Dans le cadre du consortium de séquençage du génome de la souche DAOM197198 de *Rhizophagus irregularis*, mon équipe était en charge de l'annotation des gènes impliqués dans la production d'énergie, les réactions d'oxydoréduction et la mitochondrie. Cette étude m'a alors permis de mettre en évidence la potentielle chaîne respiratoire de ces champignons.

Les génomes mitochondriaux fongiques

Chez les eucaryotes, une petite proportion des protéines mitochondriales est produite dans la mitochondrie par un génome spécifique aux mitochondries, appelé génome mitochondrial. La grande majorité des protéines mitochondriales est en fait codée par des gènes présents dans le génome nucléaire qui, une fois synthétisée dans le cytosol, est alors exportée vers la mitochondrie. Quelques rares exceptions d'eucaryotes ne possèdent pas de mitochondrie ou génome mitochondrial mais leurs cellules sont tout de même équipées de dérivés de mitochondrie (*e.g.* hydrogenosomes). Le génome mitochondrial peut prendre plusieurs conformations selon les organismes comme, par

exemple, une grande molécule linéaire chez *Tetrahymena pyriformis*, trois molécules circulaires de différentes tailles chez le champignon *Spizellomyces punctatus*, ou bien une centaine de molécules linéaires chez le protiste *Amoebidium parasiticum*. La forme la plus répandue chez les eucaryotes est une molécule unique et circulaire. La taille moyenne du génome mitochondrial d'eucaryote est comprise entre 15 kb et 60 kb mais certains extrêmes existent comme chez les parasites du genre *Plasmodium* qui possèdent un génome très réduit de 6 kb ou, à l'inverse, le riz *Oryza sativa* qui possède un des plus gros génomes mitochondriaux atteignant les 490 kb. Indépendamment de la taille du génome, le nombre de gènes hébergés par les molécules codantes mitochondriales varie, allant de 5 gènes par génome chez *Plasmodium* à 100 chez les flagellés jakobides, avec une moyenne de 40 à 50 gènes au sein des eucaryotes. La quantité de gènes que l'on peut retrouver dans un génome mitochondrial d'eucaryote n'est pas corrélée avec la taille de ce dernier et il est connu que la principale modification de taille est due à la réorganisation des régions intergéniques qui subissent des insertions de répétitions en tandem, des éléments mobiles ou bien l'intégration de plasmide. Des événements géniques sont également responsables de l'augmentation de la taille comme par exemple la prolifération d'introns autoépissables qui ont envahi les génomes mitochondriaux comme, par exemple, chez le champignon filamenteux *Podospora anserina* où la proportion de séquences introniques atteint 75% du génome.

Les champignons suivent les mêmes règles que la globalité des eucaryotes, en termes de matériel mitochondrial, avec une grande variabilité de tailles, structures, et organisations de gènes et introns. Plusieurs champignons sont des aérobies obligatoires et ont une croissance filamenteuse qui nécessite une extension apicale et forme des hyphes coenocytiques. La fusion d'hyphes par anastomose entraîne le mélange de matériel mitochondrial entre individus et permet donc la rencontre de génomes mitochondriaux différents et leur possible recombinaison (Westermann, 2002). La plupart des champignons hérite leur génome d'un seul parent : c'est l'hérédité uniparentale. Chez *Podospora anserina* et *Neurospora crassa* il a été montré que le génome est hérité maternellement (Röhr *et al.*, 1998) mais certains champignons comme la levure *S. cerevisiae* peuvent hériter des deux génomes mitochondriaux parents et il faut une vingtaine de cycle de multiplication végétative de la descendance pour voir disparaître un des deux génomes (Birky, 2001). Les génomes mitochondriaux fongiques tendent à être riches en nucléotides A-T et leur taille s'étend de 19,4 kb pour *Schizosaccharomyces*

pombe à 175 kb chez *Agaricus bisporus* (Hudspeth, 1992). La variabilité du rapport entre le nombre de gènes et la taille du génome est flagrante lorsque l'on s'attarde sur le cas des levures où *Saccharomyces cerevisiae* possède 35 gènes pour une taille de 85778 pb alors que *Schizosaccharomyces pombe* est composé de 36 gènes dans un génome de 19431 pb.

Génomique de Rhizophagus irregularis : présentation du projet international Glomus Genome

En 2004, suite au succès du projet de séquençage du génome du peuplier, le Joint Genome Institute (JGI) a initié le projet de séquencer les génomes des micro-organismes présents dans le mésocosme de ce dernier afin de développer des outils génomiques qui faciliteraient l'étude des interactions qui y surviennent (Lammers *et al.*, 2004). Parmi ces organismes, *Rhizophagus irregularis* a été sélectionné et un projet de séquençage de son génome fut initié en 2005 (Coordination Francis Martin – INRA Nancy). En Février 2009, lors du commencement de ma thèse, le projet n'avait pas abouti et le génome était très parcellaire. Quelques faits expliquent la majeure partie des problèmes d'assemblage rencontrés : tout d'abord, une estimation imprécise et fluctuante de la taille du génome nucléaire. Le projet a pris comme base, pour le calcul de profondeur attendue, la taille estimée par Hijri et Sanders (2004) grâce à des techniques de cytométrie en flux, cinétiques de réassociation et reconstruction génomique. Le génome de *R. irregularis* était alors estimé à 16 Mb. Dès lors, 344 Mb de séquences génomiques ont été générés. Toutefois, les calculs de profondeur ont permis d'estimer à partir de ces données une couverture maximale de 3,1 x. La taille du génome était alors estimée à plus de 110 Mb et ce résultat a été récemment confirmé par Sędziewska et co-auteurs (2011) qui ont calculé une taille d'environ 155 Mb pour la souche DAOM197198. Par la suite, des données obtenues sur deux souches différentes (DAOM197198 et FACE#494) ont été mélangées afin d'atteindre une profondeur appropriée à l'assemblage. Ces souches peuvent présenter des génomes différents et donc fausser le travail des assembleurs qui en résulte. Le même problème a été rencontré de manière intrasouche, incriminant plusieurs versions (haplotypes) d'une même région ou bien la présence de plusieurs répétitions d'un même gène/pseudogène générant ainsi un haut taux de polymorphisme. La présence d'un grand génome ou plusieurs génomes de tailles plus réduites (homocaryotie vs hétérocaryotie), au sein d'un même individu, divise encore la communauté et les conclusions d'articles sur le sujet (*Homocaryotie* : Pawlowska & Taylor, 2004 ;

Rosendahl & Stukenbrock, 2004; Stukenbrock & Rosendahl, 2005). *Hétérocaryotie* :(Bever & Morton, 1999; Hijri *et al.*, 1999; Kuhn *et al.*, 2001; Hijri & Sanders, 2004; Hijri & Sanders, 2005; Bever & Wang, 2005; Croll & Sanders, 2009). Les champignons MA sont d'autre part connus pour avoir des génomes qui ont un pourcentage de GC très bas (Hosny *et al.*, 1997) et celui de *R. irregularis* a été estimé à environ 30% (Martin *et al.*, 2008). Les assembleurs de séquences sont très sensibles à ce genre de paramètres et l'assemblage de certaines zones avec de tels niveaux de GC est rendu impossible. Enfin, les champignons MA étant des biotrophes obligatoires, il est impossible de les cultiver isolément sauf à les co-cultiver avec des systèmes racinaires autonomes comme les racines transformées par *Agrobacterium rhizogenes* (Bécard & Fortin, 1988). Le risque est alors la présence de matériel génomique du partenaire végétal dans les extractions fongiques pouvant fausser le bon assemblage des séquences. Toutes ces caractéristiques cumulées ont mené à la production de fragments génomiques dont le plus long contig nucléaire était de 9,6 kb (70,8 kb pour le génome mitochondrial). Afin d'accélérer et améliorer la situation, un consortium de séquençage de *R. irregularis* a été constitué et une campagne d'annotation manuelle d'EST a été initiée.

A- LE METABOLISME ENERGETIQUE MITOCHONDRIAL CHEZ RHIZOPHAGUS IRREGULARIS DEDUIT DES ETUDES GENOMIQUES

L'équipe dans laquelle j'ai effectué mes travaux de thèse s'est vu attribuer différentes tâches d'annotation et mon premier objectif de thèse a été d'annoter les gènes impliqués dans la respiration cellulaire et, au-delà, dans le métabolisme énergétique mitochondrial. Cette étude m'a conduit à décrire la chaîne respiratoire de *R. irregularis* et obtenir de nouvelles informations physiologiques pour expliquer certains traits des champignons MA, comme leur germination en deux-étapes avec une phase non symbiotique, associée à une consommation moindre d'énergie, et une phase pré-symbiotique, associée à une activation de la respiration cellulaire et à une consommation importante des réserves de carbone.

Matériels et méthodes

Les séquences utilisées pour l'annotation ont été obtenues depuis la base de données *GlomusDB* hébergée par l'INRA de Nancy. Cette base de données contient 437 411 EST, assemblées en séquences consensus et transcrits virtuels non-redondants (NRVT). Ces données sont issues d'études transcriptomiques de différents laboratoires et du projet génomique réalisé au Joint Genome Institute (JGI). Parmi ces études, j'ai fourni des ARN issus d'échantillons de la souche DAOM197198 de *R. irregularis* qui ont été traités par des exsudats bruts de racines de *M. truncatula*, ou par du GR24 (analogue de synthèse des strigolactones) à une concentration de 10^{-8} M, dans du milieu M durant 24 h.

L'annotation de ces séquences a été effectuée de manière semi-automatique. Les séquences en lien avec notre tâche d'annotation ont été sélectionnées grâce à des alignements réalisés avec BLAST contre la base de données KOG « Energy production and conversion » regroupant les gènes jouant un rôle dans la production d'énergie. Les séquences alors sélectionnées ont ensuite été annotées manuellement grâce au logiciel Artemis 13.2.7 de l'institut Wellcome Trust Sanger (Rutherford *et al.*, 2000) et l'alignement BLAST (Altschul *et al.*, 1990), contre les séquences protéiques d'un set d'espèces fongiques de référence (*Neurospora crassa*, *Phycomyces blakesleeanus* et *Saccharomyces cerevisiae*), et NCBI BLAST (Altschul *et al.*, 1997; Johnson *et al.*, 2008), contre les bases de données nr (non-redondante) et CCD v2.05 (Conserved Domain Database).

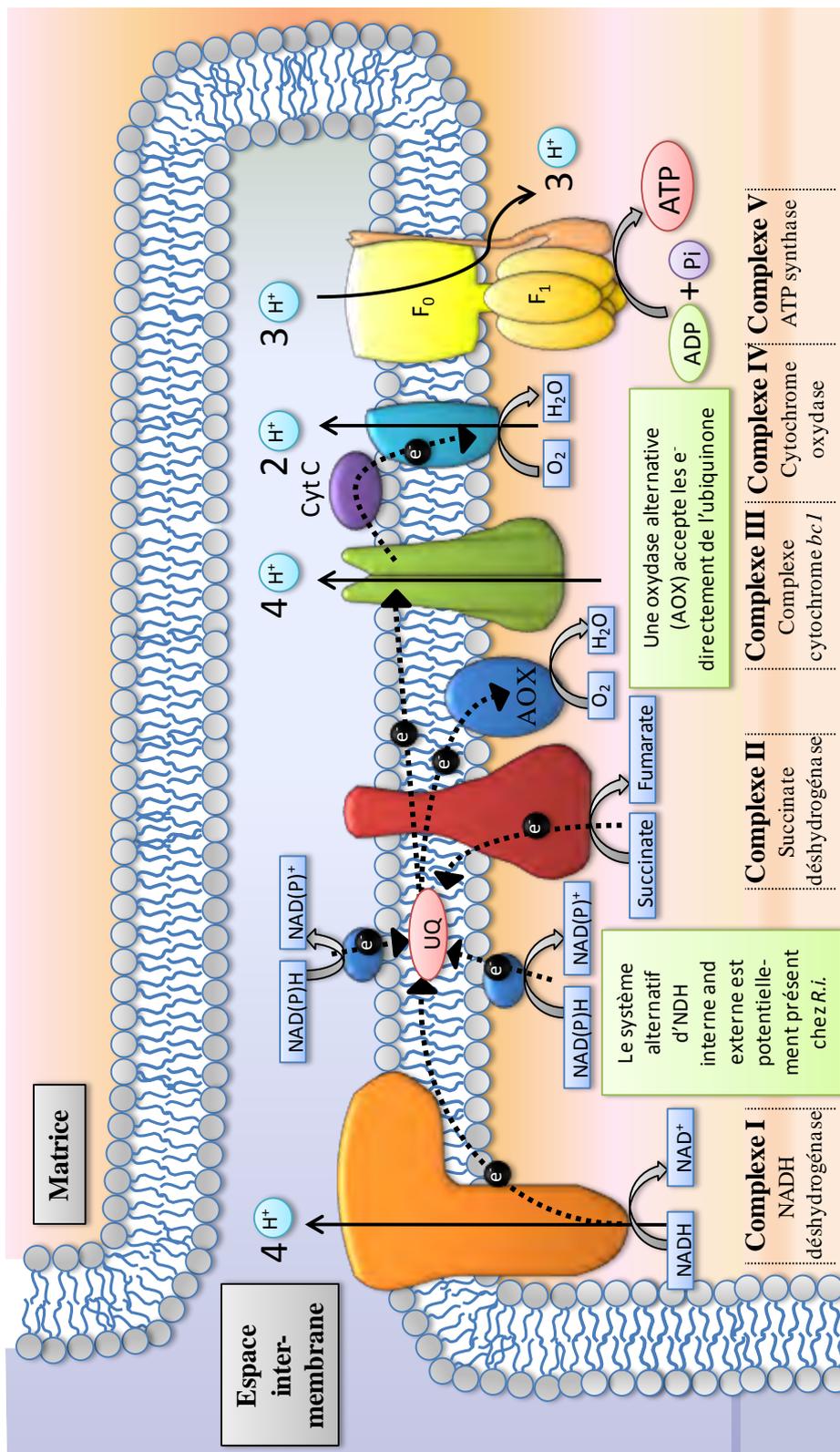


Figure 10 : Représentation schématique de la proposition de l'organisation schématique de la chaîne respiratoire de *Rhizophagus irregularis* déduite des annotations et de l'analyse génomique.

Résultats et discussion

Ces annotations m'ont permis de participer à la publication d'Emilie Tisserant et al. (2012) portant sur la description du transcriptome de Rhizophagus irregularis.

La chaîne respiratoire : voie principale et voie alternative

Par la méthode décrite dans la partie expérimentale précédente, j'ai pu annoter un total de 153 gènes impliqués dans la production d'énergie (Annexe 1). Parmi eux, 102 gènes sont représentés dans leur pleine longueur sur des NRVT et un seul d'entre eux est entier sur un contig génomique, reflétant l'assemblage très incomplet des séquences génomiques nucléaires. Dans les banques d'expression, j'ai pu identifier la totalité des gènes jouant un rôle dans le cycle de Krebs et la chaîne respiratoire classique des champignons et, notamment, des transcrits correspondant à une oxydase alternative et une NADH déshydrogénase alternative. Ces résultats me permettent de proposer un schéma de l'organisation de la chaîne respiratoire de *Rhizophagus irregularis* (Figure 10). Chez d'autres biotrophes obligatoires, comme le champignon *Blumeria graminis*, l'oomycète *Hyaloperonospora arabidopsidis*, ou l'agent de la rouille noire *Puccinia graminis*, la biotrophie obligatoire a été associée avec une perte de gènes impliqués dans les métabolismes primaires ou secondaires (Baxter *et al.*, 2010; Spanu *et al.*, 2010; Duplessis *et al.*, 2011). La chaîne respiratoire de *R. irregularis* étant complète, le statut de biotrophie obligatoire de ce champignon n'est donc pas dû à l'absence de certains gènes qui mènerait à la perte de l'aptitude à réaliser les métabolismes énergétiques durant la phase asymbiotique, et donc l'incapacité à produire de l'énergie. La faible respiration cellulaire observée lors de la germination du champignon (Besserer *et al.*, 2009) résulte donc soit d'une régulation spécifique de l'activité de ces protéines, au niveau transcriptionnel ou post-transcriptionnel, par exemple, ou, plus simplement, d'une limitation de l'activité mitochondriale par une régulation en amont du catabolisme des substrats du cycle de Krebs (Bécard *et al.*, 1991). La présence d'une oxydase alternative et d'une NADH déshydrogénase alternative confirme les résultats obtenus par Besserer et co-auteurs (2009) sur l'existence d'une voie alternative de production d'énergie chez *G. rosea* et montre également la présence de cette dernière chez *R. irregularis*. Chez *G. rosea*, la présence de ces voies a été mise en évidence grâce à des approches pharmacologiques. L'inhibition de ces voies par des traitements au SHAM (acide salicylhydroxamique A) et KCN (cyanure de potassium) était corrélée avec une diminution de 57% de l'intensité de la réponse de

ramification au GR24 et avait donc conduit à l'hypothèse d'une chaîne respiratoire ramifiée. Afin de vérifier la fonctionnalité des enzymes impliquées dans la respiration alternative chez *R. irregularis*, il serait intéressant de réaliser des études pharmacologiques similaires à celles accomplies chez *G. rosea* par Besserer *et al.* (2009) et, notamment, d'inhiber les voies alternatives, successivement ou simultanément, tout en bloquant, ou non, la voie principale de la chaîne respiratoire (Roténone pour le complexe I et Antimycine A pour le complexe III). L'inhibition spécifique de l'AOX peut être obtenue grâce au SHAM et l'inhibition des NADH déhydrogénases alternatives grâce à l'HDQ (1-hydroxy-2-dodecyl-4(1H)quinolone). Ces deux composés empêchent respectivement le transfert d'électron de la coenzyme ubiquinone à l'AOX, rendant ainsi impossible la réduction de l'O₂, et empêchant également le bon déroulement de l'oxydation du NADH par les NADH déhydrogénases alternatives.

La respiration anoxique chez *Rhizophagus irregularis*

Récemment, un gène codant pour une fumarate réductase a été retrouvé chez cet organisme montrant ainsi une potentielle respiration anaérobie permettant au champignon de survivre dans le sol, et germer spontanément, en absence d'oxygène et de plante hôte à proximité (Sędziewska *et al.*, 2012). Ce gène serait seulement exprimé lors de la phase asymbiotique mais pas lors des phases présymbiotique et symbiotique suggérant un passage d'une respiration anaérobie à une respiration aérobie lorsque le champignon perçoit les signaux de la plante. Il est à noter que ce comportement n'est pas validé dans nos expérimentations (N. Tang, communication personnelle). Dans la liste des gènes codant pour des protéines jouant un rôle dans la production d'énergie, annotés chez *R. irregularis*, j'ai pu confirmer la présence de cette fumarate réductase et nouvellement identifier une sulfide quinone oxydoréductase (SQO) qui sont toutes les deux des enzymes impliquées dans la respiration anaérobie (Theissen *et al.*, 2003). La présence d'une SQO montre la potentialité de ce champignon à utiliser l'hydrogène sulfuré comme donneur d'électron pour la production d'ATP par le biais de la chaîne respiratoire. Nous pouvons donc extrapoler que les champignons MA possèdent ces voies permettant de produire de l'énergie cellulaire, en mobilisant moins de réserves carbonées, mais qui sont moins productives en terme énergétique. La rhizosphère étant un lieu très pauvre en oxygène, le champignon doit pouvoir utiliser d'autres accepteurs d'électrons que le dioxygène pour la chaîne respiratoire. Lors des études menées par Besserer *et al.* (2009), l'inhibition des

voies de production d'énergie de la chaîne respiratoire par différents inhibiteurs n'a pas conduit à la disparition totale d'activité du champignon, lors de la stimulation par du GR24, laissant supposer que d'autres voies permettent une production d'énergie. La présence des gènes codants pour la fumarate réductase et la SQO pourrait en partie expliquer ce phénomène grâce à la capacité potentielle des enzymes produites à utiliser respectivement d'autres accepteurs ou donneurs d'électron et ainsi continuer le cycle de production d'énergie par d'autres voies. Il est intéressant de rapprocher la présence de cette voie alternative à la biologie particulière de la germination des champignons MA. En effet, les spores de champignons MA peuvent germer spontanément mais, en absence d'une racine hôte, la croissance s'arrête après quelques jours. La reprise de germination peut intervenir plusieurs fois, ce qui nécessite un mécanisme de préservation des réserves énergétiques. L'utilisation de la voie alternative de la respiration cellulaire durant la germination pourrait répondre à ce besoin d'économie des réserves carbonées de la spore ainsi qu'à un besoin de ralentir la sénescence en produisant moins d'ERO. Le passage de la voie alternative de la respiration cellulaire à la voie principale pourrait ainsi avoir un rôle dans l'adaptation à la biotrophie obligatoire. Il reste à découvrir les mécanismes impliqués dans ce changement de voie énergétique, ou métabolique, et la mise à disposition de nouveaux outils, tels que des puces transcriptomiques (Tisserant *et al.*, 2012), devrait permettre d'élucider les aspects transcriptionnels de ces changements de la physiologie asymbiotique vers la physiologie symbiotique. Des études de l'expression des enzymes de la voie respiratoire en présence, ou non, d'oxygène et en présence de différentes sources de carbone, permettront de mieux comprendre comment le champignon s'adapte à son environnement et si les mécanismes mis en jeu, et les profils d'expression, corréleront avec ceux permettant le passage de la phase asymbiotique à la phase symbiotique. La fonctionnalité des gènes identifiés grâce à ces approches génomiques et transcriptomiques va pouvoir être maintenant vérifiée dans un système hétérologue, comme la levure, qui présente un métabolisme d'anaérobie facultatif et possède donc les gènes permettant une respiration aérobie et anaérobie. La substitution d'un gène de *S. cerevisiae* par son homologue de *R. irregularis* permettra de vérifier si ce gène encode une protéine fonctionnelle qui assure le remplacement de la protéine absente, dans les conditions adéquates à l'étude de ce gène, anaérobie ou aérobie, selon l'implication du gène étudié dans les deux métabolismes.

***B- ORGANISATION ET COMPARAISON DES GENOMES MITOCHONDRIAUX
DE 5 SOUCHES DE RHIZOPHAGUS IRREGULARIS***

A l'initiation de ma thèse, aucun génome de champignon MA n'était publié ou disponible sur les bases de données publiques. Grâce à la participation de notre équipe au consortium de séquençage de *Rhizophagus irregularis*, j'ai été en charge de l'annotation du génome mitochondrial et j'ai donc eu accès au contig de 70 kb représentant ce dernier. En parallèle, le premier génome mitochondrial de *R. irregularis*, isolat FACE#494, a été publié par Lee et Young (2009). Deux génomes de différentes souches étant disponibles et annotés, j'ai pu les comparer et ainsi apprécier la répartition originale de la variabilité entre ces séquences où l'identité frôle les 100% excepté pour 3 zones de variabilité bien distinctes. Dans le cadre du projet Neofertil en collaboration avec l'entreprise Agronutrition, j'avais en charge de développer des marqueurs moléculaires permettant de discriminer et typer *R. irregularis* à l'échelle de la souche. La forte conservation de séquence alternée avec des zones de variabilité représentait alors une conformation propice au développement de tels marqueurs. La possible discrimination par le génome mitochondrial nous a donc encouragé à approfondir cette stratégie en séquençant et comparant les génomes mitochondriaux de 5 autres souches de *R. irregularis*. Grâce à ce comparatif, j'ai pu identifier les éléments générateurs de variabilité entre ces souches, les utiliser pour développer des marqueurs moléculaires, et discriminer les souches disponibles grâce à ces derniers.

Cette étude a fait l'objet d'une publication dans la revue *New Phytologist*.

Résumé de l'article

- Les champignons mycorhiziens à arbuscules (MA) sont impliqués dans une des interactions plante-champignon les plus répandues. Un certain nombre de travaux sur la dynamique des populations de champignons MA ont utilisé des séquences d'ADN mitochondrial (mt), mais, jusqu'à présent, les génomes mt des champignons MA sont peu connus. À ce jour, quatre génomes mt de trois espèces de champignons MA sont disponibles, parmi lesquels deux de *Rhizophagus irregularis*.

- Afin d'étudier la variabilité intra- et inter-souche du génome mt de *R. irregularis*, nous avons séquencé et assemblé *de novo* quatre génomes mt supplémentaires de cette espèce. Nous avons utilisé les technologies de pyroséquençage 454 et Illumina afin de directement séquencer les génomes mt à partir de l'ADN génomique total.

- Les génomes mt sont uniques dans chaque souche. Des divergences inter-souches concernant la taille du génome, dues à des séquences introniques et intergéniques, ont été observées. Le polymorphisme est provoqué par trois types d'éléments générateurs de variabilité (VGE): des homing endonucléases, des ORF contenant des domaines ADN polymérase, et des petites répétitions inversées. Basé sur le positionnement des VGE, des séquences mitochondriales et des marqueurs nucléaires, deux sous-clades de *R. irregularis* ont été caractérisés.

- La découverte des VGE met en évidence la grande plasticité intraspécifique du génome mt de *R. irregularis*. Les VGE permettent la conception de puissants marqueurs mitochondriaux pour le typage et le suivi des souches de *R. irregularis* lors d'études génétiques et populationnelles.

Comparative analysis of mitochondrial genomes of *Rhizophagus irregularis* - syn. *Glomus irregulare* - reveals a polymorphism induced by Variability Generating Elements.

Damien Formey^{(1,2,3)a}, Marion Molès^{(1,2)a}, Alexandra Haouy^(1,2), Bruno Savelli^(1,2), Olivier Bouchez⁽⁴⁾, Guillaume Bécard^(1,2) and Christophe Roux^(1,2)

¹ Université de Toulouse, UPS, UMR5546, Laboratoire de recherche en Sciences Végétales, BP42617, F-31326 Castanet-Tolosan cedex, France.

² CNRS, UMR5546, BP 42617, F-31326 Castanet-Tolosan cedex, France

³ Agro-Nutrition, Parc Activestre, 3 avenue de l'orchidée F-31390 Carbonne, France

⁴ Plateforme Génomique, Campus INRA Chemin de Borde-Rouge, F-31326 Castanet-Tolosan cedex, France

a: the first two authors contributed equally to this work

Corresponding author: Pr Christophe Roux roux@lrsv.ups-tlse.fr

Number of Tables: 4

Number of Figures: 3

Number of supporting information documents: 9

Word count: 5502

Abstract

- Arbuscular mycorrhizal (AM) fungi are involved in one of the most widespread plant-fungus interactions. A number of works on population dynamics of AM fungi have used mitochondrial (mt) DNA sequences, and yet mitochondrial AM fungus genomes are poorly known. To date, four mt genomes of three species of AM fungi are available, among which two from *Rhizophagus irregularis*.

- In order to study intra- and inter-strain mt genome variability of *R. irregularis*, we sequenced and de novo assembled four additional mt genomes of this species. We used 454 pyrosequencing and Illumina technologies to directly sequence mt genomes from total genomic DNA.

- Mt genomes are unique within each strain. Inter-strain divergences in genome size due to highly polymorphic intergenic and intronic sequences were observed. The polymorphism is brought about by three types of Variability Generating Elements (VGEs): homing endonucleases, DNA polymerase domain containing ORFs and small inverted repeats. Based on VGEs positioning, mitochondrial sequences and nuclear markers, two sub-clades of *R. irregularis* were characterized.

- The discovery of VGEs highlights the great intraspecific plasticity of *R. irregularis* mt genome. VGEs allow the design of powerful mitochondrial markers for typing and monitoring *R. irregularis* strains in genetic and population studies.

Introduction

Arbuscular Mycorrhizal (AM) symbiosis is considered as the most widespread plant-fungus interaction as it concerns around 80% of terrestrial plant species in all ecosystems (Smith & Read, 2008). AM fungi belong to a specific phylogenetic group basal to Dikarya, the Glomeromycota (Schwarzott *et al.*, 2001; Krüger *et al.*, 2011) and are of great agronomic interest due to their role in plant water and mineral assimilation (Smith & Read, 2008). A considerable amount of work dealing with the ecological incidence of AM fungi in natural and field conditions has been undertaken. A correlation was established between the diversity of AM fungal species in the soil and the plant biodiversity and productivity aboveground (Van der Heijden *et al.*, 1998). As a feedback effect, it was also reported that plant species influence the relative abundance of AM fungi by modifying the structure of the soil microorganism community (Van der Heijden *et al.*, 1998; Klironomos, 2002). Overall, these studies highlight the importance of AM fungi in structuring plant communities and increasing ecosystem productivity. The most widely studied species among Glomeromycota is *Rhizophagus irregularis* (Błaszk., Wubet, Renker & Buscot) C. Walker & A. Schüßler – Syn. *Glomus irregulare* Błaszk., Wubet, Renker & Buscot (Schüßler & Walker 2010; Blaszkowski *et al.*, 2008; Stockinger *et al.*, 2009) an ubiquitous species presents in natural and cultivated soils (Börstler *et al.*, 2008). Population and diversity analyses of *R. irregularis* need the definition of robust and reliable molecular markers. Mitochondrial (mt) DNA sequences have been mainly used, particularly the large-subunit mitochondrial ribosomal RNA genes (mtLSU) (Raab *et al.*, 2005; Börstler *et al.*, 2008, 2010; Croll *et al.*, 2008; Thiéry *et al.*, 2010). The relevance of using mitochondrial ribosomal regions as molecular markers was strengthened by the fact that a unique haplotype of mtLSU was identified within each fungal isolate (Raab *et al.*, 2005), whereas nuclear markers, like the Internal Transcribed

Spacer regions (ITS), were found to be highly polymorphic within fungal isolates (Stockinger *et al.*, 2009). Börstler and co-authors (2008) successfully designed a Restriction Fragment Length Polymorphism (RFLP) strategy on mtLSU sequences. They were able to distinguish 12 haplotypes among 16 isolates of *R. irregularis*. Surprisingly, haplotype diversity in two cultivated field sites was found to be higher than in two semi-natural sites used as references (Börstler *et al.*, 2010). These studies highlighted the utility of using mtLSU markers to investigate the diversity of *R. irregularis* strains in natural conditions. They also revealed that the sole use of mtLSU markers probably underestimates the actual strain diversity (Croll *et al.*, 2008). To sum up, these studies underlined that i) there is only one mtLSU haplotype in each *R. irregularis* strain, ii) mtLSU allows the differentiation of most strains in different haplotypes, but iii) to reach a finer scale of typing and identify individuals, it would be necessary to define more polymorphic mt regions.

Only a comprehensive analysis of several complete mitochondrial (mt) genomes of strains of the same species will reveal the occurrence of highly polymorphic regions and strain-specific markers. To date, the mt genomes from two different strains of *R. irregularis* have been sequenced - FACE#494, (Lee & Young, 2009) and DAOM197198 (HQ189519.1). This last isolate is the model AM fungal strain adopted by the scientific community because of its relatively easy propagation and *in vitro* production in large quantities (Chabot *et al.*, 1992). Recently, mt genomes of two other species of AM fungi - *Gigaspora margarita* (Pelin *et al.*, 2012) and *Gigaspora rosea* (Nadimi *et al.*, 2012) - were released. The mt genomes of these three species of AM fungi were found to possess standard fungal mt genomes including the expected classical set of genes, whereas a high interspecific variability of intergenic sequences was observed.

Table 1: Strains of *Rhizophagus irregularis* used in this study and number of reads and contigs obtained for each strain after 454 or MiSeq sequencing.

<i>Name</i>	<i>Origin of isolates</i>	<i>Reads</i>	<i>Contigs</i>
		<i>454 (MiSeq)</i>	<i>454 (MiSeq)</i>
MUCL_43204	Clarence creek - Ontario, Canada	261 046	5928
MUCL_46239	Iles-de-la-Madeleine - Québec, Canada	381 424	7311
MUCL_46240	Buckingham - Québec, Canada	87 901 (5.5 10 ⁶)	797 (1.4 10 ⁵)
MUCL_46241	Ripon - Québec, Canada	249 597	3632
DAOM197198	Pont Rouge - Québec, Canada	247 631	4360

Besides these studies on interspecific variability of mt genomes, our objective was to evaluate intraspecific polymorphism of the *R. irregularis* mt genome. Intraspecific polymorphic regions would be helpful to design markers for fundamental genetic studies *in vitro* or in the field, or for quality control of AM fungal inoculum. We therefore sequenced and *de novo* assembled new mt genomes of four different strains of *R. irregularis*. In order to evaluate the efficiency of direct sequencing from total gDNA using NGS approaches, Roche 454 GS FLX pyrosequencing and MiSeq System (Illumina) were used. We then compared these mt genomes to analyze intra- and inter-strain variability. Finally we identified useful mt genomic regions to be used as selective markers for strain typing.

Materials and methods

DNA preparation from fungal strains

Four strains of *Rhizophagus irregularis* obtained from GINCO_BEL (www.mycorrhiza.be/ginco-bel/) and strain DAOM197198 were selected for this study (Table 1). All strains were cultured on M medium (solidified with 0.3 % gellan gum; Phytigel, Sigma) with Ri T-DNA transformed carrot roots (Bécard & Fortin, 1988) in a two-compartment Petri dish system (St-Arnaud *et al.*, 1996). Mycelium and spores collected from fungal compartment after two months of cultivation (20,000 to 50,000 spores) were ground up in liquid nitrogen. The resulting powder was used for DNA extraction based on the CTAB method (Saghai-Marooof *et al.*, 1984).

Sequencing and assembling procedures

Sequencing procedures were performed at the GeT-PlaGe platform, a core facility in Toulouse (France) for genomics (<http://get.genotoul.fr/>). Approximately 0.5 µg of genomic DNA was used for construction of Roche 454 GS FLX sequencing libraries

following the manufacturer's protocol. Genomic DNA from strain DAOM197198 was first sequenced in two independent runs, each on one-eighth of the Pico Titer Plate, to test the sequencing efficiency of mt genome using Roche 454 pyrosequencing. Total genomic DNA of the four additional *R. irregularis* strains (MUCL_43204, MUCL_46239, MUCL_46240, MUCL_46241) was then sequenced. A specific index (MID, multiplex identifier) was chosen for each strain library. The libraries were then equimolarly multiplexed and sequenced on two half-plates. To complete the mt genome of MUCL_43204, we performed additional sequencing on one region of one-eighth of the Pico Titer Plate. Generated 454 reads were cleaned and filtered with Pyrocleaner 1.2 software (Mariette *et al.*, 2011), resulting in a sequence suppression of 12.8 % on average per library. Cleaned reads were assembled thanks to the GS de novo Assembler 2.5.3, from Roche, leading to a mean of 4406 contigs per library (see Table 1), each with a length of at least 500 bp. In order to complete the sequence of the mitochondrial genome of strain MUCL_46240 obtained by Roche 454 pyrosequencing and make comparison with another sequencing technique, extracts of total gDNA (0.5 µg) were sequenced using MiSeq System (Illumina). Generated reads were *de novo* assembled using CLC Genomics Workbench 5.0.1. software (CLC bio, Aarhus, Denmark) .

Sequences annotation and comparison

Mt genomes were annotated using Artemis 13.2.7 software from Wellcome Trust Sanger Institute (Rutherford *et al.*, 2000) and NCBI BLAST (Altschul *et al.*, 1997; Carver *et al.*, 2005; Johnson *et al.*, 2008) against the nr (non-redundant) and CDD v2.05 (Conserved Domain Database) databases. The transfer RNAs were annotated using tRNAscan-SE Search Server 1.21 (Lowe & Eddy, 1997; Schattner *et al.*, 2005). Small Inverted Repeats (SIR) were manually annotated thanks to NCBI BLAST and their secondary structure was predicted using RNAfold WebServer (Gruber *et al.*, 2008). The complete mt genome of

MUCL_43204, MUCL_46239 and MUCL_46240/MUCL_46241 are deposited in Genbank with the accession numbers JQ514224, JQ514223 and JQ514225, respectively.

Mt genomes were compared using Artemis Comparison Tool (ACT) v8 (Carver *et al.*, 2005) and MAUVE software 2.3.1 (Darling *et al.*, 2004) with the progressive Mauve algorithm (Darling *et al.*, 2010). The height of the similarity profile is inversely proportional to the average alignment column entropy over a region of the alignment. The percentage of nucleotide identity of the different mt genomes and Conserved Mitochondrial Genes (CMG) (ATP6, ATP8, ATP9, CYTB, COX1, COX2, COX3, NAD1, NAD2, NAD3, NAD4, NAD4L, NAD5, NAD6), compared to the FACE#494 genome, was calculated with the complete sequence of the mt genome and a generated consensus assembly of the CMG sequences using EMBOSS Stretcher online (www.ebi.ac.uk/Tools/psa/emboss_stretcher/).

Single-nucleotide polymorphism (SNP) analysis

The analyses of intra-strain polymorphism were performed on mitochondrial and nuclear contigs. For the mt genome, analyses were carried out on the consensus contig assembled by GS assembler and the analyses for nuclear genome were performed on the ten longest contigs that did not match with the mt genome. For each strain, the genome consensus contig assembly generated by GS was used as a reference and new alignment was performed using Burrows-Wheeler Align (BWA) (Li & Durbin, 2010) to generate assembly files in SAM format (<http://samtools.sourceforge.net/SAM1.pdf>). Sequence alignments were then treated with Samtools (Li *et al.*, 2009) to get a list of potential SNPs. A Perl script was developed in order to keep the SNPs that meet the following criteria: *i*) they belong to sequences without large soft-clipped portions upstream or downstream of an area of identity and with a matching quality greater than 100; *ii*) they are found at least

twice at the given read depth and are not included in, or do not generate, a nucleotide homopolymer of 3 bases or more. Data obtained from the Illumina MiSeq System were analysed using the SNP detection function of the software CLC Genomics Workbench 5.0.1.

Microsatellite analysis

Genomic DNA was extracted as described above. We used a set of 16 microsatellite markers among the 18 described by Mathimaran *et al.* (2008): Glint02 to Glint16 and Glint18. PCR products were indirectly revealed using a fluorescein amidite (FAM)-labelled M13 primer (5'→3' GTTTTCCCAGTCACGACGTTG), all forward primers possessing a M13 extension. DNA templates containing either Pfu polymerase or Taq Polymerase were amplified using the following thermal cycling conditions: 2 min at 94 °C followed by 35 cycles of 30 s at 94 °C, 30 s at 55 °C, 45 s (Pfu polymerase) or 30 s (Taq polymerase) at 72 °C and a final extension for 5 min at 72 °C. FAM-labelled M13 primer allowed visualization of the fragment length on the 48-capillary 3730 DNA Analyzer (Applied Biosystems, Inc., USA). Microsatellite data were manually analysed using Genemapper 4.0 (Applied Biosystems, Inc., USA).

Phylogenetic analyses

The alignments of mitochondrial ribosomal Large SubUnit nucleotide sequences or Conserved Mitochondrial Genes consensus amino acid sequences were performed with MAFFT online (<http://mafft.cbrc.jp/alignment/server/>) (Kato *et al.*, 2002). Alignments were automatically trimmed with the Gappyout module of the software Trimal (Capella-Gutiérrez *et al.*, 2009). The evolutionary history was inferred using both the Maximum Likelihood method based on the Tamura-Nei model (Tamura *et al.*, 2004), conducted in MEGA5 (Tamura *et al.*, 2011) and LG model of evolution performed by the software

PhyML 3.0 (Guindon *et al.*, 2010). The percentage of trees in which the associated taxa clustered together (bootstraps) was calculated based on 500 replicates. For the phylogeny conducted in MEGA5, when the number of common sites was <100 or less than one fourth of the total number of sites, the maximum parsimony method was used; otherwise the BIONJ method with MCL distance matrix was used. All positions containing gaps and missing data were eliminated. For distance analyses using data from microsatellite markers and variability generating elements, matrixes of characters were constructed by coding each marker considered as an individual character and then inferred by NJ (observed divergence as criteria of distance analysis) using PAUP software (Swofford, 1993).

Primer design and PCR

One set of primers was designed on strain-specific zones established by mt genome comparison (Table S1). PCR amplifications were carried out using Taq polymerase and the reaction buffer supplied (Promega, USA) with the following conditions: 94°C 2 min then 40 cycles at 94 °C for 30 s, 60°C for 1 min, 72°C for 1 min, followed by a final extension period at 72 °C for 10 min. Another set of primers was designed on flanking regions of short variable sequences for analyzing size polymorphism by PCR amplification (Table S1). The PCR protocol was identical except for elongation time (72 °C for 30 s).

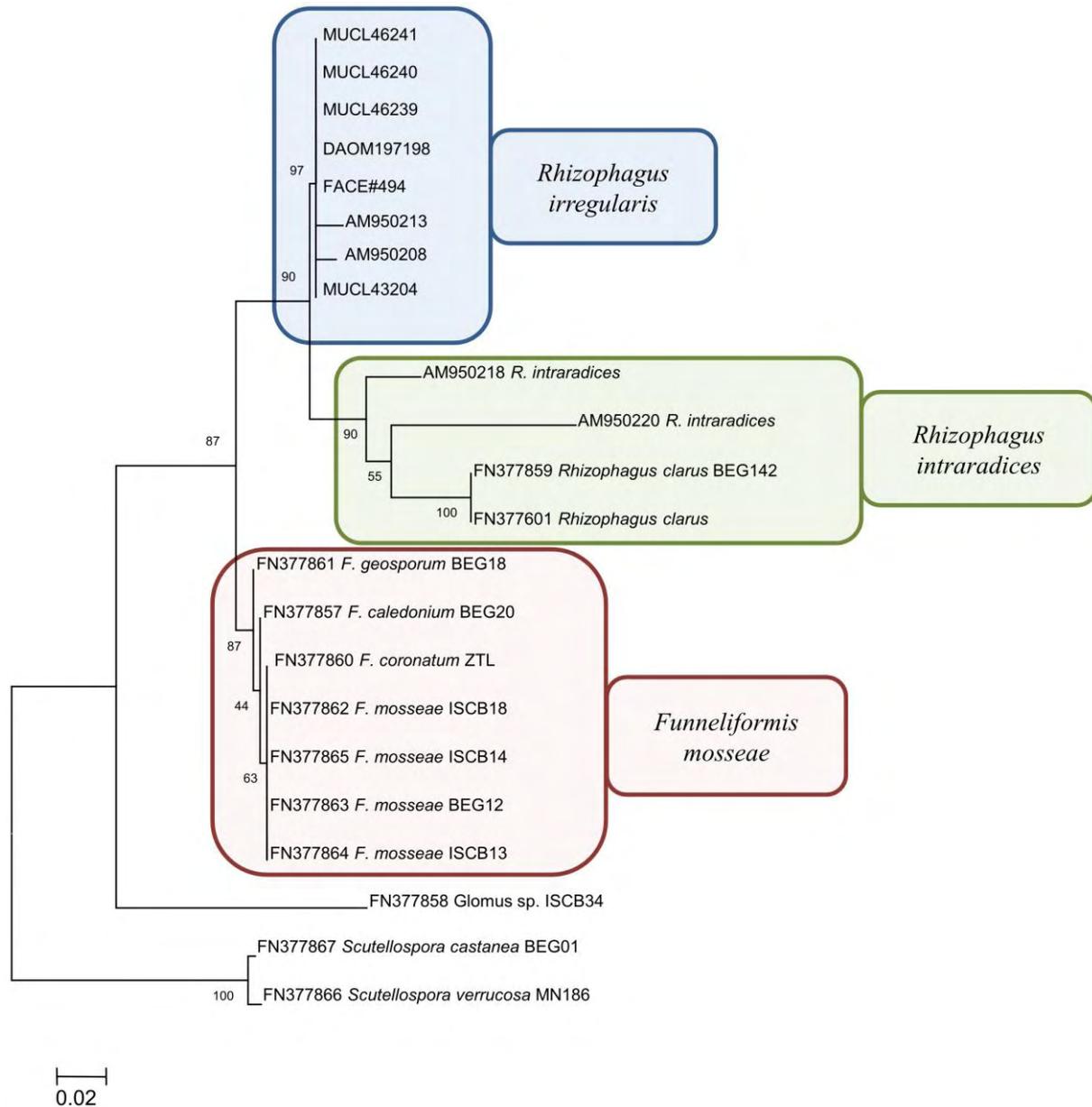


Figure 1: Phylogeny of 22 isolates of arbuscular mycorrhizal fungi based on mitochondrial ribosomal Large SubUnits (mtLSU) sequence comparison. Alignment was based on data from Thiéry *et al.* (2010) leading to 369 conserved sites of mtLSU. Phylogenetic inferences were calculated by using the Maximum Likelihood method based on the Tamura-Nei model.

The major clades are surrounded by coloured rectangles. The tree with the highest log likelihood (-1363.6593) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Branch lengths are representative of the number of substitutions per site. All positions containing gaps and missing data were eliminated.

Results

Direct sequencing of mitochondrial genomes

To evaluate the sequencing efficiency of mt genome from total gDNA, preliminary sequencing runs of 454 GS FLX Titanium were performed on strain DAOM197198, whose mt genome is already available. From two steps of one-eighth plate sequencing, 247 631 reads were obtained (Table 1). The resulting *de novo* assembled mt genome was totally identical to HQ189519.1 with a 5.97x coverage of the mt genome. From this evaluation, the genomic DNA of the four other strains was tagged in four distinct libraries and sequenced in one run of 454 GS FLX Titanium. The 343 Mb sequences obtained resulted in 979 968 reads, corresponding to an average of 244 992 reads per strain, with an average length of 436.5 bp. As mentioned in Table 1, the efficiency of sequencing among the four strains was not equivalent. Long contigs, corresponding to circular mt genomes, were obtained for the strains MUCL_46239 and MUCL_46241. Complementary 454 sequencing provided missing information for MUCL_43204. For the strain MUCL_46240, mt genome was not totally sequenced, covered at 84 % when compared to the closest mt genomes (MUCL_46241). To complete this mt genome sequence and test the sequencing efficiency of another NGS technique, the genome of MUCL_46240 was sequenced independently by the MiSeq System (Illumina), generating 1.2 Gbp. At the end, four complete genomes of strains MUCL_43204, MUCL_46239, MUCL_46240 and MUCL_46241, in addition to that of DAOM197198, were obtained with an average coverage of 11.7x, 14.8x, 46.3x and 10.3x, respectively.

SNP analysis

For all strains, the sequencing depth provided information on the intra-strain single-nucleotide polymorphism (SNP) of mt genomes. First analyses were performed on sequences obtained by Roche 454 pyrosequencing, *i.e.* for strains MUCL_43204,

Table 2: Intra-strain Single Nucleotide Polymorphism (SNP) rate estimated in nuclear and mitochondrial sequences, obtained by Roche 454 pyrosequencing, of four strains of *Rhizophagus irregularis*.

<i>Strain</i>	<i>Nuclear SNP rate</i> (SNP/10 ⁻⁵ nt)*	<i>Mitochondrial SNP rate</i> (SNP/10 ⁻⁵ nt)
DAOM197198	124.51	0.00
MUCL_46239	162.91	0.00
MUCL_46241	142.67	0.00
MUCL_43204	128.7	0.00
mean	139.7	0.00

*The nuclear contigs used correspond to the 10 longest contigs of each strain that do not match with mitochondrial sequences.

MUCL_46239, MUCL_46241 and DAOM197198. Nuclear sequences presented an average SNP rate of $139.7 \cdot 10^{-5}$ SNP/nt, whereas mitochondrial sequences did not contain any SNP (Table 2). In addition, we identified a high rate of short indels in reads when mapping them on mt and nuclear genomes (not shown). These indels could be 454 sequencing errors due to the known biases caused by homopolymeric regions and low GC content (Balzer *et al.*, 2011). In agreement with this hypothesis, no indels were found in the mt genome of the strain MUCL_46240 obtained by MiSeq System.

Phylogenetic analyses

We confirmed that the strains used belong to the clade of *Rhizophagus irregularis* species by using mitochondrial ribosomal large subunit sequences (Thiéry *et al.*, 2010 - Figure 1) and a consensus of amino acid sequences of Conserved Mitochondrial Genes (CMG) (Fig. S1). In this last tree, all CMG sequences from AM fungi (*Rhizophagus irregularis*, *Gigaspora rosea* and *Gigaspora margarita*) were grouped in a clade distinct from the other fungal taxa. Interestingly, when comparing mtLSU or CMG nucleotide sequences among the *R. irregularis* strains, the sequences split into two sub-clades (Fig. S2a, b). Similar distribution of strains in sub-clades was observed using the nuclear ITS (Fig. S3). To further confirm that nuclear sequences support this dichotomy, nuclear microsatellite markers defined for typing *R. irregularis* strains (Croll *et al.*, 2008) were used. The tree deduced from the microsatellite allele size pattern (Table S2) revealed similar clustering to that obtained from mt genome comparison: one group formed by MUCL_43204, the other formed by MUCL_46239, MUCL_46240, MUCL_46241 and DAOM197198. Intriguingly, mtLSU sequences, CMG sequences (Fig. S2) and microsatellite patterns (Table S2) of MUCL_46240 were found to be identical to those of MUCL_46241.

Table 3: Comparaison of mitochondrial genomes of five strains of *Rhizophagus irregularis* obtained in this study to that of FACE#494 (Lee and Young, 2009). CMG: Conserved Mitochondrial Genes; ORF: Open Reading Frames; HEO: Homing Endonuclease ORF; DPDCO: DNA-Polymerase Domain Containing ORF; SIR: Small Inverted Repeat.

Strain	%GC	Genome size (bp)	CMGs	Structural genes	Predicted ORFs ^a	%ID genome vs 494	%ID of CMGs vs 494	HEOs		DPDCOs	SIRs
								GIY-YIG domain	LAGLIDADG domain		
FACE#494	37.2	70606	14	28	19	-	-	2	5	5	26
DAOM197198	37.2	70783	14	28	19	92.8	99.9	2	5	6	27
MUCL_46239	37.2	70818	14	28	21	95.6	99.8	2	5	6	28
MUCL_46241 /MUCL_46240	37.3	74797	14	28	23	90.0	99.5	2	5	6	32
MUCL_43204	37.0	87754	14	27	34	64.5	65.8	2	16	5	50

^a including ORF encoding proteins with unknown function and pseudogenes.

Genome composition and comparison

Mt genomes of strains MUCL_46240 and MUCL_46241 were found totally identical confirming the above analyses (Figure 2 and Fig. S4). The four resulting mt genome sequences were compared to that of strain FACE#494 (Table 3). All genomes are circular with an average GC content of approximately 37 %. Genome size corresponds to an average size compared to other fungal mt genomes (usually less than 100 kb). Intriguingly, the genome sizes varied widely from 70 783 bp for DAOM197198, to 87 754 bp for MUCL_43204. All strains have the same classic set of proteins as already annotated in the FACE#494 strain (Fig. S4). Transfer RNAs are present in a complete set and are the same in all strains except for MUCL_43204, lacking the Pseudo-Asp-tRNA. The strains also exhibit a different number of predicted Open Reading Frames (ORF): FACE#494 having the lowest ORF content (19), MUCL_43204 the highest (34) (Table 3). This variation in number of predicted ORF could partly explain the variation in genome size. Calculation of the percentage of identity of each genome with the genome of FACE#494, using either the whole genome or CMGs only revealed (Table 3) that *i*) most divergences were found in the intergenic regions, and *ii*) MUCL_43204 is distant from the others with 65.8 % of nucleotide sequence identity (CMGs) instead of over 99% for the other strains. The synteny and identity analyses of these genomes showed conserved coding sequences, in organization and identity, but differences in sequence or position of introns and intergenic region (Figure 2). For instance, the light blue block on the left of the figure moved position (after and before the green block) and orientation (above and below the central line of each panel). This region corresponds to part of a DNA polymerase domain, like the purple boxes on the right of the figure. These purple blocks are of different sizes depending on the strain and participate in genome size variations. In MUCL_43204, this phenomenon is easily distinguishable: multiple

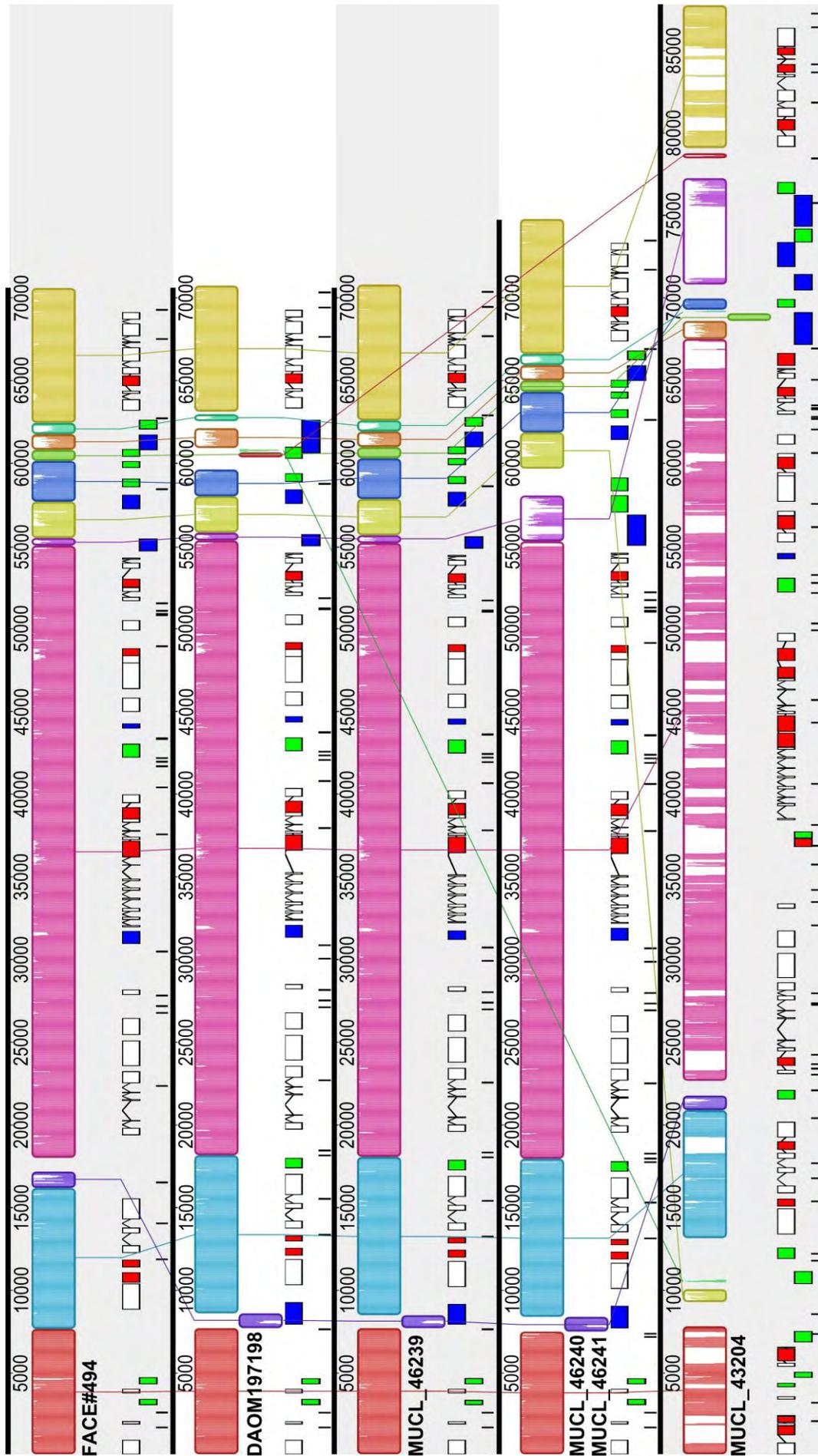


Figure 2: Multiple mitochondrial genome alignment of five *Rhizophagus irregularis* strains performed with MAUVE 2.3.1. Each genome's panel contains the name of the strain, a scale showing the sequence coordinates for that genome, and a single black horizontal center arrow representing the limit between the strands. Coloured blocks correspond to presumably homologous genome sequences presumably homologous. When a block lies above or below the center line, the aligned region is in the forward or reverse orientation, respectively, relative to the first strain. Regions without block lack detectable homology among the compared genomes. Inside each block, the height of the similarity profile corresponds to the average level of conservation in that region. Areas that are completely white were not aligned. Vertical lines are linking homologous blocks between the genomes. The boxes below the blocks represent the features of the GENBANK annotation. The upper row of represents the features on the forward strand, the middle boxes represent the features on the reverse strand, and lower boxes represent the Small Inverted Repeats. White box: Conserved Mitochondrial Genes. Blue box: DNA Polymerase Domain Containing ORF. Red box: Homing Endonuclease ORF. Green box: Predicted ORF.

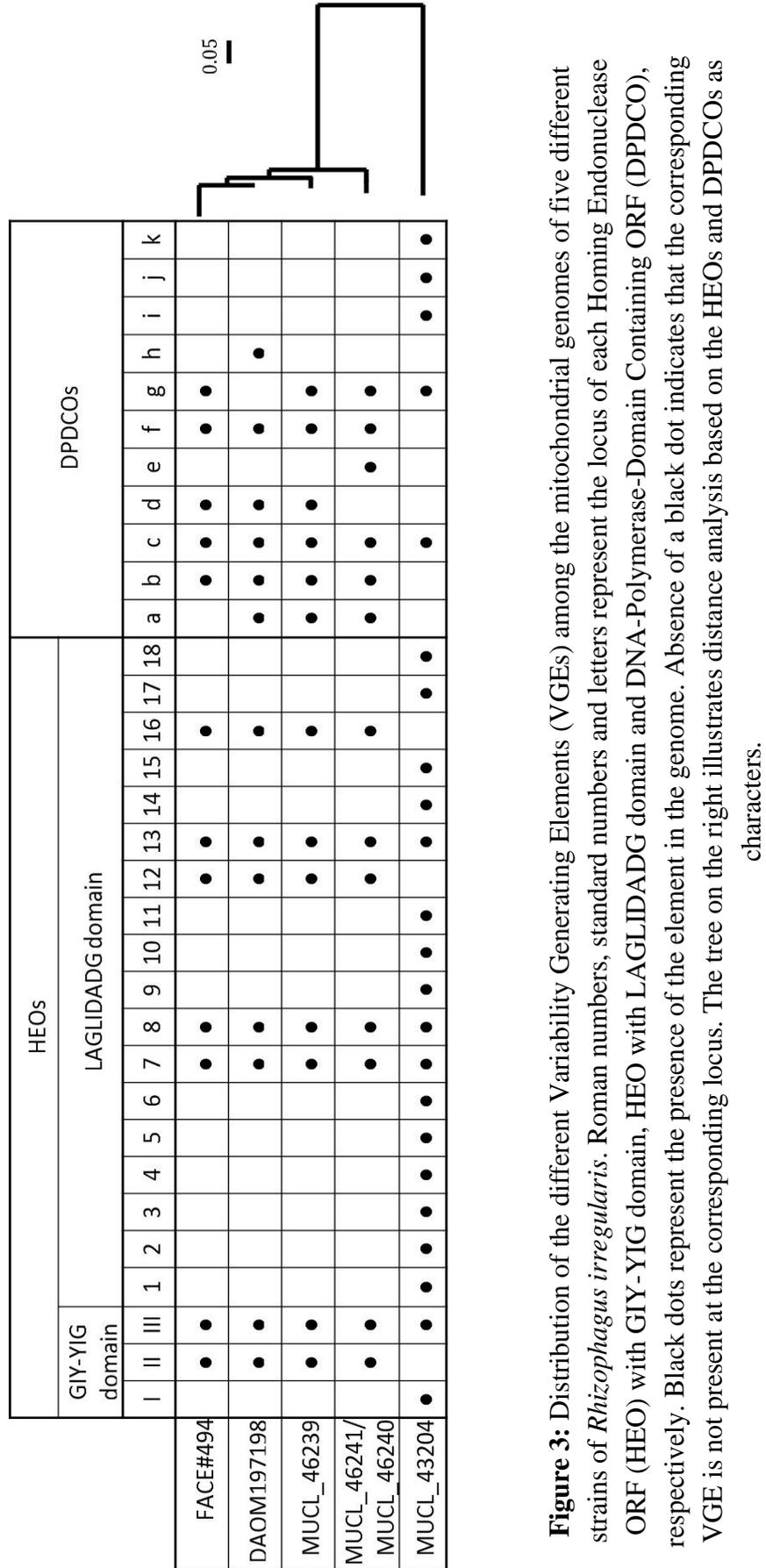


Figure 3: Distribution of the different Variability Generating Elements (VGEs) among the mitochondrial genomes of five different strains of *Rhizophagus irregularis*. Roman numbers, standard numbers and letters represent the locus of each Homing Endonuclease ORF (HEO) with GIY-YIG domain, HEO with LAGLIDADG domain and DNA-Polymerase-Domain Containing ORF (DPDCO), respectively. Black dots represent the presence of the element in the genome. Absence of a black dot indicates that the corresponding VGE is not present at the corresponding locus. The tree on the right illustrates distance analysis based on the HEOs and DPDCOs as

characters.

insertion events have led to the creation of introns that invaded the genome and increased its size. These examples indicate that the variations observed in organization and genome size are caused by insertion/deletion of genetic elements. Three types of Variability Generating Element (VGE) were identified: Homing Endonuclease ORFs (HEO), DNA-Polymerase-domain Containing ORF (DPDCO) (Figure 3) and Small Inverted Repeats (SIR) (Fig. S5). HEOs and SIRs are found in both CMG and intergenic region whereas DPDCOs are found only in the CMG intergenic region.

HEOs are enzymes encoded by ORFs embedded within introns. In the genomes studied we found two HEOs with GIY-YIG domains, two with a double type 1 LAGLIDADG domain and three with a single type 2 LAGLIDADG domain. The MUCL_43204 strain has a LAGLIDADG HEO organization different from that of the other strains. It is composed of two HEOs with a putative LAGLIDADG domain, three with single type 1 LAGLIDAG domain, four with a double type 1 LAGLIDADG domain and three with a single type-2 domain. HEO insertions are responsible for 42 % of the size increase in MUCL_43204 (not shown).

DPDCOs are long ORFs that contain complete or degenerate DNA polymerase domains and are differentially distributed in the genomes studied. We found five DPDCOs in strains FACE#494 and MUCL_43204, and six in strains MUCL_46239, MUCL_46241 and DAOM197198. One DPDCO alone explains most of the genome size difference observed between FACE#494 and MUCL_46241 (purple block in Figure 2).

SIRs are small elements (33 - 99 bp) that have a predicted structure that confers them the ability to be folded in a single hairpin (Boer & Gray, 1991). Five different types of SIRs were identified based on their sequence and folding structure (Fig. S5). They were

Table 4: Amplicon size (bp) obtained with the polymorphism length primers for different strains of *Rhizophagus irregularis*.

<i>Primers</i>	<i>*FACE</i> <i>#494</i>	<i>DAOM</i>	<i>MUCL</i>	<i>MUCL</i>	<i>MUCL</i>	<i>MUCL</i>
		197198	46239	43204	46240	46241
GenRL3	283	283	283	297	283	283
GenRL5a	152	152	152	290	256	256
GenRL5b	244	244	244	-	261	261
GenRL0a	-	415	303	-	415	415
GenRL0b	314	314	314	259	275	275

* For FACE#494, sizes are *in silico* prediction from mt genome sequence (Lee and Young, 2009).

differentially distributed in the five genomes: 26 SIRs in FACE#494, 27 in DAOM197198, 28 in MUCL_46329, 32 in MUCL_46241 and 50 in MUCL_43204.

Overall we see that the SIR number and the number and position of the HEOs/DPDCOs are the main differences between MUCL_43204 and the other strains, supporting the sub-clade formed by MUCL_43204 (Figure 3).

PCR-based strain identification strategies

As VGEs clearly differentiate the mt genomes of the different strains, these elements were targeted to define strain-specific PCR primers. A first strategy was to design primers targeting a specific region (DPDCO or HEO) for each strain, leading to a diagnostic method based on presence/absence of PCR amplification. Specific primers were successfully designed for specific DNA amplification, from one spore, of strains DAOM197198, MUCL_43204 and MUCL_46240/MUCL_46241 (Fig. S6). The strain MUCL_46239 did not present a sequence specific and long enough to design one couple of specific primers.

A second strategy was developed based on length polymorphism of amplicons. Two pairs of primers targeting the region flanking VGEs (HEO or SIR) were chosen that discriminated the five strains (Table 4 and Fig. S7). As expected, the patterns were identical for the pair of strains MUCL_46240/MUCL_46241.

Discussion

In order to analyze mt genome variability in *R. irregularis* at the strain level and provide markers for typing these strains, we sequenced the mt genomes of four new strains of *R. irregularis* in addition to DAOM197198. To our knowledge, this is the second study dealing with the intraspecific comparison of fungal mt genomes but the first ever conducted with AM fungal mt genomes. Torriani *et al.* (2008) compared the mt genomes of isolates of the Ascomycete *Mycosphaerella graminicola*, from USA and Europe, and observed little variation. We show here that comparison of mt genomes of *R. irregularis*, obtained by direct sequencing of total DNA using Next Generation Sequencing (NGS) approaches, provides new horizons to investigate the populations and genetics of this species.

Multiplexed direct sequencing of mt genomes

Recent analyses dealing with the study of mt genomes from different taxa underlined the power of NGS techniques (Horn *et al.*, 2012; Zaragoza *et al.*, 2010; Jex *et al.*, 2008). One of our goals was to estimate the ability of these techniques to generate entire mt genome sequences of AM fungi from a reasonable amount of total gDNA in a semi routine way. The first published *R. irregularis* (FACE#494) mt genome was sequenced after a preliminary step of whole genome amplification (WGA; Lee & Young, 2009). WGA is a powerful approach when the quantity of biological material is limiting, but polymerase amplification introduces sequence errors and moreover is not exhaustive enough to ensure that all regions of the mt genome will be represented (Pinard *et al.*, 2006). From our experience of using the two 454 and Illumina sequencing techniques, we learned that Illumina technique combined to user-friendly bioinformatics software (CLC Genomics Workbench 5.0.1.) was the most efficient strategy to sequence and *de novo* assemble AM fungal mt genomes from gDNA of *R. irregularis*. Total gDNA was

extracted from a reasonable amount of spores (min 10,000 *i.e.* one double compartment Petri dish). The steady increase of the effectiveness/cost ratio of NGS techniques, the constant decrease of biological material required and the improvement of bioanalytical algorithms should favour their routine use.

Mt genomes showed no variation within strains.

As indicated in Table 2, in contrast with nuclear genomes, no SNPs were detected in mt genomes. A higher SNP rate in nuclear genome than in mt genome has already been observed in yeasts (Clark-Walker, 1991), in contrast to what has been found in mammals (Saccone *et al.*, 2000). The fact that the assembly of the *R. irregularis* mt genome did not show any ambiguity, and the absence of SNP argue for a single mt genome per strain, in agreement with the results of previous studies (Kuhn *et al.*, 2001; Börstler *et al.*, 2008; Lee & Young, 2009).

Interspecific phylogenetic relationships

The phylogeny of Glomeromycota and related taxa is still under debate as the relative position of this group varies according to the sequences used (nuclear, (James *et al.*, 2006); mitochondrial, Lee and Young, 2009; nuclear and mitochondrial, Liu *et al.*, 2009) and to the number of taxa used. Recently, two phylogenetic studies based on mtDNA-encoded proteins (Pelin *et al.*, 2012; Nadimi *et al.*, 2012) proposed that Mortierellales and Glomeromycota were sister clades. Nadimi *et al.* (2012) have further defined that the robustness of this grouping was lowered when introducing in the analysis *Smittium culisetae*, a Harpellales species. After adding our sequences of mtDNA-encoded proteins to those available in Genbank, we obtained a tree (Fig. S1) where *Rhizopus oryzae* and *Mortierella verticillata* grouped together as sister clades of Glomeromycota, still forming an independent clade. This new version of the relative position of basal fungi

reinforces the comment of Nadimi *et al.* (2012) that only a broad taxon sampling of mtDNAs will allow resolving the phylogeny of basal fungi.

Intraspecific phylogenetic relationships and cosegregation of mitochondrial and nuclear genomes

By showing that the intraspecific variations were lower than interspecific variations, we confirmed that all strains belong to the same species, *Rhizophagus irregularis* (Figure 1). However, our results, both on mitochondrial (LSU, CMG, VGE number and position) and nuclear markers (ITS and microsatellites), showed concordant sub-grouping of MUCL_43204 apart from the other strains (Fig. S2-S3 and Table S2). Several studies showed that nuclear material could be exchanged between strains (Angelard *et al.*, 2010; Angelard & Sanders, 2011; Colard *et al.*, 2012) but mitochondrial genomes were not investigated because of the lack of appropriate markers. Our results suggest the existence of cosegregation mechanisms and/or coevolution of the two genomes. The mechanisms governing the transmission of mitochondria could intervene at the segregation or post-segregation levels. Indeed, mechanisms of vegetative incompatibility during anastomosis could prevent the mixing of mitochondrial material from different strains. Or, it can be envisaged that mechanisms of selection occur after hyphal fusion, driven by nuclear and mt genome interactions, as described in the oomycete *Phytophthora infestans* (Giovannetti *et al.*, 1994). Although meiotic and mating type-like genes have been identified in *Rhizophagus* spp (Halary *et al.*, 2011; Tisserant *et al.*, 2012), there is no evidence for sexual life. Clearly the mt genome homogeneity and nuclear genome heterogeneity within strains pose fundamental questions as to the mechanisms involved in the exchange of genetic material between individuals.

Mt genome plasticity is generated by three types of VGE

The comparison of five mt genomes allowed us to identify different Variability Generating Elements (VGE): Homing Endonuclease ORFs (HEOs), DNA-Polymerase-domain Containing ORF (DPDCO) and Small Inverted Repeats (SIRs).

Homing Endonuclease ORFs (HEOs)

HEOs participate in the genome dynamics thanks to the creation and reorganization of introns (Gimble, 2000). *Via* the enzymes produced, these ORFs could confer to introns the ability to be spliced and/or to be re-inserted in the genome at a target site of 10-40 bp. These sequences are considered as mobile elements that do not alter the RNA produced by the host gene thanks to their ability to self-splice (Chevalier & Stoddard, 2001). HEOs are also known to generate partial duplication of the 3' end of genes at their insertion spot, resulting in different gene versions (Paquin *et al.*, 1997). In some cases, these genes can be differently spliced giving HEOs an efficient role for neo-functionalization. In *Podospora anserina*, it has been shown that the first intron of the *COXI* mitochondrial gene plays a crucial role in the lifespan of the organism (Begel *et al.*, 1999). The *COXI* gene of MUCL_43204 has nine insertions not present in the other strains; one of them is present in the first intron and corresponds to a degenerated HEO with a LAGLIDADG domain. Metabolic activation of mitochondria is important during the presymbiotic growth of AM fungi (Besserer *et al.*, 2006; Besserer *et al.*, 2008). It would be interesting, through experiments on cell respiration, growth rate, host specificity and colonization efficiency, to examine the incidence of these HEO-induced genome rearrangements.

DNA-Polymerase-domain Containing ORF (DPDCO)

Several VGEs had partial or complete DNA polymerase domain containing ORFs. Mitochondrial DNA polymerases are largely unknown or uncharacterized in most

eukaryotes (Shutt & Gray, 2006) with the exception of human and yeast (Kaguni, 2004). In *R. irregularis*, these motifs have a size range from 307 bp to 2021 bp; some are shared between strains and others are strain specific, thereby indicating great variability and plasticity of the mt genome. In contrast to FACE#494, the other five strains have an insertion (at position 7611 in FACE#494) that corresponds to a DPDCO (1560 bp). This observation in close isolates suggests that the insertion-deletion is recent and that these elements have the ability to be mobile. It was recently proposed that the origin of DPDCOs in the ciliate *Oxytricha trifallax* comes from linear mitochondrial plasmids that act like mobile elements which invade the mt genome periodically (Swart *et al.*, 2011). In the Basidiomycete *Moniliophthora perniciosa*, the stable integration of a linear plasmid has been reported in the circular mt genome (Formighieri *et al.*, 2008). This integrated plasmid is composed of two hypothetical proteins and an RNA polymerase on the forward strand, and of one hypothetical ORF and a DNA polymerase on the reverse strand. In the mt genome of strain MUCL_46241, a sequence with a large 5 kb-ORF, the main source of genome size increase when compared to FACE#494, is also composed of two hypothetical proteins on the forward strain and a DNA polymerase on the reverse strain. This organization and the composition of this sequence suggest here the insertion of a free mitochondrial plasmid. We propose that *R. irregularis* mt genomes have been subjected to similar integration phenomena and that other strain-specific DPDCOs are ancient integrated plasmids which have degenerated. As for HEOs, the incidence of DPDCOs in metabolic activity of mitochondria has been investigated. In *Podospira anserina*, the integration of linear plasmids in the mt genome was correlated to greater longevity (Hermanns & Osiewacz, 1996).

Small Inverted Repeats (SIRs)

The third type of VGE identified are small elements of 33-99 bp that can fold in a hairpin-like structure. These SIRs were first identified in *Chlamydomonas reinhardtii* (Boer & Gray, 1991). They are assumed to contribute to genome rearrangement through a mechanism analogous to bacterial transposition or recombination (Denovan-Wright & Lee, 1994; Nedelcu & Lee, 1998). In fungi, similar elements in the same size range were also found, but folded into a double hairpin element (DHE) (Paquin *et al.*, 2000). The origin and the mechanisms of dispersion of these elements remain unclear but certain studies have reported that some of them are mobile elements (Nakazono *et al.*, 1994), probably using a transposition model like “cut and paste” (Grindley & Reed, 1985). Moreover, it has been reported that the target sites of homing endonucleases have a size range of 15-40 bp and some have a palindromic organization and thus are close to the structure of SIRs (Jurica *et al.*, 1998). In *R. irregularis*, we found SIRs in all the strains studied, either in introns, hypothetical proteins, DPDCOs, HEOs or intergenic regions, but never in exons of CMGs, suggesting the presence of selection mechanisms before or after insertion. We identified five types of SIRs differing in their sequences, able to be targeted by different kinds of homing endonucleases having different cleavage site specificities. We can speculate that a collaborative mechanism between SIRs and HEOs occurs: SIRs, containing the cleavage site, could transpose to a new position and then be cleaved by endonucleases which can insert their sequences creating new HEOs. This hypothesis is supported by the presence of different types of SIRs and different HEOs in mt genomes. However, we were not able to identify a previously described HEO cleavage site (Jurica & Stoddard, 1999) in the different types of SIRs. Several SIRs were conserved between strains, showing identical primary or secondary structures although not at the same position in the genomes. This indicates that SIR transposition is a dynamic system, still active, and greatly involved in the plasticity of the mt genome of *R. irregularis*.

Mitochondrial markers are suitable for strain differentiation and monitoring.

Mt genomes have often been used to investigate fungal population dynamics in the field (Börstler *et al.*, 2008, 2010; Sykorova *et al.*, 2011) and in genetic studies such as that investigating mitochondria inheritance after crosses in the ascomycete *Mycosphaerella graminicola* (Ware, 2006). Thanks to the development of precise mitochondrial markers, we succeeded in differentiating our five *R. irregularis* strains by using one spore only for each fungal sample and simple PCR and electrophoresis analyses. We also demonstrated that the strains MUCL_46240 / MUCL_46241 were not differentiable either by using nuclear microsatellites, by comparing their mitochondrial genome or, *a fortiori*, when using the mitochondrial markers mtLSU and CMG. MUCL_46240 and MUCL_46241 were collected in two cultivated fields in Québec (Canada), approximately 30 km apart from each other. If we exclude any experimental error during strain isolation or subculturing, we must conclude that MUCL_46240 and MUCL_46241 are two isolates of the same strain. This stresses the interest of using the mt genome as a source of stable - one genome per individual - and robust markers for typing and tracing AM fungal strains. Finally, the mitochondrial primers we defined from newly identified VGE, usable in single spore PCR assays, will permit studies of mitochondrial exchanges in populations of *R. irregularis* and will help investigate the “parental” inheritance of mitochondria in AM fungi.

In conclusion, we have identified new variable regions in mitochondrial genomes of *R. irregularis*. These regions are composed of three different types of variability-generating elements that provide new genetic markers for precise monitoring of *R. irregularis* at the strain level. These markers will also be useful for fundamental studies on genetic exchanges in AM fungi.

Acknowledgements

The authors thank Jérôme Lluch (PlaGe, Campus INRA Auzeville) and Francis Carbone (LRSV) for their role in Illumina sequencing. This work was supported by the FUI project Neofertil. MM and AH were funded by grants from Neofertil. DF was funded by Agronutrition. This work was done in the LRSV lab, which is part of the TULIP "Laboratoire d'Excellence (LABEX) (ANR -10-LABX-41)".

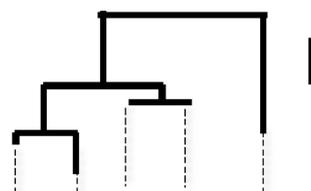
Supplementary data

Table S1: List of PCR primers used for the typing and identification of the different strains of *Rhizophagus irregularis*.

Name	Sequence of primer (5'-3')	Position of primers on mitochondrial genome (bp):						Expected amplicon size (bp)				
		DAOM197198	MUCL46239	MUCL43204	MUCL46240 / MUCL46241	DAOM197198	MUCL_46239	MUCL_43204	MUCL_46240/ MUCL_46241			
GenRL3Fb	TCAACGACCACACGCCA4AAA	66054	66089	81628	70109							
GenRL3Rb	ACGCTATCGCTAAAAGGTGG	66337	66372	81925	70392							
GenRL5Fa	CTACTATGGTCTATgTGGG	52140	52004	63116	52030							
GenRL5Ra	GGCACITTAGITTTGCATAC	52292	52156	63406	52296							
GenRL5Fb	GTTTGTCTTCTAGGGGCTT	49164	49028	/	48982							
GenRL5Rb	GTTACACCAACGGTAAGTCTG	49408	49272	/	49243							
GenRL0Fa	AATCATGAAGGACGGGAAA	8077	8079	/	7854							
GenRL0Ra	CAGAAATATACCCCTAAGGTT	8492	8382	/	8269							
GenRL0Fb	TCTTAGCTGCCCTTACATACC	70178	70213	87225	74231							
GenRL0Rb	TCTAGTACGGCCTTAGGTAA	70492	70527	87484	74506							
DAOM3Fc	ATAACTGGCCCAAGTTACGTATGG	61422	/	/	/							
DAOM8Rb	ATTAATAACAACAGAGCCCGT	62524	/	/	/							
RL3_v_Fa	TAGTTATCACGGCCCTCGTATGG	/	/	48008	/							
RL3_v_Rd	ACTGCTAGTTTGTCTTTATCCT	/	/	48574	/							
RL5Fe	CTGAGATTACTCCTCTTATCC	/	/	/	58723							
RL5_v_Rb	CTCAAGAGGATGCCTCTTCTA	/	/	/	59450							

Table S2: Summary of microsatellite allele size observed for 16 loci (Mathimaran *et al.*, 2008) from nuclear gDNA of five strains of *Rhizophagus irregularis*. Tree on the right illustrates distance analysis (observed divergence, NJ analysis) based on microsatellites characters. Bar: 0.1

	Glint02	Glint03	Glint04	Glint05	Glint06	Glint07	Glint08	Glint09	Glint10	Glint11	Glint12	Glint13	Glint14	Glint15	Glint16	Glint18
DAOM197198*	200	249	252	210	204	260	149	133	179	173	175	167	162	163	163	131
MUCL_46239	200	249	252	210	204	261	148	133	179	173	175	167	162	163	162	131
MUCL_46240	200	249	252	210	204	260	149	133	186	169	175	177	163	163	163	131
MUCL_46241	200	249	252	210	204	260	149	133	186	169	175	177	163	163	163	131
MUCL_43204	200	256	255	211	209	254	153	131	186	176	187	177	163	166	162	128



*: the differences of size observed for most microsatellites compared to data from Mathimaran and co-authors (2008) for the strain DAOM197198 could be due to the different methods of visualization used (48-capillary 3730 DNA Analyzer in our case compared to Spreadex EL800 gel).

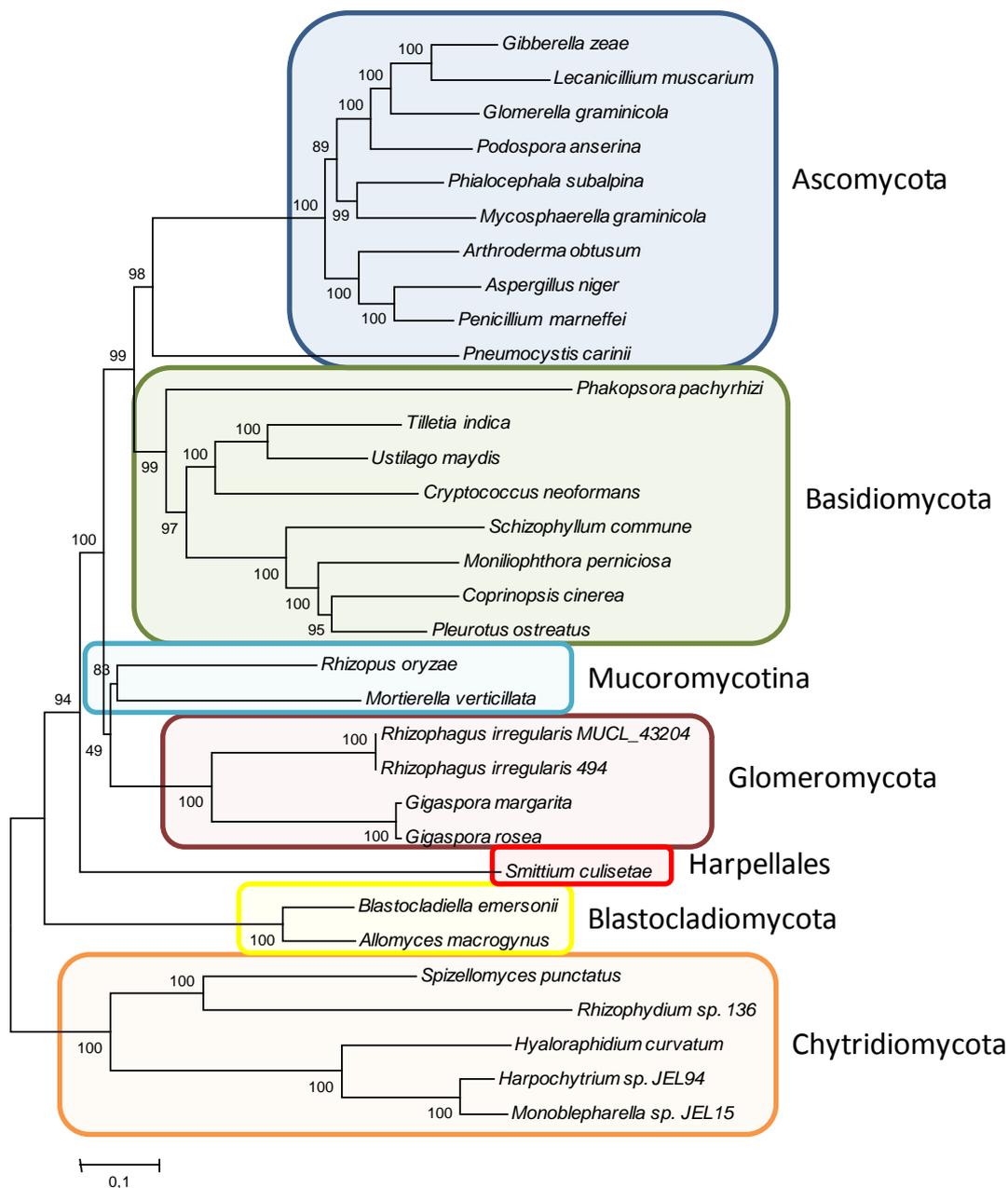


Figure S1: Fungal phylogeny based on concatenation of the 14 amino acid sequences of CMGs (conserved mitochondrial genes) comparison. Genes included in the alignment are the same as those used by Pelin *et al.* (2012) and Nadimi *et al.* (2012). We concatenated the amino acid sequences of the ATP6, ATP8, ATP9, CYTB, COX1, COX2, COX3, NAD1, NAD2, NAD3, NAD4, NAD4L, NAD5 and NAD6 proteins. Alignment procedures and phylogenetic inferences are described in Methods. The evolutionary history inferred using both the Maximum Likelihood method based on the Tamura-Nei model (shown here) and LG model of evolution gave the same grouping. The major clades are surrounded by coloured rectangles. The tree with the highest log likelihood (-90176.6958) is shown. Bootstrap values are shown next to the branches. Branch lengths are representative of the number of substitutions per site.

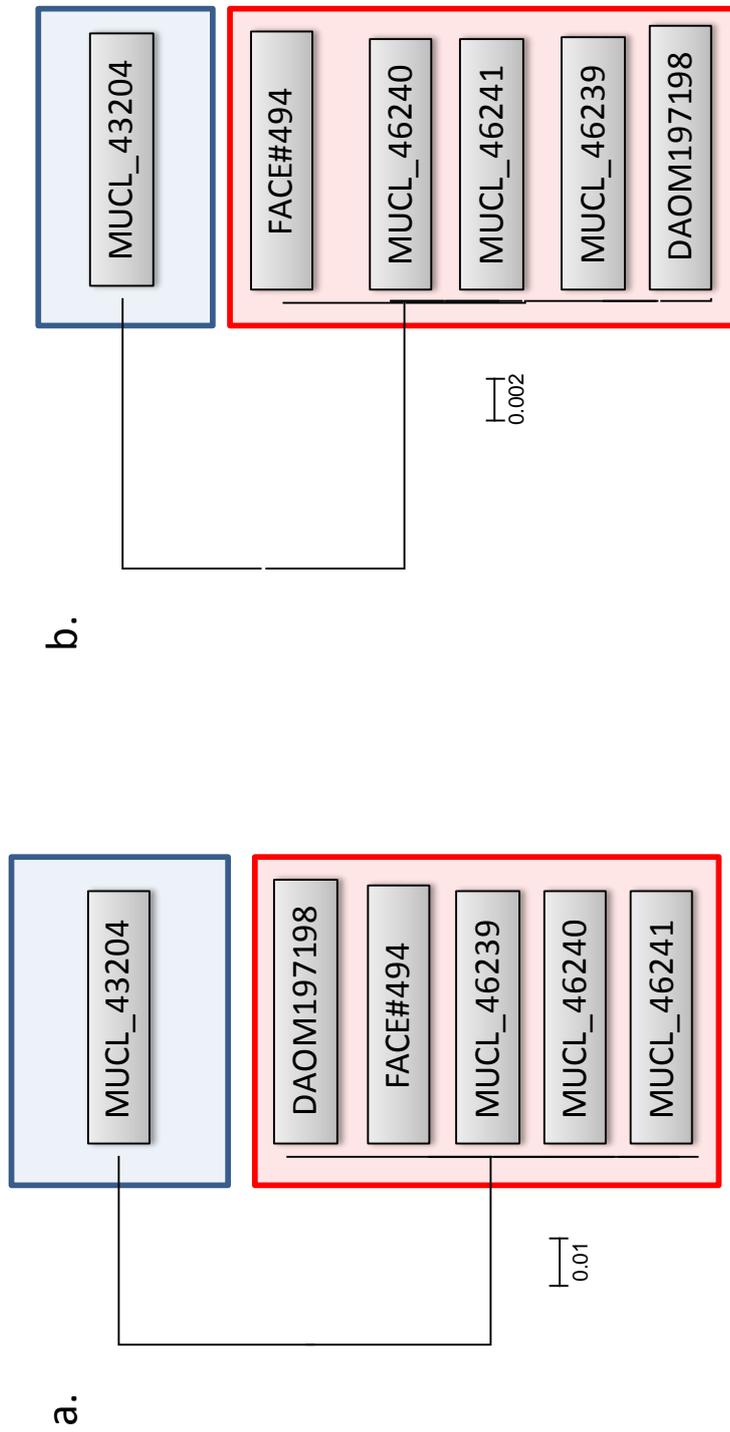


Figure S2: Unrooted phylogenetic trees based on mitochondrial ribosomal Large SubUnits (mtLSU) (a) and on Conserved Mitochondrial Genome (CMG) (b) sequences of six strains of *Rhizophagus irregularis*. Branch lengths correspond to the number of substitutions per site. Clades surrounded by colored rectangles are supported by 100% bootstrap values.

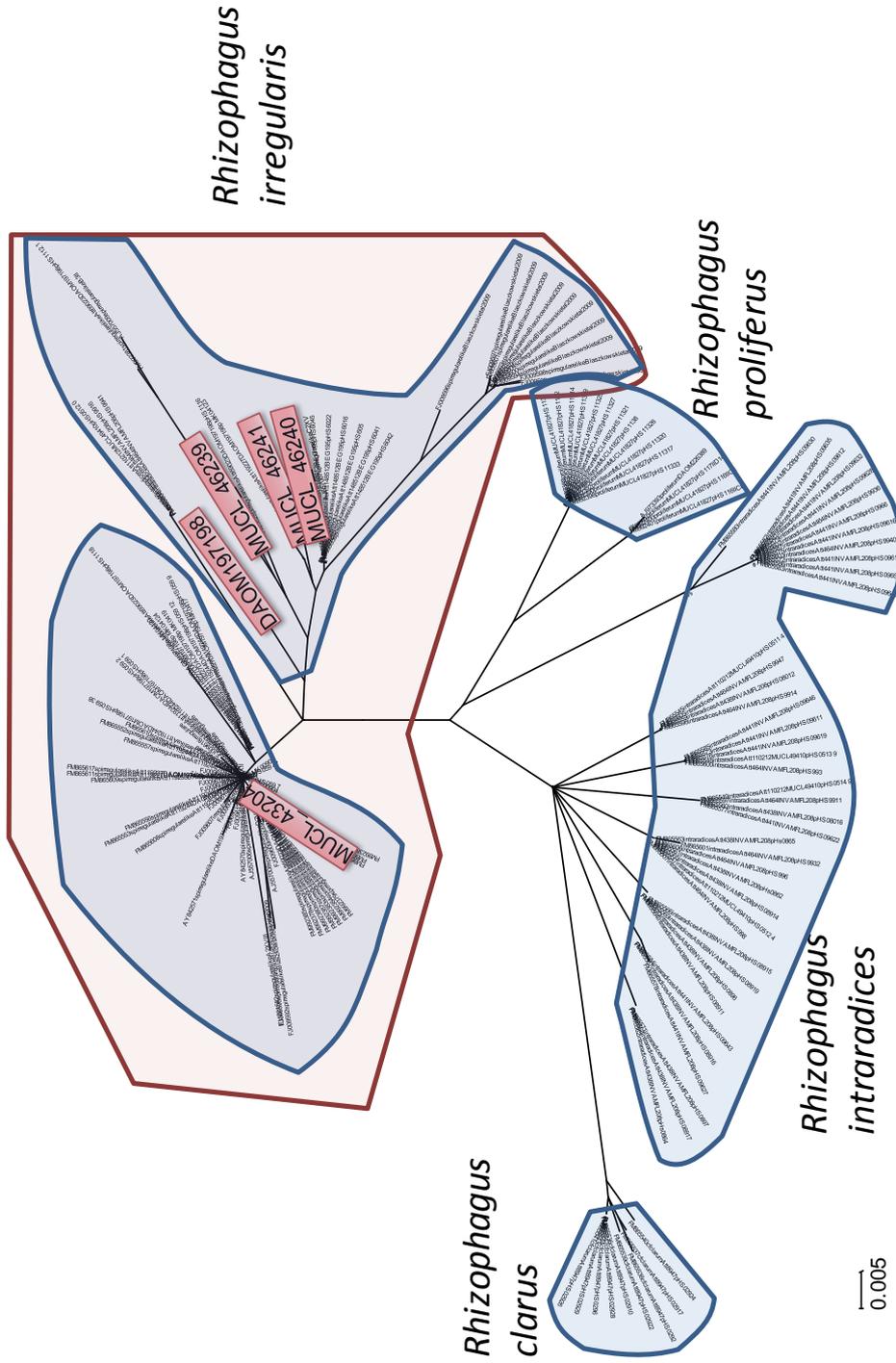


Figure S3: Phylogenetic tree of *Rhizoglyphus* species based on nuclear internal transcribed spacers using parsimony analysis. Sequences obtained in the present study were added to alignment from Stockinger *et al.* (2009). Branch lengths correspond to the number of substitutions per site. Groups are surrounded by coloured circles.

Remark: The ITS sequences added to the original tree of Stockinger *et al.* (2009) were obtained by consensus of sequencing reads. This procedure generated one chimeric ITS sequence per strain while we know that AM fungi present several alleles of ITS sequences (Stockinger *et al.*, 2009). The observed ratios of divergence of the ITS sequences we provided to this study were 0.32 %, 2.54 %, 5.45 % and 3.34 % for sequences DAOM197198, MUCL_46239, MUCL_46241 and MUCL_43204 respectively.

Figure S4: Mitochondrial genome map of the different *Rhizophagus irregularis* strains studied. Inner circle contains the small inverted repeats features. Outer circle contains all others features. Outer and inner features are on the forward and reverse strand, respectively. The origin point (0 kb) is based on the first nucleotide of the COX2 CDS. a. Map of strain DAOM197198. b. Map of strains MUCL_43204. c. Map of strain MUCL_46239. d. Map of strains MUCL_46240/46241.

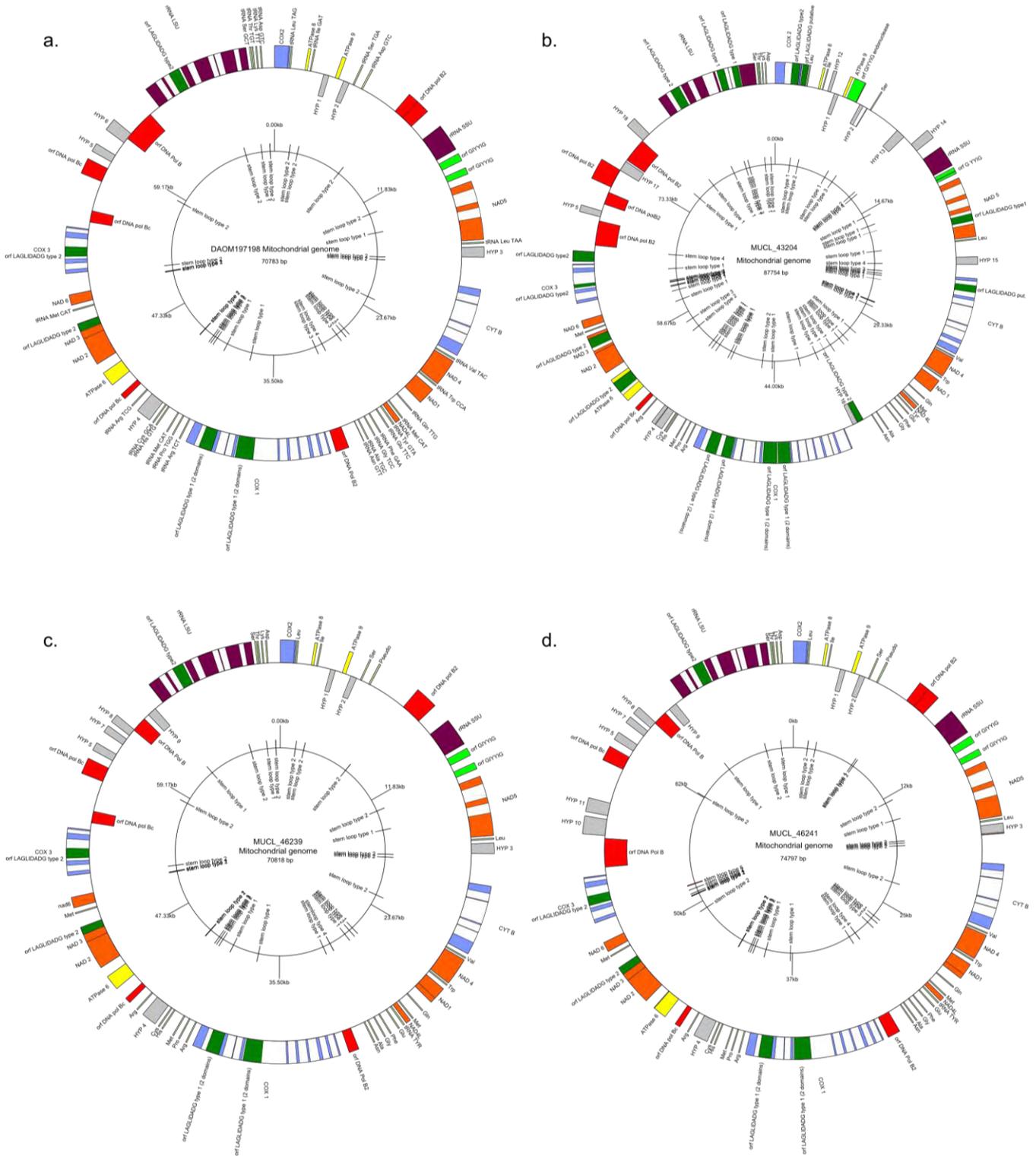
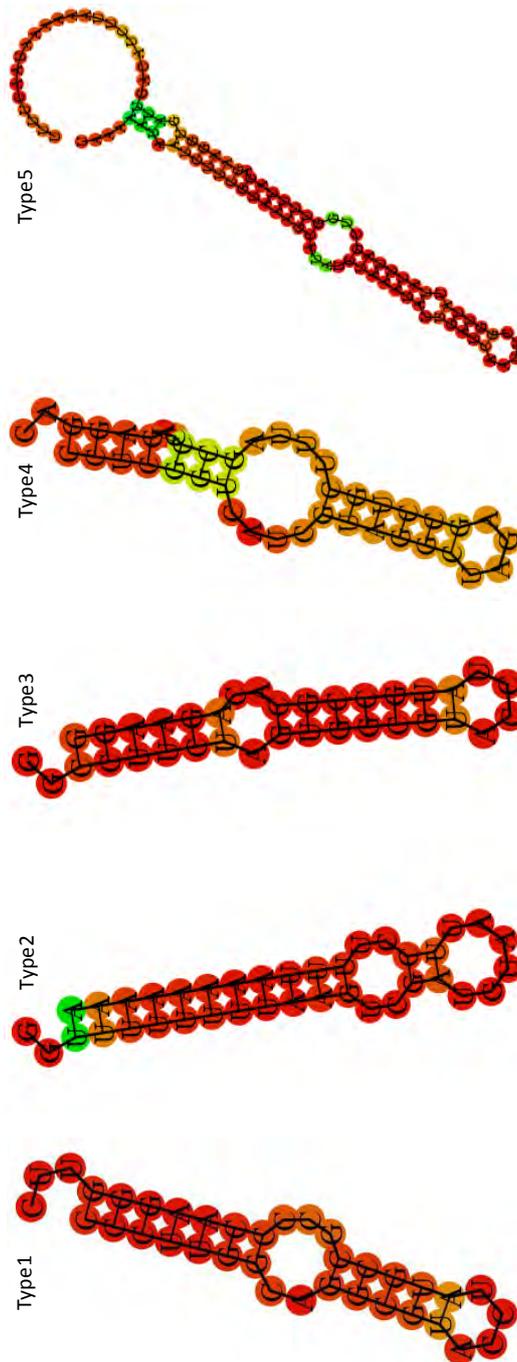


Table S2: Distribution of Small Inverted Repeats (SIR) depending on the type and insertion locus, in different strains of *Rhizophagus irregularis*. Below are represented the different types of secondary structure of SIRs.

coordinates	type	locus	coordinates	type	locus	coordinates	type	locus	coordinates	type	locus
1603	2	intergenic	1602	2	intergenic	1602	2	intergenic	1986	1	intergenic
2500	2	intergenic	2499	2	intergenic	2499	2	intergenic	3099	2	intergenic
11775	2	orf GIV-YIG endonuclease	7538	2	intergenic	7540	2	intergenic	7252	1	intron HYP2
13949	1	intron NAD5	13261	2	orf GIV-YIG endonuclease	13147	2	intergenic	8458	3	intergenic
16433	2	intergenic	15435	1	intron NAD5	15320	1	intergenic	11695	2	intergenic
22269	2	intron aCytB	18063	2	intergenic	17948	2	intron NAD5	12111	4	hypothetical protein 14
26691	1	intergenic	18361	2	intergenic	18246	2	intron NAD5	15286	1	intron GIV-YIG
27107	2	intergenic	22609	2	intron aCytB	22478	2	intergenic	16680	1	intron NAD5
27738	3	intergenic	27032	1	intergenic	26899	1	intergenic	17591	1	intron NAD5
30403	3	intergenic	27448	2	intergenic	27315	2	intergenic	20323	1	intergenic
37498	1	intron COX1	28079	3	intergenic	27946	1	intergenic	22002	4	hypothetical protein 15
40342	1	intergenic	29976	4	intergenic	29842	4	intergenic	22954	3	intergenic
41620	2	intergenic	30786	3	intergenic	30652	1	intergenic	23319	2	intron aCytB
41897	2	intergenic	37882	1	intron COX1	37747	1	intergenic	23589	1	orf LAGLADAG homing endonuclease putative
42116	1	intergenic	40726	1	intergenic	40591	1	intergenic	24153	1	intergenic
43275	1	intergenic	42004	2	intergenic	41869	2	intergenic	27186	1	intergenic
43309	2	intergenic	42281	2	intergenic	42146	2	intergenic	27271	2	intergenic
48883	2	intergenic	42500	1	intergenic	42365	1	intergenic	27878	1	intron aCytB
50838	2	intergenic	43659	1	intergenic	43524	1	intergenic	31991	1	intergenic
51025	1	intergenic	43693	2	intergenic	43558	2	intergenic	33407	1	intergenic
51481	2	intergenic	51116	1	intergenic	50980	1	intergenic	34132	1	intergenic
58392	2	intergenic	51184	2	intergenic	51048	2	intergenic	35670	1	intergenic
62682	1	intergenic	51791	2	intergenic	51655	2	intergenic	36805	1	orf LAGLADAG homing endonuclease type 2
67477	2	intron LSU	58715	2	intergenic	58565	2	intergenic	38066	1	intergenic
69247	1	intergenic	67615	2	intron LSU	62855	1	intergenic	40550	1	intron COX1
			69385	1	intergenic	67650	2	intergenic	44256	1	orf LAGLADAG homing endonuclease type 1 (2 domains)
			70301	2	intergenic	69420	1	intergenic	45630	2	intron COX1
						70336	2	intergenic	49853	1	intergenic
									50265	1	intergenic
									52105	1	intergenic
									52716	1	intron HYP4
									53201	1	intergenic
									56095	2	orf LAGLADAG homing endonuclease type 2
									57500	3	intergenic
									60582	1	intergenic
									62008	1	intergenic
									62647	1	intergenic
									62767	2	intergenic
									63234	5	intron COX3
									63304	1	intron COX3
									63465	3	intron COX3
									65014	1	intron COX3
									66908	4	intergenic
									75715	1	orf DNA polymerase B 2
									78407	1	intergenic
									81791	1	intron LSU
									83614	1	intron LSU
									84108	3	orf LAGLADAG homing endonuclease type 1
									86346	1	intergenic
									87173	1	intergenic

Figure S5 (suite) : Different types of secondary structure of SIRs.



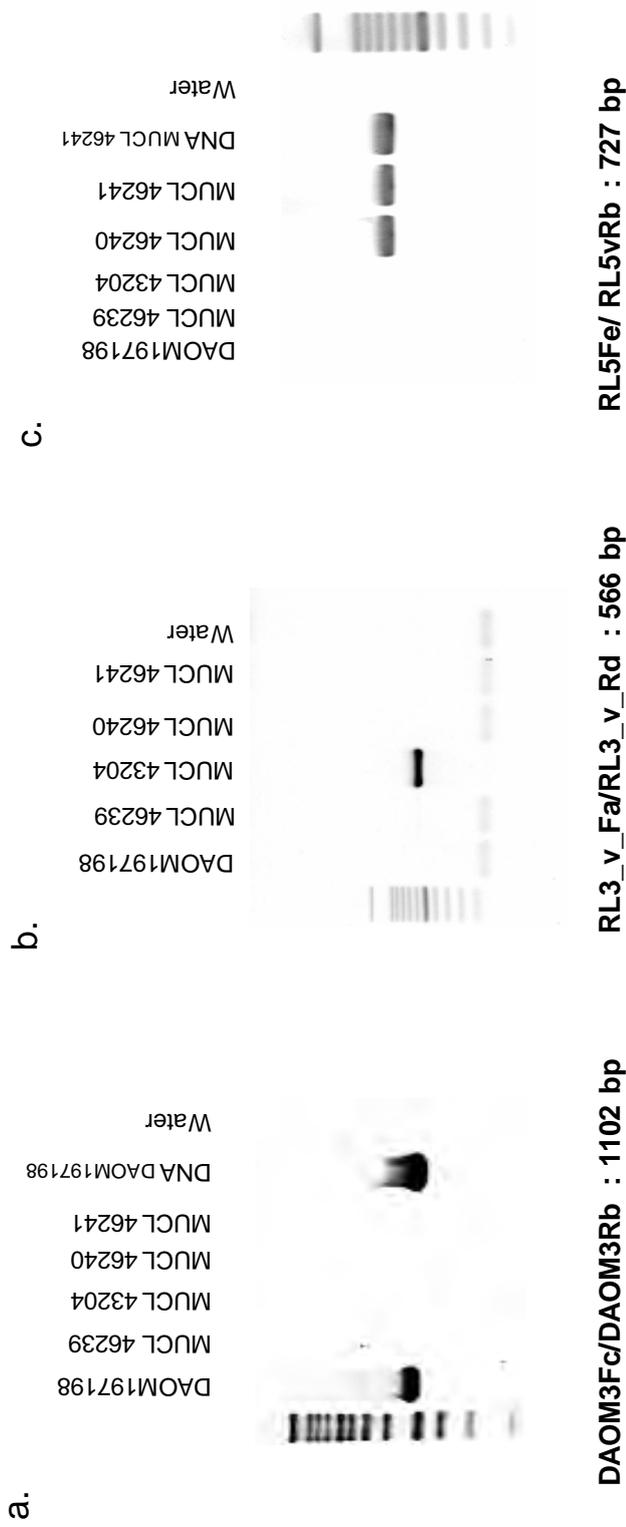


Figure S6. Electrophoresis gels of PCR amplicons from one spore of *Rhizophagus irregularis*. Positive and negative controls are performed with multisporal DNA and water matrixes. (temperature hyb : 60°C and 1 minute for elongation). a. Amplification with specific primers designed for DAOM197198. b. Amplification with specific primers designed for MUCCL_43204. c. Amplification with specific primers designed for MUCCL_46241. Each sample was tested in parallel by PCR targeting 18S region to assess the presence of DNA (not shown).

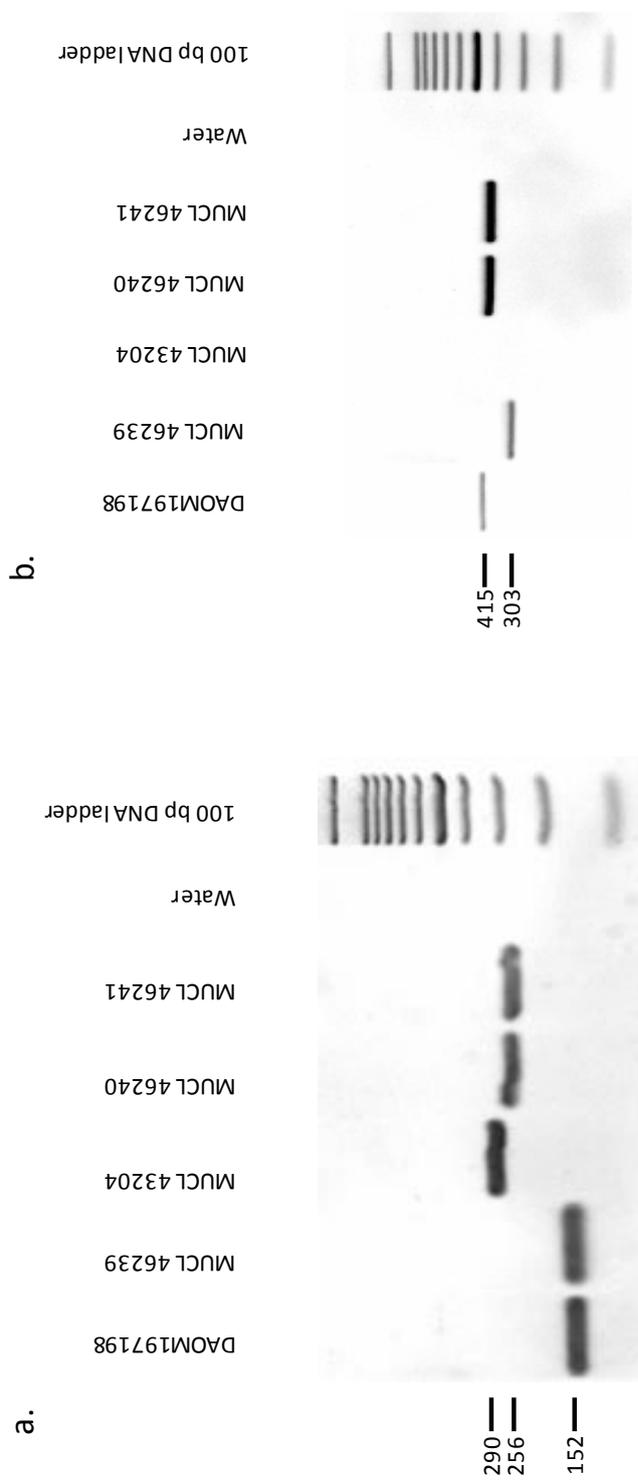


Figure S7: *Rhizophagus irregularis* strain typing by length polymorphism analysis. a. Electrophoresis gel after amplification with GenRL5Fa / GenRL5Ra primers (hybridization temperature 60 °C and elongation time 30 s). b. Electrophoresis gel after amplification with GenRL0Fa / GenRL0Ra primers (hybridization temperature 55 °C and elongation time 30 s).

Compléments aux travaux publiés

Les données incluses dans l'article précédent portent sur le comparatif de 5 génomes de *Rhizophagus irregularis*. Bien que récemment soumise et acceptée, la poursuite de nos travaux sur le séquençage de souches et sur l'expression des gènes mitochondriaux permet d'apporter deux résultats additionnels à cette publication.

Séquençage additionnel d'une souche de Rhizophagus irregularis

Le génome mitochondrial d'une souche supplémentaire de *Rhizophagus irregularis* a été obtenu. Cette souche, dénommée Col_01 dans notre équipe, provient d'un inoculum commercial. Le séquençage par stratégie Illumina Miseq a permis d'obtenir son génome mitochondrial sans ambiguïté. Il se révèle totalement identique au génome mitochondrial de la souche MUCL_43204. Cette dernière provient de Clarence Creek dans l'Ontario au Canada et la probabilité est très forte que Col_01 ait également été isolée dans ce lieu et corresponde à la même souche. Cette hypothèse est renforcée par le fait que les marqueurs microsatellites nucléaires sont également identiques. La confirmation d'une identité de séquence totale entre ces génomes grâce au séquençage complet du génome mitochondrial montre la robustesse des marqueurs que nous avons développés. Les amorces spécifiques définies pour la souche MUCL_43204 testées en PCR sur Col_01 ont généré un amplicon à la taille correspondante. L'ensemble de ces données indique que la souche utilisée dans ce produit commercial est, avec une très forte probabilité, identique à celle déjà répertoriée dans la banque GINCOBEL sous l'identifiant MUCL_43204. Nos marqueurs sont donc efficaces pour le suivi d'une souche spécifique dans un inoculum commercial et l'identification de celle-ci dans d'autres processus ou d'autres produits commercialisés. MUCL_43204 et COL_01 forment un clade qui se distingue des autres souches étudiées ici, dont DAOM197198. Cette distinction illustre probablement la division du clade actuel *Rhizophagus irregularis* en sous clades dont les taxonomistes pourront déterminer la réalité du scindement en nouvelles espèces.

Expression des gènes mitochondriaux

Des travaux d'expression génique réalisés par séquençage HiSeq Illumina sur des ARN extraits de la souche DAOM197198, réalisés par Mathilde Malbreil dans le cadre de ses travaux de thèse dans mon équipe, a permis une quantification des gènes mitochondriaux de DAOM197198 exprimés durant les étapes précoces de germination du champignon par cartographie des lectures de séquençage du transcriptome (Annexe 2).

les cadres de lecture ouverts (ORF) correspondant aux gènes mitochondriaux définis dans ce travail, et précédemment identifiés par Lee and Young (2009), sont exprimés et correspondent à la structure génique attendue. Par ailleurs, dans les conditions biologiques analysées à ce moment, tous les ORFs composant les VGE n'ont pas été détectés à la profondeur de séquençage réalisée : 4 HEO (Homing Endonucléases) sur 7 et 2 DPDCO (domaines DNA polymérase) sur 6 ont été recouverts par les lectures. A noter que l'expression diffère selon les conditions biologiques. Les séquençages à venir du transcriptome du champignon *in planta* seront instructifs pour définir si ces VGE ont une activité différentielle selon l'état physiologique du champignon.

Conclusions et perspectives

En résumé de ces travaux, par comparaison des génomes mitochondriaux de plusieurs souches/isolats de *R. irregularis*, j'ai pu répertorier trois types d'éléments générateurs de variabilité entre les souches: des éléments composés de cadres de lecture ouverts (ORF) de Homing Endonucléases (HEO), des éléments composés d'ORF contenant des domaines DNA polymérase (DPDCO), et des éléments palindromiques répétés de petite taille (SIR). Entre les génomes des souches, ces trois types d'éléments se retrouvent dans des positions variables, supposant une grande mobilité et dynamique.

Les champignons MA ont un mode de vie particulier, ce sont des champignons telluriques qui ont une dispersion des spores très limitée. Dans notre étude, nous avons pu montrer qu'il n'y a, *a priori*, pas d'hétéroplasmie mitochondriale chez ces champignons, montrant que les échanges génétiques mitochondriaux par fusion d'hyphe, ou anastomose, sont très rares entre les souches. Les VGE sont donc, potentiellement, des éléments qui sont acteurs d'un mécanisme global permettant de maintenir une certaine variabilité du génome mitochondrial. Cette variabilité étant stable au sein des souches, j'ai pu

développer des marqueurs moléculaires mitochondriaux qui permettent la différenciation de souche et le suivi spécifique de ces dernières dans un système de production.

L'ensemble de ces travaux valide le projet que la société Agronutrition m'avait confié, à savoir la description des particularités de diverses souches de *R. irregularis* et la production de marqueurs de typage.

Ces travaux sont riches de perspectives :

- i) Nous démontrons par notre schéma expérimental que, dans un avenir proche, le séquençage des génomes mitochondriaux par NGS à partir de quantité limitée de matériel biologique permettra un typage extrêmement précis des souches. En attendant les prochaines évolutions des NGS limitant le coût d'utilisation et la quantité de matériel biologique nécessaire, l'obtention de plusieurs génomes mitochondriaux au sein de cette espèce et l'identification des VGE permettent de définir de nombreuses régions de ces génomes appropriées pour différencier des souches qui ne pourraient l'être sur la base d'autres séquences cibles, telles que les régions ribosomiques mitochondriales ou nucléaires.
- ii) L'homoplasmie mitochondriale observée chez les 6 souches étudiées soulève la question des mécanismes cellulaires de compatibilité entre individus chez *R. irregularis*. Le typage précis des génomes mitochondriaux permet d'envisager un suivi des échanges de matériel génétique entre souches et de percer le système de compatibilité inter-individus chez ces champignons.
- iii) Les attentes directes de la société Agronutrition pour ce projet portaient sur la définition de régions génomiques permettant l'identification ET la quantification de souches de champignons MA par des techniques robustes. *A contrario* des ITS (trop variables au sein d'une même souche), le choix des génomes mitochondriaux et l'identification de sites hautement variables entre souches (VGE) permet de définir des amorces spécifiques d'une souche, permettant, par PCR quantitative, d'effectuer en une seule étape le typage et la quantification. Cette méthodologie permettra également de réaliser un suivi d'une souche ajoutée au champ, en ayant pris la précaution préalable de définir que le morphotype ciblé n'est pas déjà représenté sur le site.

CHAPITRE II: ANALYSE DU MICROTRANSCRIPTOME DE *MEDICAGO TRUNCATULA* LORS DE LA SYMBIOSE ENDOMYCORHIZIENNE A ARBUSCULES

INTRODUCTION

Certains éléments de cette introduction sont inclus dans un chapitre d'ouvrage à paraître :

Formey D, Jourda C, Roux C and Delaux PM, 2012. Chapter 10: What the genomics of arbuscular mycorrhizal symbiosis teaches us about root development. From: "Roots and their soil interactions: What we can learn from genomics". *Wiley*. Ed. Martin Crespi.

La symbiose MA: une co-évolution racine-champignon MA

La racine est un organe souterrain des plantes aux fonctions multiples : structure de soutien, d'assimilation de l'eau et des nutriments du sol, organe de réserve des plantes et siège d'intenses échanges avec l'environnement. Du fait de cette multiplicité de fonctions et de la nécessité d'adaptation physiologique aux différents stress biotiques et abiotiques, il faut considérer que le développement racinaire est sous contrôle de plusieurs mécanismes de régulation de l'expression génique conférant une grande plasticité et une forte réactivité de la machinerie cellulaire. Ces mécanismes sont encore globalement méconnus ou, du moins, bien moins étudiés que le développement aérien. Une raison en est que les études de génétique directe et inverse sur l'organisme modèle *Arabidopsis thaliana*, de loin l'espèce végétale la plus étudiée, élude souvent la caractérisation phénotypique du système racinaire. Outre la génétique directe et inverse, une stratégie très efficace pour étudier le développement racinaire consiste à perturber son développement et en analyser les modifications. A ce titre, l'étude de l'incidence des symbioses racinaires sur le développement racinaire est d'un grand apport, que ce soit l'incidence des champignons ectomycorhiziens, la nodulation ou la symbiose MA. En effet, cette dernière induit un grand nombre de réarrangements dans l'organisation de la racine que ce soit au niveau hormonal, architectural ou moléculaire. L'ancestralité de la symbiose MA avec les premiers végétaux terrestres, observée dès la flore de Rhynie (Remy *et al.*, 1994; Redecker *et al.*, 2000), est un argument appuyant l'hypothèse d'une co-évolution des

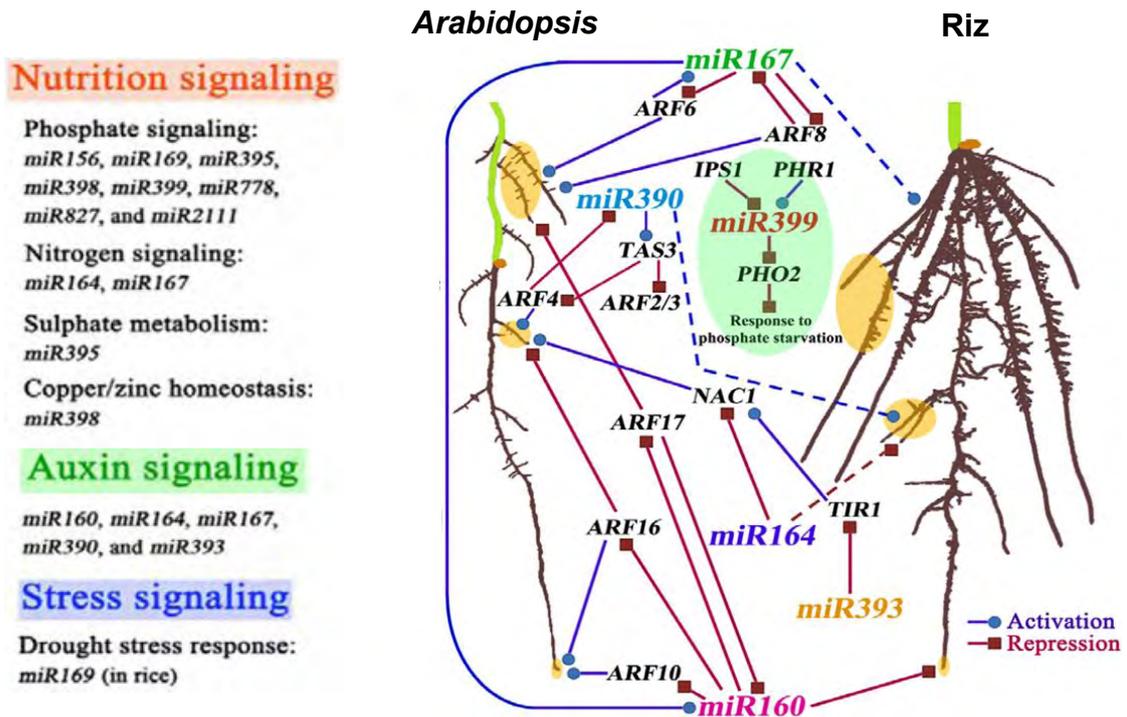


Figure 11 : Schéma des régulations régies par les microARN dans la signalisation nutritionnelle, auxinique et en réponse aux stress, chez *Arabidopsis thaliana* (à gauche) et le riz *Oryza sativa* (à droite). Adapté de Meng *et al.* (2010).

champignons MA avec les végétaux terrestres, depuis leurs premiers représentants jusqu'à leur diversité actuelle (Simon *et al.*, 1993).

Incidence des champignons MA sur le développement racinaire

L'incidence de la symbiose MA sur l'architecture racinaire, particulièrement sur le développement de racines latérales (RL), a été observée depuis de nombreuses années aussi bien sur le poireau, le mirabellier, le riz ou le maïs (Berta *et al.*, 1990, 1995; Kaldorf & Ludwig-Müller, 1999; Gutjahr *et al.*, 2009). Ces modifications ne se limitent pas à la symbiose MA : les champignons ectomycorhiziens induisent la formation de RL en produisant des analogues d'hormones végétales telles que l'auxine (Felten *et al.*, 2010) et l'éthylène (Splivallo *et al.*, 2009). Chez les champignons MA, il a été établi que la modification de la densité de RL est réalisée via des signaux diffusibles : l'application d'exsudats de spores (Kosuta *et al.*, 2003; Olah *et al.*, 2005; Mukherjee & Ané, 2011) ou de Myc-LCO (lipo-chito-oligosaccharide) synthétique (Maillet *et al.*, 2011) stimulent la formation de nouvelles RL chez *Medicago truncatula* de façon DMI-dépendante.

La superposition des mécanismes génétiques qui contrôlent l'établissement de la symbiose et la formation des RL est également illustrée par le mutant de maïs *lrt1* (Paszkowski & Boller, 2002). Ce mutant présente un phénotype racinaire particulier avec une très forte réduction du nombre de RL. Ce phénotype est partiellement rétabli par la mycorhization et également par une teneur élevée en phosphate. Le phosphate étant un des nutriments majeurs apportés par les champignons MA à leur hôte, les réversions phénotypiques partielles du mutant *lrt1* par les champignons MA et l'apport de phosphate suggèrent une superposition des mécanismes de régulation de la formation de RL liés à ces deux types de stimuli. Toutefois l'incidence du phosphate sur le développement racinaire n'est pas spécifique aux espèces végétales établissant des symbioses MA. Chez *Arabidopsis thaliana*, la carence en phosphate affecte la croissance des poils absorbant, de la racine principale et des RL par une voie dépendante de l'auxine (Nacry *et al.*, 2005; Pérez Torres *et al.*, 2009).

Derniers éléments illustrant l'incidence de la symbiose MA sur le développement racinaire, les études transcriptomiques réalisées depuis une décennie montrent que la racine hôte présente une reprogrammation génétique massive en réponse à la colonisation par les champignons MA (Journet *et al.*, 2002; Küster *et al.*, 2004; Hohnjec *et al.*, 2005;

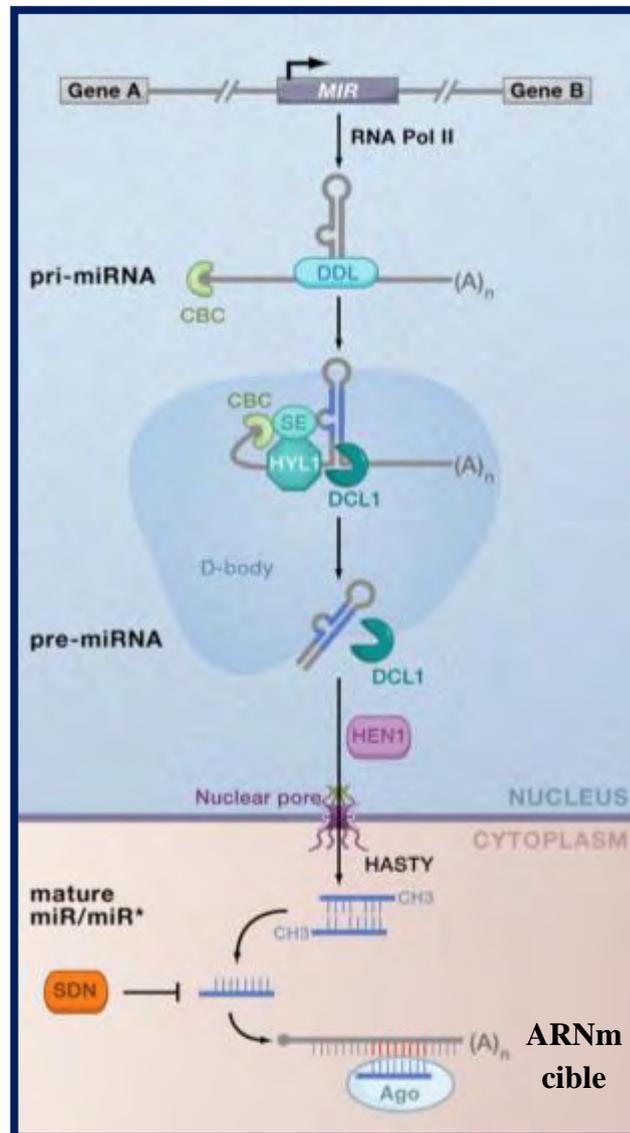


Figure 12 : Biosynthèse et reconnaissance de la cible des microARN chez les plantes. La partie supérieure (bleue) prend place dans le noyau alors que la partie basse (orange) prend place dans le cytoplasme. Adapté de Voinnet (2009).

Siciliano *et al.*, 2007; Gaude *et al.*, 2011). L'analyse comparative des profils transcriptomiques révèle que la symbiose AM entraîne l'activation des métabolismes auxiniques et d'une grande diversité de gènes impliqués dans la réorganisation de la paroi cellulaire végétale, qui sont autant de marqueurs de la modification de l'architecture racinaire sous l'incidence de la mycorhization. Ces profils transcriptomiques révèlent également une forte activation de facteurs de transcriptions membres des familles GRAS, NAC et MYB, dont le rôle dans le contrôle du développement racinaire a été décrit (Dolan, 2007; Shin *et al.*, 2007; Christiansen *et al.*, 2011). Ces gènes régulateurs clés pourraient également intervenir dans la régulation à la fois du développement racinaire et de l'acquisition de nutriments (Krajinski & Frenzel, 2007). Ces régulateurs maîtres, qui sont les facteurs de transcriptions, sont connus pour être eux-mêmes hautement régulés par des régulateurs fins de l'expression : les microARN.

Les microARN: de petits régulateurs aux grands effets

Les microARN sont des petits régulateurs d'environ 21 nucléotides (20 nt à 24 nt) qui régulent négativement chez la plupart des eucaryotes l'expression des gènes. Chez les plantes, ces petits ARN sont impliqués dans différents processus comme le développement, les voies de signalisation hormonales, la balance nutritionnelle, et les adaptations aux stress biotiques et abiotiques (Bartel, 2004; Jones-Rhoades *et al.*, 2006; Mallory & Vaucheret, 2006; Ruiz-Ferrer & Voinnet, 2009) (Figure 11). Les microARN sont codés par des séquences génomiques, dont l'organisation est toujours étudiée actuellement. La plupart sont transcrits grâce à l'ARN polymérase II, générant ainsi un transcrit primaire appelé pri-miARN (Lee *et al.*, 2004) (Figure 12). Ce dernier a des caractéristiques physico-chimiques et structurales typiques qui lui permettent de se replier sur lui-même avec des appariements imparfaits. Cette séquence est alors prise en charge par différentes protéines : DAWDLE (DDL), SERRATE (SE), HYPONASTIC LEAVES1 (HYL1), Cap-binding complex (CBC) et notamment l'enzyme Dicer-like 1 (DCL1) qui permet le clivage et la formation d'une structure en tige-boucle (épingle à cheveux) d'une taille plus réduite, appelée pre-miARN, puis d'un duplex mature de taille définitive composé du microARN et d'un brin complémentaire (miR*). Cet ARN double brin mature est alors protégé par une méthylation grâce à l'enzyme HEN1 et exporté par la protéine HASTY vers le cytoplasme. Le microARN mature, alors désolidarisé du miR* qui a été dégradé, est pris en charge et incorporé dans la protéine AGO afin d'exercer son

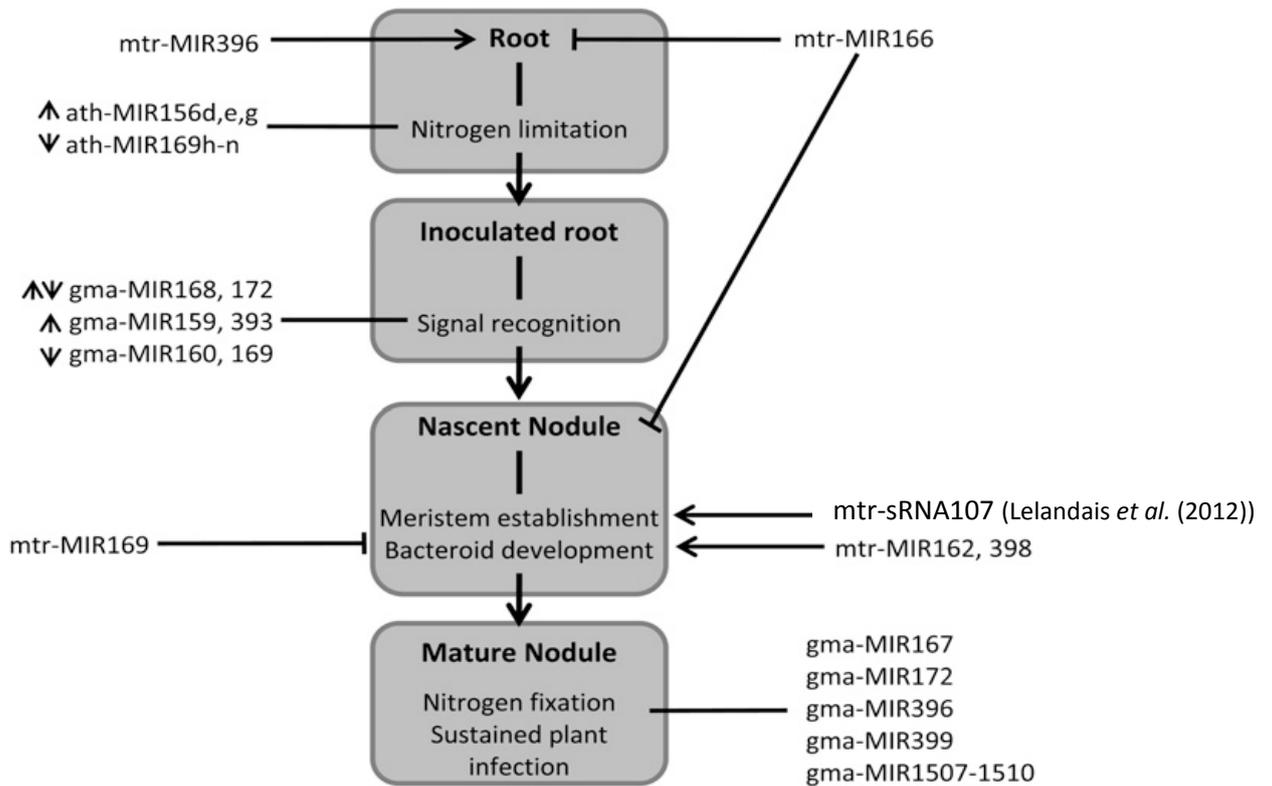


Figure 13 : Modèle de l'interaction des microARN lors de la formation des nodules. Les flèches représentent une activation alors que les barres représentent une répression. Les traits sans indications représentent une implication du microARN dans le processus ciblé dont le rôle n'a pas été caractérisé. Les flèches à gauche des noms de microARN représentent la modification de l'expression du microARN correspondant (haut: sur-expression ; bas: sous-expression). Le rôle potentiel des microARN dans l'homéostasie de l'azote est extrapolé à partir de données obtenues chez *Arabidopsis thaliana*. Le rôle fonctionnel dans ces étapes a été démontré seulement pour les miR166 et miR169. Adapté de Simon *et al.* (2009).

pouvoir de régulateur négatif (Figure 12). Il est connu que les microARN de plante reconnaissent leurs cibles par complémentarité de séquence, qu'elle soit parfaite (Jones-Rhoades *et al.*, 2006) ou comportant des mésappariements (Llave *et al.*, 2002). Les cibles peuvent être des transcrits codant pour des protéines, ou non, suggérant des régulations en cascade par les séquences non codantes, et peuvent être reconnues à divers endroits sur la totalité de la séquence selon les sites cibles des microARN (5' UTR, 3' UTR ou ORF). Une fois l'appariement réalisé entre le microARN et sa cible, la régulation peut prendre plusieurs formes. La plupart des cibles reconnues par des microARN subissent un clivage catalysé par l'enzyme AGO (Addo-Quaye *et al.*, 2008) suivie d'une dégradation. Dans certains cas, lorsque le microARN présente des mésappariements en position centrale avec sa cible, mais une complémentarité parfaite aux extrémités, il se fixe à la cible sans entraîner de clivage (Franco-Zorrilla *et al.*, 2007), mais inhibe la traduction de cette dernière (Brodersen *et al.*, 2008). Ce phénomène se retrouve également avec des interactions microARN/cible parfaites, montrant la complexité des mécanismes de régulation réalisés par ces éléments régulateurs.

Les microARN: incidences sur la symbiose MA et le développement racinaire

La découverte de l'incidence des microARN dans les interactions entre les plantes et les micro-organismes est relativement récente (Navarro *et al.*, 2006; Subramanian *et al.*, 2008). Le rôle des microARN dans la symbiose rhizobienne a notamment été décrit. Durant le développement nodulaire chez *Medicago truncatula*, le miR169 joue un rôle important pour la régulation du facteur de transcription *MtHAP2-1* (Combiér *et al.*, 2006). L'inhibition de l'expression de *MtHAP2-1* par ARN interférant entraîne un arrêt du développement des nodules et donc une symbiose incomplète et non fonctionnelle. Autre exemple, le miR166 intervient dans la régulation du développement racinaire et nodulaire chez *M. truncatula* et sa surexpression mène à un nombre réduit de racines latérales et de nodules symbiotiques (Boualem *et al.*, 2008). Ces deux exemples, ainsi que d'autres microARN candidats proposés suite à des approches de microtranscriptomiques (Simon *et al.*, 2009) (Figure 13), illustrent le rôle des microARN dans l'établissement et le développement de la symbiose rhizobienne.

Les études de l'implication des microARN dans la symbiose MA sont plus fragmentaires. Le miR399, qui régule le maintien de l'homéostasie du phosphate chez

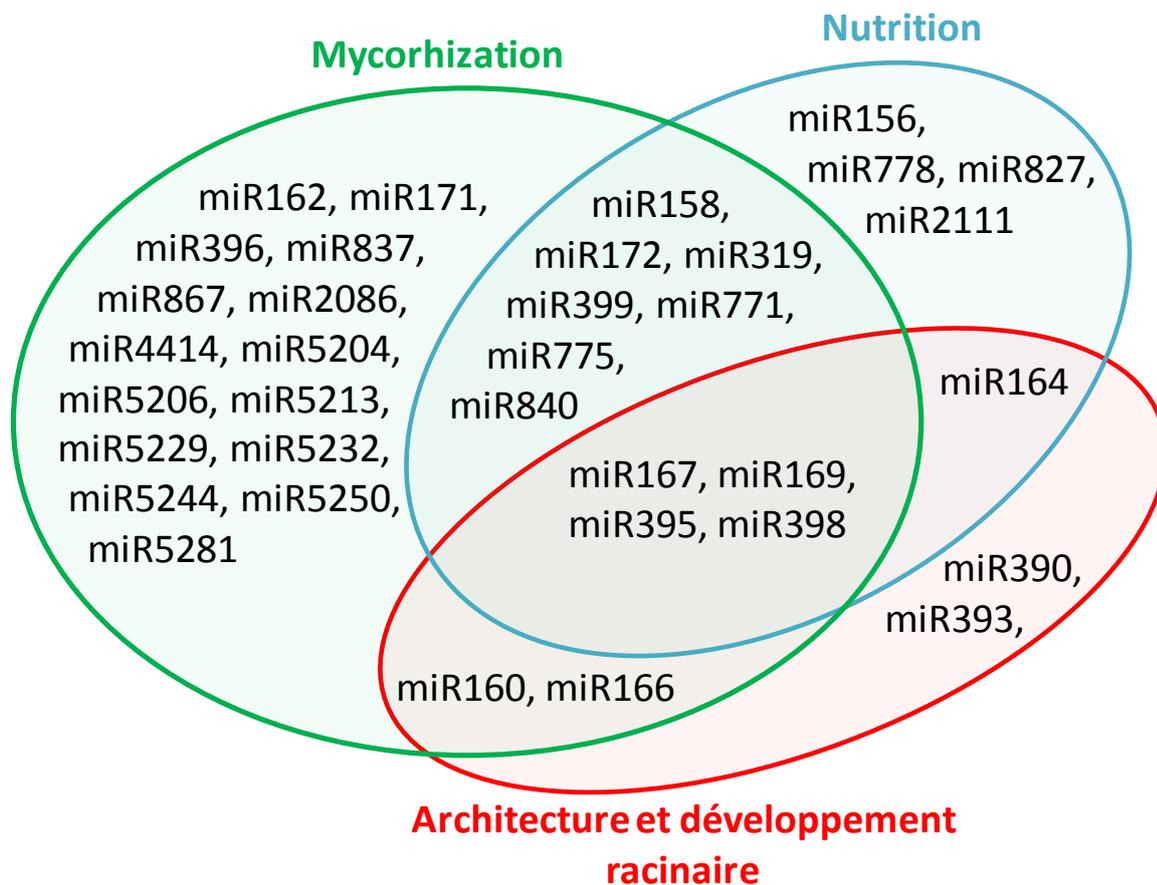


Figure 14: Diagramme de Venn représentant la répartition et les chevauchements des familles de microARN. Le cercle bleu contient les microARN impliqués dans la signalisation nutritionnelle, le cercle rouge, ceux impliqués dans l'architecture et le développement racinaire, et le cercle vert, ceux impliqués dans la mycorhization. D'après (Kawashima *et al.*, 2009; Branscheid *et al.*, 2010; Gu *et al.*, 2010; Meng *et al.*, 2010; Devers *et al.*, 2011; Khan *et al.*, 2011)

Arabidopsis thaliana (Bari *et al.*, 2006), a été décrit chez *M. truncatula* comme jouant un rôle dans la mycorhization en maintenant une faible expression et, *a fortiori*, une faible activité de *MtPho2*. Sans cette régulation par le miR399, l'expression de *MtPho2* augmenterait, en réponse à l'assimilation du phosphate apporté par le champignon, et inhiberait le bon déroulement de la symbiose (Branscheid *et al.*, 2010). Par ailleurs, l'incidence de la mycorhization sur la formation des racines latérales ayant été observée (*cf.* introduction du Chapitre II), le rôle du miR166 dans cette réponse peut être envisagé. L'hypothèse d'une telle similitude de régulation, par le miR166 et d'autres microARN intervenant dans la régulation de l'architecture racinaire, entre les symbioses rhizobienne et MA est appuyée par l'observation d'un effet similaire des facteurs Nod et des facteurs diffusibles exsudés par les champignons MA sur la formation de racines latérales (Kosuta *et al.*, 2003; Olah *et al.*, 2005; Mukherjee & Ané, 2011).

Lors de l'initiation de mon travail de thèse, les connaissances sur l'implication des microARN dans la symbiose MA étaient quasi-nulles. Durant ces trois années, plusieurs travaux sur la caractérisation de microARN jouant un rôle dans la symbiose MA ont été publiés et sont résumés dans la figure 14. Ils montrent de façon attendue un recouvrement des régulations de la symbiose MA avec le développement racinaire et la nutrition des plantes. Il faut citer particulièrement le travail de Devers *et al.* (2011) qui a porté sur des approches comparables à celles que nous avons initiées. Toutefois, les méthodologies de séquençage massif (NGS) et de traitement informatique employées pour réaliser ces cartes d'expression de microARN étaient encore au stade du développement et le travail engagé dans notre équipe a été poursuivi.

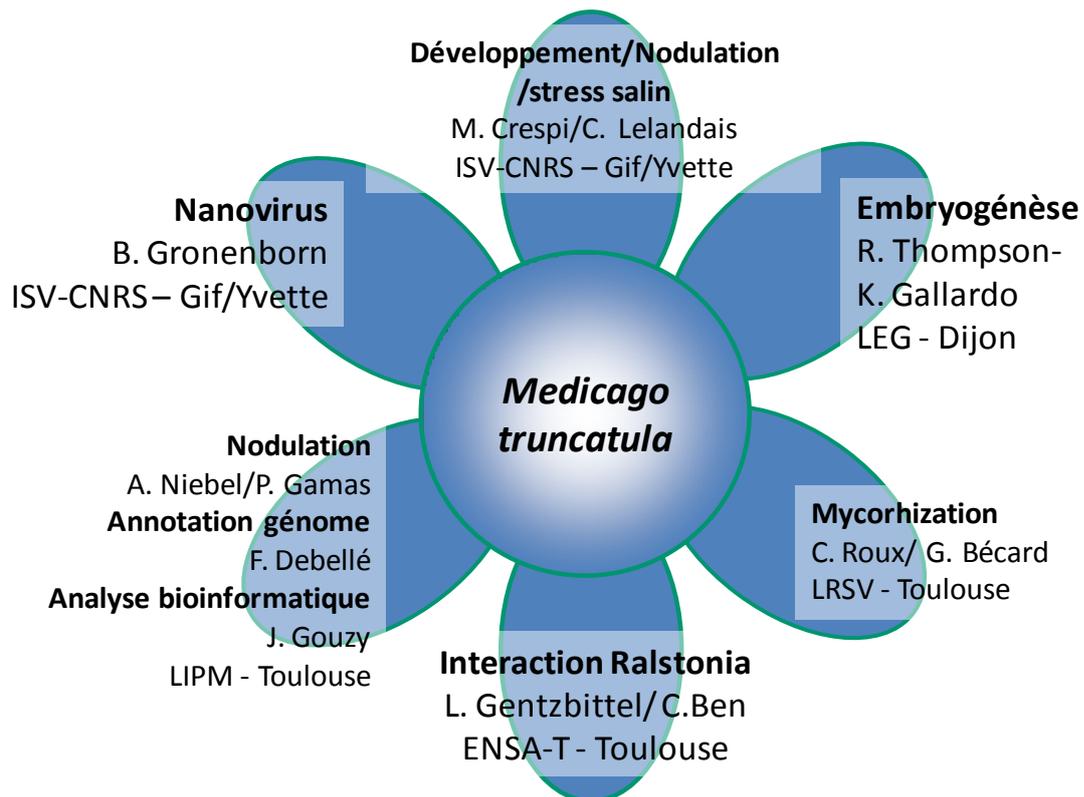


Figure 15: Schéma des différentes conditions étudiées dans le projet Mirmed ainsi que les intervenants associés et leurs établissements d'origine.

Objectif et contexte de l'étude: Le projet MIRMED

L'essentiel des travaux présentés dans le cadre de ce projet font l'objet d'un article en préparation :

Damien Formey*, **Erika Sallet***, **Christine Lelandais-Brière***, **Cécile Ben***, **Jean-Philippe Combier**, **Frédéric Debellé**, **Florian Frugier**, **B. Gronenborn**, **Caroline Hartmann**, **Andreas Niebel**, **P. Wincker**, **Christelle Reynes**, **Christophe Roux**^α, **Laurent Gentzbittel**^α, **Jérôme Gouzy**^α and **Crespi Martin**^α. Genome-wide identification of new small RNAs associated with *Medicago truncatula* root growth in abiotic and biotic environments. Article en rédaction pour soumission à *Genome biology*.

Les études de microARN ont longtemps été menées de manière ciblée mais, depuis ces dernières années, les nouvelles technologies de séquençage (NGS) ont permis de réaliser ces études à un niveau plus global et massif. Ces NGS permettent en effet d'obtenir de larges collections de séquences d'ADN et/ou d'ARN grâce à trois types de technologie : 454 (Roche), Solexa (Illumina) et SOLiD (Life Technologies) (Metzker, 2010). Leur capacité à générer des millions de séquences, mais de tailles plus réduites que les précédentes technologies, en font des techniques puissantes et adéquates pour l'analyse des petits ARN non codants (<50 nt). La profondeur de séquençage que l'on peut obtenir avec ces technologies permet de révéler la quasi-totalité des petits ARN dans un tissu donné, à un instant donné, et permet donc la découverte de nouveaux microARN. La quantification des petits ARN identifiés devient alors possible rendant accessibles les profils d'expression de chaque microARN dans les différentes conditions testées.

L'équipe de recherche au sein de laquelle j'effectue ma thèse participe au projet MIRMED (*microRNA from Medicago*, financement Génoscope, responsable scientifique Dr. Martin Crespi) dont l'objectif est de proposer un panorama le plus large possible des petits ARN, notamment les microARN, de *Medicago truncatula*, au cours de différents stades de développement et en réponse à des stress biotiques et abiotiques, comme le stress salin, l'embryogénèse, l'interaction avec la bactérie pathogène *Ralstonia*, l'interaction avec un nanovirus, la nodulation et la mycorhization (Figure 15). Ce projet se base sur un séquençage Solexa de petits ARN isolés et s'appuie sur un nouveau pipeline informatique (J. Gouzy, E. Sallet - LIPM - CNRS/INRA - Toulouse) récemment publié (Lelandais-Brière *et al.*, 2009). Dans ce catalogue de microARN, ma participation, en collaboration avec A. Niebel et F. Debellé (LIPM - CNRS/INRA - Toulouse), a porté sur l'obtention, l'analyse et la comparaison des microARN en conditions témoins, mycorhizées, nodulées et en réponse à des signaux précoces : facteurs Nod et Myc-LCO.

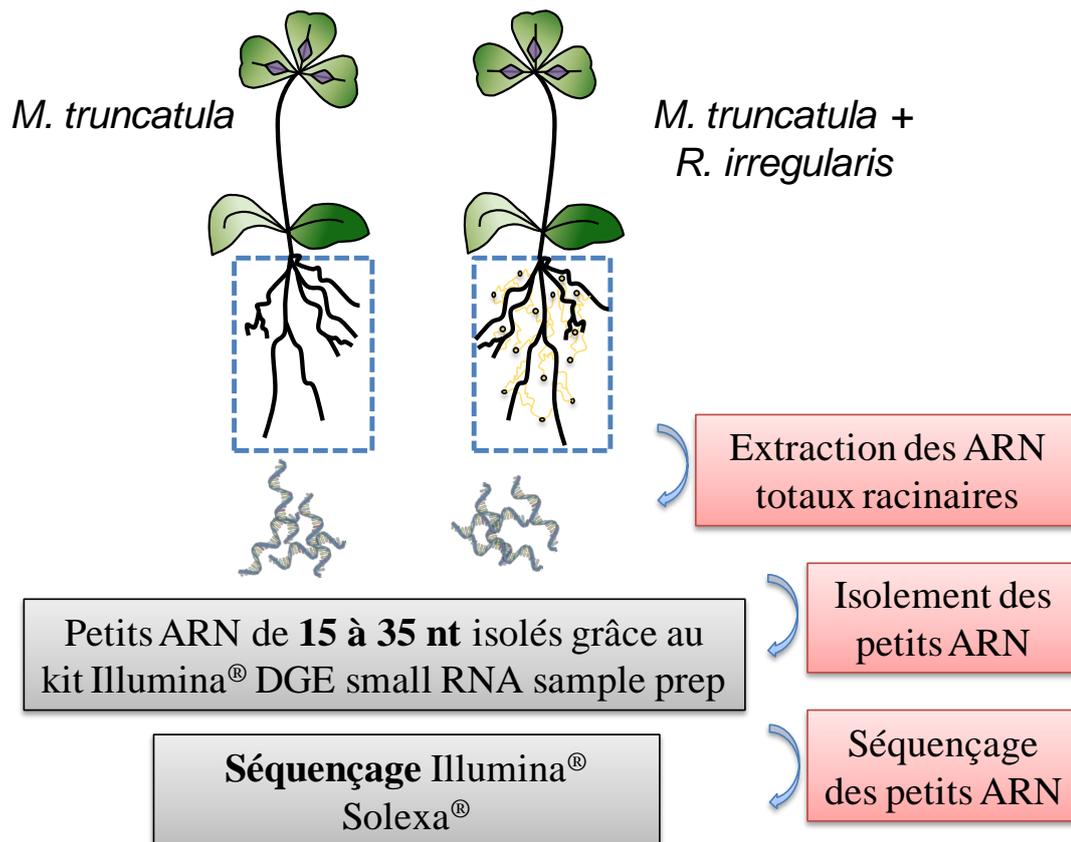


Figure 16: Protocole de production des banques de petits ARN issus de racines de *Medicago truncatula* mycorhizées.

A - ANALYSE DU MICROTRANSCRIPTOME DE MEDICAGO TRUNCATULA LORS DES ETAPES TARDIVES DE LA SYMBIOSE MYCORHIZIENNE A ARBUSCULES

Les premières banques de la seconde phase du projet MIRMED avaient pour but de dresser une liste de microARN différentiellement accumulés dans le système racinaire de *M. truncatula* après l'établissement complet des symbioses rhizobienne, d'une part, et endomycorhizienne à arbuscules, d'autre part.

Matériels et méthodes

Préparation du matériel végétal

Les graines de *Medicago truncatula* sont tout d'abord scarifiées avec du H₂SO₄ 95 % pendant 5 mn puis rincées à l'eau stérile. Elles sont ensuite stérilisées à l'eau de Javel (3 % chlore actif) pendant 1 mn 30 puis rincées à l'eau stérile où elles gonfleront durant 1 h. Les graines imbibées sont alors stockées sur un milieu agar/eau à 4 °C durant 3 jours pour synchroniser leur germination. Les graines sont ensuite mises à germer à 14 °C durant une nuit.

Mycorhization

Les plantes ont été cultivées en pots durant 7 semaines dans un milieu d'argile calcinée inoculé avec 1000 spores par litre de *R. irregularis* et arrosé avec une solution de milieu Long Ashton (Hewitt, 1966) carencée en phosphate (7,5 µM). Une partie du système racinaire a été conservée pour le comptage du taux de mycorhization et l'autre partie a été broyée à l'azote liquide et la poudre résultante homogénéisée avant de procéder à l'extraction d'ARN totaux (Figure 16). Trois répétitions biologiques ont été réalisées.

Nodulation

Les banques de petits ARN de racines nodulées que j'ai exploitées ont été générées par A. Niebel avant mon arrivée en thèse.

Pour l'obtention des plantes nodulées, le mutant sunn-sickle de *M. truncatula* a été utilisé (Penmetsa *et al.*, 2003). Ce génotype a été sélectionné car il est plus sensible que la

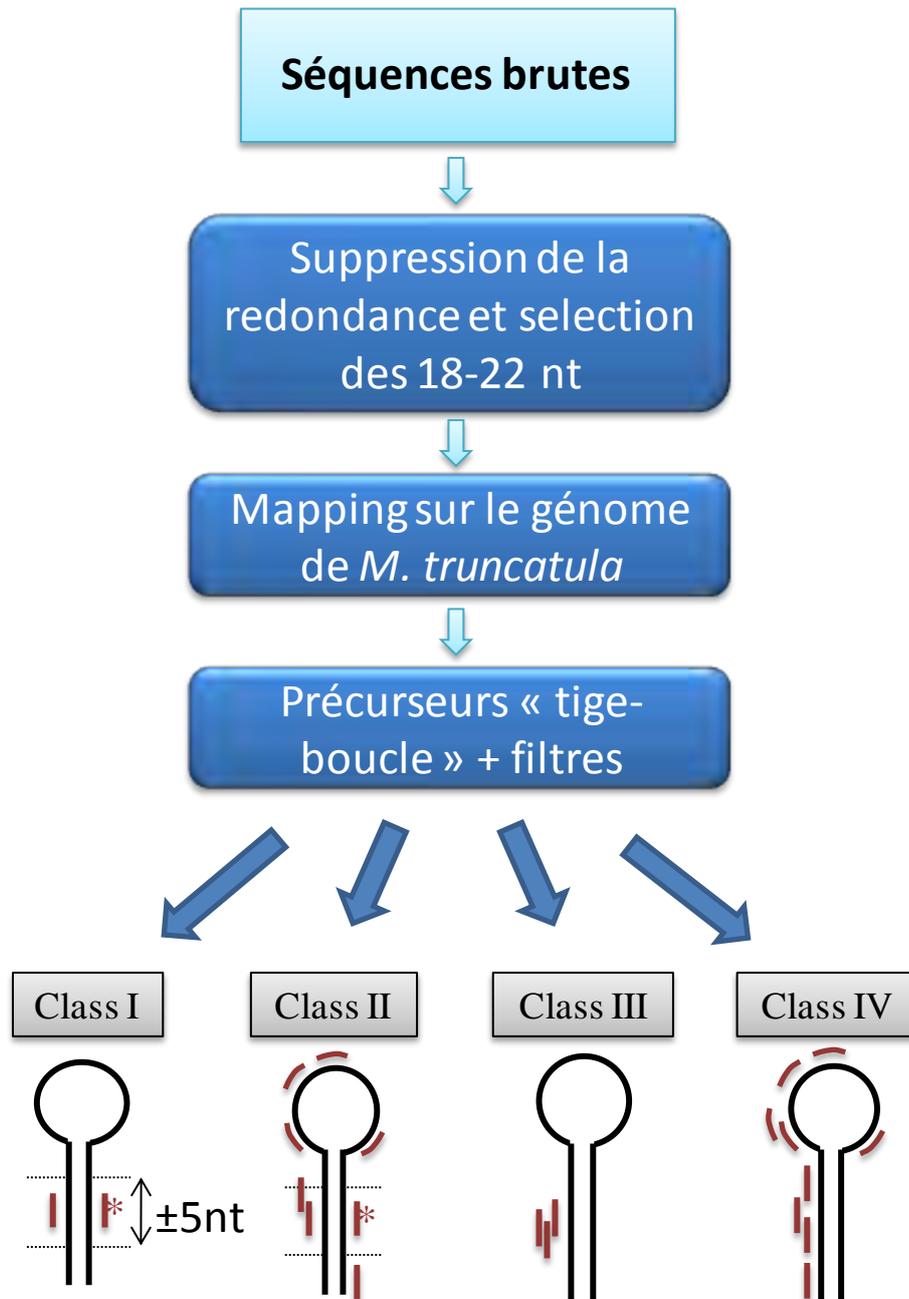


Figure 17: Etapes du pipeline permettant la caractérisation et l'annotation des précurseurs de microARN à partir des séquences brutes. Les tiges-boucles en bas de figure représentent les pre-microARN. Les traits rouges représentent les lectures de séquençage cartographiées sur ces précurseurs.

plante sauvage à l'induction par les facteurs Nod et possède une réponse transcriptomique exacerbée de la bactérie *S. meliloti*, permettant une meilleure détection des modifications d'expression de gènes (A. Niebel, communication personnelle). La culture de ces plantules a été réalisée durant 5 jours dans des caissons permettant une culture en condition aéroponique, à 22 °C, 75 % humidité et avec une photopériode de 16 h. Le milieu (Lullien *et al.* 1987) vaporisé au niveau des racines contient des bactéries *S. meliloti* et est sans azote afin de ne pas inhiber les processus de mise en place de la nodulation. Trois répétitions biologiques ont été réalisées.

Préparation des banques de petits ARN

Seules les parties racinaires des plantes ont été récoltées et broyées à l'azote liquide en fine poudre grâce au broyeur à bille Qiagen Tissuelyser. Par la suite, 100 mg de cette poudre ont été utilisés afin d'en extraire les ARN totaux grâce au kit *mirVana* miRNA isolation kit d'Ambion en suivant les instructions du fabricant. Les ARN des triplicats biologiques de chaque condition ont été regroupés afin de ne produire qu'un seul échantillon. La vérification de la bonne intégrité des ARN a été effectuée grâce au Bioanalyzer 2100 d'Agilent et seuls les échantillons ayant un RIN de plus de 8 ont été sélectionnés pour l'étape suivante. Afin d'obtenir les banques de petits ARN nous avons utilisé le kit Small RNA Sample Prep Kit d'Illumina en suivant les indications du guide « Preparing Samples for Small RNA sequencing Using the Alternative v1.5 Protocol ». Une quantité de 5 µg par échantillon a été utilisée et seuls les petits ARN ayant une taille comprise entre 15 nt et 35 nt ont été sélectionnés. Les banques ont été séquencées par le Génoscope (Evry, France) grâce au séquenceur Solexa HiSEQ d'Illumina. Les quatre banques (Mycorhizée et son témoin ainsi que Nodulée et son témoin) ont été réalisées en simplifiat, *i.e.* à raison d'une banque par condition.

Traitement informatique des séquences

Les séquences brutes obtenues par séquençage ont été nettoyées et traitées selon la méthode décrite par Lelandais-Brière *et al.* (2009) et adaptée aux données Solexa. Les microARN ont été annotés grâce au pipeline LeARN (Noirot *et al.*, 2008) adapté selon les critères établis par la communauté des « miRologues » dans Meyers *et al.* (2008) (Figure 17). Globalement, les adaptateurs des séquences brutes sont éliminés, les séquences correspondant aux ARN ribosomiques également, la redondance est comptabilisée afin de

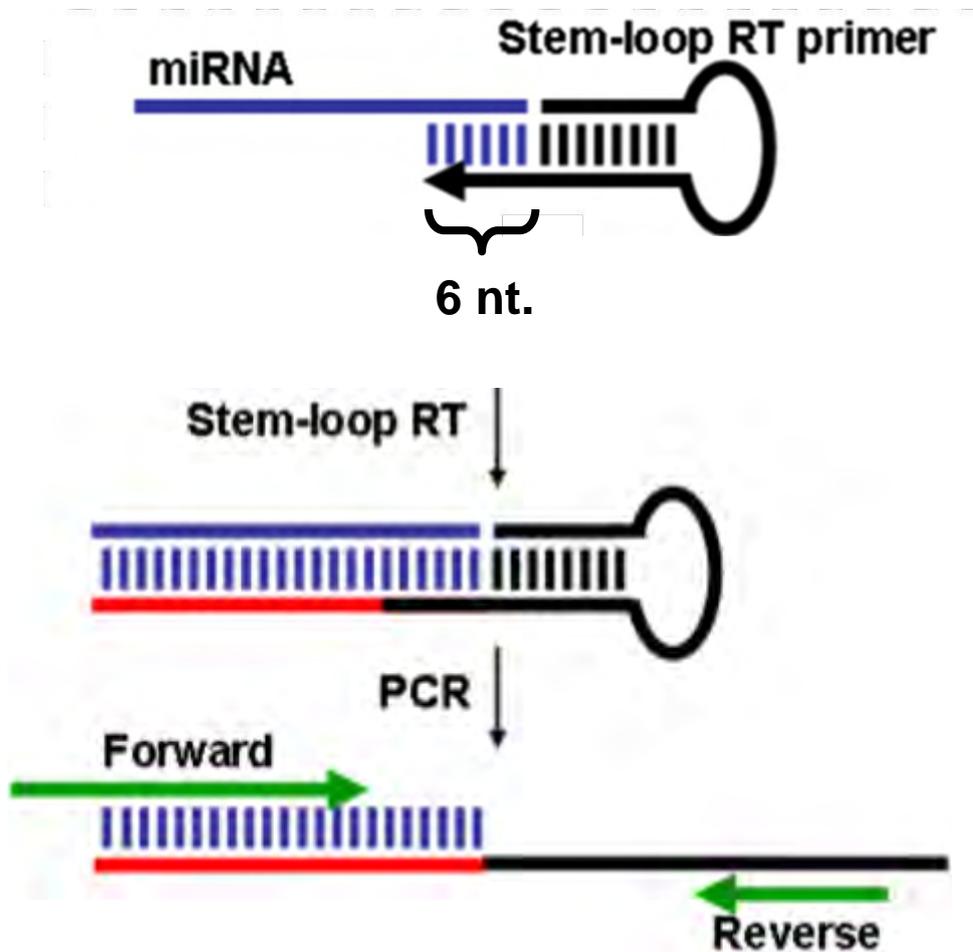


Figure 18: Schéma représentant le protocole de stem-loop qRT-PCR. Les bâtonnets représentent les nucléotides. Les brins bleus représentent le microARN. Noirs: amorce de transcription inverse. Vert: amorces de PCR. nt: nucléotides. D'après Varkonyi-Gasic *et al.* (2007).

quantifier chaque petit ARN, puis supprimée. Les séquences sont mappées sur la version Mt3.5.1 du génome de *Medicago truncatula* et une fenêtre de 400 nt autour de la zone correspondante au microRNA est analysée afin de retrouver une structure de repliement caractéristique des microARN. Les autres petits ARN issus des séquençages sont alors retrouvés sur les tiges boucles et, selon leur distribution (quantité et localisation), les précurseurs prédits sont répartis dans cinq classes selon la robustesse de la prédiction. Les microARN prédits sont alors comparés à la base de données miRbase (Release 18, www.mirbase.org, Griffiths-Jones, 2004) afin de distinguer les microARN connus des microARN nouvellement prédits. Seuls les microARN se retrouvant dans les 5 % des séquences totales brutes les plus exprimées au sein d'une même banque sont sélectionnés. L'ensemble de ces critères m'ont permis de déterminer quels microARN sont caractérisés dans chaque banque et ainsi déterminer ceux présents différemment entre les banques. Les cibles sont ensuite prédites grâce au programme miRanda v3.3a. Suite à la publication du dégradome par Devers *et al.* (2011), l'algorithme Cleaveland utilisé dans cette dernière a également été appliqué pour prédire les cibles des microARN identifiés dans notre étude. L'analyse bioinformatique de ces données, qui constituent un catalogue d'une grande richesse, a été réalisée par Erika Sallet et Jérôme Gouzy (LIPM) et également mis en forme sous interface web pour une exploitation par les membres du projet MIRMED avant ouverture à la communauté scientifique après publication.

Quantification des microARN par northern blot

Les northern blot ont été réalisés selon le protocole publié en annexe (3) avec environ 10 µg d'ARN totaux par condition. Afin de normaliser les expressions par échantillon, une sonde marquée radioactivement ciblant le petit ARN splicéosomal U6, dont l'expression ne varie pas, a été hybridée sur les hauts poids moléculaires des membranes. Des sondes ciblant les microARN miR171h, miR5214 et le microARN E8W526M04IJ5Y8 ont également été produites et hybridées sur les membranes.

Validation de l'expression de microARN candidats par la technique de Stem-loop qRT-PCR

Afin de valider l'expression des microARN obtenues grâce aux séquençages des banques de petits ARN, certains microRNA ont été sélectionnés selon leur profil

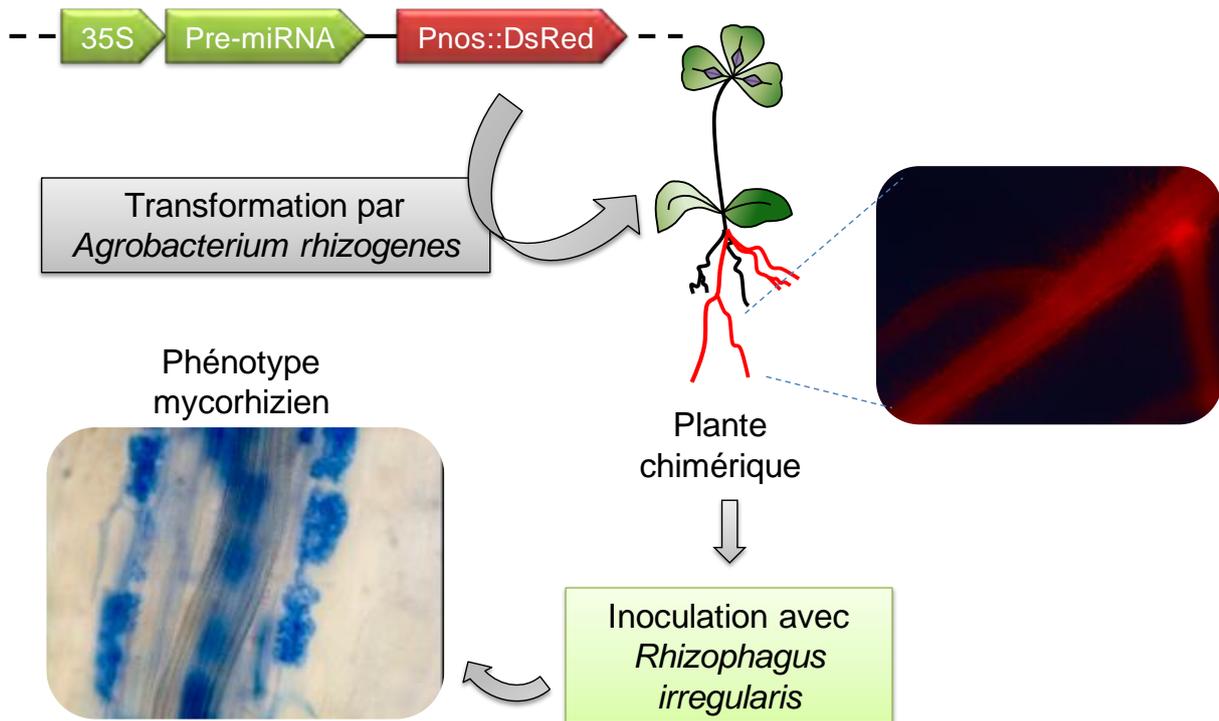


Figure 19: Stratégie de criblage par le phénotypage mycorhizien des transformants sur-exprimant le microARN d'intérêt. D'après Boisson-Dernier *et al.* (2001).

d'expression entre les banques et selon leur néo-identification, ou non, grâce à notre analyse.

Le protocole de quantification des microARN matures par stem-loop qRT-PCR est adapté de Varkonyi-Gasic *et al.* (2007) et inspiré de Chen *et al.* (2005) (Figure 18). Comparée à la traditionnelle technique de quantification qu'est le northern blot, la stem-loop qRT-PCR requiert moins de matrice (de 500 ng à 2,5 µg, contre 5 µg à 15 µg pour le northern blot), donc moins de matériel végétal, ne nécessite pas de radioactivité et permet une quantification de plusieurs candidats (jusqu'à 10) en une seule réaction. Cette technique consiste en une transcription inverse du microARN mature grâce à une amorce particulière s'hybridant sur 6 nucléotides en 3' et se repliant sur elle-même afin d'orienter la transcription (Figure 18). La conformation particulière de l'amorce de reverse transcription nécessite l'utilisation de la Superscript III d'Invitrogen pour une meilleure efficacité et une stabilité de la structure à haute température. Une fois la transcription inverse effectuée, l'amorce stem-loop se déploie et permet la fixation d'un primer universel en 3' pour l'amplification en PCR quantitative. La spécificité d'amplification se fait grâce au primer 5' qui se fixe sur la totalité du microARN (Figure 18).

Obtention de transformants pour l'analyse fonctionnelle des microARN

Des microARN d'intérêts issus de l'analyse des banques ont été sélectionnés pour observer leur incidence sur l'établissement de la symbiose MA par sur-expression. Pour ce faire, les séquences des précurseurs de microARN ont été introduites dans un vecteur de surexpression pPEX sous contrôle d'un promoteur 35S, associés à un gène rapporteur codant pour la protéine DsRed, sous contrôle d'un promoteur PNos (J.P. Combier, com. pers.). Afin d'obtenir rapidement des validations, des transformants chimériques de *Medicago truncatula* ont été obtenus (seule la partie racinaire est transformée) (Boisson-Dernier *et al.*, 2001). Les racines transformées infectées par la bactérie *Agrobacterium rhizogenes* sont sélectionnées sous loupe binoculaire en détectant la protéine fluorescente DsRed. Les plantes possédant des racines infectées par la bactérie, sur-exprimant alors le microARN, ont été transférées en pot afin d'être inoculées par le champignon (selon le protocole décrit dans la section « Mycorhization » de la partie « Préparation du matériel végétal » de ce « matériels et méthodes ») pour en définir le phénotype mycorhizien des plantes (Figure 19).

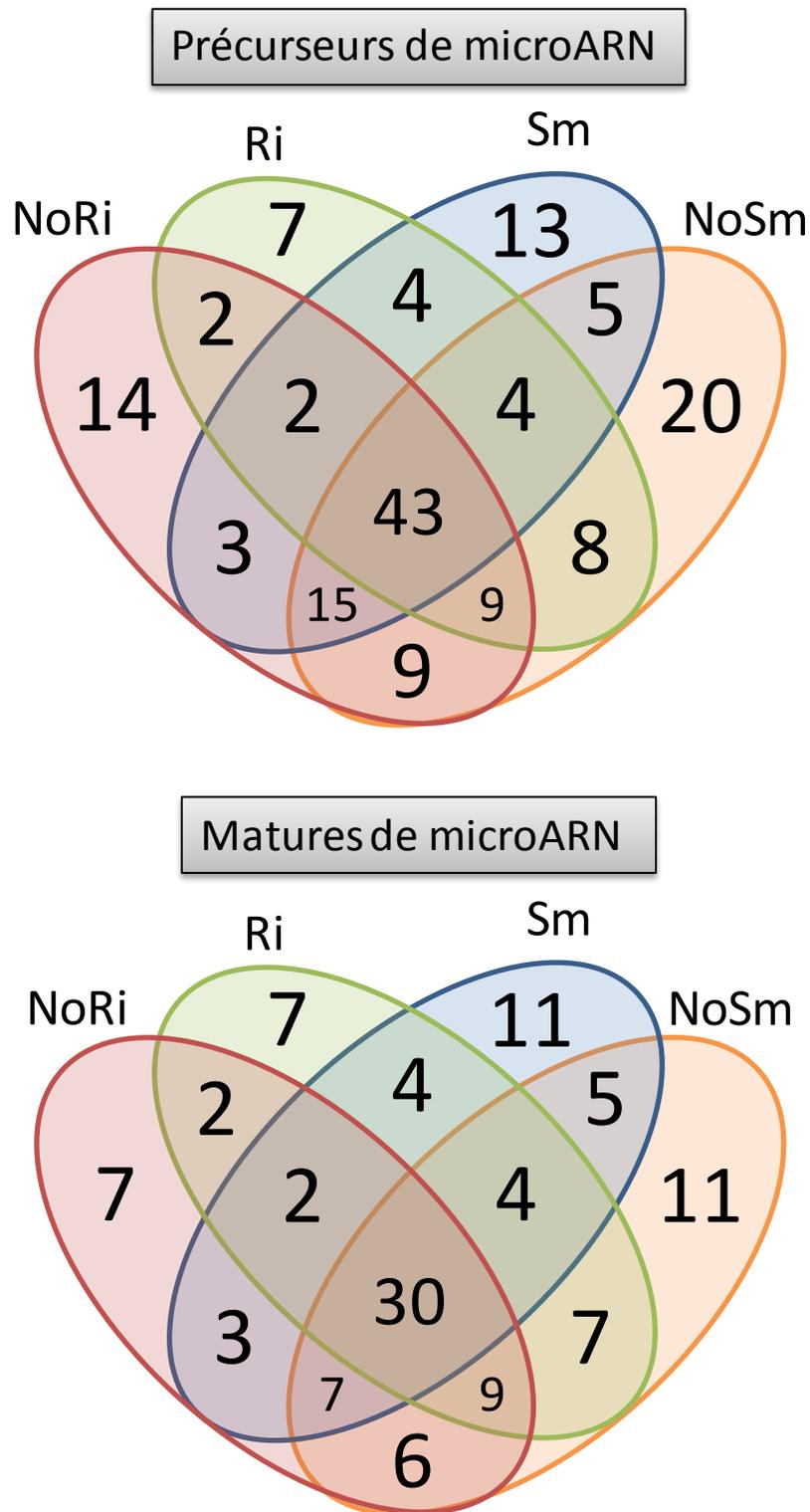


Figure 20: Diagrammes de Venn représentant le nombre de microARN caractérisés par banque et partagés par les banques. NoRi : non mycorhizée ; Ri : mycorhizée ; NoSm : non nodulée et Sm : nodulée.

Afin d'analyser spatio-temporellement l'expression des microARN miRAW et miRBQ, des transformants chimériques de *Medicago truncatula* exprimant un gène rapporteur GUS sous contrôle des promoteurs de microARN ont été produits en suivant le protocole précédent. Les racines des transformants obtenus ont été traitées avec des Myc-LCO ou des facteurs Nod à 10^{-8} M, selon la méthode décrite dans la section « Mise au point des conditions expérimentales pour la réalisation des analyses microtranscriptomiques » (page 159).

Résultats et discussion

Séquençage HiSeq des banques de petits ARN

Grâce à l'élaboration de ces banques, 20 224 799 séquences ont été obtenues, 12 666 107 après le nettoyage des adaptateurs et la suppression des séquences correspondant aux ARNt et ARNr, et enfin 7 448 464 à la suite de la suppression de la redondance. Après confrontation avec le génome de *M. truncatula* et confirmation du repliement des zones adjacentes du microARN en tige-boucle, 158 précurseurs de microARN de classe I et II ont été identifiés, générant potentiellement 98 microARN matures. Parmi ces microARN, 41 sont référencés dans miRBase 18 (dont 28 chez *M. truncatula*), donc déjà identifiés dans des études précédentes, et 57 sont des microARN nouvellement identifiés (Annexe 4) dont 7 ont des cibles validées par le dégradome réalisé par Devers *et al.* (2011). Parmi les quatre banques séquencées, 7 précurseurs de microARN ont été identifiés par notre pipeline seulement grâce aux séquences de la banque issue des racines de *M. truncatula* mycorhizées, générant 7 microARN matures, 13 identifiés seulement dans la banque issue des plantes nodulées, générant 11 microARN matures, 20 identifiés seulement dans la banque témoin de la banque issue des plantes nodulées, générant 11 microARN matures, et 14 identifiés seulement dans la banque témoin de la banque issue des plantes mycorhizées, générant 7 microARN matures (Figure 20).

La caractérisation spécifique des microARN dans chaque banque permet de faire un comparatif de cette caractérisation entre les banques inoculées et leur témoin correspondant. Grâce à cela nous avons pu voir que 22 microARN sont présents dans la banque mycorhization et absent du témoin et 23 microARN qui ont le profil inverse. Dans la banque nodulation, nous retrouvons 20 microARN absents chez le témoin et 33 microARN qui ont le profil inverse (Figure 20).

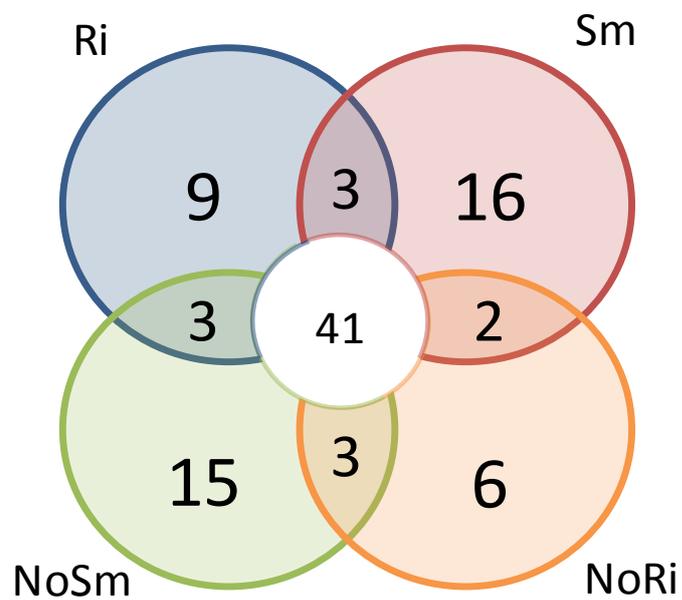


Figure 21: Diagramme de Venn représentant le nombre de microARN matures sous-exprimés et sur-exprimés significativement avec un ratio minimum de 2, selon les inductions. NoRi : non mycorhizée ; Ri : mycorhizée ; NoSm : non nodulée et Sm : nodulée.

Concernant l'expression de ces microARN matures (Annexe 5), 57 sont différentiellement représentés plus ou moins deux fois dans les banques inoculées comparées aux banques témoins correspondantes : en condition « mycorhize » 9 microARN sont sur-exprimés et 6 sous-exprimés, en condition « nodulation » 16 microARN sont sur-exprimés et 15 sous-exprimés (Figure 21). Si on compare les conditions symbiotiques entre elles : 3 microARN sont surexprimés et 3 sous-exprimés dans les deux symbioses, 2 sont sur-exprimés lors de la nodulation mais réprimés lors de la mycorhization, et 3 sont sur-exprimés lors de la mycorhization mais réprimés lors de la nodulation.

Quantification des microARN candidats dans les banques de symbiose MA tardive

Les cinq microARN candidats suivants ont été sélectionnés en raison de leur profil d'expression entre les banques, leur néo-identification, et l'intérêt de leur cible : miR171b, miR171h, miR166, miRAW et miRBQ. Le microARN miR171b cible le facteur de transcription SCARECROW (Llave *et al.*, 2002). Il est globalement peu représenté mais présente une accumulation plus importante dans les racines mycorhizées (Myc : 34 ; T : 9). Le miR166 est le microARN le plus représenté dans les banques de mycorhization (Myc : 184104 ; T : 274823) et cible un facteur de transcription HD-Zip III qui a précédemment été caractérisé comme jouant un rôle dans la formation des racines latérales et le développement des nodules (Boualem *et al.*, 2008). Le miR166 présente une accumulation moins importante lors de la mycorhization. Le miR171h est représenté dans les banques de manière intermédiaire aux deux précédents candidats (Myc : 2735 ; T : 2217) et cible un facteur de transcription GRAS *MtNsp2* qui est essentiel à la signalisation durant le développement des nodules racinaires (Oldroyd & Long, 2003). Des travaux complémentaires réalisés dans l'équipe, auxquelles j'ai participé, ont montré que le miR171h est aussi impliqué dans les processus permettant un bon établissement de la colonisation mycorhizienne (Lauressergues *et al.*, 2012). Les miRAW et BQ sont des microARN qui ont été identifiés dans notre étude, sont très peu représentés dans les banques mycorhizées (Myc : 2 ; T : 5 et Myc : 0 ; T : 5) et ciblent respectivement un gène codant pour une protéine Fbox et une kinase calmoduline dépendante.

Ces banques de petits ARN représentant les étapes tardives des symbioses MA et nodulaires ont été produites en simplifiat. Ce manque de répétition a été compensé grâce au regroupement de plusieurs échantillons de racines obtenus à des intervalles de temps

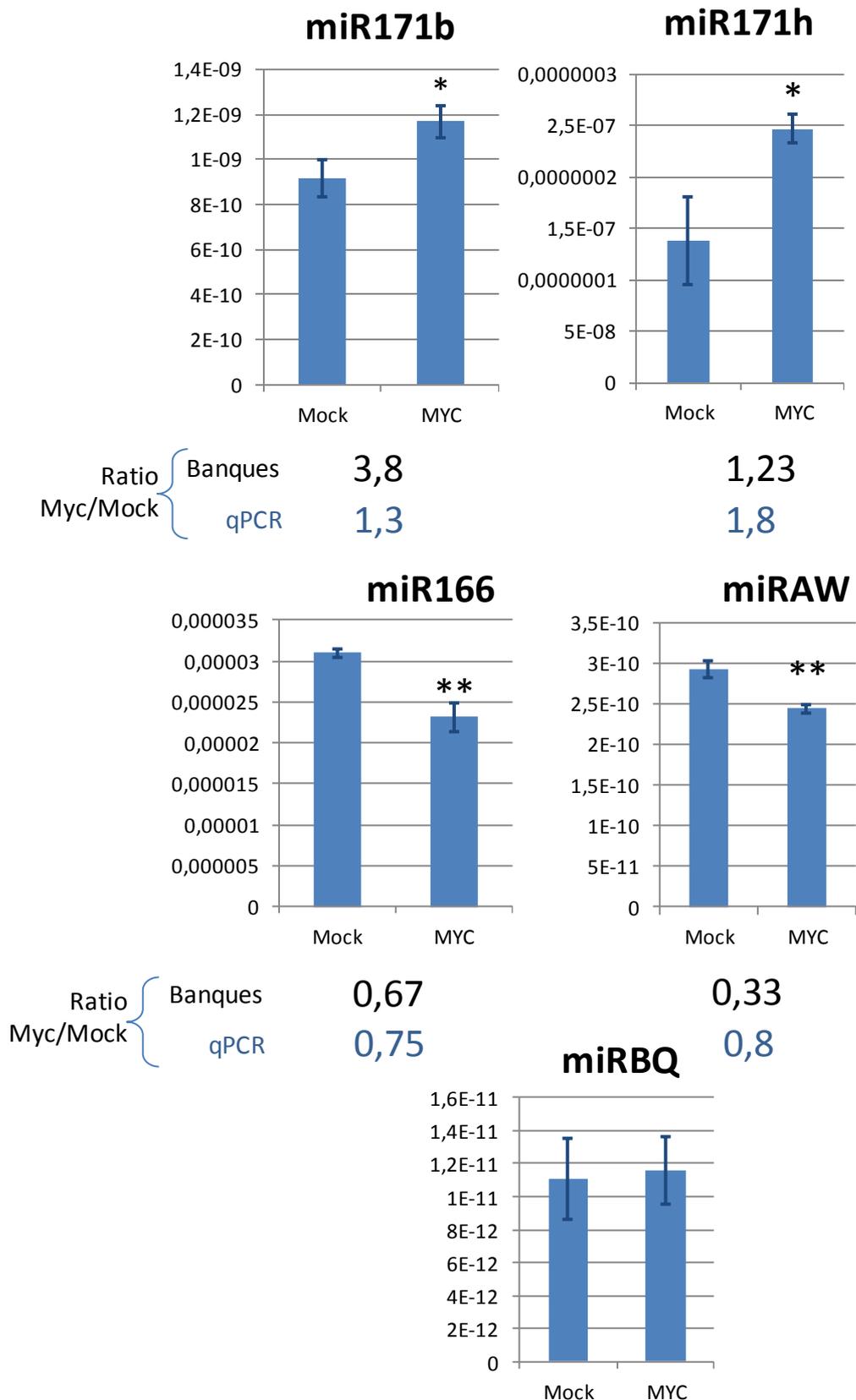


Figure 22: Histogrammes représentant l'expression normalisée de plusieurs microARN matures dans les conditions mycorhizées (MYC) et non mycorhizées (Mock) quantifiée par stem-loop qRT-PCR. Les ratios des données d'expression provenant des banques et des analyses stem-loop qRT-PCR sont indiqués sous les histogrammes.

différents. La variabilité habituellement présente entre des répétitions biologiques est donc intégrée, avec ce protocole, mais n'a pu être mesurée. La sélection des microARN présentant des profils d'expression différents entre les banques est donc plus risquée. La validation de l'accumulation des microARN par stem-loop qRT-PCR sur les différentes répétitions biologiques me permet donc d'accéder à cette variabilité et confirmer le différentiel d'expression des microARN ciblés.

Grâce à la stem-loop qRT-PCR, j'ai pu valider la tendance du profil d'expression de 4 microARN sur les 5 testés (Figure 22). Seul le miRBQ n'a pu être détecté dans nos échantillons. Les microARN AW et BQ sont très peu exprimés dans les banques et, la quantification impossible du miRBQ comparée à la détection du miRAW par stem-loop qRT-PCR, montre que ces microARN sont à la limite de détection par cette technique et/ou par le séquençage HiSeq. Bien que la tendance du profil d'expression soit conservée, les ratios d'expression entre les banques mycorhizées et témoins ne sont pas les mêmes selon la technique utilisée. Les deux méthodes ont chacune leurs biais propres. Le séquençage n'est pas ciblé mais, du fait de la limite de profondeur de séquençage et de la richesse relative de chaque microRNA, il n'est pas exhaustif. Certains microARN dans certaines banques ne sont donc peut-être pas représentés alors que dans une autre banque, avec une profondeur de séquençage différente, ils le seront. Ce biais de représentativité au sein des différentes banques est susceptible d'entraîner des biais dans la quantification relative ou absolue, et donc dans la révélation de faux-positifs et faux-négatifs. Ce biais pourrait être compensé par une normalisation réalisée grâce à des témoins internes « spike-in », équimolaires dans les banques, présents à différentes abondances. En ce qui concerne la stem-loop qRT-PCR, la normalisation de la quantité d'ARN se fait grâce aux transcrits de haut poids moléculaire (Gènes de ménage : *e.g.* Ubiquitine, EF1a) comme lors d'une PCR quantitative classique, rendant ainsi possible la normalisation par échantillon pour chaque expérimentation. La spécificité de la stem-loop qRT-PCR rend cette technique exhaustive car chacun des microARN matures présents dans l'échantillon va être comptabilisé et la comparaison entre banque devient alors réalisable. Bien que cette spécificité soit un atout, elle peut générer des biais lorsque le microARN ciblé est un membre d'une famille (maximum 3 nt d'écart entre les membres) où lorsque certains membres sont très proches (une base de différence au niveau du cœur du microARN) rendant la quantification aspécifique.

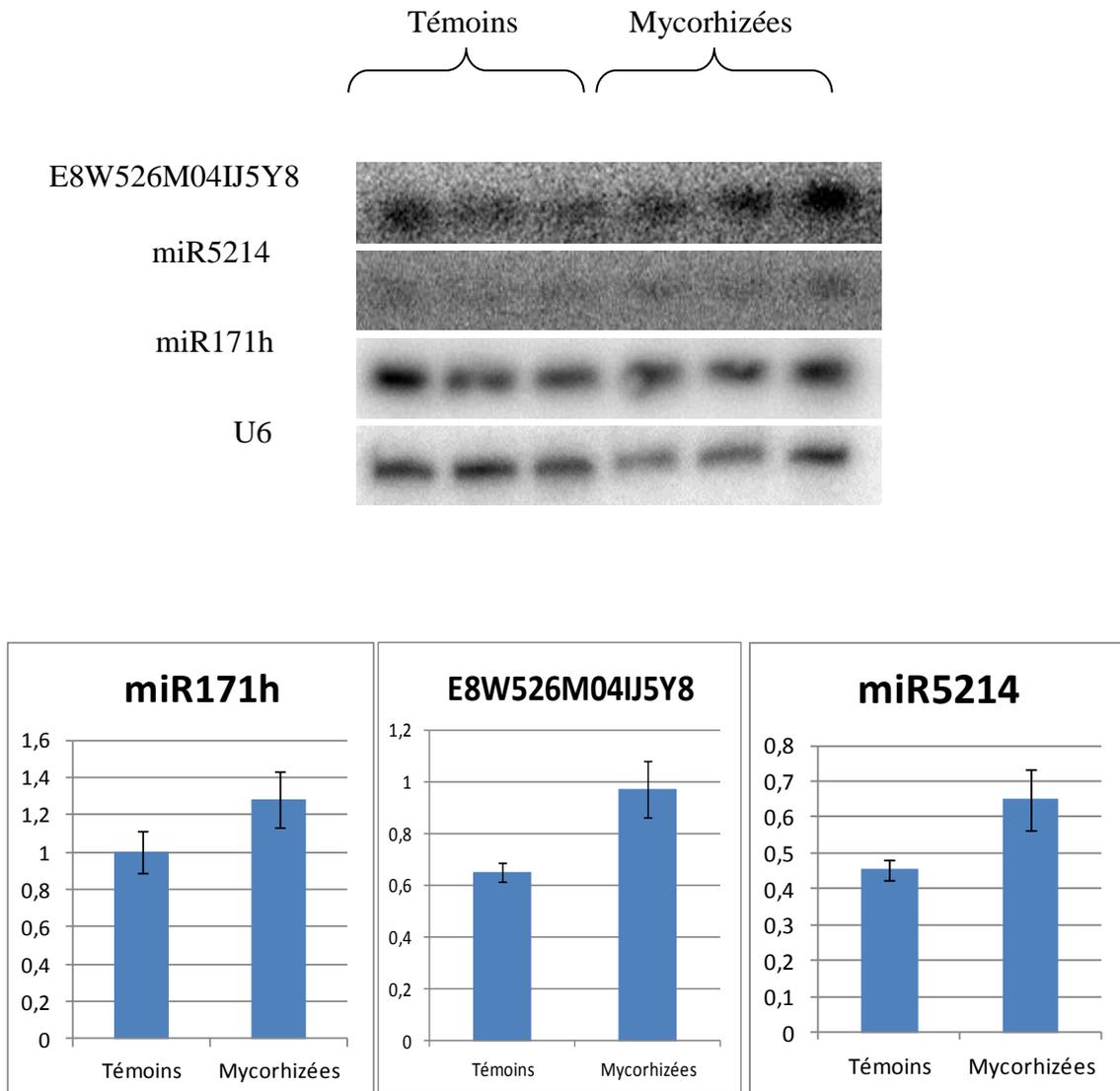


Figure 23: Northern blot réalisés avec les sondes ciblant le miR171h, le miR5214, le microARN E8W526M04IJ5Y8 et le snRNA U6 sur les ARN totaux de triplicats biologiques témoins et mycorhizés. Les histogrammes représentent les ratios obtenus grâce à la numérisation des signaux. Les barres d'erreur représentent les erreurs standards à la moyenne.

Afin de valider les résultats obtenus grâce à ces deux techniques innovantes, des Northern blot, qui représentent une technique plus traditionnelle, plus robuste et reconnue par la communauté scientifique, ont été réalisés. Cette technique a été utilisée pour analyser l'accumulation des miR171h, miR5214 et le microARN E8W526M04IJ5Y8 sur les échantillons mycorhizés. La tendance des profils d'expression de ces trois microARN a été confirmée (Figure 23). Les quantifications par Hiseq et stem-loop qRT-PCR sont donc solides et gagnent en robustesse avec l'augmentation du nombre de répétitions biologiques utilisées et lorsque l'abondance des microARN étudiés est très importante.

Phénotypage du caractère mycorhizien de plantes sur-exprimant des microARN d'intérêt

Une première étape de validation fonctionnelle des microARN candidats a été réalisée par sur-expression. Les plantes sur-exprimant le miR166 n'ont pas présenté de phénotype dans la colonisation des racines par le champignon. Le manque de phénotype particulier par ce type d'approche n'est pas une preuve de la non-implication du microARN dans le processus étudié. Afin de vérifier le bon fonctionnement de la sur-expression, l'accumulation des précurseurs du miR166 a été confirmée. Cependant, on peut imaginer que le précurseur ne produit pas le mature correspondant, dont l'action est alors inexistante. L'absence de phénotype racinaire tend vers cette hypothèse. En effet, comme décrit dans Boualem *et al.* (2008) le miR166 joue un rôle dans l'architecture racinaire et sa sur-expression entraîne une diminution du nombre de racines latérales. Les transformants sur-exprimant ce microARN ne possèdent pas moins de racines latérales, l'action du microARN mature n'a probablement pas eu lieu.

A contrario, les plantes transformées sur-exprimant les microARN miRAW ou miRBQ ont un phénotype particulier de colonisation par le champignon (Figure 24). Les plantes sur-exprimant le miRAW ne développent pas d'arbuscules et seuls des hyphopodia ont été retrouvés à la surface de l'épiderme suggérant un blocage de la formation du PPA (Appareil de Pré-Pénétration) ou des hyphes intra-racinaires. Le miRAW a pour cible prédite une protéine kinase dépendante du Ca^{2+} (Cpk) qui est connue pour jouer un rôle dans des cascades de signalisation. La sur-expression du miRAW entrainerait une inhibition de la production de la protéine, par le biais de la répression des transcrits, et empêcherait le bon déroulement des cascades de signalisation qui régissent la pénétration des hyphes du champignon à l'intérieur de la racine. Chez le riz, le gène *OsCPK18*, codant

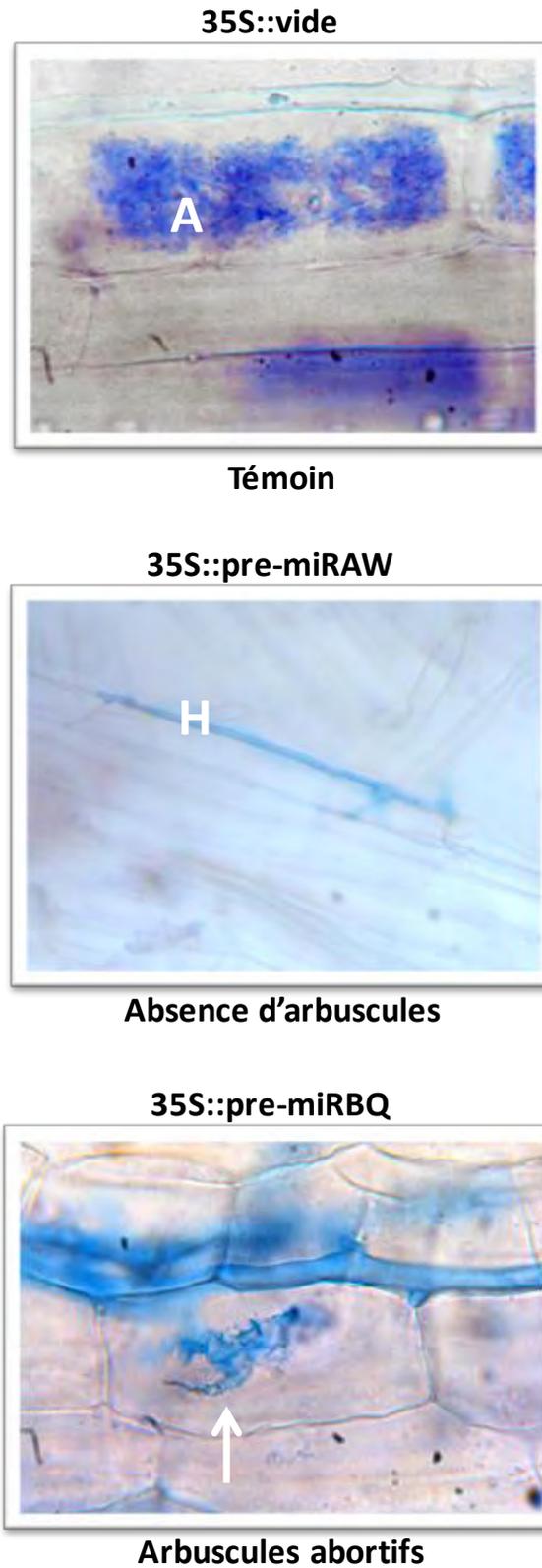
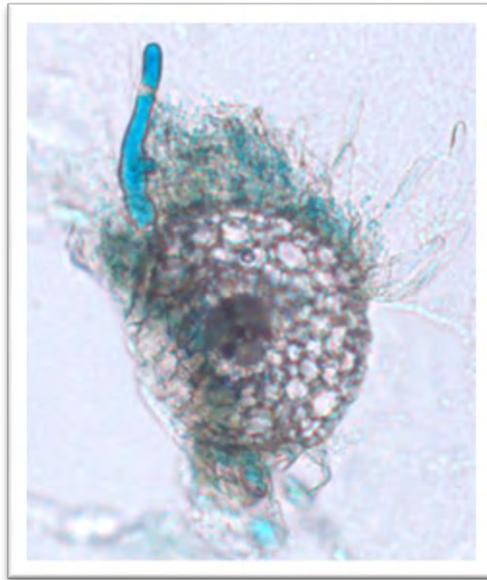


Figure 24: Phénotypes mycorhiziens des constructions sur-exprimant les microARN miRAW et miRBQ. Le champignon est coloré à l'encre bleue. A : arbuscule. H : hyphopode. Photos JP Combier.

pour une protéine kinase dépendante du Ca^{2+} , est surexprimé dans les cellules épidermiques lors de la colonisation par les champignons endomycorhiziens à arbuscules afin de préparer la plante à la pénétration du champignon (Campos-Soriano *et al.*, 2011b). Si cette surexpression n'intervient pas à cause de l'inhibition par le microARN, la pénétration des hyphes du champignon est alors impossible. Dans la banque de petits ARN correspondant aux échantillons mycorhizés, le miRAW est moins représenté que dans le témoin suggérant, par extrapolation, une augmentation des transcrits de sa cible potentielle. La Cpk correspondante serait alors sur-exprimée et permettrait la cascade de signalisation nécessaire à l'entrée du champignon dans la racine. Malgré le rôle de sa potentielle cible, le miRAW n'est pas différentiellement exprimé lors de l'induction par les Myc-LCO, que ce soit dans les banques de petits ARN correspondantes ou grâce à la construction rapportrice GUS. Le miRAW est probablement régulé par d'autres signaux directement fongiques ou venant de la plante hôte générés lors de la mise en place des hyphopodia.

Le sur-expresseur miRBQ a lui aussi été analysé dans son phénotype mycorhizien. Ce sur-expresseur possède aussi un phénotype particulier et, comparé à la colonisation du sur-expresseur miRAW, le champignon pénètre dans la racine et va jusqu'à développer, très sporadiquement, des arbuscules abortifs (Figure 24). Le miRBQ cible une protéine Fbox qui appartient à une famille de protéines connues pour jouer un rôle dans la transduction des signaux, notamment hormonaux (Yu *et al.*, 2007). Elles sont largement régulées par des microARN (Jones-Rhoades *et al.*, 2006). Chez la tomate, il a été montré que le bon développement et la fonctionnalité des arbuscules est régulé par l'acide abscissique (Herrera-Medina *et al.*, 2007) qui est connu pour être régulé par une protéine Fbox (Koops *et al.*, 2011). On peut imaginer une régulation faisant intervenir ces trois éléments: Le miRBQ régule cette protéine Fbox, qui régule elle-même une hormone, telle que l'acide abscissique, qui régit alors le bon développement des arbuscules. La sous-expression du microARN lors de la mycorhization entraînerait une augmentation de l'expression de la protéine Fbox qui permettrait une plus grande concentration d'acide abscissique et donc une meilleure colonisation par le champignon. Afin d'étudier l'expression spatio-temporelle du miRBQ, des transformants portant une construction rapportrice GUS soumise à la régulation par le promoteur du miRBQ a été réalisée. Lors de l'induction par les Myc-LCO, le miRBQ semble s'exprimer exclusivement dans les poils absorbants de la racine (Figure 25) suggérant un rôle de régulation de la Fbox cible.

A.



B.



Figure 25: Localisation de l'expression du miRBQ grâce à une construction pro-miRBQ::GUS dans des racines de *Medicago truncatula* traitées aux Myc-LCO. A. Coupe transversale. B. Racine entière.

Les Fbox pouvant également jouer un rôle dans la transduction du signal, on peut imaginer que la Fbox cible du miRBQ permet la transduction du signal généré par les Myc-LCO perçus par les poils absorbants. On peut faire l'hypothèse qu'en absence de cette Fbox, dans le cas de la sur-expression, la transduction du signal ne se ferait pas complètement et mènerait au mauvais développement des structures fongiques et notamment des arbuscules. Aucune coloration GUS n'est révélée après induction de ces transformants par les facteurs Nod. L'expression du miRBQ n'est donc pas induite par les facteurs Nod et la régulation qui en découle semble s'appliquer à l'établissement de la symbiose MA et non rhizobienne. La Fbox ciblée par le miRBQ est probablement une protéine spécifique de la symbiose MA, comparée à la symbiose rhizobienne, et constitue potentiellement une des clés dans la distinction des facteurs diffusibles fongiques et bactériens par la plante. Ces résultats ne sont que les prémisses d'une nouvelle étude qui devrait porter sur la validation de la cible de ce microARN par des approches telles que la RACE-PCR (Rapid Amplification of cDNA-Ends) ainsi que sur l'expression et l'implication de cette cible dans la symbiose MA.

Conclusion

Au moment de l'obtention des résultats générés par les banques de mycorhization tardive, une étude comparant le catalogue des microARN entre une condition mycorhizée et son témoin non mycorhizé chez *Medicago truncatula*, associée à l'étude du dégradome correspondant, a été publiée (Devers *et al.*, 2011). Les données obtenues par cette approche ont été générées par la même technologie de séquençage que celle utilisée pour l'élaboration de nos banques en condition mycorhizée mais le pipeline informatique permettant l'identification des microARN (miRDeep *vs.* LeARN pour nos banques) ainsi que la version du génome de *Medicago truncatula* (3.5.1 pour nos banques) étaient différents. Nous avons donc pu découvrir 57 microARN matures nouveaux en plus de ceux retrouvés dans l'étude de Devers *et al.* (2011), et non identifiés auparavant dans d'autres études, dont 16 sont différentiellement exprimés lors de la mycorhization (6 sous-exprimés et 10 sur-exprimés). Ces microARN sont autant de candidats potentiels pour une analyse fonctionnelle complète de leur implication dans la symbiose MA.

B - EFFET DES MYC-LCO ET DES FACTEURS NOD SUR LE MICROTRANSCRIPTOME DE MEDICAGO TRUNCATULA

1) Mise au point des conditions expérimentales pour la réalisation des analyses microtranscriptomiques

Lors de l'initiation de mes travaux de thèse, les signaux symbiotiques fongiques et les gènes de *M. truncatula* régulés dans les étapes précoces de la symbiose MA n'étaient pas identifiés. En parallèle de mon projet, des travaux portant sur la caractérisation de composés fongiques diffusibles agissant sur la mise en place de la symbiose, les Myc-LCO, étaient initiés dans une collaboration entre notre équipe de recherche (V. Puech, G. Bécard), l'équipe de Jean Dénarié (LIPM - INRA/CNRS) et l'équipe de chimiste de Véréna Poinot (IMRCP - UMR 5623 - CNRS/UPS). Ma participation à ces travaux m'ont permis d'accéder à ces Myc-LCO et d'initier des inductions de plantules de *M. truncatula* par ces composés afin de générer des banques de petits ARN de cette plante dans ces conditions. Les analyses microtranscriptomiques qui ont été réalisées au cours de cette thèse sont des expérimentations onéreuses qui ont nécessité un matériel végétal précieux. La mise au point des conditions expérimentales en amont de ces expérimentations a donc été primordiale. La caractérisation des Myc-LCO et leur implication dans l'établissement de la symbiose MA ont nécessité la vérification de l'induction de certains gènes marqueurs de la symbiose MA par ces composés, et ce, d'une manière voie-DMI-dépendante. Mes essais ont donc eu pour double avantage 1) la validation du dispositif expérimental et de la bonne induction de mon matériel végétal, et 2) la validation de l'effet inducteur de gènes marqueurs de la symbiose MA par les Myc-LCO grâce à des approches transcriptomiques. Mon étude est partie intégrante de l'article Maillet *et al.* (2011).

Matériel et méthode

Préparation du matériel végétal

Afin d'obtenir le matériel biologique nécessaire pour les hybridations sur puces à oligonucléotides, le mutant super nodulant de *Medicago truncatula*, *sunn-sickle*, a été utilisé (Penmetsa *et al.*, 2003). Les cinétiques de mycorhization de ce génotype montrent

une colonisation par *R. irregularis* plus rapide (Annexe 6) que celle obtenue chez le sauvage correspondant, reflétant une capacité plus grande à la mycorhization et probablement des réponses transcriptomiques plus importantes, comme attendu (A. Niebel, communication personnelle).

Les prérequis *sine qua non* antérieurement établis dans la littérature par la communauté scientifique (Catoira *et al.*, 2000; Riely *et al.*, 2004; Parniske, 2008), pour qu'un facteur diffusible sécrété par le champignon soit considéré comme une molécule signale, comporte un volet concernant la dépendance à la voie DMI. Afin de vérifier si l'expression différentielle des gènes lors de l'induction par les Myc-LCO dépend de la voie DMI, les expérimentations ont été réalisées en parallèle sur le génotype sauvage A17 et le mutant *dmi3*.

Les plantules ont été cultivées selon la méthode de culture en aéroponie décrite dans la partie « Analyse du microtranscriptome de *Medicago truncatula* lors des étapes tardives de la symbiose MA» (page 139). Après 5 jours de croissance, les plantules sont alors transférées dans des tubes contenant un milieu Farhaeus liquide (Boisson-Dernier *et al.*, 2001), ajusté de la même manière que le milieu utilisé pour la culture en aéroponie, afin d'effectuer les inductions par les différents composés. Ces derniers ont été rajoutés au milieu à une concentration de 10^{-8} M et les plantes ont été récoltées après 4 h et 24 h d'induction. Les facteurs Nod et Myc-LCO sulfaés sont des molécules très proches au niveau structural et, afin de s'affranchir des potentielles régulations croisées qu'il peut y avoir entre ces deux composés, toutes les inductions ont été menées avec des Myc-LCO non sulfaés. Trois répétitions biologiques ont été réalisées.

Préparation des échantillons et hybridations

Seules les parties racinaires des plantules ont été récoltées et broyées à l'azote liquide en fine poudre grâce au broyeur à bille Qiagen Tissuelyser. Par la suite, 100 mg de cette poudre ont été utilisés afin d'en extraire les ARN grâce au kit Qiagen RNeasy plant mini kit, selon les instructions du fabricant. La qualité des ARN requise pour l'hybridation des échantillons sur une puce à oligonucléotides, devant atteindre un RIN (RNA Integrity Number) supérieur à 8, a été contrôlée grâce aux puces Agilent RNA nano sur un Agilent Bioanalyzer 2100. La transcription inverse des ces échantillons a été réalisée grâce au kit Superscript III d'Invitrogen en utilisant des amorces poly T et le protocole du fabricant.

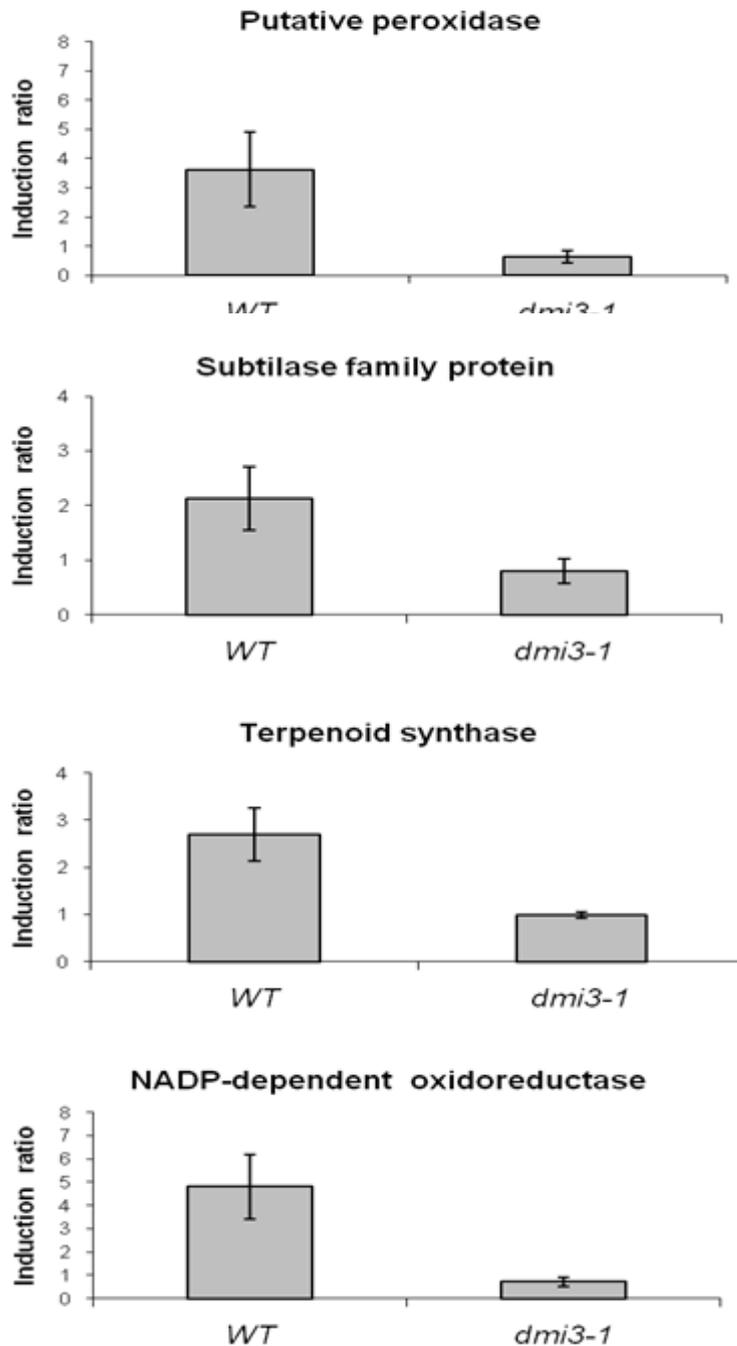


Figure 26: Analyse par qRT-PCR de l'expression de quatre marqueurs de la mycorhization après traitement aux Myc-LCO sur des plantules de *Medicago truncatula* sauvages (WT) et mutant *dmi3-1* (*dmi3-1*). Les histogrammes représentent les moyennes, et les erreurs standards à la moyenne, des ratios des valeurs d'expression des plantes traitées sur celles des plantes non-traitées. (Maillet *et al.*, 2011)

Les hybridations ont été effectuées sur des puces Affymetrix GeneChip *Medicago* Genome Arrays, selon le protocole du fabricant GeneChip 3' IVT Express. Les puces ont été scannées sur l'Affymetrix GeneChip scanner 3000 7G et analysées grâce au logiciel Affymetrix Expression Console.

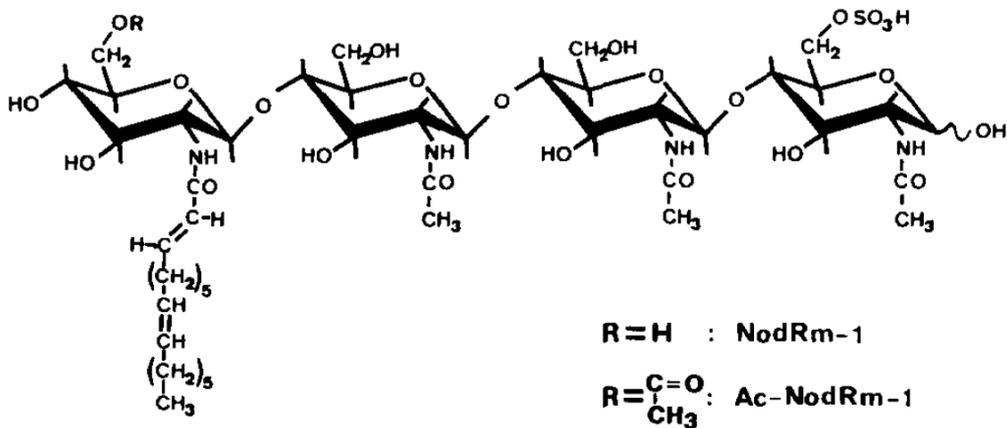
Validation par PCR en temps réel

L'expression des gènes candidats sélectionnés a été validée par PCR quantitative sur le thermocycleur LightCycler LC480 384 puits de Roche dans le mix réactionnel suivant : 4 μ L de SYBR green master I Roche 2X ; 2 μ L du mélange d'amorces 5' et 3' à 0,5 μ M et 2 μ L d'ADNc dilués au 1/25^{ème}. L'efficacité des amorces a été contrôlée pour atteindre un minimum de 93% et les amplifications ont été effectuées en duplicat technique sur les triplicats biologiques. La normalisation des données d'expression a été faite par rapport aux gènes de référence *Ubp* et *Efla* selon la méthode décrite dans Vandesompele et co-auteurs (2002). La quantification relative a été réalisée selon la méthode des $2^{-\Delta\Delta Ct}$.

Résultats et discussions

Parmi les gènes dérégulés lors de l'induction par les Myc-LCO chez le double mutant *sunnsickle*, vingt sont surexprimés significativement après 4 h d'induction et onze après 24 h d'induction (Annexe 7). Dix candidats ayant les plus grands différentiels d'expression entre les échantillons induits et témoins ont été sélectionnés afin de valider leur expression en PCR quantitative sur les génotypes sauvage A17 et mutant *dmi3* (MTGI9-TC135762, MTGI9-TC139394, MTGI9-TC122080, MTGI9-TC125698, MTGI9-TC128211, MTGI9-TC127442, MTGI9-TC139078, MTGI9-TC120449, MTGI9-TC136093 et MTGI9-TC114681). Dans cet échantillonnage de gènes candidats, quatre sont sur-exprimés significativement dans le génotype A17, lors de l'induction par les Myc-LCO comparé aux échantillons non induits, alors que leur expression reste inchangée lors de cette même induction dans le génotype mutant *dmi3* (Figure 26). Cette absence de réponse chez le mutant *dmi3* montre que ces gènes jouent un rôle dépendant de cette voie symbiotique et sont donc potentiellement impliqués dans l'établissement de la symbiose. Etant également induits par les Myc-LCO après 24 h, ces gènes sont des marqueurs de la mise en place de la réponse de la plante aux facteurs diffusibles du champignon. Ces quatre gènes ont été retrouvés dans des précédentes études comme également sur-

A.



B.

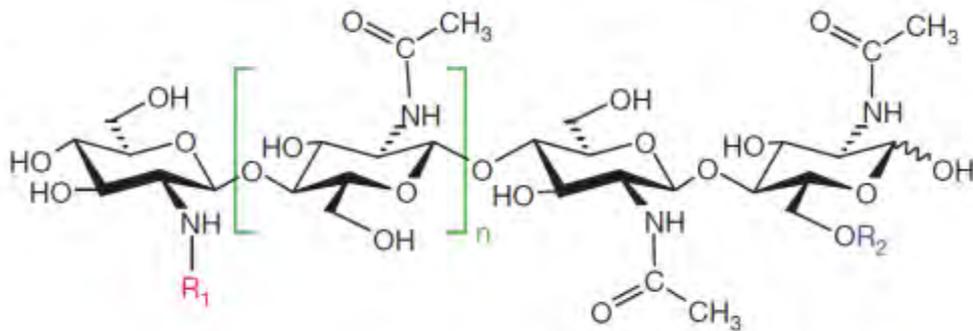


Figure 27: Structure proposée des molécules sécrétées par bactéries nodulantes et les champignons mycorhiziens à arbuscules (MA). A. Les facteurs Nod diffusés lors de la nodulation (Roche *et al.*, 1991). B. Les Myc-LCO diffusés lors de la symbiose MA (Maillet *et al.*, 2011).

exprimés lors des étapes tardives de la mycorhization. Comme le processus de colonisation des racines par les champignons MA est asynchrone, le champignon prépare constamment la plante à accueillir ses structures au sein de la racine. Il est donc normal de retrouver les transcrits de ces gènes sur-accumulés même après le début de la colonisation racinaire du champignon.

Cette analyse a été réalisée afin de vérifier et valider mon approche expérimentale d'induction des plantules de *Medicago truncatula* par les Myc-LCO. Des gènes marqueurs connus comme jouant un rôle dans la mycorhization voient leur expression modifiée lors de cette induction. Cette approche est donc validée me permettant d'initier la production de matériel biologique similaire afin d'observer les variations du microtranscriptome, en parallèle des modifications transcriptomiques déjà observées, induites par les Myc-LCO.

2) Analyse du microtranscriptome de *Medicago truncatula* en réponse aux Myc-LCO et facteurs Nod

Les symbioses MA et nodulaires partagent une voie de régulation commune chez la plante hôte, appelée la voie CSP, pour établir une association fonctionnelle. Les composés diffusés dans le sol par ces symbiontes, qui permettent l'activation de cette voie, sont des lipochitoooligosaccharides (Myc-LCO et facteurs Nod) qui ont des structures très proches (Figure 27). Au commencement de ma thèse, les gènes, notamment les facteurs de transcriptions, qui constituent cette voie de signalisation chez la plante hôte, étaient soupçonnés d'être finement régulés par des microARN lors de la perception des composés diffusibles des symbiontes. Certains des gènes régulés chez la plante hôte seraient donc probablement soumis aux mêmes régulateurs durant les symbioses rhizobienne et MA alors que certains doivent subir une régulation différente durant les deux symbioses afin de permettre une distinction de ces dernières. Récemment, il a été montré que le miR171h régule le facteur de transcription *NSP2*, présent dans la CSP, lors de la mycorhization et l'induction par les Myc-LCO alors que cette régulation est absente lors de la nodulation (Lauressergues *et al.*, 2012). La caractérisation des microARN jouant un rôle dans la perception des Myc-LCO et/ou des facteurs Nod permettrait d'identifier les gènes permettant cette perception ainsi qu'une meilleure compréhension des régulations qui régissent les phénomènes qui en découlent. Afin d'avoir une vision globale de ces microARN, dans le cadre du projet MIRMED, j'ai réalisé un séquençage du

microtranscriptome de *M. truncatula* lors de l'induction par les Myc-LCO, d'une part, et les facteurs Nod, d'autre part.

Matériels et methods

Préparation du matériel végétal

La préparation du matériel végétal a été réalisé de la même manière que dans la partie «Mise au point des conditions expérimentales pour la réalisation des analyses microtranscriptomiques» (page 159), avec trois répétitions biologiques induites par les myc-LCO, trois répétitions biologiques induites par les facteurs Nod, et leurs trois témoins non-induits correspondants. Le génotype de *M. truncatula* utilisé ici est le double mutant super nodulant *sun-sickle* comme expliqué dans la partie « Mise au point des conditions expérimentales pour la réalisation des analyses microtranscriptomiques ».

Préparation des banques de petits ARN

Les banques de petits ARN ont été obtenues selon le protocole précédemment décrit dans la section « Matériels et méthodes » de la partie «Analyse du microtranscriptome de *Medicago truncatula* lors des étapes tardives de la symbiose endomycorhizienne à arbuscules» (page 139). Les trois répétitions biologiques de chaque condition ont été séquencées donnant lieu à neuf banques de petits ARN (3 induites par les Myc-LCO, 3 par les facteurs Nod et 3 témoins). L'expression de certains microARN a été validée grâce à la technique des Northern blot décrite dans la partie «Analyse du microtranscriptome de *M. truncatula* lors des étapes tardives de la symbiose endomycorhizienne à arbuscules».

Résultats et discussion

Grâce à l'élaboration de ces banques, 273 395 064 séquences ont été obtenues, 90 404 032 après le nettoyage des adaptateurs et la suppression des séquences correspondant aux ARNt et ARNr, et enfin 46 049 327 à la suite de la suppression de la redondance. Après confrontation avec le génome de *M. truncatula* et confirmation du repliement des zones adjacentes du microARN en tige-boucle, 358 précurseurs de microARN de classe I et II ont été identifiés, générant potentiellement 228 microARN matures. Parmi ces

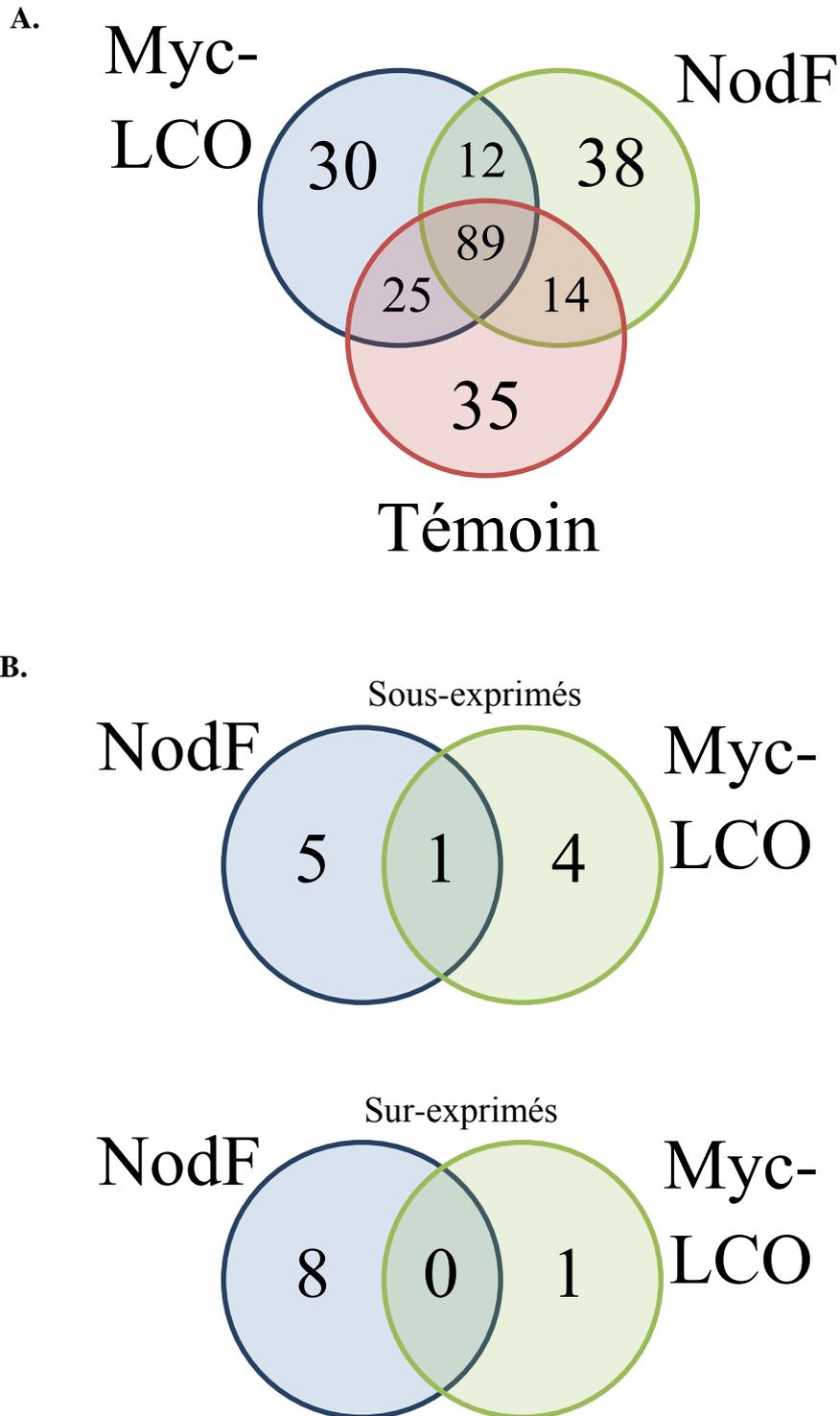


Figure 28: Diagrammes de Venn représentant A. Le nombre des microARN matures trouvés dans les différentes conditions : témoin, traitée au facteur Nod (NodF), traitée aux Myc-LCO (Myc-LCO). B. La répartition des microARN matures sous-exprimés et sur-exprimés significativement avec un ratio minimum de 2 dans les différentes conditions.

microARN, 50 sont référencés dans miRBase 18 (dont 37 chez *M. truncatula*), donc déjà identifiés dans des études précédentes, et 178 sont nouvellement identifiés (Annexe 8) dont 18 ont des cibles qui ont été par la suite confirmées par le dégradome produit par Devers *et al.* (2011). Parmi les neuf banques séquencées, 46 ont été identifiés par notre pipeline seulement grâce aux séquences des trois banques induites par les myc-LCO, générant 30 microARN matures, 57 identifiés seulement dans les banques induites par les facteurs Nod, générant 38 microARN matures, et 48 identifiés seulement dans les banques témoins non induites, générant 35 microARN matures (Figure 28 A.).

Lorsque l'on s'intéresse à la quantification de ces microARN matures (Annexe 9), on peut voir que 19 présentent des moyennes significativement différentes, avec un ratio supérieur à 2 entre les conditions. Cinq sont sous-exprimés et 8 sur-exprimés lors de l'induction aux facteurs Nod, 4 sont sous-exprimés et 1 sur-exprimé lors de l'induction aux Myc-LCO, et 1 est sous-exprimé lors des deux inductions (Figure 28 B.).

Comme démontré plus tard par Maillet *et al.* (2011), le facteur de transcription *MtNsp2* intervient dans la régulation de la perception des Myc-LCO. Et ce facteur de transcription est régulé par le microARN miR171h (Devers *et al.*, 2011) ce qui laisse également présumer une régulation du miR171h et donc une modification de son expression lors de l'induction par les Myc-LCO. Récemment, Laressergues *et al.* (2012) ont confirmé que le miR171h régulait *MtNsp2* et ont montré que ce microARN était sur-exprimé lors de l'induction par les Myc-LCO. Dans nos banques induites grâce aux Myc-LCO, le miR171h ne se retrouve pas différenciellement accumulé entre la condition témoin et la condition induite. Cette absence de différentiel a été confirmée grâce au northern blot (Figure 29). Le miR171h peut être considéré comme témoin de la bonne induction des racines par les Myc-LCO et l'absence de modification de son expression dans nos banques peut être expliquée par plusieurs hypothèses. Premièrement, les champignons endomycorhiziens sécrètent un mélange de deux types de molécules Myc-LCO, des molécules sulfatées et non sulfatées. Les banques ont été réalisées après traitement avec des Myc-LCO non sulfatés, les sulfatés étant trop semblables aux facteurs Nod, pour simplifier les interprétations. Or, les différences d'expression observées par Laressergues *et al.* (2012) ont été obtenues grâce à une induction réalisée avec un mélange de Myc-LCO comprenant les formes sulfatées. D'autre part, les inductions réalisées par Laressergues *et al.* (2012) ont été réalisées sur des génotypes sauvages A17 alors que nos banques ont été réalisées en utilisant le double mutant *sunnsickle*. Il a été

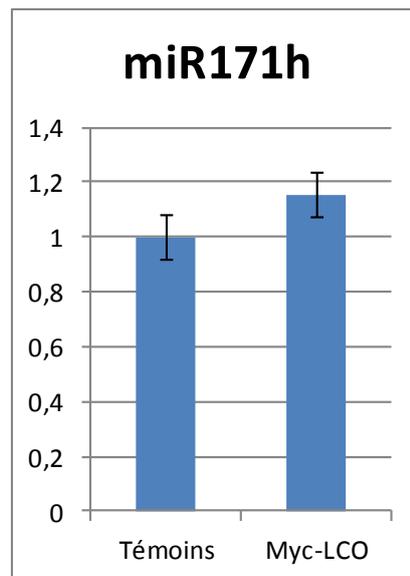
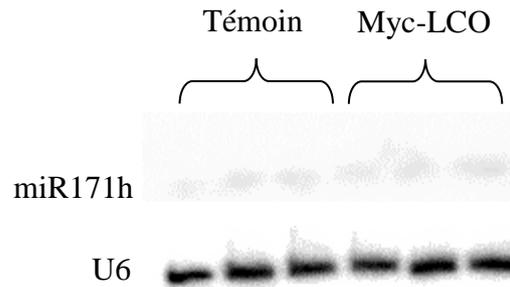


Figure 29: Northern blot réalisés avec les sondes ciblant le miR171h et le snRNA U6 sur les ARN totaux de triplicats biologiques témoins et induits avec les Myc-LCO. Les histogrammes représentent les ratios obtenus grâce à la numérisation des signaux. Les barres d'erreurs représentent les erreurs standards à la moyenne.

montré que ces deux géotypes ont les mêmes profils transcriptomiques lors de la nodulation avec seulement un différentiel d'expression plus fort pour le double mutant que le sauvage (A. Niebel, communication personnelle). Les modifications du microtranscriptome, notamment pour la condition de mycorhization tardive, n'ont pas été testées et il est possible que pour le double mutant *sun-sickle* la réponse se fasse différemment. Enfin, les inductions réalisées par Laressergues *et al.* (2012) ont été effectuées à une concentration de 10^{-7} M alors que nos plantes ayant permis la production des banques ont été induites grâce à des Myc-LCO concentrés à 10^{-8} M.

Nous retrouvons tout de même des microARN différemment exprimés lors de l'induction par les Myc-LCO non sulfatés dans nos conditions d'expérimentation. Parmi ces microARN, le miR172c est un des microARN les plus exprimés dans les banques, ceci facilitant son analyse et augmentant la solidité des données de quantification observées. Ce microARN, sur-exprimé dans les banques traitées aux Myc-LCO, cible le facteur de transcription APETALA2 dont le clivage a été confirmé dans le dégradome, issu de banques mycorhizées, généré par Devers *et al.* (2011). Ce microARN est également sur-exprimé dans les nodules de soja (Subramanian *et al.*, 2008) et de *M. truncatula* (Lelandais-Brière *et al.*, 2009), et lors de la mycorhization chez la tomate (Gu *et al.*, 2010), suggérant un rôle important dans l'établissement de ces symbioses. Lelandais *et al.* (2009) ont également montré une localisation spécifique du miR172 dans la zone d'infection par la bactérie nodulante, suggérant un rôle de ce microARN dans la différenciation cellulaire lors de l'infection par la bactérie. Le facteur de transcription APETALA2, cible du miR172c, est également connu pour jouer un rôle dans le développement floral (Aukerman, 2003) et il est sur-exprimé lors de la carence en phosphate suggérant une inactivation du microARN dans ces conditions (Nilsson *et al.*, 2010). Ces résultats fragmentaires suggèrent que le miR172c joue un rôle dans la différenciation cellulaire en contrôlant le facteur de transcription APETALA2 et que cette différenciation cellulaire est modifiée lors de la mise en place des symbioses rhizobienne et endomycorhizienne. L'étude fonctionnelle du miR172c par des approches d'inactivation et de surexpression permettra de voir son incidence réelle sur la mise en place de la symbiose endomycorhizienne à arbuscules.

Conclusion

Ces microARN constituent le premier catalogue de petits ARN dont la régulation est affectée par des signaux diffusibles produit par un champignon endomycorhizien à arbuscules. Cette liste constitue un premier crible qui permettra par la suite de réaliser des études fonctionnelles sur ces microARN et mieux comprendre leur rôle dans la régulation de l'établissement de la symbiose MA. L'étude des cibles de ces microARN permettra également de disséquer un peu plus en profondeur les gènes impliqués dans la perception des facteurs diffusibles du champignon et mieux comprendre les régulateurs qui régissent le dialogue moléculaire s'établissant entre les deux partenaires.

C - LIMITATION DES APPROCHES DE SEQUENÇAGE MASSIF POUR LA DETECTION ET LA QUANTIFICATION DES MICROARN

Les études de biologie moléculaires actuelles tendent de plus en plus à utiliser des approches massives générant de grandes quantités de résultats traités avec des filtres basés sur des paramètres globaux permettant d'isoler un maximum de candidats tout en minimisant les faux-positifs et les faux-négatifs. L'étude des petits ARN est également rentrée dans cette ère en commençant par les puces à oligonucléotides permettant la quantification ciblée, et avec *a priori*, des petits ARN d'une condition donnée. Plus récemment, les séquençages massifs tel que le pyroséquençage ont permis de développer des approches sans *a priori* permettant la découverte et la quantification des petits ARN. Ces approches, bien que générant des résultats abondants, sont soumises à des biais. Récemment, il a été montré, dans une étude portant sur l'influence de l'utilisation de différents protocoles de production de banques de petits ARN pour le séquençage massif, que selon le protocole et/ou la technologie de séquençage utilisés, les résultats différaient sur la caractérisation des microARN, mais plus encore sur leur quantification (Toedling *et al.*, 2012). En effet, lors de cette étude, des banques réalisées avec des protocoles et technologies identiques, sur des petits ARN issus de lignées cellulaires différentes, comportaient des différences mineures comparées à des banques générées à partir de microARN issus d'une même lignée cellulaire mais grâce à des technologies différentes. Dans notre étude, les banques tardives et précoces ont été produites grâce à la même technologie mais des protocoles légèrement différents (v1.0 vs. v1.5), elles sont donc comparables entre elles au sein d'une étape de la symbiose mais pas entre les étapes.

Lorsque le projet MIRMED débuta, les techniques de séquençage massif, pour l'identification et la quantification des petits ARN, commençaient juste à émerger chez les plantes et le projet représentait alors un travail pionnier dans le domaine. Les biais techniques n'étaient pas tous connus et la profondeur de séquençage annonçait une quantification absolue de la totalité des petits ARN présents (Creighton *et al.*, 2009). Face au coût important de ces techniques, il a donc été décidé de séquencer les banques de petits ARN sans répétitions techniques, en mélangeant les répétitions biologiques de manière équimolaire afin d'intégrer la variabilité biologique. Les protocoles de production des banques passent par des étapes de ligation des adaptateurs, des transcriptions inverses et/ou des amplifications pouvant entraîner des biais (McCormick *et al.*, 2011). Les

adaptateurs ne se fixant vraisemblablement pas de manière homogène sur toutes les séquences selon les caractères physico-chimiques des séquences, cela peut entraîner une représentation différente dans les banques. Lors de la suppression *in silico* des adaptateurs, il est très courant que ces derniers ne soient pas totalement supprimés à cause de modifications de la séquence après amplification et la reconnaissance plus ou moins strict de cette dernière, entraînant ainsi la formation de séquences artéfactuelles (McCormick *et al.*, 2011). Dans les séquences brutes générées par nos banques, ce phénomène est très visible avec les microARN les plus abondants. Des séquences correspondant à un nucléotide près aux microARN les plus abondants se retrouvent très présentes dans les banques. La plupart de ces séquences ne sont pas retrouvées à 100 % d'identité sur le génome, donc éliminées, et sont probablement le produit d'ARN dégradés, d'erreurs systématiques de la polymérase ou de résidus d'adaptateur. Certains microARN font partis de familles très vastes (*e.g.* mtr-miR2592 comportant 63 membres ; la plus grande trouvée chez *Medicago truncatula* dans miRBase 18.0) et chaque membre est éloigné des autres par 3 nucléotides de différences au maximum. Il est donc possible que des séquences produites artéfactuellement par des biais systématiques, ou non, lors de la génération des banques soient 100 % identiques à un membre de la famille d'un microARN très représenté. Les séquences correspondant à ce membre sont donc comptabilisées et le profil d'expression de ce membre est donc modifié. La quantification des microARN grâce aux banques sans réplicat est donc soumise à des biais, mais elle peut être utilisée de manière indicative, pour un criblage globale, et non absolu. Afin de d'éviter au maximum ces biais, le coût des séquençages diminuant et la profondeur de séquençage augmentant, on pourra plus facilement réaliser les répétitions techniques et biologiques nécessaires, si possible en les multiplexant sur une même plaque afin de limiter le coût et la variabilité technique.

Grâce au séquençage de nos banques, des centaines de million de séquence ont été générées. Le filtre des séquences retenues présentant une réalité biologique se fait de manière *in silico* et le choix des paramètres pour définir ce qu'est un microARN est particulièrement important et difficile. Ces petits ARN se présentent sous des tailles (matures de 19 nt à 24 nt et précurseurs de 54 nt à 518 nt chez *M. truncatula* ; miRBase 18) et des formes de tige-boucles différentes (nombre de mésappariements variables), et des distributions et abondances variables des lectures de séquençage sur les précurseurs (présence ou non du miRstar, décalage d'appariement entre le miR et le miRstar). Les critères de caractérisation des microRNA ont été établis par la communauté des

« mécanisticiens » de ce domaine (Meyers *et al.*, 2008) mais leur conclusion fut que ces critères sont en constante évolution au fur et à mesure de la découverte de nouveaux microARN. En appliquant ces critères, plusieurs microARN déjà connus chez *M. truncatula*, et conservés chez d'autres espèces, ont été rejetés par notre analyse. Les critères ont donc été ajustés pour obtenir un maximum des microARN déjà connus mais force est de constater que des faux-positifs et des faux-négatifs se sont tout de même retrouvés présents ou éjectés dans nos banques.

La prédiction *in silico* des cibles des microARN, dont la cible n'a pas été validée par RACE (Rapid Amplification of CDNA-Ends) ou une autre technique, est aussi soumise à des erreurs. Il a été longtemps proposé que le mécanisme de reconnaissance des cibles chez les plantes se fait directement par une complémentarité inverse parfaite des séquences du microARN et de la cible, entraînant une dégradation de cette dernière. Les outils de prédiction de cibles de microARN dédiés aux plantes ont donc été développés dans ce sens laissant de côté les cibles comportant des mésappariements avec le microARN correspondant, considéré alors comme le mécanisme majeur chez les animaux. Par la suite, cette vision a été réfutée avec des contre-exemples. En 2005, une étude a montré que chez les animaux, les microARN pouvaient également agir sur l'accumulation des transcrits cibles en les clivant (Bagga *et al.*, 2005) et, inversement, en 2008, il a été montré chez les plantes que les microARN peuvent réprimer la traduction du transcrit ciblé sans clivage de ce dernier (Dugas & Bartel, 2008). Les interactions entre les microARN et les cibles peuvent donc prendre des formes très différentes chez les plantes, la prédiction *in silico* devient donc plus difficile à paramétrer. Le croisement entre différents prédicteurs de cible permet d'être plus confiant dans les résultats. Dans notre étude, l'utilisation de prédicteurs multiples a été couplée aux données issues du dégradome obtenu chez *M. truncatula* dans des conditions mycorhizée et non mycorhizée (Devers *et al.*, 2011), renforçant ainsi nos prédictions par des preuves physiques d'un clivage éventuel. Malgré cela, la prédiction des cibles non clivées reste très aléatoire et permet seulement un criblage des transcrits dont la possible inhibition de la traduction pourra être vérifiée par la diminution d'accumulation des protéines correspondantes.

En conclusion, les approches massives de séquençage permettent le dressage de catalogues de microARN dont la présence varie en fonction de conditions données, la prédiction des cibles correspondantes, la découverte de nouveaux microARN avec l'appui du génome correspondant et la quantification de ces derniers en multipliant les répétitions, qu'elles soient techniques ou biologiques. Notre étude a donc permis le criblage de plusieurs microARN affectés lors des étapes présymbiotiques et symbiotiques de la symbiose MA, mais leur réelle implication physiologique ne pourra donc se faire qu'au travers d'une analyse plus précise et fonctionnelle en s'attardant sur la relation microARN/cible (colocalisation, clivage, changement de l'accumulation de transcrits/protéines) ainsi que sur l'effet de la suppression et la surexpression du microARN et de sa(ses) cible(s) correspondantes.

CONCLUSION GENERALE: APPORT SCIENTIFIQUE DES TRAVAUX DE THESE, DEVELOPPEMENTS ET APPLICATIONS ENVISAGEES

Mes travaux de thèse ont porté sur deux volets d'étude distincts, répondant à deux objectifs spécifiques d'intérêts à la fois scientifique et d'exploitation pour la société Agronutrition.

Le premier volet a consisté à étudier le mitochondriome du partenaire fongique, *Rhizophagus irregularis*, d'une part en identifiant et annotant les gènes impliqués dans le métabolisme mitochondrial dans le cadre de l'étude du génome de ce champignon, et, d'autres parts, d'apprécier la variabilité intra-spécifique du génome mitochondrial par comparaison de génomes mitochondriaux de plusieurs souches de *R. irregularis*. L'étude du métabolisme énergétique a essentiellement un objectif scientifique : définir les mécanismes de régulation du métabolisme du champignon. L'étude des génomes mitochondriaux s'inscrivait dans un contexte d'étude plus général (projet Neofertil, impliquant la société Agronutrition et l'équipe de recherche du LRSV) pour lequel mon objectif était de définir une stratégie de typage et de quantification des souches fongiques.

Le second volet d'étude a porté sur l'identification, chez le partenaire végétal, *Medicago truncatula* de petits éléments régulateurs de l'expression génique, les microARN, potentiellement impliqués dans les étapes précoces et tardives de la symbiose MA. L'objectif de cette approche était de mieux comprendre la régulation de cette symbiose, avec la possibilité de définir des lignées végétales plus efficaces pour la propagation des souches fongiques *in vitro*.

De ces deux volets d'étude, nous pouvons tirer plusieurs lignes directrices pour des travaux à venir.

La mitochondrie : un organite révélateur de l'état physiologique du champignon

L'accès aux données du projet international de séquençage du génome de *R. irregularis* (Lammers *et al.*, 2004; Martin *et al.*, 2008) m'a permis d'identifier les gènes impliqués dans la production d'énergie et la respiration chez ce champignon, aboutissant à une proposition d'organisation de la chaîne respiratoire. J'ai pu confirmer la présence

d'une voie alternative de cette chaîne et mettre en évidence une probable respiration anoxygène grâce à la présence de gènes impliqués dans cette voie. Il faut désormais définir à quelles étapes du cycle biologique du champignon ces différentes voies sont mises en place, impliquant des changements de régulation énergétique et métabolique. Ce projet, poursuivi en thèse au laboratoire (projet de thèse de Mathilde Malbreil) devrait permettre d'élucider les mécanismes qui sous-tendent les changements de la physiologie asymbiotique vers la physiologie symbiotique. Des études de l'expression des enzymes de la voie respiratoire en présence, ou non, d'oxygène et en présence de différentes sources de carbone, permettront de mieux comprendre comment le champignon s'adapte à son environnement et si les mécanismes mis en jeu, et les profils d'expression des gènes, corréleront avec ceux permettant le passage de la phase asymbiotique à la phase symbiotique.

Le génome mitochondrial : une cible de choix pour réaliser des empreintes génétiques et définir un protocole de quantification du champignon

En séquençant le génome mitochondrial de six nouvelles souches de *R. irregularis*, j'ai pu répertorier trois types d'éléments générateurs de variabilité (VGE) génomique entre ces souches: des éléments composés de cadres de lecture ouverts (ORF) de Homing Endonucléases (HEO), des éléments composés d'ORF contenant des domaines ADN polymérase (DPDCO), et des éléments palindromiques répétés de petite taille (SIR). Ces trois types d'éléments se retrouvent dans des positions variables selon les génomes, supposant une grande mobilité et dynamique. Les VGE sont donc, potentiellement, des éléments qui sont acteurs d'un mécanisme global permettant de maintenir une certaine variabilité du génome mitochondrial de ces champignons dont les échanges de matériels génétiques, notamment mitochondriaux, sont très rares entre les souches. Cette variabilité n'étant pas observée au sein de toutes les souches, il est possible de développer des marqueurs moléculaires mitochondriaux permettant d'identifier chacune d'entre elles. Ceci est illustré par le fait que, parmi les 6 isolats étudiés, nous avons pu définir que 4 d'entre eux étaient, selon toute probabilité, identiques deux à deux (MUCL_46240 / MUCL_46241 ; MUCL43204 / Col_01). Les VGE sont ainsi des régions de choix pour définir des marqueurs spécifiques des souches qui permettraient d'analyser les échanges de génome mitochondriaux par anastomose, lors d'expérimentation de co-inoculation par

plusieurs souches, et ainsi rentrer dans l'analyse génétique des marqueurs de compatibilité végétative ou sexuelle de ces champignons. Une autre application serait l'identification et la quantification d'une souche en une seule étape grâce à la définition d'amorces oligonucléotidiques de PCR en temps réel sur des VGE spécifiques de cette souche. Cette méthode permettra d'assurer le suivi de souches d'intérêt dans un système de production, remplissant ainsi la mission qui m'incombait au sein de l'entreprise Agronutrition. Ces marqueurs pourraient également permettre un suivi des souches au champ, dans la mesure où une analyse préalable définit la spécificité des marqueurs de la souche ajoutée par rapport à la flore fongique initiale.

Les microARN : des éléments régulateurs de la symbiose permettant de définir des lignées végétales supermycorhiziennes

Ma participation au projet Géoscope MIRMED (responsable scientifique Dr. Martin Crespi) a permis de caractériser le microtranscriptome de *M. truncatula* durant les étapes présymbiotiques et symbiotiques de la mycorhization. La validation du rôle biologique des microARN candidats identifiés, dans l'établissement de la symbiose, devra se poursuivre par une analyse fonctionnelle permettant de préciser leur implication dans ce processus. Au delà de l'intérêt scientifique de ces approches, ces études ne sont pas déconnectées d'une application industrielle pour l'entreprise Agronutrition. La modification d'expression de certains de ces microARN peut conduire à une hypermycorhization, et/ou une colonisation plus rapide, de l'hôte par dé-répression des mécanismes régulant sa colonisation (Lauressergues *et al.*, 2012). Ces transformants pourraient donc présenter un potentiel industriel de grand intérêt en augmentant le rendement des systèmes de production d'inoculum *in vitro* par accélération du cycle de production et/ou augmentation de la production de spores de champignons MA par unité de production.

A n'en pas douter, plusieurs de ces perspectives seront poursuivies dans l'équipe et par la société Agronutrition avec lesquelles j'ai réalisé ces travaux.

BIBLIOGRAPHIE

Addo-Quaye C, Eshoo TW, Bartel David P, Axtell MJ. 2008. Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. *Current biology : CB* **18**: 758–62.

Akiyama K, Matsuzaki K, Hayashi H. 2005. Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature* **435**: 824–7.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. BASIC LOCAL ALIGNMENT SEARCH TOOL. *Journal of Molecular Biology* **215**: 403–410.

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. National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.* altschul@ncbi.nlm.nih.gov, 3389–402.

Ané J-M, Kiss GB, Riely BK, Penmetsa RV, Oldroyd Giles E D, Ayax C, Lévy J, Debelle F, Baek J-M, Kalo P, et al. 2004. Medicago truncatula DMI1 required for bacterial and fungal symbioses in legumes. *Science (New York, N.Y.)* **303**: 1364–7.

Arikawa Y, Enomoto K, Muratsubaki H, Okazaki M. 1998. Soluble fumarate reductase isoenzymes from *Saccharomyces cerevisiae* are required for anaerobic growth. *FEMS microbiology letters* **165**: 111–6.

Aukerman M. 2003. Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. *The Plant Cell Online* **15**: 2730–2741.

Bagga S, Bracht J, Hunter S, Massirer K, Holtz J, Eachus R, Pasquinelli AE. 2005. Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell* **122**: 553–63.

Balzer S, Malde K, Jonassen I. 2011. Systematic exploration of error sources in pyrosequencing flowgram data. *Bioinformatics* **27**: 304–9.

Balergue C, Puech-Pagès V, Bécard G, Rochange SFC-P. 2011. The regulation of arbuscular mycorrhizal symbiosis by phosphate in pea involves early and systemic signalling events. *Journal of Experimental Botany* **62**: 1049–60.

Bari R, Pant BD, Stitt M, Scheible W-R. 2006. PHO2, microRNA399, and PHR1 define a phosphate-signaling pathway in plants. *Plant Physiology* **141**: 988–999.

- Bartel D P. 2004.** MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **116**: 281–297.
- 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 (New York, N.Y.)* **330**: 1549–51.
- Begel O, Boulay J, Albert B, Dufour E, Sainsard-Chanet A. 1999.** Mitochondrial Group II Introns, Cytochrome c Oxidase, and Senescence in *Podospora anserina*. *Molecular and Cellular Biology* **19**: 4093–4100.
- Berta G, Fusconi A, Trotta A, Scannerini S. 1990.** Morphogenetic modifications induced by the mycorrhizal fungus *Glomus* strain E3 in the root system of *Alium porrum* L. *New Phytologist* **114**: 207–215.
- Berta G, Trotta a, Fusconi a, Hooker JE, Munro M, Atkinson D, Giovannetti M, Morini S, Fortuna P, Tisserant B, et al. 1995.** Arbuscular mycorrhizal induced changes to plant growth and root system morphology in *Prunus cerasifera*. *Tree physiology* **15**: 281–293.
- Besserer A, Bécard Guillaume, Jauneau Alain, Roux Christophe, Séjalon-Delmas Nathalie. 2008.** GR24, a synthetic analog of strigolactones, stimulates the mitosis and growth of the arbuscular mycorrhizal fungus *Gigaspora rosea* by boosting its energy metabolism. *Plant physiology* **148**: 402–13.
- Besserer A, Bécard Guillaume, Roux Christophe, Séjalon-delmas N. 2009.** Role of mitochondria in the response of arbuscular mycorrhizal fungi to strigolactones. *Plant Signaling and Behavior* **4**: 75–77.
- Besserer A, Puech-Pagès Virginie, Kiefer P, Gomez-Roldan V, Jauneau Alain, Roy Sébastien, Portais J-C, Roux Christophe, Bécard Guillaume, Séjalon-Delmas Nathalie. 2006.** Strigolactones stimulate arbuscular mycorrhizal fungi by activating mitochondria. *PLoS biology* **4**: e226.
- Bever J, Morton J. 1999.** Heritable variation and mechanisms of inheritance of spore shape within a population of *Scutellospora pellucida*, an arbuscular mycorrhizal fungus. *American Journal of Botany* **86**: 1209–1216.
- Bever JD, Wang M. 2005.** Arbuscular mycorrhizal fungi: hyphal fusion and multigenomic structure. *Nature* **433**: E3–4; discussion E4.
- Birky CW. 2001.** The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annual Review of Genetic* **35**: 125-48.

- Blaszkowski J, Czerniawska B, Wubet T. 2008.** Glomus irregulare, a new arbuscular mycorrhizal fungus in the Glomeromycota. *Mycotaxon* **106**: 247–267.
- Boer PH, Gray M W. 1991.** Short dispersed repeats localized in spacer regions of Chlamydomonas reinhardtii mitochondrial DNA. *Curr Genet* **19**: 309–12.
- Boisson-Dernier A, Chabaud M, Garcia F, Bécard G, Rosenberg C, Barker D G. 2001.** Agrobacterium rhizogenes-transformed roots of Medicago truncatula for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. *Molecular plant-microbe interactions : MPMI* **14**: 695–700.
- Bonfante Paola, Genre A. 2010.** Mechanisms underlying beneficial plant-fungus interactions in mycorrhizal symbiosis. *Nature communications* **1**: 48.
- Bonfante Paola, Requena Natalia. 2011.** Dating in the dark: how roots respond to fungal signals to establish arbuscular mycorrhizal symbiosis. *Current opinion in plant biology* **14**: 451–7.
- Boualem A, Laporte P, Jovanovic M, Laffont C, Plet J, Combier J-P, Niebel A, Crespi M, Frugier F. 2008.** MicroRNA166 controls root and nodule development in Medicago truncatula. *Plant Journal* **54**: 876–887.
- Branscheid A, Sieh D, Pant BD, May P, Devers EA, Elkrog A, Schauser L, Scheible W-R, Krajinski Franziska. 2010.** Expression Pattern Suggests a Role of MiR399 in the Regulation of the Cellular Response to Local Pi Increase During Arbuscular Mycorrhizal Symbiosis. *Molecular Plant-Microbe Interactions* **23**: 915–926.
- Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer Patrice, Yamamoto YY, Sieburth L, Voinnet Olivier. 2008.** Widespread translational inhibition by plant miRNAs and siRNAs. *Science (New York, N.Y.)* **320**: 1185–90.
- Bárzana G, Aroca R, Paz JA, Chaumont F, Martinez-Ballesta MC, Carvajal M, Ruiz-Lozano JM. 2012.** Arbuscular mycorrhizal symbiosis increases relative apoplastic water flow in roots of the host plant under both well-watered and drought stress conditions. *Annals of botany* **109**: 1009–1017.
- Bécard G, Doner L, Rolin D, Douds D, Pfeffer P. 1991.** Identification and quantification of trehalose in vesicular-arbuscular mycorrhizal fungi by in vivo ¹³C NMR and HPLC analyses. *New Phytologist* **118**: 547–552.
- 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, Roux C, Séjalon-Delmas N, Puech V, Roy S. 2004 (UPS-CNRS).** Modulateurs de développement des champignons mycorhiziens à arbuscules, et applications. FR2865897
- Börstler B, Raab PA, Thiéry O, Morton JB, Redecker D. 2008.** Genetic diversity of the arbuscular mycorrhizal fungus *Glomus intraradices* as determined by mitochondrial large subunit rRNA gene sequences is considerably higher than previously expected. *New Phytologist* **180**: 452–65.
- Börstler B, Thiéry O, Sýkorová Z, Berner A, Redecker D, Borstler B, Thiery O, Sykorova Z. 2010.** Diversity of mitochondrial large subunit rDNA haplotypes of *Glomus intraradices* in two agricultural field experiments and two semi-natural grasslands. *Molecular Ecology* **19**: 1497–511.
- Campos-soriano L, García-martínez J, Segundo BSAN. 2011.** The arbuscular mycorrhizal symbiosis promotes the systemic induction of regulatory defence-related genes in rice leaves and confers resistance to pathogen infection. *Molecular Plant Pathology* **13**:579-92.
- Campos-Soriano L, Gómez-Ariza J, Bonfante Paola, San Segundo B. 2011.** A rice calcium-dependent protein kinase is expressed in cortical root cells during the presymbiotic phase of the arbuscular mycorrhizal symbiosis. *BMC plant biology* **11**: 90.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009.** trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics (Oxford, England)* **25**: 1972–3.
- Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J. 2005.** ACT: the Artemis Comparison Tool. *Bioinformatics* **21**: 3422–3.
- Catoira R, Galera C, De Billy F, Penmetsa R V, Journet EP, Maillet F, Rosenberg C, Cook D, Gough C, Dénarié J. 2000.** Four genes of *Medicago truncatula* controlling components of a nod factor transduction pathway. *The Plant cell* **12**: 1647–66.
- Chabaud Mireille, Genre A, Sieberer BJ, Faccio A, Fournier J, Novero M, Barker David G, Bonfante Paola. 2011.** Arbuscular mycorrhizal hyphopodia and germinated spore exudates trigger Ca²⁺ spiking in the legume and nonlegume root epidermis. *The New phytologist* **189**: 347–55.
- Chabot M, Bécard G, Piché Y. 1992.** Life cycle of *Glomus intraradices* in root organ culture. *Mycologia* **84**: 315–321.
- Charpentier M, Bredemeier R, Wanner G, Takeda N, Schleiff E, Parniske M. 2008.** Lotus japonicus CASTOR and POLLUX are ion channels essential for perinuclear calcium spiking in legume root endosymbiosis. *The Plant cell* **20**: 3467–79.

Chen C, Ridzon D a, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR, et al. 2005. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic acids research* **33**: e179.

Chevalier BS, Stoddard BLC-P. 2001. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Research* **29**: 3757–74.

Christiansen MW, Holm PB, Gregersen PL. 2011. Characterization of barley (*Hordeum vulgare* L.) NAC transcription factors suggests conserved functions compared to both monocots and dicots. *BMC research notes* **4**: 302.

Clark-Walker GD. 1991. Contrasting mutation rates in mitochondrial and nuclear genes of yeasts versus mammals. *Current Genetics* **20**: 195–198.

Combier J-P, Frugier F, De Billy Françoise, Boualem A, El-Yahyaoui F, Moreau S, Vernie T, Ott T, Gamas P, Crespi M, et al. 2006. MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. *Genes & Development* **20**: 3084–3088.

Creighton CJ, Reid JG, Gunaratne PH. 2009. Expression profiling of microRNAs by deep sequencing. *Briefings in bioinformatics* **10**: 490–7.

Croll D, Wille L, Gamper HA, Mathimaran N, Lammers PJ, Corradi N, Sanders I R. 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**: 672–87.

Croll Daniel, Sanders Ian R. 2009. Recombination in *Glomus intraradices*, a supposed ancient asexual arbuscular mycorrhizal fungus. *BMC evolutionary biology* **9**: 13.

Czaja LF, Hogekamp C, Lamm P, Maillet F., Andres Martinez E, Samain E, Denarie J., Kuster H, Hohnjec N. 2012. Transcriptional responses towards diffusible signals from symbiotic microbes reveal MtNFP- and MtDMI3-dependent reprogramming of host gene expression by AM fungal LCOs. *Plant Physiology* **159**:1672-85.

Dangeard P. 1898. Une maladie du peuplier dans l'ouest de la France. *Le Botaniste* **5**: 38–43.

Dangeard P. 1901. Le “Rhizophagus populinus” Dangeard. *Le Botaniste* **7**: 285–292.

Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* **14**: 1394–403.

Darling AE, Mau B, Perna N. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* **5**: e11147.

- Denovan-Wright EM, Lee Robert W. 1994.** Comparative Structure and Genomic Organization of the Discontinuous Mitochondrial Ribosomal RNA Genes of *Chlamydomonas eugametos* and *Chlamydomonas reinhardtii*. *Journal of Molecular Biology* **241**: 298–311.
- Derelle D, Declerck S, Genet P, Dajoz I, Van Aarle IM. 2012.** Association of highly and weakly mycorrhizal seedlings can promote the extra- and intraradical development of a common mycorrhizal network. *FEMS Microbiology Ecology* **79**: 251–259.
- Devers EA, Branscheid A, May P, Krajinski Franziska. 2011.** Stars and Symbiosis: MicroRNA- and MicroRNA*-Mediated Transcript Cleavage Involved in Arbuscular Mycorrhizal Symbiosis. *Plant Physiology* **156**: 1990–2010.
- Dolan L. 2007.** SCARECROWs at the border. *Science* **316**: 377–378.
- Dugas D V, Bartel B. 2008.** Sucrose induction of Arabidopsis miR398 represses two Cu/Zn superoxide dismutases. *Plant molecular biology* **67**: 403–17.
- Duplessis S, Cuomo C a, Lin Y-C, Aerts A, Tisserant Emilie, Veneault-Fourrey C, Joly DL, Hacquard S, Amselem Joëlle, Cantarel BL, et al. 2011.** From the Cover: Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 9166–71.
- Endre G, Kereszt A, Kevei Z, Mihacea S. 2002.** A receptor kinase gene regulating symbiotic nodule development. *Nature* **417**: 962–966.
- Felten J, Legué V, Ditengou FA. 2010.** Lateral root stimulation in the early interaction between *Arabidopsis thaliana* and the ectomycorrhizal fungus *Laccaria bicolor*: is fungal auxin the trigger? *Plant signaling & behavior* **5**: 864–867.
- Fester T, Strack D, Hause B. 2001.** Reorganization of tobacco root plastids during arbuscule development. *Planta* **213**: 864–868.
- Formighieri EF, Tiburcio RA, Armas ED, Medrano FJ, Shimo H, Carels N, Góes-Neto A, Cotomacci C, Carazzolle MF, Sardinha-Pinto N, et al. 2008.** The mitochondrial genome of the phytopathogenic basidiomycete *Moniliophthora perniciosa* is 109 kb in size and contains a stable integrated plasmid. *Mycological Research* **112**: 1136–1152.
- Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, García JA, Paz-Ares J. 2007.** Target mimicry provides a new mechanism for regulation of microRNA activity. *Nature genetics* **39**: 1033–7.

Garriock M, Peterson R, Ackerley C. 1989. Early stages in colonization of *Allium porrum* (leek) roots by the vesicular-arbuscular mycorrhizal fungus, *Glomus versiforme*. *New phytologist* **112**: 85-92.

Gaude N, Bortfeld S, Duensing N, Lohse M, Krajinski Franziska. 2011. Arbuscule-containing and non-colonized cortical cells of mycorrhizal roots undergo extensive and specific reprogramming during arbuscular mycorrhizal development. *The Plant journal : for cell and molecular biology* **69**: 510–28.

Genre A, Chabaud Mireille, Timmers T, Bonfante Paola, Barker David G. 2005. Arbuscular Mycorrhizal Fungi Elicit a Novel Intracellular Apparatus in *Medicago truncatula* Root Epidermal Cells before Infection. *American Society of Plant Biologists* **17**: 3489–3499.

Gerdemann J, Trappe J. 1974. *The Endogonaceae in the Pacific Northwest*. Bronx, N.Y.: the New York Botanical Garden in collaboration with the Mycological Society of America.

Gimble FS. 2000. Invasion of a multitude of genetic niches by mobile endonuclease genes. *FEMS Microbiology Letters* **185**: 99–107.

Gleason C, Chaudhuri S, Yang T, Muñoz A, Poovaiah BW, Oldroyd Giles E D. 2006. Nodulation independent of rhizobia induced by a calcium-activated kinase lacking autoinhibition. *Nature* **441**: 1149–52.

Gomez-Roldan V, Roux Christophe, Girard D, Becard Guillaume, Puech Virginie. 2007. Strigolactones: Promising Plant Signals. *Plant signaling and behaviour and behaviour* **2**: 163–164.

Graham JH, Leonard RT, Menge J a. 1981. Membrane-mediated decrease in root exudation responsible for phosphorus inhibition of vesicular-arbuscular mycorrhiza formation. *Plant physiology* **68**: 548–52.

Griffiths-Jones S. 2004. The microRNA Registry. *Nucleic acids research* **32**: D109–11.

Grindley NDF, Reed RR. 1985. Transpositional Recombination in Prokaryotes. *Annual Review of Biochemistry* **54**: 863–896.

Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. 2008. The Vienna RNA websuite. *Nucleic Acids Reserch* **36**: W70–4.

Gu M, Xu K, Chen A, Zhu Yiyong, Tang G, Xu G. 2010. Expression analysis suggests potential roles of microRNAs for phosphate and arbuscular mycorrhizal signaling in *Solanum lycopersicum*. *Physiologia Plantarum* **138**: 226–237.

- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010.** New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* **59**: 307–21.
- Gutjahr C, Novero M, Guether M, Montanari O, Udvardi M, Bonfante Paola. 2009.** Presymbiotic factors released by the arbuscular mycorrhizal fungus *Gigaspora margarita* induce starch accumulation in *Lotus japonicus* roots. *The New phytologist* **183**: 53–61.
- 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 Evolution* **3**: 950–8.
- Harrison M J, Dewbre GR, Liu JY. 2002.** A phosphate transporter from *Medicago truncatula* involved in the acquisition of phosphate released by arbuscular mycorrhizal fungi. *Plant Cell* **14**: 2413–2429.
- Van der Heijden MGA, Klironomos JN, Ursic M, Moutoglis P, Streitwolf-Engel R, Boller Thomas, Wiemken Andres, Sanders Ian R. 1998.** Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* **396**: 69–72.
- Helber Nicole, Wippel K, Sauer N, Schaarschmidt S, Hause B, Requena Natalia. 2011.** A Versatile monosaccharide transporter that operates in the arbuscular mycorrhizal fungus *Glomus* sp is crucial for the symbiotic relationship with plants. *The Plant cell* **23**: 3812–23.
- Hermanns J, Osiewacz H. 1996.** Induction of longevity by cytoplasmic transfer of a linear plasmid in *Podospora anserina*. *Current Genetics* **29**: 250–256.
- Herrera-Medina MJ, Steinkellner S, Vierheilig H, Ocampo Bote JA, García Garrido JM. 2007.** Abscisic acid determines arbuscule development and functionality in the tomato arbuscular mycorrhiza. *The New phytologist* **175**: 554–64.
- Hewitt E. 1966.** Sand and Water Culture Methods Used in the Study of Plant Nutrition. Sand and Water Culture Methods Used in the Study of Plant Nutrition. London: Commonwealth Agricultural Bureau, .
- Hijri M, Hosny M, Van Tuinen D, Dulieu H. 1999.** Intraspecific ITS polymorphism in *Scutellospora castanea* (Glomales, Zygomycota) is structured within multinucleate spores. *Fungal genetics and biology : FG & B* **26**: 141–51.
- Hijri M, Sanders I R. 2004.** The arbuscular mycorrhizal fungus *Glomus intraradices* is haploid and has a small genome size in the lower limit of eukaryotes. *Fungal Genetics and Biology* **41**: 253–261.

Hijri M, Sanders I. 2005. Low gene copy number shows that arbuscular mycorrhizal fungi inherit genetically different nuclei. *Nature* **433**: 160–163.

Hohnjec Natalija, Vieweg MF, Pühler A, Becker A, Küster Helge. 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 Physiology* **137**: 1283–1301.

Horn S, Durka W, Wolf R, Ermala A, Stubbe A, Stubbe M, Hofreiter M. Mitochondrial Genomes Reveal Slow Rates of Molecular Evolution and the Timing of Speciation in Beavers (Castor), One of the Largest Rodent Species. *PLoS ONE* **6**: e14622.

Hosny M, Païs de Barros JP, Gianinazzi-Pearson V, Dulieu H. 1997. Base composition of DNA from glomalean fungi: high amounts of methylated cytosine. *Fungal genetics and biology: FG & B* **22**: 103–11.

Hudspeth M. 1992. The fungal mitochondrial genome - a broader perspective. In: Dekker M, ed. Handbook of applied Mycology. New-York: Fungal Biotechnology, 213–242.

Ivanov S, Fedorova EE, Limpens E, Mita S De, Genre A, Bonfante Paola. 2012. Rhizobium – legume symbiosis shares an exocytotic pathway required for arbuscule formation. *Proceedings of the National Academy of Sciences of the United States of America* doi:10.1073/pnas.1200407109.

James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Gueidan C, Fraker E, Miadlikowska J, et al. 2006. Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* **443**: 818–22.

Javot H, Penmetsa RV, Breuillin F, Bhattarai KK, Noar RD, Gomez S Karen, Zhang Q, Cook DR, Harrison Maria J. 2011. Medicago truncatula mtpt4 mutants reveal a role for nitrogen in the regulation of arbuscule degeneration in arbuscular mycorrhizal symbiosis. *The Plant journal : for cell and molecular biology* **68**: 954–65.

Javot H, Penmetsa RV, Terzaghi N, Cook DR, Harrison Maria J. 2007. A Medicago truncatula phosphate transporter indispensable for the arbuscular mycorrhizal symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 1720–5.

Jex AR, Hu MD, Littlewood TJ, Waeschenbach A, Gasser RB. 2008. Using 454 technology for long-PCR based sequencing of the complete mitochondrial genome from single Haemonchus contortus (Nematoda). *BMC Genomics* **9**: 11.

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. 2008. NCBI BLAST: a better web interface. *Nucleic Acids Res* **36**: W5–9.

Jones-Rhoades MW, Bartel David P, Bartel B. 2006. MicroRNAs and their regulatory roles in plants. *Annual Review of Plant Biology* **57**: 19–53.

Joseph-horne T, Hollomon DW, Wood PM. 2001. Fungal respiration: a fusion of standard and alternative components. *Science* **1504**: 179-95.

Journet E-P, Van Tuinen Diederik, Gouzy Jérôme, Crespeau H, Carreau V, Farmer M-J, Niebel A, Schiex T, Jaillon O, Chatagnier O, et al. 2002. Exploring root symbiotic programs in the model legume *Medicago truncatula* using EST analysis. *Nucleic acids research* **30**: 5579–5592.

Jurica M S, Stoddard B L. 1999. Homing endonucleases: structure, function and evolution. *Cell Molecular and Life Science* **55**: 1304–26.

Jurica Melissa S, Monnat Jr RJ, Stoddard Barry L. 1998. DNA Recognition and Cleavage by the LAGLIDADG Homing Endonuclease I-Cre I. *Molecular Cell* **2**: 469–476.

Kaguni LS. 2004. DNA POLYMERASE γ , THE MITOCHONDRIAL REPLICASE 1. *Annual Review of Biochemistry* **73**: 293–320.

Kaldorf M, Ludwig-Müller J. 1999. AM fungi might affect the root morphology of maize by increasing indole-3-butyric acid biosynthesis. *Physiologia plantarum* **109**: 58–67.

Kaló P, Gleason C, Edwards A, Marsh J, Mitra RM, Hirsch S, Jakab J, Sims S, Long Sharon R, Rogers J, et al. 2005. Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. *Science (New York, N.Y.)* **308**: 1786–9.

Katoh K, Misawa K, Kuma K, Miyata TC-135756. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**: 3059–66.

Klironomos JN. 2002. Feedback with soil biota contributes to plant rarity and invasiveness in communities. *Nature* **417**: 67–70.

Koops P, Pelsler S, Ignatz M, Klose C, Marrocco-Selden K, Kretsch T. 2011. EDL3 is an F-box protein involved in the regulation of abscisic acid signalling in *Arabidopsis thaliana*. *Journal of experimental botany* **62**: 5547–60.

Kosuta S, Chabaud Mireille, Gough Clare, De J, Barker David G. 2003. A Diffusible Factor from Arbuscular Mycorrhizal Fungi Induces Symbiosis-Specific MtENOD11 Expression in Roots of *Medicago truncatula* 1. *Plant Physiology* **131**: 952–962.

- Kosuta S, Hazledine S, Sun J, Miwa H, Morris RJ, Downie JA, Oldroyd Giles E D. 2008.** Differential and chaotic calcium signatures in the symbiosis signaling pathway of legumes. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 9823–8.
- Krajinski F, Frenzel A. 2007.** Towards the elucidation of AM-specific transcription in *Medicago truncatula*. *Phytochemistry* **68**: 75–81.
- Krüger M, Krüger C, Walker Christopher, Stockinger Herbert, Schüßler Arthur. 2011.** Phylogenetic reference data for systematics and phylotaxonomy of arbuscular mycorrhizal fungi from phylum to species level. *New Phytologist*: **193**: 970-84.
- Kuhn G, Hijri M, Sanders I R. 2001.** Evidence for the evolution of multiple genomes in arbuscular mycorrhizal fungi. *Nature* **414**: 745–8.
- Kuhn H, Küster Helge, Requena Natalia. 2010.** Membrane steroid-binding protein 1 induced by a diffusible fungal signal is critical for mycorrhization in *Medicago truncatula*. *The New phytologist* **185**: 716–33.
- Küster Helge, Hohnjec Natalija, Krajinski Franziska, El Yahyaoui F, Manthey K, Gouzy Jérôme, Dondrup M, Meyer F, Kalinowski J, Brechenmacher L, et al. 2004.** Construction and validation of cDNA-based Mt6k-RIT macro- and microarrays to explore root endosymbioses in the model legume *Medicago truncatula*. *Journal of Biotechnology* **108**: 95–113.
- Labidi S, Ben Jeddi F, Tisserant Benoit, Debiane D, Rezgui S, Grandmougin-Ferjani A, Lounès-Hadj Sahraoui A. 2011.** Role of arbuscular mycorrhizal symbiosis in root mineral uptake under CaCO₃ stress. *Mycorrhiza* **22**: 337-45.
- Lammers P, Tuskan G a, DiFazio SP, Podila GK, Martin F. 2004.** Mycorrhizal symbionts of *Populus* to be sequenced by the United States Department of Energy's Joint Genome Institute. *Mycorrhiza* **14**: 63–4.
- Lauressergues D, Delaux P-M, Formey Damien, Lelandais-Brière C, Fort S, Cottaz S, Bécard Guillaume, Niebel A, Roux Christophe, Combier J-P. 2012.** The microRNA miR171h modulates arbuscular mycorrhizal colonization of *Medicago truncatula* by targeting NSP2. *The Plant journal : for cell and molecular biology* **72**: 512-22.
- Lee J, Young JP. 2009.** The mitochondrial genome sequence of the arbuscular mycorrhizal fungus *Glomus intraradices* isolate 494 and implications for the phylogenetic placement of *Glomus*. *New Phytologist* **183**: 200–11.
- Lee Y, Kim M, Han J, Yeom K-H, Lee S, Baek SH, Kim VN. 2004.** MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal* **23**: 4051–60.

- Lelandais-Brière C, Naya L, Sallet E, Calenge F, Frugier F, Hartmann C, Gouzy Jérôme, Crespi M. 2009.** Genome-wide *Medicago truncatula* small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. *The Plant cell* **21**: 2780–96.
- Li H, Durbin RC-P. 2010.** Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**: 589–95.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009.** The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–9.
- Liao J, Singh S, Hossain MS, Andersen SU, Ross L, Bonetta D, Zhou Y, Sato S, Tabata S, Stougaard J, et al. 2012.** Negative regulation of CCaMK is essential for symbiotic infection. *The Plant journal : for cell and molecular biology* **72**: 572–584.
- Liu Y, Leigh JW, Brinkmann H, Cushion MT, Rodriguez-Ezpeleta N, Philippe H, Lang B F. 2009.** Phylogenomic analyses support the monophyly of Taphrinomycotina, including *Schizosaccharomyces* fission yeasts. *Molecular Biology and Evolution* **26**: 27–34.
- Llave C, Xie Z, Kasschau KD, Carrington JC. 2002.** Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science (New York, N.Y.)* **297**: 2053–6.
- Logi C, Sbrana C, Giovannetti M. 1998.** Cellular events involved in survival of individual arbuscular mycorrhizal symbionts growing in the absence of the host. *Applied and environmental microbiology* **64**: 3473–9.
- Lullien V, Barker D G, Lajudie P De, Huguet T. 1987.** Plant gene expression in effective and ineffective root nodules of alfalfa (*Medicago sativa*). **478**: 469–478.
- Lévy J, Bres C, Geurts René, Chalhoub B, Kulikova O, Duc G, Journet E-P, Ané J-M, Lauber E, Bisseling T, et al. 2004.** A putative Ca²⁺ and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. *Science (New York, N.Y.)* **303**: 1361–4.
- Maillet Fabienne, Poinot V, Andre O, Puech-Pages V, Haouy A, Gueunier M, Cromer L, Giraudet D, Formey Damien, Niebel A, et al. 2011.** Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature* **469**: 58–U1501.
- Malloch D, Pirozynski K. 1980.** Ecological and evolutionary significance of mycorrhizal symbioses in vascular plants (a review). *Proceedings of the National Academy of Science* **77**: 2113–2118.

- Mallory AC, Vaucheret H. 2006.** Functions of microRNAs and related small RNAs in plants. *Nature Genetics* **38**: S31–S36.
- Margulis L, Fester R. 1991.** *Symbiosis as Source of Evolutionary Innovation: Speciation and Morphogenesis*. Cambridge: MIT Press.
- Mariette J, Noirot C, Klopp CC-P, Jerome M, Jérôme M. 2011.** Assessment of replicate bias in 454 pyrosequencing and a multi-purpose read-filtering tool. *BMC Research Notes* **4**: 149.
- Martin F, Gianinazzi-Pearson V, Hijri M, Lammers P, Requena N, Sanders I R, Shachar-Hill Y, Shapiro H, Tuskan GA, Young JPW. 2008.** The long hard road to a completed *Glomus intraradices* genome. *New Phytologist* **180**:747-50.
- Mathimaran N, Falquet L, Ineichen K, Picard C, Redecker D, Boller T, Wiemken A. 2008.** Microsatellites for disentangling underground networks: strain-specific identification of *Glomus intraradices*, an arbuscular mycorrhizal fungus. *Fungal Genetics and Biology* **45**: 812–7.
- McCormick KP, Willmann MR, Meyers BC. 2011.** Experimental design, preprocessing, normalization and differential expression analysis of small RNA sequencing experiments. *Silence* **2**: 2.
- Metzker ML. 2010.** Sequencing technologies - the next generation. *Nature reviews. Genetics* **11**: 31–46.
- Meyers BC, Axtell MJ, Bartel B, Bartel David P, Baulcombe D, Bowman JL, Cao X, Carrington JC, Chen X, Green PJ, et al. 2008.** Criteria for annotation of plant MicroRNAs. *The Plant cell* **20**: 3186–90.
- Mosse B. 1988.** Some studies relating to “independent” growth of vesicular–arbuscular endophytes. *Canadian Journal of Botany* **66**: 2533–2540.
- Mukherjee A, Ané J-M. 2011.** Germinating spore exudates from arbuscular mycorrhizal fungi: molecular and developmental responses in plants and their regulation by ethylene. *Molecular plant-microbe interactions : MPMI* **24**: 260–270.
- Nacry P, Canivenc G, Muller B, Azmi A, Van Onckelen H, Rossignol M, Dumas P. 2005.** A Role for Auxin Redistribution in the Responses of the Root System Architecture to Phosphate Starvation in *Arabidopsis*. *Plant Physiology* **138**: 2061–2074.
- Nadimi M, Beaudet D, Forget L, Hijri Mohamed, Lang B Franz. 2012.** Group I intron-mediated trans-splicing in mitochondria of *Gigaspora rosea*, and a robust phylogenetic affiliation of arbuscular mycorrhizal fungi with Mortierellales. *Molecular biology and evolution* **29**:2199-210.

- Nakazono M, Kanno A, Tsutsumi N, Hirai A. 1994.** Palindromic repeated sequences (PRSs) in the mitochondrial genome of rice: evidence for their insertion after divergence of the genus *Oryza* from the other Gramineae. *Plant Molecular Biology* **24**: 273–281.
- Navarro L, Dunoyer P, Jay F, Arnold B, Dharmasiri N, Estelle M, Voinnet O, Jones JDG. 2006.** A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. *Science* **312**: 436–439.
- Nedelcu AM, Lee R W. 1998.** Short repetitive sequences in green algal mitochondrial genomes: potential roles in mitochondrial genome evolution. *Molecular Biology and Evolution* **15**: 690–701.
- Nilsson L, Müller R, Nielsen TH. 2010.** Dissecting the plant transcriptome and the regulatory responses to phosphate deprivation. *Physiologia plantarum* **139**: 129–43.
- Noirot Céline, Gaspin C, Schiex T, Gouzy Jérôme. 2008.** LeARN: a platform for detecting, clustering and annotating non-coding RNAs. *BMC bioinformatics* **9**: 21.
- Oehl F, Sieverding E, Palenzuela J, Ineichen Kurt, Alves da silva G. 2011.** Advances in Glomeromycota taxonomy and classification. *IMA Fungus* **2**: 191–199.
- Olah B, Briere C, Becard G, Denarie J, Gough C. 2005.** Nod factors and a diffusible factor from arbuscular mycorrhizal fungi stimulate lateral root formation in *Medicago truncatula* via the DMI1/DMI2 signalling pathway. *Plant Journal* **44**: 195–207.
- Oldroyd G E D, Long S R. 2003.** Identification and characterization of nodulation-signaling pathway 2, a gene of *Medicago truncatula* involved in Nod factor signaling. *Plant Physiology* **131**: 1027–1032.
- Paquin B, Laforest M-J, Forget L, Roewer I, Wang Z, Longcore J, Lang B F. 1997.** The fungal mitochondrial genome project: evolution of fungal mitochondrial genomes and their gene expression. *Current Genetics* **31**: 380–395.
- Parniske M. 2008.** Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nature reviews. Microbiology* **6**: 763–75.
- Paszkowski U, Boller T. 2002.** The growth defect of *Irt1*, a maize mutant lacking lateral roots, can be complemented by symbiotic fungi or high phosphate nutrition. *Planta* **214**: 584–590.
- Pawlowska T, Taylor J. 2004.** Organization of genetic variation in individuals of arbuscular mycorrhizal fungi. *Nature* **427**: 733–7.

- Pelin A, Pombert J-F, Salvioli A, Bonen L, Bonfante Paola, Corradi Nicolas. 2012.** The mitochondrial genome of the arbuscular mycorrhizal fungus *Gigaspora margarita* reveals two unsuspected trans-splicing events of group I introns. *The New phytologist* **194**:836-45.
- Penmetsa R V, Frugoli JA, Smith LS, Long S R. 2003.** Dual Genetic Pathways Controlling Nodule Number in *Medicago truncatula*. *Plant* **131**: 998–1008.
- Pinard R, De Winter A, Sarkis GJ, Gerstein MB, Tartaro KR, Plant RN, Egholm M, Rothberg JM, Leamon JH. 2006.** Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC genomics* **7**: 216.
- Pumplin N, Mondo SJ, Topp S, Starker CG, Gantt JS, Harrison Maria J. 2010.** *Medicago truncatula* Vapyrin is a novel protein required for arbuscular mycorrhizal symbiosis. *The Plant journal : for cell and molecular biology* **61**: 482–94.
- Pérez Torres CA, López Bucio J, Herrera Estrella L. 2009.** Low phosphate signaling induces changes in cell cycle gene expression by increasing auxin sensitivity in the *Arabidopsis* root system. *Plant signaling & behavior* **4**: 781–783.
- Pérez-Tienda J, Valderas a, Camañes G, García-Agustín P, Ferrol N. 2012.** Kinetics of NH₄⁺ uptake by the arbuscular mycorrhizal fungus *Rhizophagus irregularis*. *Mycorrhiza* **22**: 485-91.
- Raab PA, Brennwald A, Redecker D. 2005.** Mitochondrial large ribosomal subunit sequences are homogeneous within isolates of *Glomus* (arbuscular mycorrhizal fungi, Glomeromycota). *Mycological Research* **109**: 1315–22.
- Radutoiu S, Madsen LH, Madsen EB, Felle HH, Umehara Y, Grønlund M, Sato S, Nakamura Y, Tabata S, Sandal N, et al. 2003.** Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* **425**: 585–92.
- Redecker D, Kodner R, Graham LE. 2000.** Glomalean fungi from the Ordovician. *Science* **289**: 1920–1.
- Remy W, Taylor TN, Hass H, Kerp HC-P. 1994.** Four hundred-million-year-old vesicular arbuscular mycorrhizae. *Proceedings of the National Academy of Science U S A* **91**: 11841–3.
- Riely BK, Ané J-M, Penmetsa RV, Cook DR. 2004.** Genetic and genomic analysis in model legumes bring Nod-factor signaling to center stage. *Current opinion in plant biology* **7**: 408–13.

- Rinaldi A, Comandini O, Kuyper T. 2008.** Ectomycorrhizal fungal diversity: separating the wheat from the chaff. *Fungal Divers* **33**: 1-45.
- Rosendahl S, Stukenbrock E. 2004.** Community structure of arbuscular mycorrhizal fungi in undisturbed vegetation revealed by analyses of LSU rDNA sequences. *Molecular ecology* **13**: 3179–86.
- Ruiz-Ferrer V, Voinnet Olivier. 2009.** Roles of Plant Small RNAs in Biotic Stress Responses. *Annual Review of Plant Biology* **60**: 485–510.
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000.** Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944–5.
- Röhr H, Kües U, Stahl U. 1998.** Organelle DNA of plants and fungi: inheritance and recombination. *Progress in Botany* **60**: 39–87.
- Saccone C, Gissi C, Lanave C, Larizza A, Pesole G, Reyes A. 2000.** Evolution of the mitochondrial genetic system: an overview. *Gene* **261**: 153–159.
- Saghai-Marroof MA, Soliman KM, Jorgensen RA, Allard RW. 1984.** Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Science U S A* **81**: 8014–8.
- Schwarzott D, Walker C, Schüssler A. 2001.** Glomus, the largest genus of the arbuscular mycorrhizal fungi (Glomales), is nonmonophyletic. *Mol Phylogenet Evol* **21**: 190–7 ST – Glomus, the largest genus of the arbus.
- Schüßler A, Walker C. 2010.** *The Glomeromycota: a species list with new families and new genera*. Gloucester: The Royal Botanic Garden Edinburgh, The Royal Botanic Garden Kew, Botanische Staatssammlung Munich, and Oregon State University.
- Schüßler A, Schwarzott Daniel, Walker Christopher. 2001.** A new fungal phylum, the Glomeromycota: phylogeny and evolution. *Mycological Research* **105**: 1413–1421.
- Selosse M a, Le Tacon F. 1998.** The land flora: a phototroph-fungus partnership? *Trends in ecology & evolution* **13**: 15–20.
- Shin R, Burch AY, Huppert K a, Tiwari SB, Murphy AS, Guilfoyle TJ, Schachtman DP. 2007.** The Arabidopsis transcription factor MYB77 modulates auxin signal transduction. *The Plant cell* **19**: 2440–2453.
- Shutt TE, Gray Michael W. 2006.** Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends in Genetics* **22**: 90–95.

- Siciliano V, Genre A, Balestrini Raffaella, Cappellazzo G, DeWit PJGM, Bonfante Paola. 2007.** Transcriptome analysis of arbuscular mycorrhizal roots during development of the prepenetration apparatus. *Plant physiology* **144**: 1455–1466.
- Sieberer BJ, Chabaud Mireille, Timmers AC, Monin A, Fournier J, Barker David G. 2009.** A nuclear-targetedameleon demonstrates intranuclear Ca²⁺ spiking in *Medicago truncatula* root hairs in response to rhizobial nodulation factors. *Plant physiology* **151**: 1197–206.
- Simon L, Bousquet J, Lévesque RC, Lalonde M. 1993.** Origin and diversification of endomycorrhizal fungi and coincidence with vascular land plants. *Nature* **363**: 67–69.
- Simon S a, Meyers BC, Sherrier DJ. 2009.** MicroRNAs in the rhizobia legume symbiosis. *Plant physiology* **151**: 1002–8.
- Smith FA, Grace EJ, Smith SE. 2009.** More than a carbon economy: nutrient trade and ecological sustainability in facultative arbuscular mycorrhizal symbioses. *The New phytologist* **182**: 347–58.
- Smith S, Read D. 2008.** *Mycorrhizal symbiosis. 3rd edition.* Academic press.
- Snoek ISI, Steensma HY. 2007.** Factors involved in anaerobic growth of *Saccharomyces cerevisiae*. *Yeast*: 1–10.
- Spanu PD, Abbott JC, Amselem Joelle, Burgis T a, Soanes DM, Stüber K, Ver Loren van Themaat E, Brown JKM, Butcher S a, Gurr SJ, et al. 2010.** Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science (New York, N.Y.)* **330**: 1543–6.
- Splivallo R, Fischer U, Göbel C, Feussner I, Karlovsky P. 2009.** Truffles regulate plant root morphogenesis via the production of auxin and ethylene. *Plant physiology* **150**: 2018–2029.
- St-Arnaud M, Hamel C, Vimard B, Caron M, Fortin JA. 1996.** Enhanced hyphal growth and spore production of the arbuscular mycorrhizal fungus *Glomus intraradices* in an in vitro system in the absence of host roots. *Mycological Research* **100**: 328–332.
- Stockinger H, Walker C, Schüssler A. 2009.** “*Glomus intraradices* DAOM197198”, a model fungus in arbuscular mycorrhiza research, is not *Glomus intraradices*. *New Phytologist* **183**: 1176–87.
- Stracke S, Kistner C, Yoshida S, Mulder L, Sato S, Kaneko T, Tabata S, Sandal N, Stougaard J, Szczyglowski K, et al. 2002.** A plant receptor-like kinase required for both bacterial and fungal symbiosis. *Nature* **417**: 959–962.

- Stukenbrock E, Rosendahl S. 2005.** Clonal diversity and population genetic structure of arbuscular mycorrhizal fungi (*Glomus* spp.) studied by multilocus genotyping of single spores. *Molecular ecology* **14**: 743–52.
- Subramanian S, Fu Y, Sunkar R, Barbazuk WB, Zhu J-K, Yu O. 2008.** Novel and nodulation-regulated microRNAs in soybean roots. *Bmc Genomics* **9**: 160.
- Swart EC, Nowacki M, Shum J, Stiles H, Higgins BP, Doak TG, Schotanus K, Magrini VJ, Minx P, Mardis ER, et al. 2011.** The *Oxytricha trifallax* mitochondrial genome. *Genome Biology and Evolution* **4**: 136–54.
- Swofford DL. 2003.** Phylogenetic Analysis Using Parsimony (PAUP).
- Sykorova Z, Borstler B, Zvolenska S, Fehrer J, Gryndler M, Vosatka M, Redecker D. 2011.** Long-term tracing of *Rhizophagus irregularis* isolate BEG140 inoculated on *Phalaris arundinacea* in a coal mine spoil bank, using mitochondrial large subunit rDNA markers. *Mycorrhiza* **22**: 69–80.
- Sędziewska K a, Vetter K, Bode R, Baronian Keith, Watzke Roland, Kunze Gotthard. 2012.** GiFRD encodes a protein involved in anaerobic growth in the arbuscular mycorrhizal fungus *Glomus intraradices*. *Fungal genetics and biology : FG & B.* **49**: 313–21.
- Sędziewska KA, Fuchs J, Temsch EM, Baronian K, Watzke R, Kunze G. 2011.** Estimation of the *Glomus intraradices* nuclear DNA content. *New Phytologist* **192**: 794–7.
- Tamasloukht M, Sejalou-Delmas N, Kluever A, Jauneau A, Roux C, Becard G, Franken PC-166906. 2003.** Root factors induce mitochondrial-related gene expression and fungal respiration during the developmental switch from asymbiosis to presymbiosis in the arbuscular mycorrhizal fungus *Gigaspora rosea*. *Plant Physiol* **131**: 1468–78.
- Tamura K, Nei M, Kumar S. 2004.** Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Science U S A* **101**: 11030–5.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011.** MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**: 2731–9.
- Theissen U, Hoffmeister M, Grieshaber M, Martin W. 2003.** Single eubacterial origin of eukaryotic sulfide:quinone oxidoreductase, a mitochondrial enzyme conserved from the early evolution of eukaryotes during anoxic and sulfidic times. *Molecular biology and evolution* **20**: 1564–74.

- Thiéry O, Börstler B, Ineichen K, Redecker D. 2010.** Evolutionary dynamics of introns and homing endonuclease ORFs in a region of the large subunit of the mitochondrial rRNA in *Glomus* species (arbuscular mycorrhizal fungi, Glomeromycota). *Molecular Phylogenetic and Evolution* **55**: 599–610.
- Tisserant E, Kohler A, Dozolme-Seddas P, Balestrini R, Benabdellah K, Colard A, Croll D, Da Silva C, Gomez S K, Koul R, et al. 2012.** The transcriptome of the arbuscular mycorrhizal fungus *Glomus intraradices* (DAOM 197198) reveals functional tradeoffs in an obligate symbiont. *New Phytologist* **193**: 755–69
- Toedling J, Servant N, Ciaudo C, Farinelli L, Voinnet Olivier, Heard E, Barillot E. 2012.** Deep-Sequencing Protocols Influence the Results Obtained in Small-RNA Sequencing. *PloS one* **7**: e32724.
- Torriani SFF, Goodwin SB, Kema GHJ, Pangilinan JL, McDonald BA. 2008.** Intraspecific comparison and annotation of two complete mitochondrial genome sequences from the plant pathogenic fungus *Mycosphaerella graminicola*. *Fungal Genetics and Biology* **45**: 628–637.
- Trappe J. 1987.** Phylogenetic and ecologic aspects of mycotrophy in the angiosperms from an evolutionary standpoint. *Ecophysiology of va mycorrhizal plants*. Boca Raton, FL, USA: CRC Press, 5–25.
- Umehara M, Hanada A, Yoshida S, Akiyama Kohki, Arite T, Takeda-Kamiya N, Magome H, Kamiya Y, Shirasu K, Yoneyama K, et al. 2008.** Inhibition of shoot branching by new terpenoid plant hormones. *Nature* **455**: 195–200.
- Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. 2002.** Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology* **3**: 34.
- VanEtten H, Matthews D, Matthews PS. 1989.** Phytoalexin detoxification: importance for pathogenicity and practical implications. *Annual Review of Phytopathology* **27**: 143–164.
- Varkonyi-Gasic E, Wu R, Wood M, Walton EF, Hellens RP. 2007.** Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. *Plant methods* **3**: 12.
- Wais RJ, Galera C, Oldroyd G, Catoira R, Penmetsa R V, Cook D, Gough C, Denarié J, Long S R. 2000.** Genetic analysis of calcium spiking responses in nodulation mutants of *Medicago truncatula*. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 13407–12.

Ware SB. 2006. *Aspects of sexual reproduction in Mycosphaerella species on wheat and barley: genetic studies on specificity, mapping, and fungicide resistance.* Wageningen Universiteit.

Weidmann S, Sanchez L, Descombin J, Chatagnier O, Gianinazzi Silvio, Gianinazzi-Pearson Vivienne. 2004. Fungal elicitation of signal transduction-related plant genes precedes mycorrhiza establishment and requires the *dmi3* gene in *Medicago truncatula*. *Molecular plant-microbe interactions : MPMI* **17**: 1385–93.

Westermann B. 2002. Merging mitochondria matters: cellular role and molecular machinery of mitochondrial fusion. *EMBO reports* **3**: 527–31.

Young N, Oldroyd GED, Geurts Rene. 2011. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**: 520–524.

Yu H, Wu J, Xu N, Peng M. 2007. Roles of F-box Proteins in Plant Hormone Responses. *Acta Biochimica et Biophysica Sinica* **39**: 915–922.

Zaragoza Joseph Diegoli, Marta Lin, Dawei MVF, Arbustini E. Mitochondrial DNA Variant Discovery and Evaluation in Human Cardiomyopathies through Next-Generation Sequencing. *PLoS ONE* **5**: e12295.

Zhang Q, Blaylock L a, Harrison Maria J. 2010. Two *Medicago truncatula* half-ABC transporters are essential for arbuscule development in arbuscular mycorrhizal symbiosis. *The Plant cell* **22**: 1483–97.

Zhu YG, Smith S, Barritt A. 2001. Phosphorus (P) efficiencies and mycorrhizal responsiveness of old and modern wheat cultivars. *Plant and Soil*: 249–255.

ANNEXES

Annexe 1 : Liste des annotations réalisées dans le cadre de la tâche d'annotation
«Production d'énergie et mitochondrie» chez *Rhizophagus irregularis*.

Gènes annotés	
2-oxoisovalerate dehydrogenase subunit beta, mitochondrial precursor	cytochrome c heme lyase
3-isopropylmalate dehydratase large subunit	Cytochrome c oxidase polypeptide IV, mitochondrial precursor
3-oxoadipate CoA-transferase subunit B	Cytochrome c oxidase polypeptide V, mitochondrial precursor
Acetaldehyde dehydrogenase 2	cytochrome c oxidase polypeptide VI
Acetyl-CoA hydrolase	Cytochrome c oxidase polypeptide VIa
Aconitate hydratase	Cytochrome c oxidase polypeptide VIb
Acyl carrier protein mito precursor	cytochrome c1 heme lyase
ADP ATP carrier prot	Cytochrome C1, heme protein, mitochondrial precursor
Aldehyde dehydrogenase	D-arabinose 1-dehydrogenase
Alternative oxidase	Dihydrolipoyl dehydrogenase
ATP synthase protein 9, mitochondrial precursor	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex
ATP synthase subunit 4, mitochondrial precursor	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial precursor
ATP synthase subunit 5, mitochondrial precursor	D-lactate dehydrogenase [cytochrome] 2
ATP synthase subunit alpha, mitochondrial precursor	DNA repair protein RAD51 homolog 2
ATP synthase subunit beta mito precursor	electron transfer flavoprotein subunit alpha, mitochondrial precursor
ATP synthase subunit d, mitochondrial	Electron transfer flavoprotein subunit beta (Beta-ETF)
ATP synthase subunit delta, mitochondrial precursor	Electron transfer flavoprotein-ubiquinone oxidoreductase, mitochondrial precursor
ATP synthase subunit epsilon	External NADH-ubiquinone oxidoreductase 1
ATP synthase subunit gamma, mitochondrial precursor	External NADH-ubiquinone oxidoreductase 2
ATP-citrate synthase subunit 1	Flavo-hemoprotein
BAG family molecular chaperone regulator 1	Fumarate hydratase, mitochondrial precursor
band 7 family protein R614	fumarate reductase
Calcium-binding mitochondrial carrier protein Aralar2	Fumarate reductase iron-sulfur subunit
Carboxyvinyl-carboxyphosphonate phosphorylmutase	Galactose-1-phosphate uridylyltransferase
Carrier protein YMC1, mitochondrial precursor	Glycerol-3-phosphate dehydrogenase
CBP3-like protein	glycerol-3-phosphate dehydrogenase, mitochondrial precursor
Citrate synthase	Glyoxylate reductase/hydroxypyruvate reductase
Citrate synthase, mitochondrial precursor	Hydroxymethylglutaryl-CoA lyase, mitochondrial precursor
cytochrome b2, mito precursor	Iron sulfur cluster assembly protein 1, mitochondrial precursor
cytochrome b5	isocitrate lyase 113032
cytochrome b5 B11H24.095	Kynurenine 3-monooxygenase
Cytochrome b5 reductase 4	Lactobacillus shifted protein
Cytochrome b-c1 complex subunit 2, mitochondrial precursor	Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex
Cytochrome b-c1 complex subunit 5	Malate dehydrogenase
Cytochrome b-c1 complex subunit 6	Malate synthase, glyoxysomal
Cytochrome b-c1 complex subunit 7	Mitochondrial 2-oxoglutarate-malate carrier protein
Cytochrome b-c1 complex subunit 9	Mitochondrial acidic protein MAM33
Cytochrome c	Mitochondrial carnitine-acylcarnitine carrier protein
	mitochondrial carrier C1682.09c

Annexe 1 (suite): Liste des annotations réalisées dans le cadre de la tâche d'annotation «Production d'énergie et mitochondrie» chez *Rhizophagus irregularis*.

Gènes annotés	
mitochondrial carrier C4G9.20c	norsolorinic acid reductase
mitochondrial carrier C688.09	Nuclear protein SNF4
mitochondrial carrier C8C9.12c	oxidoreductase KIAA1576
Mitochondrial carrier protein LEU5	oxidoreductase yajO
mitochondrial carrier protein YHM1-SHM1	Peroxisomal membrane protein PMP47B
mitochondrial carrier YIL006W	protein C1556.08c
mitochondrial carrier YPR011C	protein C4D7.02c
Mitochondrial phosphate carrier protein	Pyruvate carboxylase
mitochondrial phosphate carrier protein bis	Pyruvate dehydrogenase E1 component subunit alpha
Mitochondrial substrate carrier family protein H	Pyruvate dehydrogenase E1 component subunit beta
Mitochondrial thiamine pyrophosphate carrier 1	quinone oxidoreductase
NAD(P)H-quinone oxidoreductase chain 2	Quinone oxidoreductase-like protein 2 homolog
NADH dehydrogenase 1 alpha subcomplex subunit 12	Replication factor C subunit 5
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 13	Reticulon-4-interacting protein 1 homolog, mitochondrial precursor
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 2	Retinal dehydrogenase 1
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 6	S-adenosylmethionine mitochondrial carrier protein
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 7	SCO1 mitochondrial precursor
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 9 (NADH-ubiquinone oxidoreductase B22 subunit) (Complex I-B22) (CI-B22)	SDS23
NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial precursor	SHY1
NADH dehydrogenase [ubiquinone] iron-sulfur protein 7, mitochondrial precursor	Solute carrier family 25 member 35
NADH-cytochrome b5 reductase	Spermidine synthase
NADH-ubiquinone oxidoreductase 10.5 kDa subunit	Stomatin-like protein 2 (SLP-2)
NADH-ubiquinone oxidoreductase 12 kDa subunit, mitochondrial precursor	Succinate dehydrogenase [ubiquinone] cytochrome b small subunit, mitochondrial precursor
NADH-ubiquinone oxidoreductase 14.8 kDa subunit	Succinate dehydrogenase [ubiquinone] iron-sulfur subunit
NADH-ubiquinone oxidoreductase 19.3 kDa subunit, mitochondrial precursor	Succinate dehydrogenase flavoprotein subunit
NADH-ubiquinone oxidoreductase 20.8 kDa subunit	Succinate-fumarate mitochondrial transporter
NADH-ubiquinone oxidoreductase 21 kDa subunit	succinate-semialdehyde dehydrogenase
NADH-ubiquinone oxidoreductase 23 kDa subunit	succinyl-CoA ligase subunit alpha, mitochondrial precursor
NADH-ubiquinone oxidoreductase 29.9kDa subunit	Succinyl-CoA ligase subunit beta, mitochondrial precursor
NADH-ubiquinone oxidoreductase 30.4 kDa subunit, mitochondrial precursor	Sulfide quinone oxidoreductase, mitochondrial precursor
NADH-ubiquinone oxidoreductase 40 kDa subunit, mitochondrial precursor	Sulfite reductase [NADPH] flavoprotein component
NADH-ubiquinone oxidoreductase 51 kDa subunit	Surfeit locus protein 1
NADH-ubiquinone oxidoreductase 78 kDa subunit, mitochondrial precursor	trans-2-enoyl-CoA reductase, mitochondrial precursor
NADP-dependent malic enzyme	tRNA wybutosine-synthesizing protein 1 homolog
NADPH dehydrogenase C23G7.10c	UDP-glucuronosyltransferase 1-4 precursor
NADPH--cytochrome P450 reductase	Voltage-gated potassium channel subunit beta-1
NADPH-dependent diflavin oxidoreductase 1	Zinc-type alcohol dehydrogenase-like protein C1773.06c
Nitrate reductase	Zinc-type alcohol dehydrogenase-like protein C2E1P3.01

Annexe 2: Cartographie des lectures de séquençage (Traits colorés en partie basse de la figure) sur le génome mitochondrial de DAOM197198 (Trait noir en haut). Le diagramme rose représente le niveau d'expression des gènes.



Annexe 3 : Protocole de réalisation des Northern blot.
(H. Vaucheret - A. Mallory - IJPB Versailles)

LMW RNA gel blot

Prepare the gel (TBE-15% Acryl-Urea 7M)

Clean the plates (ethanol)

Coat plates with Sigma-cote (some drops and spread with sopalin)

Assemble plates together

Weight 20.25 g urea in a falcon 50 mL

Add 16.8 mL 19:1acrylamide 40%

Add 4.5 mL TBE 10X

Heat 20 sec in microwave

Just after, agitate the volume manually until urea looks fully dissolved

Add 7.5 mL water and agitate the volume until solution appears homogeneous

Load the volume into a 50 mL syringe and filter through a 0.45 µm filter into a new tube

Add 150 µL APS 10%. Agitate } Avoid bubbles formation
Add 45 µL TEMED. Agitate }

Pour between plates, place the comb and tighten with forceps

Use after 45 mn

Running the gel

Assemble gel onto the gel running apparatus with an aluminum plate just below the wells

Clean urea out of wells with a syringe

Remove air bubbles at the bottom of the gel with a syringe

Prerun the gel at 12W (about 450V and 25 mA) for 45 min (0.5X TBE running buffer)

Prepare samples of 5-30 µg of total RNA (volume of 15 µL or less, less is better!). Bring all samples to the same volume with water and add an equal volume of 2X loading dye (LD)

LD contains:

0.025% BPB

0.025% XC

25 mM EDTA pH8

8M urea

Aliquot and store at 4°C

Heat samples in 85°C heat block for 2 min, spin down. While the samples are heating, stop gel prerun and clean urea out of the wells again with a syringe. Use a beveled tip to load onto warm gel. Run gel at 2W (about 280V and 7 mA) for 45 minutes. Then increase running speed to 12W (about 600V and 20 mA) and run gel until the BPB dye just runs off the gel (Xylene cyanol dye should still be in the middle of the gel) – about 1.5h

Separate gel onto saran wrap, cut away any excess gel (remove the wells of the gel). Stain the gel with EtBr in 0.5X TBE for 5 min, and look for tRNA and 5S rRNA bands (78 and 120 nt) for quality of RNA prep and for visual loading control.

Transfer gel to membrane

Cut nylon membrane (Hybond NX exclusively!!!) and 4 sheets of 3MM Whatman chromatography filter paper slightly larger than the size of the gel. Pre-wet membrane and filter paper in 0.5X TBE. Cut one corner of the membrane to orient the loading of the gel.

Place soaked membrane on the gel that sits on saran wrap, and position membrane to make sure that overlaps all lanes containing samples. Place the cut edge of the membrane on the side of the gel that has the first lane to be able to orient the membrane later. Put 2 sheets of prewet whatman paper on top of the membrane. Flip the sandwich over to remove the saran wrap. Place 2 pieces of prewet whatman paper on the gel. Attach transfer apparatus to a Hoefer EPS 2A200 power supply that handles high current (Amersham Pharmacia). Transfer at constant current (2 mA/cm²) for 1 hour. The voltage may get higher towards end of run.

IMPORTANT: make sure that the gel and membrane are oriented properly during the transfer so that the RNA in the gel is transferred to the membrane !!!

Remove gel sandwich from apparatus, and flip sandwich back over so that membrane sits on top of gel. Write on the top of the membrane with pencil the date to orient the front and back of the membrane. Save the gel and restain with EtBr for 5 min. Take a picture of restained gel, all the RNA < 200nt should be transferred, while larger RNAs just at the well limit will not transfer.

EDC-mediated crosslinking-Will immobilize RNA with 5' or 3' phosphate groups on the membrane.

Prepare in a casserole dish one piece of whatman membrane that has been prewetted with the following solution (PREPARE THIS SOLUTION FRESH EACH TIME).

Solution: 122.5 μ L of 12.5M 1-methylimidazole Sigma cat.336092 to 11 mL of water. pH to 8.0 with 200 μ L 1N HCl. Add 0.373g EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide Sigma cat.E7750 and bring the final volume to 12 mL with water. (EDC is very sensitive to temperature changes, so work quickly to measure the correct amount and return it to the freezer promptly). Mix well!

Once the whatman paper is saturated with the solution pour off the excess liquid. Lay the membrane on the whatman paper with the RNA side facing upward (away from whatman). Cover the dish with plastic wrap and place at 60°C for 1.5h.

After 1.5h, remove the whatman paper and wash the membrane with water 3 times to remove excess EDC

Bake the membrane for 30 min at 80°C to dry for storage. Store dry blot in a sealed bag or wrapped in saran wrap at -20°C until you are ready for probing

Oligo probe for LMW RNA in BB buffer

Prehybridization, labeling and hybridization

Note: you can prehyb, label the probe and start the hybridization all in one day by prehybing in the morning, labeling the probe in the afternoon and starting hybridization for overnight, or you can prehyb the night before and do labeling and hyb the following day. Also, the volume amounts listed here are for small hyb bottles (50 mL capacity). If doing large northern requiring larger bottles, double all volumes.

PreHyb/Hyb solution

<u>Final:</u>	<u>start</u>	<u>50 mL</u>	<u>400mL</u>
5X SSC	20X	12.5 mL	100 mL
20mM Na ₂ HPO ₄ pH7.2	1M	1 mL	8 mL
7% SDS	10%	35 mL	280 mL
2X Denhardt's solution	100X	1.5 mL	12 mL

For each blot, prepare 25 mL of warm pre-hyb solution (50°C). Denature 1 mg of sheared salmon sperm DNA (100µL of 10mg/mL) by boiling for 5 min. Either immediately chill on ice or add to warm pre-hyb, and add to prewarmed hybridization tube. After the solution has coated the tube, add blot so that the RNA side is facing the hyb solution. Place the tube back in hyb-oven and rotate at 50°C, making sure membrane is not creased. Prehyb in oven for at least 3 hours, but can go overnight.

Labeling

Labeling reaction in 20µL (T4 euromedex cat.09-10311)

dH ₂ O	9 µL
10 µM oligo	2 µL
10 X reaction buffer B	2 µL
[γ- ³² P]-ATP (6000Ci/mmol)	2 µL (150 µCi/µL)
24% PEG 6000 solution	4 µL
T4 kinase (10U/µL)	1 µL

Incubate for 1 hour at 37°C

Add 2.5 µL 0.5M EDTA (pH8) and 30 µL water to labeling reaction and purify probe from unincorporated label with G-25 microspin columns (GE). Add eluted radiolabeled oligo to bottle.

Washing

Non stringent wash solution

<u>Final</u>	<u>start</u>	<u>200 mL</u>	<u>500 ML</u>
3X SSC	20X	30 mL	75 mL
25 mM NaH ₂ PO ₄ pH7.5	1 M	5 mL	12.5 mL
5% SDS	10%	100 mL	250 mL
dH ₂ O		45 mL	115 mL

stringent wash solution

<u>Final</u>	<u>start</u>	<u>200 mL</u>	<u>500 ML</u>
1X SSC	20X	10 mL	25 mL
1% SDS	10%	20 mL	50 mL
dH ₂ O		170 mL	425 mL

dump hot hybridization solution in a designated radioactive liquid waste container. Add 20 mL of non stringent wash to each bottle. Cap and rotate in Hyb Oven for 10 min at 50°C. Dump this first wash also in radioactive liquid waste container. Repeat non stringent wash, 20 mL, 10 min.

Finally, wash once with stringent wash, 20 mL, 10 min et 50°C. Can repeat stringent wash if there is a lot of background signal.

Now remove blot from hyb tube, wrap securely with saran wrap and expose to phosphoimaging screen.

Scan screen overnight or several days for signal optimization

Annexe 4 : Liste des microARN caractérisés dans les banques de symbioses tardives.

ID précurseur	ID Mature	Séquence du mature	Cleave land	Taille du précurseur	Taille du mature	Nb de reads
MicroARN nouvellement identifiés						
Mt3.5.1s03_r879	204DVAAXX:1:121:505:205	UCAGAAGAGUACA UAAACUGC		196	21	273
Mt3.5.1s05_r167	204DVAAXX:1:15:558:319	CUAUGUUAGGCUCAUUCGGUU		205	21	462
Mt3.5.1Chr6_r814	204DVAAXX:1:155:837:171	UUUCAAAAAGAUUUAGUUGGGA		147	21	1060
Mt3.5.1Chr1_r135	204DVAAXX:1:17:363:490	UCAUGAACAGCUUGAUCGGAG		181	21	1093
Mt3.5.1Chr6_r618	204DVAAXX:1:18:385:724	UCAGGAUUUGGGUCAUAACAA		329	21	1145
Mt3.5.1Chr3_r2073	204DVAAXX:1:198:375:549	AAAGUGAA CAAGAUUUCGGUA		217	21	31
Mt3.5.1Chr7_r930	204DVAAXX:1:204:797:733	UGUGUGUGAAACUUGGCAUUGG		107	22	22
Mt3.5.1s07_r59	204DVAAXX:1:23:876:465	UAGAAGAAGCUGUCAGAGAAG		300	21	178
Mt3.5.1Chr8_r822	204DVAAXX:1:24:103:868	UCCGUUGGAGAAAAAGGUUUU		212	21	607
Mt3.5.1Chr6_r55	204DVAAXX:1:24:665:363	CAGGAGAAGAAUGUAUGGCUU		210	21	274
Mt3.5.1Chr4_r255	204DVAAXX:1:26:850:133	UUGUUUGUGGAUGGCAGAAGGA		213	21	698
Mt3.5.1s05_r1131	204DVAAXX:1:3:796:213	UGAGGGGCGUUUGAAAGCGG		123	21	28039
Mt3.5.1Chr5_r2574	204DVAAXX:1:32:540:822	UUUUUCUUUCGUUGGUGCAAU		392	21	114
Mt3.5.1Chr3_r2219	204DVAAXX:1:32:914:588	UGGAAUGAAGACCAUGGUAUA		168	21	202
Mt3.5.1Chr7_r219	204DVAAXX:1:38:207:882	UCAGAUCUGGGAUUGCUGGA		409	21	140
Mt3.5.1Chr7_r217	204DVAAXX:1:38:891:306	AGAUGACUAUUGGGCAUGACAU		225	22	130
Mt3.5.1Chr0_r631	204DVAAXX:1:4:984:557	AUGGGAUAUGGGAGGACUGGU		125	21	3269
Mt3.5.1s02_r438	204DVAAXX:1:44:263:792	UUUGAAUGGAAGUACCUCCGU		413	21	915
Mt3.5.1Chr3_r9	204DVAAXX:1:48:383:785	UUAACGUAGAAGUCCUCCUAG		320	21	92
Mt3.5.1Chr0_r53	204DVAAXX:1:5:372:183	UUAGGAUUGGAUGGUGAUUAU		326	21	505
Mt3.5.1Chr4_r1455	204DVAAXX:1:5:777:691	UACAGAACGUUGAAUCAGGAA		98	21	1622
Mt3.5.1s05_r995	204DVAAXX:1:64:887:276	UACUGGACAAGUGGUGAGAAG	ok	108	21	514
Mt3.5.1s02_r1292	204DVAAXX:1:7:295:638	UUGGUGUCGUGUCGGUGUCAG		81	21	35
Mt3.5.1s07_r664	204DVAAXX:1:81:753:107	AA CAGCUUGAUCGGAGAGACA		167	21	308
Mt3.5.1Chr4_r1032	204DVAAXX:1:87:827:131	UUCUAAAUUGUUGGGAUUUC		138	21	42
Mt3.5.1Chr1_r542	204DVAAXX:2:127:348:645	AACGGUCGGAUUGAACGUGGC		124	21	277
Mt3.5.1Chr3_r1180	204DVAAXX:2:235:688:205	CGCUGCUGUAGUUCUUCGGC		306	21	120
Mt3.5.1s02_r906	204DVAAXX:3:189:935:39	UGAAGGUCGUCACUAUUUUA		251	21	25
Mt3.5.1Chr3_r2394	207G6AAXX:1:71:231:644	UGUGUCGGAGUACAAGUUAG		105	20	143
Mt3.5.1Chr7_r1614	207G6AAXX:2:27:597:723	UCAGGCGGGGAGCAAUAUG		125	21	10
Mt3.5.1s06_r1318	E4D3Z3Y01BILN1	UUCUUAUACGUUGACUCUUUC		130	21	2283
Mt3.5.1s06_r728	E4D3Z3Y01BRB91	UUACAAUCCGCAUUGAUUAU		107	21	3524
Mt3.5.1s05_r227	E4D3Z3Y01BZ8WO	UCCUAAGGAGCAGAUAGCAUU		96	21	12293
Mt3.5.1s05_r226	E4D3Z3Y01CTAU8	UUUUCGGGAAAGUGGAGUACA		399	21	19610
Mt3.5.1Chr4_r158	E4D3Z3Y01EO1V5	AAUAUGACGGAGUGUAAAUGC		117	21	3605
Mt3.5.1Chr5_r472	E8W526M03F18QX	UGCGUAUACUGUGCGGUAAG		145	21	439
Mt3.5.1Chr5_r2538	E8W526M03FP075	UGAGAUGAGCAGAAUUUCUU		437	21	2874
Mt3.5.1Chr3_r428	E8W526M03G9DZL	AUGAUUGGACAGUUAUGGG		182	21	1421
Mt3.5.1Chr5_r1771	E8W526M03GTK4Y	AUCAUGCUAUCCUUUGGAUU		142	21	824
Mt3.5.1Chr2_r861	E8W526M03GTL95	CAACAGGAGAAUUCGUCGGU		129	21	1566
Mt3.5.1Chr2_r1167	E8W526M04I757V	UACGACAGGAGAGAUGAUGCCG		109	21	707
Mt3.5.1s02_r1351	E8W526M04IB8GH	UCGGACCAGGCUUCGUUCCCC	ok	103	21	4904
Mt3.5.1s05_r393	E8W526M04IV7SK	CAAACUGA GUUUGCAAUCGUC		235	21	2900
Mt3.5.1Chr5_r86	E8W526M04JBO5D	UAGUUUGCAGAGUUUUGGUU		171	21	351
Mt3.5.1Chr6_r595	E8W526M04JPUFZ	UUGGUACUUCGUGAUUUGAU		107	21	336
Mt3.5.1Chr0_r966	E8W526M04JU7CT	UUGAAUUAUGAGAACUUGAAG	ok	126	21	39064
Mt3.5.1s05_r971	EVSUSMY01BFXHY	UCGUGGGAUUGUUAUUGGAC		102	21	5161
Mt3.5.1Chr5_r658	EVSUSMY01BO06G	UGGAAUGGAGGAUCAGGUAAA		151	21	59859
Mt3.5.1Chr2_r1059	EVSUSMY01C3BIZ	UUGAGCCGCGUCAAUAUCUUG	ok	127	21	679
Mt3.5.1s03_r137	EVSUSMY01C43QP	CGCCAAAGGAGAGUCGCCUG	ok	97	21	158
Mt3.5.1Chr7_r165	EVSUSMY01CBKZO	CAACAGGACUCGAGCAUUUCG		239	21	1033
Mt3.5.1Chr4_r2629	EVSUSMY01D5U7X	UGGGUCAUUGCUUGAACUGG		225	21	2186
Mt3.5.1s01_r761	EVSUSMY01D6WRK	UUGGGUGACAUGAUUGGGA		127	21	577
Mt3.5.1Chr6_r1111	EVSUSMY01D8S12	UUAUUGGUUCAAACGGUAGA		130	21	4907
Mt3.5.1s03_r234	EVSUSMY01E17TU	UCUCGUUCCAUAUUCGUCUA	ok	118	21	2642
Mt3.5.1s01_r23	EVSUSMY01E1RYV	AACCAGGCUCUGAUACCAUGA		122	21	105381
Mt3.5.1s07_r535	EVSUSMY01EAALD	UUAUCAGGGAAAUCACAGUU	ok	118	21	139059

Annexe 4 (suite) : Liste des microARN caractérisés dans les banques de symbioses tardives.

ID précurseur	ID Mature	Séquence du mature	Cleave land	Taille du précurseur	Taille du mature	Nb de reads
MicroARN identifiés dans miRBase 18 (autres espèces)						
Mt3.5.1Chr5_r15	ath-miR168a	UCGCUUGGUGCAGGUCGGGAA		122	21	234953
Mt3.5.1s04_r966	ath-miR172a	AGAAUCUUGAUGAUGCUGCAU	ok	132	21	51725
Mt3.5.1s06_r1087	ath-miR172c	AGAAUCUUGAUGAUGCUGCAG	ok	120	21	54680
Mt3.5.1s06_r1098	ath-miR319a	UUGGACUGAAGGGAGCUCCU	ok	213	21	32516
Mt3.5.1Chr2_r1846	bdi-miR166f	UCUCGGACCAGGCUUCAUUC	ok	195	21	240863
Mt3.5.1Chr1_r1415	bna-miR156a	UGACAGAAGAGAGUGAGCACA	ok	113	21	6269
Mt3.5.1Chr0_r165	gma-miR159b	AUUGGAGUGAAGGGAGCUCCA	ok	184	21	20800
Mt3.5.1Chr3_r1253	gma-miR171b-3p	CGAGCCGAAUCAUAUCACUC	ok	302	21	35793
Mt3.5.1s06_r459	osa-miR167d	UGAAGCUGCCAGCAUGAUCUG		107	21	61975
Mt3.5.1Chr0_r298	ptc-miR160h	UGCUCUGGCUCCUGCAUGCCA	ok	108	21	144
Mt3.5.1s01_r1112	ptc-miR167f	UGAAGCUGCCAGCAUGAUCUU	ok	112	21	147703
Mt3.5.1Chr4_r1775	vvi-miR171f	UUGAGCCGCGCCAAUAUCACU	ok	118	21	11749
Mt3.5.1s02_r609	vvi-miR2111	UAAUCUGCAUCCUGAGGUCUA	ok	105	21	4254
MicroARN identifiés dans miRBase 18 (<i>M. truncatula</i>)						
Mt3.5.1Chr6_r606	mtr-miR156g	UUGACAGAAGAUAGAGGGCAC	ok	158	21	589737
Mt3.5.1s06_r1296	mtr-miR159a	UUUGGAUUGAAGGGAGCUCUA	ok	197	21	6094474
Mt3.5.1s04_r658	mtr-miR160	UGCUCUGGCUCCUGUAUGCCA	ok	118	21	25428
Mt3.5.1Chr2_r1815	mtr-miR160c	UGCUCUGGCUCCUGAAUGCCA	ok	111	21	531
Mt3.5.1Chr4_r1690	mtr-miR162	UCGAUAAACCUCUGCAUCCAG		129	21	16112
Mt3.5.1Chr1_r792	mtr-miR164a	UGGAGAAGCAGGGCACGUGCA	ok	125	21	258566
Mt3.5.1Chr8_r1611	mtr-miR166	UCGGACCAGGCUUCAUUC	ok	133	21	1609096
Mt3.5.1s02_r240	mtr-miR167	UGAAGCUGCCAGCAUGAUCUA	ok	147	21	28122
Mt3.5.1Chr2_r214	mtr-miR169c	CAGCCAAAGGGUGAUUUGCCGG	ok	116	21	1728
Mt3.5.1Chr2_r218	mtr-miR169d	AAGCCAAGGAUGACUUGCCGG	ok	272	21	2787
Mt3.5.1Chr2_r216	mtr-miR169q	UGAGCCAGGAUGACUUGCCGG	ok	130	21	1611
Mt3.5.1s05_r591	mtr-miR171b	UGAUUGAGCCGCGUCAUAUC	ok	104	21	389
Mt3.5.1Chr5_r1717	mtr-miR172	AGAAUCCUGAUGAUGCUGCAG	ok	136	21	18486
Mt3.5.1Chr2_r1326	mtr-miR2086	GACAUGAAUGCAGAACUGGAA		127	21	49360
Mt3.5.1Chr8_r284	mtr-miR2088b	UCAAUGUAAUUCUAGGUCUAC		115	21	13607
Mt3.5.1Chr7_r362	mtr-miR2111h	UAAUCUGCAUCCUGAGGUUUA	ok	103	21	12676
Mt3.5.1Chr5_r670	mtr-miR2119	UCAAAGGGAGGUGUGGAGUAG	ok	137	21	955
Mt3.5.1s07_r931	mtr-miR2592b	AAAUGCUUGAGUCCUGUUGUU		199	21	684
Mt3.5.1Chr4_r1210	mtr-miR2634	UUUAUUCUCAGUUUGUUGCUC		103	21	2419
Mt3.5.1Chr2_r1325	mtr-miR2651	UUUGAUUGGUAUGCCUGCAUU		147	21	618
Mt3.5.1Chr4_r1299	mtr-miR390	AAGCUCAGGAGGGAUAGCGCC		125	21	6769
Mt3.5.1Chr4_r1781	mtr-miR396b	UUCACAGCUUUCUUGAACUG	ok	121	21	193745
Mt3.5.1Chr7_r820	mtr-miR5204	GCUGGAAGGUUUUGUAGGAAC		96	21	17408
Mt3.5.1s04_r498	mtr-miR5232	UACAUGUCGCUCACCCUGAA		157	21	30219
Mt3.5.1s06_r834	mtr-miR5239	UGGGAGAAAAGAUGAAUGUG		123	21	88061
Mt3.5.1Chr5_r734	mtr-miR5558	UUUCCAAUUCUAAGUCUAUC		103	21	7659
Mt3.5.1Chr6_r34	mtr-miR5559	UACUUGGUGAAUUGUUGAUC		151	21	2917
Mt3.5.1Chr7_r1483	mtr-miR5561	CAUUUGGAGAGACAUAGACAA		109	21	3948

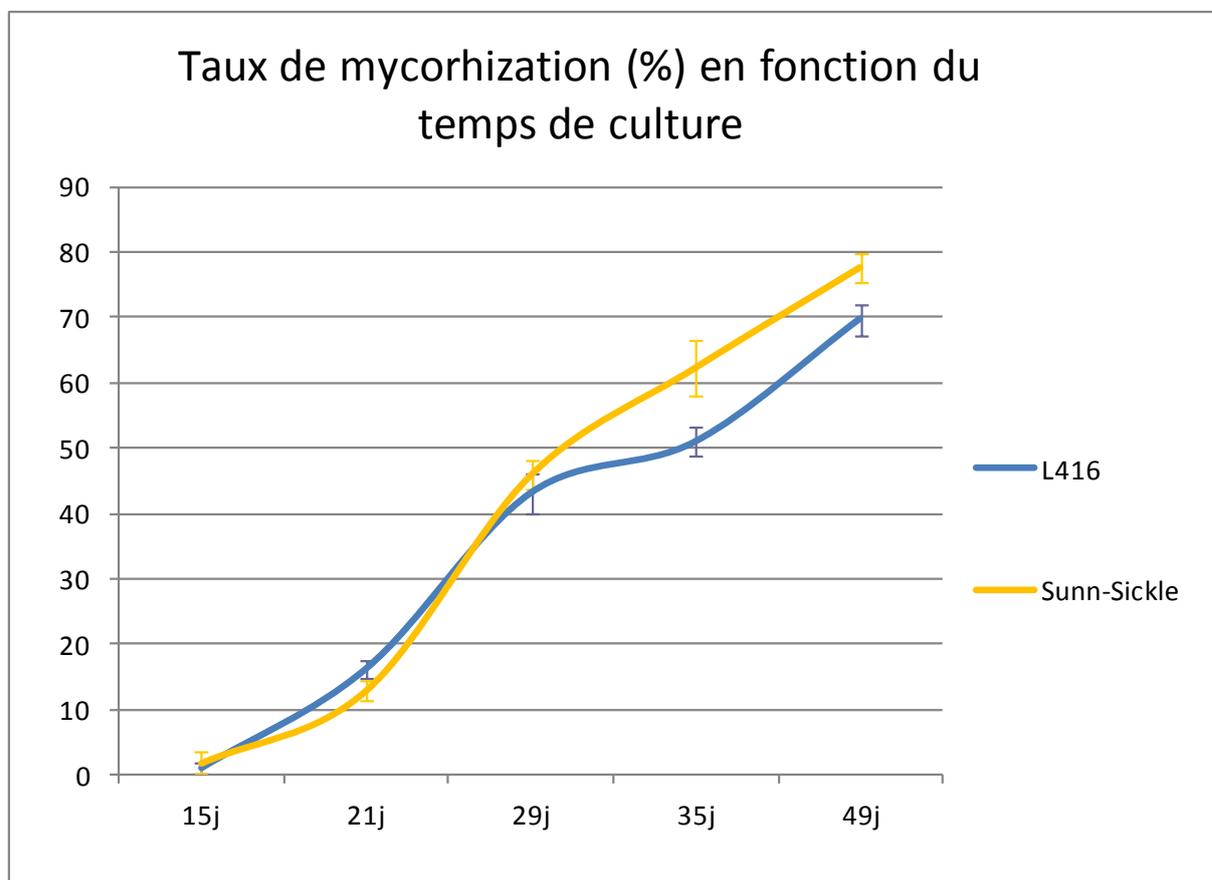
Annexe 5 : Données de quantification normalisées des microARN caractérisés dans les banques de symbioses tardives.

ID Mature	Expression normalisée	Expression normalisée	Expression normalisée	Expression normalisée
	R.i.	No R.i.	S.m.	No S.m.
MicroARN nouvellement identifiés				
204DVAAXX:1:121:505:205	12,36	15,71	22,79	3,55
204DVAAXX:1:15:558:319	35,31	22,7	68,37	85,09
204DVAAXX:1:155:837:171	141,25	52,38	116,8	56,73
204DVAAXX:1:17:363:490	158,9	158,88	293,42	93,95
204DVAAXX:1:18:385:724	60,03	73,33	37,03	104,59
204DVAAXX:1:198:375:549	15,89	8,73	0	3,55
204DVAAXX:1:204:797:733	9,34	41,42	28,24	36,49
204DVAAXX:1:23:876:465	0	3,49	25,64	8,86
204DVAAXX:1:24:103:868	17,66	8,73	25,64	30,14
204DVAAXX:1:24:665:363	65,33	47,14	62,67	53,18
204DVAAXX:1:26:850:133	95,34	59,36	105,4	108,14
204DVAAXX:1:3:796:213	775,1	721,08	1447,16	1485,54
204DVAAXX:1:32:540:822	14,12	8,73	11,39	17,73
204DVAAXX:1:32:914:588	49,44	29,68	125,34	44,32
204DVAAXX:1:38:207:882	8,83	13,97	48,43	23,05
204DVAAXX:1:38:891:306	233,6	62,12	232,94	262,73
204DVAAXX:1:4:984:557	180,09	106,5	153,83	131,18
204DVAAXX:1:44:263:792	45,91	43,65	37,03	30,14
204DVAAXX:1:48:383:785	30,02	22,7	22,79	14,18
204DVAAXX:1:5:372:183	86,51	90,79	199,41	51,41
204DVAAXX:1:5:777:691	81,22	90,79	88,31	154,23
204DVAAXX:1:64:887:276	47,67	78,57	54,13	76,23
204DVAAXX:1:7:295:638	8,83	15,71	2,85	5,32
204DVAAXX:1:81:753:107	82,98	31,43	59,82	58,5
204DVAAXX:1:87:827:131	10,59	8,73	5,7	1,77
204DVAAXX:2:127:348:645	14,12	1,75	5,7	14,18
204DVAAXX:2:235:688:205	0	0	28,49	14,18
204DVAAXX:3:189:935:39	8,83	24,44	2,85	5,32
207G6AAXX:1:71:231:644	50,02	11,42	177,82	36,14
207G6AAXX:2:27:597:723	0	1,75	22,79	0
E4D3Z3Y01BILN1	22,95	24,44	25,64	70,91
E4D3Z3Y01BRB91	10,59	8,73	145,29	101,04
E4D3Z3Y01BZ8WO	2305,87	1257,09	720,73	1118,58
E4D3Z3Y01CTAU8	912,81	817,11	225,05	744,54
E4D3Z3Y01EO1V5	10,59	1,75	14,24	15,95
E8W526M03F18QX	28,25	36,67	150,98	88,64
E8W526M03FP075	358,42	207,77	472,89	313,77
E8W526M03G9DZL	26,48	22,7	5,7	60,27
E8W526M03GTK4Y	19,42	5,24	14,24	17,73
E8W526M03GTL95	120,06	343,95	150,98	76,23
E8W526M04I757V	19,42	10,48	11,39	21,27
E8W526M04IB8GH	324,87	436,49	259,23	315,54
E8W526M04IV7SK	234,82	193,8	188,02	200,32
E8W526M04JBO5D	21,19	10,48	42,73	47,86
E8W526M04JPUFZ	37,08	31,43	22,79	54,95
E8W526M04JU7CT	1,77	1,75	19,94	15,95
EVSUSMY01BFXHY	413,15	501,09	193,71	198,54
EVSUSMY01BO06G	5955,36	11703,17	11637,08	12607,56
EVSUSMY01C3BIZ	28,25	8,73	11,39	14,18
EVSUSMY01C43QP	26,48	20,95	76,92	17,73
EVSUSMY01CBKZO	5,3	24,44	22,79	14,18
EVSUSMY01D5U7X	79,45	45,39	56,97	56,73
EVSUSMY01D6WRK	130,65	122,22	34,18	90,41
EVSUSMY01D8S12	775,1	462,68	524,17	429
EVSUSMY01E17TU	24,72	40,16	51,28	90,41
EVSUSMY01E1RYV	381,37	200,79	903,05	1249,77
EVSUSMY01EAALD	1338,32	1272,8	1734,88	2292,12

Annexe 5 (suite) : Données de quantification normalisées des microARN caractérisés dans les banques de symbioses tardives.

ID Mature	Nb de reads	Expression	Expression	Expression	Expression
		normalisée <i>R.i.</i>	normalisée <i>No R.i.</i>	normalisée <i>S.m.</i>	normalisée <i>No S.m.</i>
MicroARN identifiés dans miRBase 18 (autres espèces)					
ath-miR168a	234953	59214,56	59144,38	36896,82	49136,16
ath-miR172a	51725	13155,46	8445,21	6646,1	5254,33
ath-miR172c	54680	616,19	415,54	3068,09	499,91
ath-miR319a	32516	21,19	3,49	42,73	12,41
bdi-miR166f	240863	60817,72	35594,87	18251,85	86359,66
bna-miR156a	6269	247,18	284,59	424,46	159,54
gma-miR159b	20800	1020,51	743,78	2210,62	327,95
gma-miR171b-3p	35793	2734,91	2217,37	1478,49	382,91
osa-miR167d	61975	2817,89	4625,05	5646,19	7370,96
ptc-miR160h	144	10,59	1,75	0	0
ptc-miR167f	147703	18141,5	17683,08	10802,4	3797,16
vvi-miR171f	11749	90,05	96,03	17,09	219,82
vvi-miR2111	4254	10,59	68,09	25,64	365,18
MicroARN identifiés dans miRBase 18 (<i>M. truncatula</i>)					
mtr-miR156g	589737	1267,7	1073,77	12212,53	7755,63
mtr-miR159a	6094474	1774,42	2002,62	2324,57	2646,67
mtr-miR160	25428	323,1	375,38	304,81	356,32
mtr-miR160c	531	30,02	17,46	0	39
mtr-miR162	16112	194,22	298,56	250,69	553,09
mtr-miR164a	258566	47605,76	63610,54	12101,43	9494,67
mtr-miR166	1609096	184103,99	274822,74	84641,61	206083,61
mtr-miR167	28122	7860,43	11095,57	8309,76	4823,56
mtr-miR169c	1728	7,06	8,73	179,47	0
mtr-miR169d	2787	0	5,24	28,49	56,73
mtr-miR169q	1611	3,53	1,75	2,85	26,59
mtr-miR171b	389	33,55	8,73	25,64	14,18
mtr-miR172	18486	123,59	165,87	828,98	1845,4
mtr-miR2086	49360	5789,39	6255,77	11232,56	8120,81
mtr-miR2088b	13607	782,16	591,88	786,25	1549,35
mtr-miR2111h	12676	26,48	122,22	42,73	1161,13
mtr-miR2119	955	26,48	52,38	244,99	285,41
mtr-miR2592b	684	8,83	12,22	8,55	17,73
mtr-miR2634	2419	3,53	3,49	5,7	17,73
mtr-miR2651	618	61,8	94,28	59,82	76,23
mtr-miR390	6769	363,71	420,78	3258,95	829,63
mtr-miR396b	193745	2265,26	1330,42	1019,85	1169,99
mtr-miR5204	17408	33,55	5,24	5,7	24,82
mtr-miR5232	30219	303,68	632,04	94,01	546
mtr-miR5239	88061	86,51	139,68	173,77	292,5
mtr-miR5558	7659	5,3	10,48	14,24	3,55
mtr-miR5559	2917	8,83	24,44	1370,24	2302,76
mtr-miR5561	3948	229,53	289,83	247,84	313,77

Annexe 6 : Graphique représentant le taux de mycorhization selon le temps de culture du double mutant *sun-sickle* (jaune) et le sauvage correspondant L416 (bleu) de *Medicago truncatula*. Les barres d'erreurs représentent les erreurs standards à la moyenne.



Annexe 7: Liste des gènes de *M. truncatula* sur-exprimés lors de l'induction par les Myc-LCO après 4h et 24h d'induction.

Sonde	Gène	Taux d'induction
Significativement induits à 4h d'induction aux Myc-LCO		
Mtr.35830.1.S1_at	hypothetical protein	27,6
Mtr.19174.1.S1_at	hypothetical protein	14,84
Mtr.35511.1.S1_at	Proteinase inhibitor I13, potato inhibitor	11,95
Mtr.44069.1.S1_at	hypothetical protein	11,16
Mtr.12845.1.S1_at	Auxin-induced protein 5NG4	8,88
Mtr.26489.1.S1_at	pectate lyase family proteinVirulence factor	8,74
Mtr.10704.1.S1_at	MtN24 gene	8,64
Mtr.18031.1.S1_at	hypothetical protein	6,85
Mtr.12833.1.S1_at	hypothetical protein	5,54
Mtr.25451.1.S1_s_at	alanyl-tRNA synthetase	5,37
Mtr.2221.1.S1_at	hypothetical protein	5,23
Mtr.25451.1.S1_at	CDNA-AFLP fragment	5,11
Mtr.49589.1.S1_at	expressed protein;Embryo-specific	5
Mtr.40336.1.S1_at	hypothetical protein	4,78
Mtr.26415.1.S1_at	hypothetical protein	4,53
Mtr.10522.1.S1_s_at	alanyl-tRNA synthetase	4,33
Mtr.6719.1.S1_at	hypothetical protein	3,84
Mtr.26206.1.S1_at	hypothetical protein	3,58
Mtr.17298.1.S1_at	expressed protein	3,2
Mtr.42256.1.S1_at	Permease of the major facilitator superfamily	3,12
Sonde	Gène	Taux d'induction
Significativement induits à 24h d'induction aux Myc-LCO		
Mtr.23516.1.S1_at	Protein MAM3.IMP dehydrogenase	10
Mtr.37912.1.S1_at	NADP-dependent oxidoreductase	6,7
Mtr.7546.1.S1_at	Low-molecular-weight cysteine-rich protein	5,71
Mtr.52092.1.S1_s_at	pectinesterase family protein	5,56
Mtr.28094.1.S1_at	RWP-RK domain-containing protein	4,88
Mtr.10626.1.S1_at	Flavonol synthase/flavanone 3-hydroxylase	4,59
Mtr.11288.1.S1_at	Pectinesterase inhibitor	4,24
Mtr.47198.1.S1_s_at	hypothetical protein	4,16
Mtr.39333.1.S1_at	expressed protein	4,12
Mtr.2844.1.S1_at	hypothetical protein	4,1
Mtr.50087.1.S1_at	expressed protein	4,08

Annexe 8: Liste des microARN caractérisés dans les banques induites par les Myc-LCO, les facteurs Nod et leur témoin correspondant.

ID précurseur	ID Mature	Séquence du mature	Cleav eland	Taille du précurseur	Taille du mature	Nb de reads
MicroARN nouvellement identifiés						
Mt3.5.1s05_r448	204DVAAXX:1:100:511:243	UGGACAGCUUGAAGCAAUUAU		121	21	193745
Mt3.5.1Chr0_r1168	204DVAAXX:1:104:568:153	UCAUUGGAUAGAAUUAAGGC		302	21	35793
Mt3.5.1s04_r1181	204DVAAXX:1:115:305:539	CGAGGAAAAUUAUGAGUGCAG		154	21	28478
Mt3.5.1Chr4_r222	204DVAAXX:1:117:892:119	GAUACGUUUUGAGGAAUGGGC		115	21	13607
Mt3.5.1Chr8_r683	204DVAAXX:1:117:974:986	UAAGAUCUGUUGGAGUUGGCG		96	21	12293
Mt3.5.1s04_r110	204DVAAXX:1:12:274:48	GAAAAACAUGAAUGUCGGC		105	21	8028
Mt3.5.1Chr5_r2532	204DVAAXX:1:121:104:684	AGUUGACUGAGUUGGCGAGGC		105	21	8028
Mt3.5.1Chr5_r1413	204DVAAXX:1:125:952:169	AUAGAGAAGAGCGUGCAGAGC		104	21	8028
Mt3.5.1s05_r733	204DVAAXX:1:13:549:498	CGAUUGUAUGAGCAAUGUAC		103	21	7659
Mt3.5.1Chr6_r1269	204DVAAXX:1:13:767:169	UAAAGGAUUGACAGUUGGGGU		183	22	6414
Mt3.5.1Chr4_r2506	204DVAAXX:1:137:442:324	UUUGGAAAGUAGACGAUGACA		110	21	5779
Mt3.5.1Chr6_r730	204DVAAXX:1:14:142:991	UGUGACUGAAUGGAAAGAGUGC		130	21	4907
Mt3.5.1Chr5_r2476	204DVAAXX:1:14:326:719	AAGAUGUUGCAUGAAUGGUU		257	21	4803
Mt3.5.1s06_r1137	204DVAAXX:1:144:276:71	UUGGAUUUGGAAUGCAUGA		122	21	4777
Mt3.5.1Chr7_r163	204DVAAXX:1:149:148:548	CAUAUGUUUGAUUCUUUGGC	ok	105	21	4254
Mt3.5.1s05_r167	204DVAAXX:1:15:558:319	CUAUGUUAGGCUCAUUCGGUU		180	21	3629
Mt3.5.1s04_r1114	204DVAAXX:1:15:684:513	UGAAGCACGGAUACGGACACU		138	21	3535
Mt3.5.1Chr6_r513	204DVAAXX:1:155:659:976	CAUCAUUUCAAGGUUCUGU	ok	121	21	3283
Mt3.5.1Chr6_r814	204DVAAXX:1:155:837:171	UUUCAAAAAGUUUAGUUGGGA		151	21	2917
Mt3.5.1s04_r796	204DVAAXX:1:159:171:199	AUGAUUUUAUUUGUAGGGGC		235	21	2900
Mt3.5.1Chr2_r1820	204DVAAXX:1:163:744:763	UUUAUGAUGCAUGUAAUUGC		437	21	2874
Mt3.5.1s05_r608	204DVAAXX:1:167:105:210	UGGAUGGGGGAAUUAUUUGAG		272	21	2787
Mt3.5.1Chr2_r1506	204DVAAXX:1:168:733:996	UCGAGAGUCCGUAUAAAAGA		130	21	2283
Mt3.5.1Chr1_r135	204DVAAXX:1:17:363:490	UCAUGAACAGCUUGAUCGGAG		138	21	1988
Mt3.5.1s05_r1055	204DVAAXX:1:17:744:233	UUGUUGCGGUGAUUGUUGGGA		70	21	1928
Mt3.5.1s04_r321	204DVAAXX:1:175:654:403	GUGGACUCGGAAAGAAAGGCA		111	21	1808
Mt3.5.1s05_r573	204DVAAXX:1:199:744:741	AUGAAGGUGAUGGAACAUGGC		208	21	1803
Mt3.5.1Chr8_r447	204DVAAXX:1:20:314:948	UCAUUGACGAUGAUUAGAAGC		98	21	1622
Mt3.5.1Chr7_r270	204DVAAXX:1:204:660:188	GCUUUUAGGAUUAUUUGUUAAG		130	21	1611
Mt3.5.1Chr6_r624	204DVAAXX:1:207:20:105	UAGAUGAAGGACUGGUUUGGC		129	21	1566
Mt3.5.1s01_r728	204DVAAXX:1:21:388:55	UUGUGCUGAUGAUGAUCGUUC		134	21	1560
Mt3.5.1Chr5_r349	204DVAAXX:1:22:877:222	CCAGAUACAGAUUUGUUGCUC		320	21	1444
Mt3.5.1s01_r384	204DVAAXX:1:220:955:912	UGUUUACAAUGAUUGUGGAUU		125	21	1321
Mt3.5.1s07_r59	204DVAAXX:1:23:876:465	UAGAAGAAGCUGUCAGAGAAAG		100	21	1207
Mt3.5.1Chr7_r102	204DVAAXX:1:24:383:906	UGAUUGAAAAUUAAAUGCGUC		329	21	1145
Mt3.5.1Chr6_r732	204DVAAXX:1:25:697:853	UAUAGGAAUUUGGGGGAUGGG		141	21	1090
Mt3.5.1s01_r490	204DVAAXX:1:26:842:487	UGGAAGUUGAUGAUGACGGUG		141	21	1090
Mt3.5.1Chr6_r739	204DVAAXX:1:263:344:432	CGAAGAAUGUGGAGCAGGGUG		145	21	1090
Mt3.5.1Chr5_r2539	204DVAAXX:1:263:634:705	CAAGAAUUGAAGACUUAUGGA		141	21	1090
Mt3.5.1s07_r199	204DVAAXX:1:268:706:434	CAAGCGAGAGAUCAAGGUCGG		141	21	1090
Mt3.5.1s02_r585	204DVAAXX:1:279:268:92	UUGGAUUGAUGAGAUUUUAG		393	21	1060
Mt3.5.1Chr2_r1139	204DVAAXX:1:289:693:586	CAUCAUCGUUGAAGUUCGGC		393	21	1060
Mt3.5.1Chr4_r2510	204DVAAXX:1:294:247:201	UCUAAAAGAAUGUUCGCGUG		255	21	1033
Mt3.5.1s05_r14	204DVAAXX:1:298:21:746	UGGAUGUGUGACAAAAGAGAAC		251	21	1033
Mt3.5.1s05_r1131	204DVAAXX:1:3:796:213	UGAGGGGCGUUUGAAAGCGG		64	21	1005
Mt3.5.1Chr2_r840	204DVAAXX:1:3:921:219	UUGGAACCGGCGUCAGUUAAC		103	21	943
Mt3.5.1Chr4_r1932	204DVAAXX:1:305:496:350	UGAUGAUUGGAAUGGAAAGGC		95	21	920
Mt3.5.1Chr3_r925	204DVAAXX:1:323:187:649	UGAGACACUGAGAUUGGGC		117	21	912
Mt3.5.1s02_r678	204DVAAXX:1:33:869:538	GCAUGUGAUGAGUUACUCGGA		117	21	844
Mt3.5.1Chr7_r1995	204DVAAXX:1:37:241:436	GAGAGUAGGAUAUGAAGAGGC		142	21	824
Mt3.5.1s06_r505	204DVAAXX:1:38:438:58	UUGGAGAUUUGUCGUUGAG		117	21	747
Mt3.5.1Chr3_r716	204DVAAXX:1:39:940:940	UGUGGAUUAUGAUUAGAUCU		104	21	722
Mt3.5.1Chr6_r135	204DVAAXX:1:4:247:604	UAAGAUGAGACUUGUAUGGC		114	21	710
Mt3.5.1s04_r1340	204DVAAXX:1:4:295:688	AUUCAGUGACUUUAGGGCGGG		109	21	707
Mt3.5.1Chr0_r631	204DVAAXX:1:4:984:557	AUGGGAUUGGGAGGAUCGGU		251	21	707
Mt3.5.1s02_r438	204DVAAXX:1:44:263:792	UUUGAAUUGGAAUGUACUCGU		106	21	700
Mt3.5.1Chr4_r1036	204DVAAXX:1:57:433:454	CGGAAAAGCAGAAUUCUUGA		213	21	698
Mt3.5.1s05_r995	204DVAAXX:1:64:887:276	UACUGGACAAGUGGUGAGAAG	ok	255	21	684

Annexe 8 (suite): Liste des microARN caractérisés dans les banques induites par les Myc-LCO, les facteurs Nod et leur témoin correspondant.

ID précurseur	ID Mature	Séquence du mature	Cleav eland	Taille du précurseur	Taille du mature	Nb de reads
Mt3.5.1Chr7_r1965	204DVAAXX:1:69:700:457	UGGGACUAAAUGAGACGGUU		152	21	684
Mt3.5.1Chr7_r1977	204DVAAXX:1:70:355:305	ACGAUGGUCAGUGUCAGAUGC		117	21	684
Mt3.5.1Chr8_r283	204DVAAXX:1:71:147:94	UUAAUUUGGAAUCAUUAGAAGC		131	21	674
Mt3.5.1Chr3_r2207	204DVAAXX:1:76:768:743	UUGGAAGGAAUUGUGGUAA		100	21	657
Mt3.5.1Chr0_r918	204DVAAXX:1:82:176:232	UGGAUUAUUUUUGAAACGG		100	21	657
Mt3.5.1Chr5_r2792	204DVAAXX:1:83:959:886	UCUUAAAAAGGAGAAAAGAGU		147	21	618
Mt3.5.1Chr0_r1057	204DVAAXX:1:86:163:286	AUAUGGACAGCUUGAAGACAA	ok	212	21	607
Mt3.5.1Chr1_r1434	204DVAAXX:1:89:774:610	UGAACAGAAAGAACAUUUGGC		190	21	576
Mt3.5.1Chr1_r542	204DVAAXX:2:127:348:645	AACGGUCGGAUUGAACGUGGC		111	22	573
Mt3.5.1s04_r860	204DVAAXX:2:128:734:86	UUUCUGUGAUUCUCUGCGCA		284	21	566
Mt3.5.1Chr3_r2186	204DVAAXX:2:185:210:515	UGACUCAGUUUGAAGAUGG		140	21	557
Mt3.5.1Chr2_r1766	204DVAAXX:2:212:278:171	UGACCUAAUGGGACAGUCAGG		326	21	505
Mt3.5.1Chr7_r86	204DVAAXX:2:217:939:819	UUUGGUACCAA GCAGAGAUCA		129	21	495
Mt3.5.1Chr4_r1448	204DVAAXX:2:24:551:888	UGGUACUUGUGAUGUAGUUGC		130	21	495
Mt3.5.1s02_r744	204DVAAXX:2:247:865:72	UUUAUUUAGGGACGGAUGGA		129	21	495
Mt3.5.1Chr6_r770	204DVAAXX:2:32:528:805	UAUAGGAAUUUGUGGAAUGGG		116	21	439
Mt3.5.1s04_r703	204DVAAXX:2:7:379:788	UUUGAAUAGGAGAUCUGAUG		127	21	439
Mt3.5.1Chr1_r1394	204DVAAXX:2:99:754:737	CUUCUCUUGCAGGUGACGACG		220	21	434
Mt3.5.1s06_r1180	204DVAAXX:3:133:725:372	UCUCUUUAAAAUAGGACCGGA		107	21	431
Mt3.5.1Chr7_r127	204DVAAXX:3:149:618:723	AGGACCAAAACUUAAACGGGC		198	21	419
Mt3.5.1Chr3_r802	204DVAAXX:3:180:606:873	UUGGGGGCCAAAACUGCUACA		240	21	414
Mt3.5.1s01_r694	204DVAAXX:3:26:584:753	CUUAUAAUAGGAACGGAGGG		437	21	405
Mt3.5.1Chr5_r1911	204DVAAXX:3:44:498:486	CGACAUAUUUUAGAAAUCG		95	21	363
Mt3.5.1Chr6_r240	204DVAAXX:3:51:558:941	CUGCGCAAGAUGAAGGUUUGG		122	21	362
Mt3.5.1s05_r828	204DVAAXX:3:70:427:398	UCUUUGAUGAUGUGGUGGUCUG		171	21	351
Mt3.5.1Chr3_r715	204DVAAXX:3:74:607:111	UAAAUGUGUUUGGAAUUAAGGU		402	21	351
Mt3.5.1s04_r115	204DVAAXX:3:86:216:807	GGAAAACAUAGAAUGCGGGGC		225	21	343
Mt3.5.1s04_r188	204DVAAXX:3:89:618:955	ACACAACCCACAAAACCGGC		107	21	336
Mt3.5.1Chr0_r1105	204DVAAXX:3:92:621:167	CUGGUGAUAGAAAUAAAUUGC		93	21	336
Mt3.5.1s02_r1141	207G6AAXX:1:108:628:401	CAUUUCGGUUCUCAGCAGUGGC		167	21	308
Mt3.5.1Chr5_r391	207G6AAXX:1:118:359:697	AUGCUUAGAUUAGGGACGGA		255	21	277
Mt3.5.1Chr7_r1764	207G6AAXX:1:181:100:135	CUAACGGAAUGUUGAUGUGGC		137	21	276
Mt3.5.1Chr1_r1467	207G6AAXX:1:187:868:382	AGGGACGAAAUUCAAUUGGC		210	21	274
Mt3.5.1Chr4_r228	207G6AAXX:1:220:851:531	CAGUAGGAAGGAUGCGGCCAC		196	21	273
Mt3.5.1Chr7_r665	207G6AAXX:1:58:480:677	UGUUUUUAGGGUCAGUCGGC		144	21	262
Mt3.5.1Chr7_r667	207G6AAXX:1:67:740:719	AUAGAGUUGGGUGACAUUGGC		123	21	255
Mt3.5.1Chr3_r2394	207G6AAXX:1:71:231:644	UGUGUCGGAGUACAAUUGAG		67	21	253
Mt3.5.1Chr5_r2525	207G6AAXX:1:80:823:494	UUCAGCAUAAACCGGAACCGC		165	21	233
Mt3.5.1s07_r1069	207G6AAXX:2:145:549:392	GAAUGUCGGAAGUUCAGAUGC		165	21	233
Mt3.5.1s04_r953	207G6AAXX:2:170:330:668	UGACGUGGUGGAAUAGGGACC		153	22	228
Mt3.5.1Chr8_r1647	207G6AAXX:2:203:798:720	GCUUAUAAUAGUGAUCGGAGGG		264	21	228
Mt3.5.1Chr5_r1054	207G6AAXX:2:51:73:687	CGGGCGAGUGUACAAUACGU		267	21	228
Mt3.5.1s05_r342	207G6AAXX:3:267:286:450	AUAUGGACAACUUGAAGACAA		140	21	224
Mt3.5.1Chr7_r257	207G6AAXX:3:278:221:528	GGGAAAACAUGAGGUGUUUGGC		125	21	219
Mt3.5.1Chr6_r326	207G6AAXX:4:135:532:575	UAGAACAUCUCCAGAAGCAAC		159	21	217
Mt3.5.1s06_r986	207G6AAXX:4:137:563:742	UGGAUGUGUACUGAACGUGCG		140	21	217
Mt3.5.1Chr4_r1670	207G6AAXX:4:279:910:301	UACUAUGAAGCACAGAUACGG		216	21	213
Mt3.5.1s06_r22	207G6AAXX:4:68:259:473	UCAUCGAAUUAUGAGCUGGC	ok	149	21	213
Mt3.5.1Chr5_r372	207G6AAXX:4:84:387:471	CUUAUAAUUGAGACGGAGGUA		168	21	209
Mt3.5.1Chr7_r926	20F1DAAXX:6:17:613:233	AUGCGAUCGAAAAAUGUAGC		143	21	209
Mt3.5.1Chr3_r1838	20F1DAAXX:6:200:383:227	GCUUAUAAUAGUGAUCGGAGGA		130	21	204
Mt3.5.1Chr0_r55	20F1DAAXX:7:236:104:187	UGACAAAAAUGCAAAGAAGAG		168	21	202
Mt3.5.1Chr4_r2438	20F1DAAXX:7:8:702:97	CGACAUUGGACAAAAAAGGG		326	21	202
Mt3.5.1s06_r1211	20F1DAAXX:8:199:47:427	ACUCAGUUUUUGGGACGGA		412	21	198
Mt3.5.1s02_r959	20F1DAAXX:8:36:70:928	CUUAUAAUUAAGGACGGAGGG		412	21	198
Mt3.5.1Chr8_r1477	31_HELIUM_0000:7:100:12088:9250	AUGGACUAGAGAUUUGGGCU		412	21	198
Mt3.5.1s05_r514	32_HELIUM_0000:1:62:16450:13727	GGUCAUAAUUGCAGGGACGA		153	21	186
Mt3.5.1Chr7_r1875	AZOTE_0001_61FVGAAXX:1:1:15719:14450	UAGAUGUACCCAGUUUUAGGC		153	21	186
Mt3.5.1Chr7_r833	AZOTE_0001_61FVGAAXX:1:10:1676:21066	CCAGAAAACACCCUAAUAGG		150	21	173
Mt3.5.1Chr1_r303	AZOTE_0001_61FVGAAXX:1:106:14376:20179	GUCGAGUUCAAUAGGAGGCA		293	21	168

Annexe 8 (suite): Liste des microARN caractérisés dans les banques induites par les Myc-LCO, les facteurs Nod et leur témoin correspondant.

ID précurseur	ID Mature	Séquence du mature	Cleav eland	Taille du précurseur	Taille du mature	Nb de reads
Mt3.5.1Chr6_r612	AZOTE_0001_61FVGAAXX:1:12:18184:16367	GGAGAUGACUGUGACUUUGGA		275	21	167
Mt3.5.1s05_r775	AZOTE_0001_61FVGAAXX:1:16:17126:14193	UGGUUCGGCUCGAAACUCGGC		298	21	167
Mt3.5.1Chr4_r2643	AZOTE_0001_61FVGAAXX:1:38:18284:4909	CCGAGUUUAGGCUUUGCAAGGC		97	21	158
Mt3.5.1Chr8_r1081	AZOTE_0001_61FVGAAXX:1:5:1908:11829	GGCAAACAUUAAUUGAAGACA		143	21	156
Mt3.5.1s06_r913	AZOTE_0001_61FVGAAXX:1:58:9347:21235	AGGAGGUAGGACUAAAAUCGC		189	21	156
Mt3.5.1Chr8_r1369	AZOTE_0001_61FVGAAXX:1:65:6993:13934	UUUAGGGUUUUCUGAGCCAG		267	21	152
Mt3.5.1Chr7_r592	AZOTE_0001_61FVGAAXX:1:69:9330:6857	GUAGAGGCAUGUGCACUGGC		123	21	146
Mt3.5.1Chr6_r981	AZOTE_0001_61FVGAAXX:1:82:6313:2449	AGGACAAACUCUCAACACA		123	21	146
Mt3.5.1Chr2_r1019	AZOTE_0001_61FVGAAXX:1:9:2363:15857	UAUUGUCGGUUGAGUUUAUUC		222	21	146
Mt3.5.1Chr4_r761	AZOTE_0001_61FVGAAXX:3:62:7832:20773	AUUGAACCGAUUUAAAAGUG		222	21	146
Mt3.5.1Chr3_r1732	AZOTE_0001_61FVGAAXX:3:93:12933:13709	GAUGAACUAAAAGUUUUGGC		221	21	146
Mt3.5.1Chr7_r816	E4D3Z3Y01A14UP	AACCCUCAAGGCUUCCAUUGG	ok	219	21	146
Mt3.5.1Chr5_r664	E4D3Z3Y01BHN7L	UUACCUAUUCCACCAAUUCCA	ok	122	21	146
Mt3.5.1s06_r728	E4D3Z3Y01BRB91	UUACAAUUCGCAAUUGAUUUAU		108	21	144
Mt3.5.1s02_r1120	E4D3Z3Y01C9R5R	UUUUAUGAAUUAUCGGUAGUA		296	21	144
Mt3.5.1s05_r1241	E4D3Z3Y01CE5US	UUCGGUGUAGCAUUUGACUAC		116	21	141
Mt3.5.1s02_r607	E4D3Z3Y01CK6W2	UAAUCUGCAUCCUGCGGUCUA	ok	409	21	140
Mt3.5.1s05_r226	E4D3Z3Y01CTAU8	UUUUCGGGAAAGUGGAGUACA		225	22	130
Mt3.5.1s06_r956	E4D3Z3Y01EDUEZ	CAAAUGGACAGGAAUUAUCGG		116	21	130
Mt3.5.1Chr6_r788	E4D3Z3Y01EH4T3	AUUUCUAGAGAUAGCAUUAUA		124	21	129
Mt3.5.1Chr4_r158	E4D3Z3Y01EO1V5	AAUAUGACGGAGUGUAAAUGC		124	21	129
Mt3.5.1Chr7_r442	E8W526M03F3P9L	UUGACUAUGAGGAUGCUGAUG		190	21	125
Mt3.5.1Chr5_r472	E8W526M03F18QX	UGCGUAUACUGUGGGGUAAG		232	21	123
Mt3.5.1Chr5_r516	E8W526M03G7ZSK	UGAAGUAUUUGGGGAAACUCU	ok	196	21	123
Mt3.5.1Chr3_r428	E8W526M03G9DZL	AUGAUUGUGACAGUUAUGGG		102	21	123
Mt3.5.1s07_r895	E8W526M03GL4RU	UUAGGACUAUGGUUUGGACAA		209	21	122
Mt3.5.1Chr2_r573	E8W526M03GVXF3	CUCAACAGGACGGAUCACUAC		306	21	120
Mt3.5.1s02_r1351	E8W526M04I88GH	UCGACAGGCUUUGUUCUCCC	ok	121	22	115
Mt3.5.1Chr3_r287	E8W526M04I15Y8	UUAAUUGUAGAUAGACUAGA		97	21	114
Mt3.5.1s06_r835	E8W526M04U7CT	UUGAAUUAUGAGAACUUGAAG	ok	234	21	113
Mt3.5.1Chr0_r488	EVSUSMY01AFYO4	UGCAUUUGCACCGCACUUUC		234	21	113
Mt3.5.1s02_r608	EVSUSMY01B76KK	UAAUCUGCAUCCUGAGGUGUA	ok	120	21	112
Mt3.5.1s05_r971	EVSUSMY01BFXHY	UCGUGGGAUUGUUAUUGGAC		122	21	111
Mt3.5.1Chr5_r520	EVSUSMY01BLMNF	UGAAGUGUUUGGGGAAACUCU	ok	178	21	111
Mt3.5.1Chr5_r658	EVSUSMY01BO06G	UGGAAUGGAGGAUCAGGUAUA		178	21	109
Mt3.5.1Chr8_r303	EVSUSMY01BQ22N	UGACAGGUUUGAAUUUAUCUC		336	21	106
Mt3.5.1Chr4_r1118	EVSUSMY01BTVXS	UCUCGGAGGACUGGAAGAUUC		141	21	105
Mt3.5.1Chr2_r1059	EVSUSMY01C3BIZ	UUGAGCCGCGUCAUAUUCUUG	ok	142	21	105
Mt3.5.1s07_r413	EVSUSMY01CBKZO	CAACAGGACUCGAGCAUUUCG		262	21	105
Mt3.5.1Chr5_r671	EVSUSMY01CRPC9	UUGUGUUCUCAGGUCACCCCU		140	21	105
Mt3.5.1Chr4_r2629	EVSUSMY01D5U7X	UGGGUCAUUUGCUUGAACUGG		135	21	103
Mt3.5.1s04_r858	EVSUSMY01D67WF	UCUCGAUCAUUAAGUGCAU		108	21	102
Mt3.5.1s01_r413	EVSUSMY01DB76E	GGAAUGGGUUCUUUGGAAAGA		114	21	102
Mt3.5.1s03_r234	EVSUSMY01E17TU	UCUCGUUCCAUAUUCGUCUA	ok	113	21	102
Mt3.5.1s01_r23	EVSUSMY01E1RYV	AACAGGCUUGAUACCAUGA		106	21	102
Mt3.5.1s07_r535	EVSUSMY01EAALD	UUAAUCAGGGAAUUCACAGUU	ok	91	21	97
Mt3.5.1s06_r1141	EVSUSMY01EGIU5	UUUGGCAUUCUGUCCACCUCC	ok	195	21	96
Mt3.5.1s07_r602	HELIUM_0001:7:11:321:1639	UUUGUGACUGAUGAAUGAUGC		195	21	96
Mt3.5.1Chr7_r1999	HELIUM_0001:7:11:743:1692	AAGAAAGUGUCACUUUGAGCU		195	21	96
Mt3.5.1Chr8_r523	HELIUM_0001:7:23:475:1686	UGGGUGUACCGUAGAAUUUGC		195	21	96
Mt3.5.1Chr5_r1460	HELIUM_0001:7:25:232:1311	UGGAACAGAAUCAAGAACUG		118	21	96
Mt3.5.1Chr5_r2263	HELIUM_0001:7:8:969:1451	AGUAAAAGUUUGCAUUUAGA		197	21	96
Mt3.5.1Chr8_r731	HELIUM_0001:7:88:1132:320	AUGACUGA GAUUGAGAAUUUU	ok	236	21	96
Mt3.5.1s07_r231	HELIUM_0001:7:9:478:686	UAGACUGA UUGUUGAUGUUGG		206	21	94
Mt3.5.1Chr7_r1798	HELIUM_0001:7:93:1273:277	UCAACAAAAGACUCUGAAGC		344	21	93
Mt3.5.1s02_r263	HELIUM_0001:8:103:136:442	UUUUGUAAGGAUGAAUUAGGC		320	21	92
Mt3.5.1Chr7_r1532	HELIUM_0001:8:29:1307:1004	GGAAUGAACCAUUAUGAUGAC		310	21	88
Mt3.5.1Chr7_r151	HELIUM_0001:8:5:827:565	ACGUGUCACAUUGUCAUUGGC		105	21	84
Mt3.5.1s05_r1141	HELIUM_0001:8:68:1739:1137	AACGGACAUAACAACUUUGGC		175	21	82
Mt3.5.1Chr4_r2182	HELIUM_0001:8:92:696:10	AUAAGGCUAGAAACACUAGU		422	21	77

Annexe 8 (suite): Liste des microARN caractérisés dans les banques induites par les Myc-LCO, les facteurs Nod et leur témoin correspondant.

ID précurseur	ID Mature	Séquence du mature	Cleav eland	Taille du précurseur	Taille du mature	Nb de reads
MicroARN identifiés dans miRBase 18 (autres espèces)						
Mt3.5.1s02_r144	ahy-miR156c	UUGACAGAAGAGAGAGAGCAC	ok	326	21	74
Mt3.5.1Chr5_r15	ath-miR168a	UCGCUUGGUGCAGGUCGGGAA		328	21	74
Mt3.5.1s04_r966	ath-miR172a	AGAAUCUUGAUGAUGCUGCAU	ok	142	21	73
Mt3.5.1s06_r1087	ath-miR172c	AGAAUCUUGAUGAUGCUGCAG	ok	109	21	72
Mt3.5.1s06_r1098	ath-miR319a	UUGGACUGAAGGGAGCUCCU	ok	201	21	70
Mt3.5.1Chr1_r725	ath-miR408	AUGCACUGCCUUCUCCUGGC		222	21	70
Mt3.5.1Chr7_r957	bdi-miR166f	UCUCGGACCAAGGCUUCAUCC	ok	223	21	70
Mt3.5.1Chr1_r1415	bnm-miR156a	UGACAGAAGAGAGUGAGCAC	ok	114	21	65
Mt3.5.1Chr0_r165	gma-miR159b	AUUGGAGUGAAGGGAGCUCCA	ok	125	21	65
Mt3.5.1s06_r459	osa-miR167d	UGAAGCUGCCAGCAUGAUCUG		171	22	64
Mt3.5.1s01_r1112	ptc-miR167f	UGAAGCUGCCAGCAUGAUCUU	ok	171	22	64
Mt3.5.1Chr4_r1775	vvi-miR171f	UUGAGCCGCGCCAAUAUCAU	ok	112	21	63
Mt3.5.1s01_r1139	zma-miR171f	UUGAGCCGUGCCAAUAUCA	ok	396	21	62
MicroARN identifiés dans miRBase 18 (<i>M. truncatula</i>)						
Mt3.5.1Chr5_r659	mtr-miR1510a*	UUGUCUUAACCAUCCUCCCA	ok	155	21	61
Mt3.5.1Chr8_r1355	mtr-miR156e	UUGACAGAAGAUAGAGAGCAC	ok	130	21	59
Mt3.5.1Chr6_r606	mtr-miR156g	UUGACAGAAGAUAGAGGGCAC	ok	107	21	59
Mt3.5.1s06_r1296	mtr-miR159a	UUUGGAUUAGAAGGGAGCUCUA	ok	122	21	59
Mt3.5.1s04_r658	mtr-miR160	UGCCUGGCUCUCCUGUAUGCCA	ok	122	21	59
Mt3.5.1Chr2_r1815	mtr-miR160c	UGCCUGGCUCUCCUGAAUGCCA	ok	107	21	59
Mt3.5.1Chr4_r1690	mtr-miR162	UCGAUAAACUCUGCAUCCAG		107	21	59
Mt3.5.1Chr7_r1619	mtr-miR164a	UGGAGAAGCAGGGCAGUGCA	ok	427	21	59
Mt3.5.1Chr8_r1611	mtr-miR166	UCGGACCAGGCUUCAUCCCC	ok	123	21	59
Mt3.5.1s02_r240	mtr-miR167	UGAAGCUGCCAGCAUGAUCUA	ok	242	21	58
Mt3.5.1Chr2_r214	mtr-miR169c	CAGCCAAGGGUGAUUUGCCGG	ok	380	21	57
Mt3.5.1Chr7_r14	mtr-miR169g	CAGCCAAGGAUGACUUGCCGG	ok	366	21	57
Mt3.5.1Chr2_r219	mtr-miR169m	GAGCCAAGGAUGACUUGCCGG	ok	388	21	57
Mt3.5.1s03_r846	mtr-miR171	UGAUUGAGUCGUGCCAAUAUC	ok	365	21	57
Mt3.5.1s05_r591	mtr-miR171b	UGAUUGAGCCGUGCAUAUAUC	ok	364	21	57
Mt3.5.1Chr4_r2289	mtr-miR171f	UUGAGCCGUGCCAAUAUCA	ok	224	21	55
Mt3.5.1Chr5_r1717	mtr-miR172	AGAAUCCUGAUGAUGCUGCAG	ok	399	21	54
Mt3.5.1Chr2_r1326	mtr-miR2086	GACAUUGAAUGCAGAACUGGAA		240	21	53
Mt3.5.1Chr7_r363	mtr-miR2111h	UAAUCUGCAUCCUGAGGUUUA	ok	240	21	53
Mt3.5.1Chr5_r670	mtr-miR2119	UCAAAGGGAGGUGUGGAGUA	ok	167	21	51
Mt3.5.1s07_r931	mtr-miR2592b	AAAUGCUUGAGUCCUUGUUU		100	21	49
Mt3.5.1s05_r561	mtr-miR2592bi	UGGAACAUGGGAAUGCCGGU		339	21	48
Mt3.5.1Chr4_r812	mtr-miR2612a	UGAUAGUGUCAACUAGUACAG	ok	339	21	48
Mt3.5.1Chr5_r1605	mtr-miR2623	UCGGCUGUACUGUCCUUAUG		339	21	48
Mt3.5.1Chr5_r1185	mtr-miR2633	UGACAUUUGCUCCAGAUUCA		166	21	48
Mt3.5.1Chr4_r1210	mtr-miR2634	UUUAUUCUCAUUUGUUGCUC		369	21	48
Mt3.5.1Chr5_r2290	mtr-miR2643	UUUGGGAUCAGAAAUAAGAGA	ok	118	21	47
Mt3.5.1Chr5_r2049	mtr-miR2648	UAGCCAAUGGGAAUAACAGAU		168	21	45
Mt3.5.1Chr4_r1299	mtr-miR390	AAGCUCAGGAGGGAUAGCGCC		177	21	44
Mt3.5.1Chr3_r683	mtr-miR393	UCCAAAGGGAUUCGCAUUGAUC	ok	298	21	44
Mt3.5.1Chr4_r1780	mtr-miR396a	UUCCAAGGCUUUCUUAACUU	ok	209	21	44
Mt3.5.1Chr7_r1456	mtr-miR398b	UGUGUUCUCAGGUCGCCUUG	ok	148	21	44
Mt3.5.1Chr7_r820	mtr-miR5204	GCUGGAAGGUUUUGUAGGAAC		176	21	44
Mt3.5.1s04_r498	mtr-miR5232	UACAUGUCGUCUCACCCUGAA	ok	161	21	43
Mt3.5.1s06_r834	mtr-miR5239	UGGGAGAAAAAGAUAGAAUGUG	ok	171	21	43
Mt3.5.1Chr5_r1358	mtr-miR5250	UGAGAAGUUAAGAUACGGAAC		160	21	43
Mt3.5.1Chr7_r1483	mtr-miR5561	CAUUUGGAGAGACAUGACAA		138	21	42

Annexe 9: Données de quantification normalisées des microARN caractérisés dans les banques induites par les Myc-LCO, les facteurs Nod et leur témoin correspondant.

ID Mature	Expr norm Myc-LCO 1	Expr norm Myc-LCO 2	Expr norm Myc-LCO 3	Expr norm Nod-F 1	Expr norm Nod-F 2	Expr norm Nod-F 3	Expr norm témoin 1	Expr norm témoin 2	Expr norm témoin 3
MicroARN nouvellement identifiés									
204DVAAXX:1:100:511:243	1,19	2,1	1,18	3,66	3,54	4,09	4,94	6,42	4,2
204DVAAXX:1:104:568:153	22,64	39,99	22,36	20,13	42,5	42,26	21	28,43	29,4
204DVAAXX:1:115:305:539	4,77	2,1	1,18	1,83	0	2,73	7,41	7,34	4,2
204DVAAXX:1:117:892:119	28,6	41,05	14,12	27,45	29,51	40,89	29,65	23,84	28
204DVAAXX:1:117:974:986	7,15	7,37	8,24	1,83	2,36	9,54	7,41	5,5	2,8
204DVAAXX:1:12:274:48	48,85	135,77	64,72	60,4	59,03	81,79	48,18	80,7	62,99
204DVAAXX:1:121:104:684	30,98	27,36	27,06	25,62	38,96	46,35	16,06	27,51	21
204DVAAXX:1:125:952:169	14,3	25,26	31,77	25,62	18,89	17,72	21	16,51	22,4
204DVAAXX:1:13:549:498	11,91	8,42	14,12	12,81	12,99	5,45	12,35	16,51	12,6
204DVAAXX:1:13:767:169	36,94	52,62	56,48	40,27	68,47	79,06	42	66,03	44,79
204DVAAXX:1:137:442:324	21,45	21,05	25,89	10,98	7,08	10,9	19,77	29,35	23,8
204DVAAXX:1:14:142:991	202,55	227,33	282,39	215,97	406,11	380,31	200,13	216,42	229,57
204DVAAXX:1:14:326:719	7,15	10,52	16,47	9,15	5,9	17,72	8,65	10,09	15,4
204DVAAXX:1:144:276:71	16,68	7,37	7,06	5,49	8,26	8,18	3,71	12,84	12,6
204DVAAXX:1:149:148:548	0	6,31	1,18	1,83	2,36	1,36	1,24	1,83	0
204DVAAXX:1:15:558:319	4,77	3,16	3,53	9,15	8,26	6,82	2,47	5,5	4,2
204DVAAXX:1:15:684:513	2,38	12,63	7,06	0	0	1,36	6,18	3,67	2,8
204DVAAXX:1:155:659:976	7,15	4,21	0	7,32	4,72	5,45	1,24	2,75	1,4
204DVAAXX:1:155:837:171	20,25	23,15	23,53	9,15	25,97	17,72	27,18	27,51	16,8
204DVAAXX:1:159:171:199	4,77	12,63	9,41	3,66	8,26	17,72	9,88	5,5	4,2
204DVAAXX:1:163:744:763	0	24,83	8,21	0	9,01	8,96	0	7,29	0
204DVAAXX:1:167:105:210	13,11	3,16	7,06	3,66	4,72	12,27	11,12	6,42	9,8
204DVAAXX:1:168:733:996	4,77	7,37	1,18	12,81	7,08	2,73	6,18	4,59	7
204DVAAXX:1:17:363:490	33,36	16,84	18,83	20,13	14,17	14,99	22,24	14,67	14
204DVAAXX:1:17:744:233	2,38	1,05	4,71	5,49	2,36	1,36	3,71	3,67	12,6
204DVAAXX:1:175:654:403	9,53	13,68	9,41	7,32	10,62	14,99	9,88	5,5	14
204DVAAXX:1:199:744:741	22,64	35,78	5,88	9,15	55,49	27,26	12,35	20,18	14
204DVAAXX:1:20:314:948	599,3	941,94	387,11	426,45	1138,1	992,35	455,86	500,71	440,95
204DVAAXX:1:204:660:188	14,3	12,63	11,77	5,49	11,81	19,08	8,65	17,42	1,4
204DVAAXX:1:207:20:105	13,11	17,89	16,47	21,96	16,53	28,63	27,18	13,76	15,4
204DVAAXX:1:21:388:55	19,06	10,52	104,72	23,79	21,25	98,14	13,59	20,18	354,16
204DVAAXX:1:22:877:222	3,57	5,26	3,53	5,49	8,26	2,73	3,71	11	8,4
204DVAAXX:1:220:955:912	22,64	17,89	11,77	10,98	16,53	16,36	18,53	17,42	15,4
204DVAAXX:1:23:876:465	17,87	20	23,53	10,98	0	0	12,35	18,34	7
204DVAAXX:1:24:383:906	7,15	6,31	4,71	7,32	11,81	8,18	8,65	5,5	9,8
204DVAAXX:1:25:697:853	11,91	18,94	15,3	16,47	15,35	12,27	24,71	21,09	25,2
204DVAAXX:1:26:842:487	23,83	44,2	35,3	29,28	36,6	28,63	39,53	38,52	28
204DVAAXX:1:263:344:432	0	1,05	1,18	0	0	0	1,24	5,5	4,2
204DVAAXX:1:263:634:705	5,96	8,42	3,53	5,49	1,18	2,73	4,94	4,59	8,4
204DVAAXX:1:268:706:434	19,06	17,89	7,06	1,83	10,62	6,82	11,12	10,09	11,2
204DVAAXX:1:279:268:92	2,38	2,1	9,41	12,81	3,54	6,82	8,65	4,59	4,2
204DVAAXX:1:289:693:586	8,34	2,1	2,35	1,83	2,36	4,09	3,71	6,42	1,4
204DVAAXX:1:294:247:201	17,87	26,31	22,36	25,62	11,81	19,08	13,59	29,35	25,2
204DVAAXX:1:298:21:746	94,13	83,14	191,79	89,68	118,05	192,2	101,3	146,73	197,38
204DVAAXX:1:3:796:213	1588,2	1397,7	1422,6	1638,1	1449,7	1544,4	1447,9	1395,7	1359,2
204DVAAXX:1:3:921:219	15,49	11,58	11,77	9,15	0	1,36	18,53	11	12,6
204DVAAXX:1:305:496:350	26,21	32,63	25,89	29,28	49,58	39,53	35,83	33,93	39,2
204DVAAXX:1:323:187:649	10,72	11,58	5,88	7,32	12,99	17,72	7,41	8,25	12,6
204DVAAXX:1:33:869:538	14,3	4,21	11,77	10,98	15,35	25,9	9,88	7,34	19,6
204DVAAXX:1:37:241:436	30,98	35,78	37,65	31,11	61,39	47,71	37,06	20,18	28
204DVAAXX:1:38:438:58	22,64	18,94	20	25,62	18,89	17,72	23,47	23,84	32,2
204DVAAXX:1:39:940:940	35,74	44,2	43,54	64,06	48,4	32,71	30,88	35,75	53,19
204DVAAXX:1:4:247:604	349,1	356,78	460,07	371,54	425	421,2	337,26	320,97	368,15
204DVAAXX:1:4:295:688	425,35	389,41	416,53	417,3	421,45	426,66	450,92	409,92	408,75
204DVAAXX:1:4:984:557	206,12	257,85	210,62	248,92	252,64	321,7	192,72	187,08	247,77
204DVAAXX:1:44:263:792	82,21	70,51	63,54	78,7	20,07	36,8	74,12	78,87	67,19
204DVAAXX:1:57:433:454	42,89	43,15	48,24	29,28	53,12	76,33	25,94	47,69	51,79

Annexe 9 (suite): Données de quantification normalisées des microARN caractérisés dans les banques induites par les Myc-LCO, les facteurs Nod et leur témoin correspondant.

ID Mature	Expr norm Myc-LCO 1	Expr norm Myc-LCO 2	Expr norm Myc-LCO 3	Expr norm Nod-F 1	Expr norm Nod-F 2	Expr norm Nod-F 3	Expr norm témoin 1	Expr norm témoin 2	Expr norm témoin 3
204DVAAXX:1:64:887:276	8,34	9,47	21,18	20,13	34,24	29,99	21	11	33,6
204DVAAXX:1:69:700:457	44,08	82,09	71,78	40,27	105,07	122,68	54,36	55,94	50,39
204DVAAXX:1:70:355:305	20,25	26,31	25,89	25,62	16,53	27,26	13,59	11,92	18,2
204DVAAXX:1:71:147:94	72,68	147,34	76,48	73,21	166,46	194,93	60,53	100,88	76,99
204DVAAXX:1:76:768:743	20,25	13,68	23,53	29,28	14,17	21,81	19,77	27,51	28
204DVAAXX:1:82:176:232	0	1,05	2,35	3,66	2,36	6,82	6,18	6,42	5,6
204DVAAXX:1:83:959:886	5,96	11,58	2,35	5,49	2,36	10,9	3,71	4,59	2,8
204DVAAXX:1:86:163:286	2,38	4,21	14,12	1,83	5,9	2,73	4,94	10,09	12,6
204DVAAXX:1:89:774:610	46,47	45,26	28,24	43,93	22,43	10,9	44,47	33,01	30,8
204DVAAXX:2:127:348:645	26,21	25,26	17,65	16,47	14,17	31,35	33,36	11,92	22,4
204DVAAXX:2:128:734:86	2,38	0	7,06	0	3,54	0	0	0,92	0
204DVAAXX:2:185:210:515	4,77	2,1	2,35	1,83	4,72	6,82	2,47	2,75	4,2
204DVAAXX:2:212:278:171	4,77	8,42	9,41	7,32	20,07	16,36	8,65	4,59	7
204DVAAXX:2:217:939:819	69,1	29,47	29,42	65,89	33,06	36,8	49,42	51,35	53,19
204DVAAXX:2:24:551:888	2,38	3,16	5,88	1,83	4,72	8,18	1,24	2,75	2,8
204DVAAXX:2:247:865:72	5,96	6,31	10,59	1,83	1,18	1,36	2,47	6,42	4,2
204DVAAXX:2:32:528:805	9,53	4,21	9,41	7,32	15,35	2,73	7,41	5,5	5,6
204DVAAXX:2:7:379:788	46,47	20	74,13	98,83	46,04	32,71	64,24	50,44	81,19
204DVAAXX:2:99:754:737	3,57	1,05	8,24	9,15	1,18	2,73	3,71	1,83	11,2
204DVAAXX:3:133:725:372	1,19	0	8,24	0	5,9	5,45	2,47	1,83	2,8
204DVAAXX:3:149:618:723	13,11	18,94	23,53	14,64	27,15	36,8	13,59	28,43	21
204DVAAXX:3:180:606:873	4,77	5,26	2,35	14,64	3,54	2,73	7,41	1,83	4,2
204DVAAXX:3:26:584:753	7,15	1,05	7,06	5,49	0	2,73	7,41	6,42	7
204DVAAXX:3:44:498:486	35,79	22,07	9,9	5,34	18,49	13,76	15,22	55,78	18,67
204DVAAXX:3:51:558:941	7,15	6,31	18,83	3,66	8,26	10,9	7,41	14,67	16,8
204DVAAXX:3:70:427:398	3,57	3,16	4,71	3,66	4,72	5,45	1,24	4,59	5,6
204DVAAXX:3:74:607:111	0	0	3,53	3,66	1,18	8,18	0	0	2,8
204DVAAXX:3:86:216:807	166,8	81,04	61,19	186,69	79,1	47,71	268,08	133,89	82,59
204DVAAXX:3:89:618:955	2,38	3,16	2,35	1,83	5,9	1,36	1,24	2,75	1,4
204DVAAXX:3:92:621:167	16,68	31,57	23,53	21,96	24,79	42,26	9,88	25,68	29,4
207G6AAXX:1:108:628:401	53,68	99,32	72,59	101,55	206,48	137,59	95,13	74,38	89,6
207G6AAXX:1:118:359:697	0	1,05	0	0	2,36	0	6,18	0,92	2,8
207G6AAXX:1:181:100:135	7,15	2,1	8,24	7,32	5,9	2,73	6,18	4,59	4,2
207G6AAXX:1:187:868:382	2,38	1,05	1,18	0	0	1,36	8,65	0,92	0
207G6AAXX:1:220:851:531	2,38	5,26	2,35	3,66	1,18	5,45	4,94	0,92	0
207G6AAXX:1:58:480:677	3,57	7,37	2,35	1,83	2,36	2,73	2,47	0	2,8
207G6AAXX:1:67:740:719	3,57	8,42	1,18	3,66	4,72	6,82	6,18	0,92	2,8
207G6AAXX:1:71:231:644	36,32	82,76	98,52	105,67	63,08	26,88	94,47	51,05	63,01
207G6AAXX:1:80:823:494	14,3	10,52	3,53	7,32	4,72	5,45	7,41	3,67	7
207G6AAXX:2:145:549:392	5,96	0	3,53	1,83	7,08	10,9	3,71	5,5	1,4
207G6AAXX:2:170:330:668	3,58	11,04	16,5	5,34	6,16	6,88	0	7,97	11,2
207G6AAXX:2:203:798:720	7,16	13,79	29,7	26,72	27,74	34,4	34,25	21,25	41,07
207G6AAXX:2:51:73:687	2,38	5,26	0	3,66	3,54	1,36	3,71	5,5	2,8
207G6AAXX:3:267:286:450	2,38	1,05	7,06	0	1,18	0	1,24	2,75	1,4
207G6AAXX:3:278:221:528	0	8,28	3,3	0	9,25	3,44	7,61	18,59	14,93
207G6AAXX:4:135:532:575	7,15	6,31	10,59	9,15	8,26	9,54	9,88	14,67	7
207G6AAXX:4:137:563:742	22,64	20	14,12	18,3	17,71	13,63	12,35	25,68	14
207G6AAXX:4:279:910:301	5,96	22,1	4,71	7,32	27,15	21,81	1,24	4,59	2,8
207G6AAXX:4:68:259:473	5,96	6,31	5,88	5,49	9,44	10,9	11,12	7,34	2,8
207G6AAXX:4:84:387:471	17,89	5,52	19,8	26,72	9,25	6,88	22,83	15,94	29,87
20F1DAAXX:6:17:613:233	5,96	5,26	2,35	7,32	8,26	8,18	2,47	7,34	4,2
20F1DAAXX:6:200:383:227	7,16	5,52	6,6	16,03	12,33	34,4	11,42	10,63	7,47
20F1DAAXX:7:236:104:187	10,72	7,37	9,41	23,79	3,54	6,82	18,53	28,43	12,6
20F1DAAXX:7:8:702:97	4,77	4,21	8,24	7,32	2,36	1,36	11,12	3,67	5,6
20F1DAAXX:8:199:47:427	0	6,31	1,18	0	0	0	0	0,92	0
20F1DAAXX:8:36:70:928	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
31_HELIUM_0000:7:100:12088:9250	0	0	8,21	0	0	26,88	0	7,29	0
32_HELIUM_0000:1:62:16450:13727	0	0	0	35,22	0	17,92	0	7,29	0
AZOTE_0001_61FVGAAXX:1:1:15719:14450	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
AZOTE_0001_61FVGAAXX:1:10:1676:21066	10,72	7,37	15,3	14,64	8,26	8,18	4,94	8,25	4,2
AZOTE_0001_61FVGAAXX:1:106:14376:20179	27,24	8,28	0	46,96	18,02	26,88	0	7,29	31,5

Annexe 9 (suite): Données de quantification normalisées des microARN caractérisés dans les banques induites par les Myc-LCO, les facteurs Nod et leur témoin correspondant.

ID Mature	Expr norm Myc-LCO 1	Expr norm Myc-LCO 2	Expr norm Myc-LCO 3	Expr norm Nod-F 1	Expr norm Nod-F 2	Expr norm Nod-F 3	Expr norm témoin 1	Expr norm témoin 2	Expr norm témoin 3
AZOTE_0001_61FVGAAXX:1:12:18184:16367	10,72	8,42	8,24	7,32	8,26	10,9	7,41	9,17	12,6
AZOTE_0001_61FVGAAXX:1:16:17126:14193	1,19	6,31	2,35	0	3,54	5,45	1,24	3,67	2,8
AZOTE_0001_61FVGAAXX:1:38:18284:4909	3,57	3,16	3,53	0	5,9	2,73	1,24	0,92	1,4
AZOTE_0001_61FVGAAXX:1:5:1908:11829	0	5,26	4,71	5,49	9,44	10,9	0	2,75	2,8
AZOTE_0001_61FVGAAXX:1:58:9347:21235	8,34	2,1	1,18	3,66	4,72	2,73	9,88	0,92	0
AZOTE_0001_61FVGAAXX:1:65:6993:13934	18,16	0	16,42	35,22	0	0	18,89	0	7,88
AZOTE_0001_61FVGAAXX:1:69:9330:6857	18,16	24,83	16,42	23,48	9,01	8,96	28,34	21,88	15,75
AZOTE_0001_61FVGAAXX:1:82:6313:2449	0	24,83	0	0	0	0	9,45	7,29	23,63
AZOTE_0001_61FVGAAXX:1:9:2363:15857	17,87	7,37	15,3	10,98	14,17	13,63	19,77	15,59	22,4
AZOTE_0001_61FVGAAXX:3:62:7832:20773	5,96	2,1	2,35	0	1,18	1,36	0	1,83	1,4
AZOTE_0001_61FVGAAXX:3:93:12933:13709	9,08	16,55	0	0	0	26,88	9,45	0	7,88
E4D3Z3Y01A14UP	7,15	2,1	5,88	3,66	2,36	4,09	4,94	2,75	0
E4D3Z3Y01BHN7L	56	50,52	35,3	64,06	70,83	73,61	53,12	64,19	69,99
E4D3Z3Y01BRB91	71,49	49,47	31,77	62,23	62,57	51,8	58,06	39,43	64,39
E4D3Z3Y01C9R5R	17,87	11,58	3,53	7,32	7,08	13,63	9,88	11,92	11,2
E4D3Z3Y01CESUS	4,77	8,42	2,35	12,81	4,72	4,09	6,18	1,83	0
E4D3Z3Y01CK6W2	8,34	8,42	8,24	10,98	14,17	24,54	2,47	11	14
E4D3Z3Y01CTAU8	1525,1	785,13	1584,9	1491,7	704,79	940,55	1630,7	1486,5	1250,1
E4D3Z3Y01EDUEZ	2666,5	1295,6	3244	3545,2	1424,9	2355,5	2436,2	2421,9	3324,6
E4D3Z3Y01EH4T3	107,23	185,23	157,67	131,78	118,05	139,04	140,83	141,23	225,37
E4D3Z3Y01EO1V5	264,5	290,48	350,64	309,31	207,78	223,55	277,96	264,11	313,56
E8W526M03F3P9L	8,34	3,16	11,77	16,47	5,9	5,45	11,12	11	9,8
E8W526M03F18QX	9,53	8,42	17,65	29,28	8,26	6,82	14,82	12,84	9,8
E8W526M03G7ZSK	9,53	21,05	14,12	20,13	33,06	10,9	12,35	49,52	18,2
E8W526M03G9DZL	20,25	32,63	28,24	12,81	60,21	39,53	19,77	23,84	29,4
E8W526M03GL4RU	965,08	593,58	1057,8	1253,7	815,76	689,74	925,3	868,44	1030,3
E8W526M03GVXF3	126,29	98,93	215,33	120,8	22,43	46,35	128,48	106,38	162,38
E8W526M04IB8GH	95,32	95,77	105,9	76,87	145,21	130,86	58,06	63,28	99,39
E8W526M04IJSY8	408,67	408,35	360,05	722,95	806,31	687,01	287,84	362,23	429,75
E8W526M04JU7CT	1405,9	741,98	1294,3	913,3	950,34	1023,7	929,01	1489,3	1272,4
EVSUSMY01AFYO4	2,38	5,26	3,53	1,83	3,54	5,45	1,24	4,59	8,4
EVSUSMY01B76KK	53,62	50,52	40,01	51,25	83,82	79,06	38,3	49,52	65,79
EVSUSMY01BFXHY	45,28	101,04	45,89	53,08	123,96	96,78	61,77	66,03	64,39
EVSUSMY01BLMNF	2,38	7,37	8,24	1,83	3,54	8,18	9,88	11,92	7
EVSUSMY01BO06G	636,24	364,15	577,73	486,85	187,71	170,39	737,52	763,9	579,53
EVSUSMY01BQ22N	5,96	4,21	1,18	0	12,99	6,82	4,94	2,75	5,6
EVSUSMY01BTVXS	3,57	6,31	4,71	0	16,53	19,08	4,94	4,59	1,4
EVSUSMY01C3BIZ	35,74	30,52	25,89	36,61	41,32	42,26	35,83	25,68	32,2
EVSUSMY01CBKZO	11,91	7,37	11,77	12,81	15,35	6,82	12,35	14,67	11,2
EVSUSMY01CRPC9	13,11	8,42	12,94	10,98	10,62	16,36	19,77	4,59	5,6
EVSUSMY01DSU7X	115,57	110,51	80,01	111,65	140,48	106,32	90,18	107,29	96,59
EVSUSMY01D67WF	34,55	22,1	20	34,77	25,97	13,63	21	22,01	35
EVSUSMY01DB76E	1401,2	1575,5	1749,7	1345,2	1350,5	1203,6	1454,1	1717,6	1411
EVSUSMY01E17TU	98,89	81,04	62,36	76,87	107,43	87,24	65,48	75,2	58,79
EVSUSMY01E1RYV	3164,5	3331	3417	2331,8	3584,1	3203,3	2311,4	3181,2	3419,8
EVSUSMY01EAALD	4905,2	6266,3	3794,7	3964,4	6834,2	6817	3964,4	3759	3906,9
EVSUSMY01EGIU5	50,04	57,88	56,48	54,91	68,47	84,51	37,06	55,94	58,79
HELIUM_0001:7:11:321:1639	7,15	8,42	11,77	5,49	7,08	9,54	7,41	14,67	8,4
HELIUM_0001:7:11:743:1692	4,77	9,47	5,88	7,32	11,81	2,73	9,88	5,5	5,6
HELIUM_0001:7:23:475:1686	5,96	2,1	4,71	7,32	9,44	5,45	4,94	3,67	2,8
HELIUM_0001:7:25:232:1311	14,3	23,15	17,65	14,64	40,14	31,35	18,53	19,26	7
HELIUM_0001:7:8:969:1451	13,11	5,26	2,35	0	4,72	4,09	0	7,34	4,2
HELIUM_0001:7:88:1132:320	8,34	2,1	8,24	1,83	2,36	5,45	4,94	8,25	0
HELIUM_0001:7:9:478:686	1,19	1,05	1,18	1,83	7,08	2,73	1,24	2,75	1,4
HELIUM_0001:7:93:1273:277	4,77	3,16	5,88	1,83	4,72	4,09	3,71	3,67	4,2
HELIUM_0001:8:103:136:442	2,38	4,21	4,71	5,49	4,72	6,82	4,94	6,42	8,4
HELIUM_0001:8:29:1307:1004	5,96	5,26	1,18	0	4,72	1,36	2,47	2,75	1,4
HELIUM_0001:8:5:827:565	4,77	5,26	5,88	12,81	8,26	6,82	2,47	4,59	2,8
HELIUM_0001:8:68:1739:1137	0	1,05	3,53	3,66	0	9,54	1,24	0,92	2,8
HELIUM_0001:8:92:696:10	3,57	0	3,53	5,49	1,18	1,36	2,47	5,5	7

Annexe 9 (suite): Données de quantification normalisées des microARN caractérisés dans les banques induites par les Myc-LCO, les facteurs Nod et leur témoin correspondant.

ID Mature	Expr norm Myc-LCO 1	Expr norm Myc-LCO 2	Expr norm Myc-LCO 3	Expr norm Nod-F 1	Expr norm Nod-F 2	Expr norm Nod-F 3	Expr norm témoin 1	Expr norm témoin 2	Expr norm témoin 3
MicroARN identifiés dans miRBase 18 (autres espèces)									
ahy-miR156c	394,37	354,68	307,1	754,07	277,43	189,47	235,96	450,27	482,94
ath-miR168a	817,34	1339,8	817,76	1096,3	1147,5	1184,6	857,36	751,06	723,71
ath-miR172a	189,44	253,64	162,38	201,33	206,6	194,93	168,01	157,73	208,57
ath-miR172c	5431,9	7610,3	820,12	3534,2	6397,4	2558,6	3455,4	1921,2	1252,9
ath-miR319a	1625,2	1079,8	1594,4	1755,2	1628	1333,1	1531,9	1607,6	1552,4
ath-miR408	544,5	452,55	407,12	733,93	570,2	527,53	588,04	398	433,95
bdi-miR166f	5694	5177	5538,4	6528,5	7169,5	6455,7	6163,3	5075,9	4805,6
bna-miR156a	303,82	212,6	224,74	285,52	25,97	50,44	287,84	243,02	250,57
gma-miR159b	2014,8	1596,6	1648,5	2178	1664,6	916,02	2320,1	2282,5	1535,6
osa-miR167d	701,77	715,67	676,57	684,52	191,25	212,65	639,93	712,54	566,93
ptc-miR167f	20,25	32,63	35,3	18,3	14,17	38,17	34,59	16,51	46,19
vvi-miR171f	630,28	547,27	529,49	887,68	630,41	584,78	646,11	519,97	615,92
zma-miR171f	59,57	61,04	48,24	76,87	68,47	76,33	53,12	60,53	72,79
MicroARN identifiés dans miRBase 18 (<i>M. truncatula</i>)									
mtr-miR1510a*	744,66	1107,2	788,35	638,76	1034,2	1006	721,46	794,16	782,5
mtr-miR156e	10136	7831,3	7355,2	9651	5809,5	5106,2	7245,5	7771,1	8229,6
mtr-miR156g	15297	27569	12532	19703	8837,6	9580	16655	11304	13811
mtr-miR159a	253428	188074	262745	238016	227708	230060	288757	254228	218145
mtr-miR160	407,48	636,73	325,93	558,23	854,71	881,94	353,32	298,96	390,55
mtr-miR160c	7,15	17,89	4,71	3,66	15,35	23,17	3,71	7,34	7
mtr-miR162	652,92	343,1	645,98	561,89	487,57	437,56	597,93	567,65	550,13
mtr-miR164a	17514	26552	16487	26522	17193	20075	21761	19318	24134
mtr-miR166	53060	47462	46568	47279	68884	63095	52892	44717	43463
mtr-miR167	21,45	22,1	14,12	16,47	15,35	23,17	32,12	22,93	19,6
mtr-miR169c	40,51	45,26	36,48	40,27	70,83	36,8	33,36	55,94	36,4
mtr-miR169g	13,11	8,42	2,35	23,79	18,89	6,82	9,88	8,25	11,2
mtr-miR169m	3,57	7,37	5,88	0	2,36	8,18	1,24	1,83	9,8
mtr-miR171	0	9,47	2,35	1,83	11,81	5,45	0	4,59	1,4
mtr-miR171b	10,72	16,84	7,06	23,79	22,43	12,27	18,53	5,5	18,2
mtr-miR171f	26,21	47,36	36,48	64,06	43,68	39,53	17,3	24,76	47,59
mtr-miR172	490,88	554,64	290,63	452,07	303,4	288,98	591,75	414,5	386,35
mtr-miR2086	489,69	484,13	417,71	442,92	149,93	169,03	433,62	515,38	363,96
mtr-miR2111h	540,92	779,87	474,19	669,88	994,02	1079,6	668,34	425,51	414,35
mtr-miR2119	20,25	13,68	10,59	14,64	0	0	25,94	9,17	12,6
mtr-miR2592b	29,79	41,05	25,89	40,27	51,94	50,44	24,71	25,68	41,99
mtr-miR2592bi	15,49	14,73	15,3	14,64	23,61	20,45	14,82	18,34	15,4
mtr-miR2612a	64,34	76,83	42,36	86,02	81,46	51,8	48,18	58,69	50,39
mtr-miR2623	5,96	16,84	8,24	18,3	21,25	20,45	2,47	4,59	4,2
mtr-miR2633	67,91	52,62	37,65	64,06	36,6	46,35	42	54,11	51,79
mtr-miR2634	9,53	11,58	11,77	9,15	9,44	9,54	2,47	9,17	5,6
mtr-miR2643	139,4	99,98	88,25	139,1	142,85	111,78	144,54	153,15	90,99
mtr-miR2648	59,57	86,3	51,77	53,08	103,89	111,78	56,83	41,27	46,19
mtr-miR390	227,57	334,68	204,74	287,35	207,78	290,34	224,84	194,41	229,57
mtr-miR393	2,38	3,16	5,88	3,66	4,72	6,82	1,24	1,83	1,4
mtr-miR396a	526,62	774,6	515,37	475,87	814,58	936,46	495,39	539,22	590,73
mtr-miR398b	316,93	336,78	190,62	415,47	435,62	323,06	247,08	278,78	279,97
mtr-miR5204	98,89	16,84	10,59	126,29	35,42	14,99	177,9	25,68	19,6
mtr-miR5232	1197,4	1025,1	897,78	1107,3	1416,7	1296,3	1194,6	971,15	881,89
mtr-miR5239	5390,2	3761,5	7711,7	5251	1216	1288,2	5627,2	7223,6	6713,6
mtr-miR5250	2148,2	2796,4	1275,5	1797,3	3561,7	3301,5	1397,2	1494,8	1692,4
mtr-miR5561	52,42	47,36	47,07	32,94	54,31	34,08	40,77	37,6	50,39

Damien FORMEY DE SAINT LOUVENT

« La symbiose mycorhizienne à arbuscules *Rhizophagus irregularis* / *Medicago truncatula* : données génomiques, transcriptomiques et microtranscriptomiques »

Thèse dirigée par le Pr. Christophe ROUX

Correspondant scientifique pour l'entreprise Agronutrition: Dr. Faouzi Attia

Soutenue le 19 décembre 2012, LRSV, Castanet-Tolosan

RESUME

La symbiose mycorhizienne à arbuscules (MA) est une association entre les racines des plantes et des champignons du groupe des Gloméromycètes. Ces champignons ubiquistes assurent aux végétaux hôtes un apport minéral et hydrique essentiel à leur développement et augmentent leur résistance à certains pathogènes. Près de 80 % des espèces végétales peuvent s'associer aux champignons MA, parmi lesquelles la plupart des espèces d'intérêt agronomique. Malgré son intérêt pour établir des pratiques agronomiques nécessitant moins d'intrants (eau, engrais, produits phytosanitaires), l'exploitation agronomique de la symbiose MA est rendue difficile par le peu de connaissances génétiques que nous en avons. En raison de ses activités industrielles sur le marché des biofertilisants, l'entreprise Agronutrition a soutenu mes travaux dont les objectifs principaux sont la définition des marqueurs de typage du champignon MA modèle *Rhizophagus irregularis*, ainsi que la caractérisation d'éléments régulateurs de la symbiose chez la plante hôte modèle *Medicago truncatula*.

Le premier objectif a été abordé par l'étude du mitochondriome fongique. Le projet international de séquençage du génome de *Rhizophagus irregularis*, auquel notre équipe de recherche participe, a permis l'identification des gènes impliqués dans la production d'énergie et la respiration chez ce champignon et, ainsi, de proposer un modèle d'organisation de la chaîne respiratoire de *Rhizophagus irregularis*. Nous confirmons la présence d'une voie alternative de production d'énergie et mettons en évidence la probable capacité de ce champignon à réaliser une respiration cellulaire en condition anoxique. Une approche de séquençage des génomes mitochondriaux de différentes souches de cette espèce nous a permis d'analyser la variabilité intraspécifique du génome mitochondrial. Ces travaux aboutissent à la définition de marqueurs de typage moléculaire spécifiques des souches étudiées.

La caractérisation d'éléments régulateurs de la symbiose a été réalisée dans le cadre du projet Génoscope MIRMED portant sur une étude du microtranscriptome de *Medicago truncatula* soumis à différents stress abiotiques et biotiques. La microtranscriptomique consiste à analyser l'expression des petits ARN, notamment les microARN qui sont de courtes séquences nucléotidiques capables de moduler l'expression des gènes. L'analyse des microtranscriptomes obtenus durant les étapes présymbiotiques et symbiotiques de la mycorhization ont permis d'identifier des microARN différenciellement exprimés dans ces conditions, dont certains sont nouvellement décrits. Des phénotypes non-mycorhiziens associés à la surexpression de deux d'entre eux ont été obtenus, soulignant l'intérêt de poursuivre l'analyse du rôle, dans l'établissement de la symbiose, de ces nombreux éléments régulateurs révélés par notre analyse.

Mots-clés : *Rhizophagus irregularis*, *Medicago truncatula*, microARN, Mitochondrie, marqueurs