Université de Montréal

La génomique évolutive mitochondriale révèle des échanges génétiques et la ségrégation chez les Gloméromycètes

par

Denis Beaudet

Département de Sciences Biologiques

Institut de Recherche en Biologie Végétale - Centre sur la Biodiversité

Faculté des Arts et des Sciences

Thèse présentée à la Faculté des Arts et des Sciences

en vue de l'obtention du grade de Philosophiae Doctor (PhD)

en Sciences Biologiques

Juin 2014

Université de Montréal
Faculté des Études Supérieures et Postdoctorales

Cette thèse intitulée:

La génomique évolutive mitochondriale révèle des échanges génétiques et la ségrégation chez les Gloméromycètes

Présentée par:
Denis Beaudet

a été évaluée par un jury composé des personnes suivantes:

Dr. Marc St-Arnaud, président-rapporteur
Dr. Mohamed Hijri, directeur de recherche
Dr. Sophie Breton, membre du jury
Dr. Louis Bernier, examinateur externe
Dr. Marc St-Arnaud, représentant du doyen de la FESP

Résumé

Les champignons mycorhiziens à arbuscules (CMA) sont des organismes microscopiques du sol qui jouent un rôle crucial dans les écosystèmes naturels et que l'on retrouve dans tous les habitats de la planète. Ils vivent en relation symbiotique avec la vaste majorité des plantes terrestres. Ils sont des biotrophes obligatoires, c'est-à-dire qu'ils ne peuvent croître qu'en présence d'une plante hôte. Cette symbiose permet entre autres à la plante d'acquérir des nutriments supplémentaires, en particulier du phosphore et du nitrate. Malgré le fait que cette symbiose apporte des services importants aux écosystèmes, la richesse des espèces, la structure des communautés, ainsi que la diversité fonctionnelle des CMA sont mal connues et l'approfondissement des connaissances dans ces domaines dépend d'outils de diagnostic moléculaire. Cependant, la présence de polymorphisme nucléaire intra-isolat combiné à un manque de données génomiques dans différents groupes phylogénétique de ces champignons complique le développement de marqueurs moléculaires et la détermination de l'affiliation évolutive à hauts niveaux de résolution (c.a.d. entre espèces génétiquement similaires et/ou isolats de la même espèce).

.

Pour ces raisons, il semble une bonne alternative d'utiliser un système génétique différent en ciblant le génome mitochondrial, qui a été démontré homogène au sein d'un même isolat de CMA. Cependant, étant donné le mode de vie particulier de ces organismes, une meilleure compréhension des processus évolutifs mitochondriaux est nécessaire afin de valoriser l'utilisation de tels marqueurs dans des études de diversité et en génétique des populations. En ce sens, mon projet de doctorat consistait à étudier: i) les vecteurs de divergences inter-isolats et -espèces phylogénétiquement apparentées, ii) la plasticité des génomes mitochondriaux, iii) l'héritabilité mitochondriale et les mécanismes potentiels de ségrégation, ainsi que iv) la diversité mitochondriale intra-isolat *in situ*.

À l'aide de la génomique mitochondriale comparative, en utilisant le séquençage nouvelle génération, on a démontré la présence de variation génétique substantielle inter-isolats et -espèces, engendrées par l'invasion d'éléments mobiles dans les génomes mitochondriaux des

CMA, donnant lieu à une évolution moléculaire rapide des régions intergéniques. Cette variation permettait de développer des marqueurs spécifiques à des isolats de la même espèce. Ensuite, à l'aide d'une approche analytique par réseaux de gènes sur des éléments mobiles, on a été en mesure de démontrer des évènements de recombinaisons homologues entre des haplotypes mitochondriaux distincts, menant à des réarrangements génomiques. Cela suggère une coexistence de différents haplotypes mitochondriaux dans les populations naturelles et que les cultures monosporales pourraient induirent une sous-estimation de la diversité allélique mitochondriale. Cette apparente contradiction avec l'homogénéité mitochondriale intra-isolat généralement observée, a amené à étudier les échanges génétiques à l'aide de croisements d'isolats génétiquement distincts. Malgré l'observation de quelques spores filles hétéroplasmiques, l'homoplasmie était le statut par défaut dans toutes les cultures monosporales, avec un biais en faveur de l'un des haplotypes parentaux. Ces résultats suggèrent que la ségrégation opère durant la formation de la spore et/ou le développement ddu mycélium. De plus, ils supportent la présence d'une machinerie protéique de ségrégation mitochondriale chez les CMA, où l'ensemble des gènes impliqués dans ce mécanisme ont été retrouvé et sont orthologues aux autres champignons. Finalement, on a étudié le polymorphisme mitochondrial intra-isolat à l'aide d'une approche conventionnelle de PCR en utilisant une Taq polymérase de haute fidélité, suivie de clonage et de séquençage Sanger, sur deux isolats de *R. irregularis*. Cela a permis l'observation d'hétéroplasmie *in situ*, ainsi que la co-expression de variantes d'ARNm dans une souche *in vitro.* Les résultats suggèrent que d'autres études basées sur le séquençage nouvelle génération aurait potentiellement ignoré cette variation, offrant ainsi plusieurs nouveaux arguments permettant de considérer les CMA comme des organismes possédant une population de génomes mitochondriaux et nucléaires distincts.

Mots-clés : champignons mycorhiziens à arbuscules, génomique évolutive, génome mitochondrial, marqueurs moléculaires, éléments génétiques mobiles, transfert horizontal de gènes, recombinaison homologue, réarrangements génomiques, réseaux de gènes, mécanismes de ségrégation, anastomoses, homoplasmie, heteroplasmie, séquençage nouvelle génération

Abstract

The association between arbuscular mycorrhizal fungi (AMF) and plant roots is one of the most widespread symbioses, and thus has an important role in terrestrial ecosystems. In exchange for carbohydrates, AMF improve plant fitness by enhancing mineral nutrient uptake, in particular phosphate and nitrate. Although this symbiosis provides important services to ecosystems, the species richness, community structure and functional diversity of AMF is not well understood due to a lack of reliable molecular tools. The intra-isolate genetic polymorphism of nuclear DNA observed in AMF, combined with a lack of genomic data in a broad range of phylogenetic groups, has made it difficult to develop molecular markers and to determine evolutionary relatedness at high levels of resolution (i.e. between genetically-similar species and/or isolates).

For these reasons, it seems a good alternative to use a different genetic system by targeting the mitochondrial genome, which have been shown to be homogeneous within AMF isolates. However, given the peculiar lifestyle of these organisms, a better understanding of the mitochondrial evolutionary processes and dynamics is necessary in order to validate the usefulness of such markers in diversity and population genetics studies. In that regard, the objectives of my PhD project were to investigate: i) the divergence between closely related species and isolates, ii) mitochondrial genomes plasticity, iii) mitochondrial heritability and potential segregation mechanisms and iv) *in situ* mitochondrial intra-isolate allelic diversity.

Using comparative mitochondrial genomics using and next generation sequencing (NGS) , we found substantial sequence variation in intergenic regions caused by the invasion of mobile genetic elements. This variation contributes to rapid mitochondrial genome evolution among closely related isolates and species, which makes it possible to design reliable intra- and inter-specific markers. Also, an extensive gene similarity network-based approach allowed us to provide strong evidence of inter-haplotype recombination in AMF, leading to a reshuffled mitochondrial genome. These findings suggest the coexistence of distinct mtDNA haplotypes in natural populations and raise questions as to whether AMF single spore cultivation

artificially underestimates mitochondrial genetic diversity. This apparent contradiction with the intra-isolate mtDNA homogeneity usually observed in these fungi, led to the investigation of mitochondrial heritability in the spore progeny resulting from crossed-cultures. Although an heteroplasmic state was observed in some daughter spores, we found that homoplasmy was the dominant state in all monosporal cultures, with an apparent bias towards one of the parental haplotypes. These results strongly support the presence of a putative mitochondrial segregation proteic machinery in AMF, whose complete set of genes were orthologous with those found in other fungi. Our findings suggest that segregation takes place either during spore formation or mycelium development. Finally, we performed a conventional PCR based approach with a high fidelity Taq polymerase, followed by downstream cloning and Sanger sequencing using the model organism *Rhizophagus irregularis*. We found *in situ* heteroplasmy along with substantial intra-isolate allelic variation within the mtDNA that persists in the transcriptome. Our study also suggest that genetic variation in *Glomeromycota* is higher than meets the eye and might be critically underestimated in most NGS based-AMF studies both in nuclei and mitochondria.

Keywords : arbuscular mycorrhizal fungi, evolutionary genomics, mitochondrial genome, molecular markers, mobile genetic elements, horizontal gene transfer, homologous recombination, genome rearrangements, gene similarity network, segregation mechanism, anastomosis, homoplasmy, heteroplasmy, next generation sequencing

Table des matières

Liste des tableaux

Liste des figures

Liste des abréviations

ADN	acide désoxyribonucléique
ADNc	ADN complémentaire
ADNr	ADN ribosomal
AMF	arbuscular mycorrhizal fungi
ARN	acide ribonucléique
ARNm	ARN messager
ATP	adenosine triphosphate
atp	ATP synthetase
BiP	binding protein
BLAST	basic local alignment search tool
cDNA	complementary DNA
CDS	coding sequences
CMA	champignons mycorhiziens à arbuscules
CMN	common mycorrhizal network
cob	cytochrome b
COG	cluster of orthologous gene
cox	cytochrome C oxydase
C-terminal	carboxy-terminal
DNA	desoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
dpo	plasmid-related DNA polymerase gene
EDTA	ethylenediaminetetraacetic acid
FISH	hybridation en fluorescence *in situ*
GBS	genotyping by sequencing
GFP	green fluorescent protein
HEG	homing endonuclease gene
HGT	horizontal gene transfer
HMM	hidden markov model
indel	insertion/deletion
ITS	internal transcribed spacer
LLM	low level mutation
LSU	large ribosomal subunit
MER	mycélium extra-racinaire
mORF	mobile endonuclease encoding ORF

mRNA	messenger RNA
MSA	mitochondrial segregation apparatus
mt	Mitochondrial
nad	NADH dehydrogenase
NADH	nicotinamide adenine dinucleotide
NGS	next generation sequencing
N-terminal	amino-terminal
numts	nuclear mitochondrial DNA
ORF	open reading frame
PCG	protein coding gene
PCR	polymerase chain reaction
PLS	polymerase-like sequence
qPCR	quantitative PCR
rDNA	ribosomal DNA
RMC	réseau mycorhizien commun
RNA	ribonucleic acid
rnl	ribosomal large subunit
rns	ribosomal small subunit
rpo	RNA polymerase gene
rps	ribosomal protein subunit
RT-PCR	reverse transcriptase PCR
SIR	small inverted repeat
SNP	single nucleotide polymorphism
TIR	terminal inverted repeat
tRNA	transfer RNA
WGA	whole genome amplification

Cette thèse de doctorat est dédiée à ma copine Carolane Bisson, qui a eu la patience de partager ma présence d'esprit avec la recherche pendant cinq ans. Merci et je t'aime.

Remerciements

"*Maybe it meant something. Maybe not, in the long run... but no explanation, no mix of words or music or memories can touch that sense of knowing that you were there and alive in that corner of time and the world. Whatever it meant.*"
— Hunter S. Thompson

Je termine ce travail envahi de sentiments partagés, un mélange de fierté et de nostalgie, puisque sa complétion représente un passeport vers l'inconnu. Entreprendre un doctorat c'est choisir une destination sans connaître le chemin et se rendre compte, une fois rendu, que la destination était le chemin. Sur cette route, j'ai fait face à plusieurs obstacles et pris plusieurs détours, mais j'ai rencontré des gens qui m'ont relevé, remis sur la bonne voie et permis de garder le cap. Il me fait immensément plaisir de prendre le temps de reconnaître leur rôle dans l'aboutissement de cette thèse.

En premier lieu, je tiens à remercier mon directeur de recherche Mohamed Hijri. Je crois que je n'aurais pas pu faire un doctorat dans un autre laboratoire, sans la confiance, la liberté et le support que tu m'as fourni tout au long de mon cheminement. J'espère seulement que tu ne sois pas trop déçu que je ne sois jamais venu faire du jogging avec toi... Je dois en grande partie les connaissances que j'ai acquises en génomique mitochondriale au Pr Franz Lang, du département de biochimie de l'Université de Montréal. Merci de m'avoir accueilli dans ton laboratoire en début de parcours, où j'ai grandement bénéficié de ta patience pour l'enseignement et ton dévouement à la rigueur scientifique, qui m'ont donné des bases solides sur lesquelles m'appuyer pour le reste de ma thèse. La recherche scientifique est en soit un domaine complexe, mais lorsque l'organisme étudié est considéré un intra-terrestre, elle peut complexer. Grâce à Eva Boon qui m'a pris sous son aile à mon arrivée et sa passion pour la philosophie des concepts biologiques, j'ai su apprivoiser cette bête et me défaire du syndrome de l'imposteur. Où que tu sois mes pensées sont avec toi et je t'en suis reconnaissant.

Certaines contributions dépassent le cadre de ce monde prenant qu'est la recherche. En ce sens, Sébastien Halary, a été un mentor, un grand frère, mais surtout un ami, dans les bons comme dans les mauvais moments. Yves Terrat, qui n'a jamais vraiment apprécié, à tort, mes talents d'infographiste, merci de m'avoir sans cesse relevé le moral et tiré la pipe. Vos encouragements m'ont forgé la confiance nécessaire pour affronter toutes les intempéries. Il n'y aurait pas eu de troisième chapitre à cette thèse sans votre implication. Ivan de la Providencia avec qui j'ai eu la chance de travailler sur quatre articles et bien d'autres projets, pour son travail acharné, son dynamisme et sa passion contagieuse. Alice Roy-Bolduc pour sa joie de vivre, son perfectionnisme assumé et sa sagesse. Adriana Almeida-Rodriguez et Jocelyne Ayotte qui n'ont jamais trop rechigné à ce que j'habite pratiquement dans leur bureau.

Je tiens à remercier Laurence Daubois, Manuel Labridy, Stéphanie Berthiaume et Bachir Iffis pour votre contribution à plusieurs des chapitres de cette thèse, ainsi que Maryam Nadimi pour sa collaboration dans deux articles et sur le projet de génomique mitochondriale. Dr Marc St-Arnaud pour son soutient et les discussions enrichissantes. Terrence Bell pour avoir toujours généreusement accepté de relire mes ébauches d'articles et d'en réviser l'anglais. Cristina Micali pour son soutient dans les moments importants, ainsi qu'à tous les membres de mon laboratoire et les stagiaires qui m'ont assisté en cours de route. Aussi, rien n'aurait été possible sans le soutient financier de la compagnie PremierTech par l'entremise du programme Mitacs-Accélération. La contribution du FQRNT a aussi été cruciale au bon cheminement de mon parcours doctoral. Les bourses du CSBQ, ainsi que celle de l'IRBV, m'ont aussi amené des expériences formatrices en me permettant d'assister à des congrès internationaux.

Finalement, mais non les moindres, je veux souligner l'apport de mes amis à Montréal et ailleurs qui ont crû en moi et m'ont encouragé dans la poursuite un peu folle d'un doctorat. Un immense merci à ma famille, mon frère Mathieu, ma sœur Gabrielle, ma mère Marie-Line et mon père Fabien, ainsi qu'à ma copine Carolane Bisson qui m'ont offert une fondation sur laquelle m'appuyer, un support inconditionnel et l'énergie de poursuivre dans les moments les plus difficiles. Carolane, je n'aurais jamais terminé cette thèse sans ton support et tes encouragements, merci du fond du cœur.

Chapitre 1 - Introduction générale

1.1 Potentiel d'application des CMA en agriculture

Le milieu agricultural du 21ième siècle fait face à plusieurs défis, afin de subvenir à la demande d'une population mondiale sans cesse croissante. De ces défis on note particulièrement la prise de conscience de plus en plus généralisée du méfait que causent les engrais chimiques à l'environnement, ainsi que le stress induit par le réchauffement climatique sur les plantations en ce qui a trait à la sécheresse, aux espèces invasives et aux parasites. Dans ce contexte alarmant, l'utilisation des champignons mycorhiziens à arbuscules (CMA) dans les pratiques agricoles peut permettre le développement de solutions intégrées aux problèmes à plus long terme liés à l'épuisement des ressources naturelles disponibles pour l'agriculture. Le phosphore, un élément dont l'absorption par la plante est particulièrement favorisée par la mycorhization des racines, est utilisé de façon intensive dans les grandes cultures. Avec la demande qui ne cesse d'augmenter et les réserves mondiales qui diminuent, il a été estimé que les sources de phosphore inorganique disponibles seraient épuisées dans une cinquantaine d'années (Gilbert, 2009). C'est à ce niveau que les multiples contributions de la symbiose plante-CMA, autre que l'augmentation de la croissance, peuvent être d'une grande importance dans le développement d'une agriculture durable et saine. La symbiose mycorhizienne présente de nombreux avantages pour le développement de la plante hôte, avec une panoplie d'applications en horticulture et en agriculture (Zimmerman *et al.*, 2010). L'utilisation de champignons mycorhiziens dans les grandes cultures permet ainsi aux agriculteurs de réduire les intrants (engrais, pesticides) requis pour améliorer la fertilité des sols et augmenter les rendements (Roy-Bolduc & Hijri, 2011).

1.2 Physiologie des champignons mycorhiziens à arbuscules

Les champignons mycorhiziens à arbuscules sont des organismes fascinants tant au niveau de leur mode de vie que de leur génétique particulière. Ce sont des champignons microscopiques du sol qui jouent un rôle crucial dans les écosystèmes naturels et que l'on retrouve dans tous les habitats de la planète. Ils vivent en relation symbiotique avec la grande majorité (plus de 80%) des plantes vasculaires (James, 1998; Parniske, 2008; Smith & Read, 2008). Les CMA font partie du phylum des Glomeromycota, dont les plus anciens fossiles datent d'au moins 420 millions d'années (Redecker *et al.*, 2000). Ils sont des biotrophes obligatoires, c'est-à-dire qu'ils ne peuvent croître qu'en présence d'une plante hôte qui leur fournit les carbohydrates dont ils ont besoin pour survivre (Strack *et al.*, 2003). Cette symbiose aide la plante à acquérir des nutriments supplémentaires (en particulier du phosphore et du nitrate), tout en contribuant à augmenter sa résistance aux pathogènes présents dans le sol (St-Arnaud & Vujanovic, 2007). Cette augmentation de la résistance peut se faire entre autres via la modulation de l'expression des gènes de toxicité du pathogène (Ismail *et al.*, 2011) ou par stimulation des mécanismes de défense de la plante (Ismail & Hijri, 2012). De plus, il a été démontré que les CMA jouent un rôle dans l'augmentation et le maintient de la biodiversité des plantes (Van Der Heijden *et al.*, 1998). Les CMA ont également un impact majeur sur les propriétés microbiologiques et physiques du sol (Rillig *et al.*, 2002; Rillig *et al.*, 2010).

1.3 Symbiose et mode de vie

L'établissement de la symbiose mycorhizienne est le résultat d'un dialogue moléculaire spécialisé (Bécard *et al.*, 2004; Bucher *et al.*, 2009) entre le partenaire fongique et les racines

de la plante hôte. Cet échange entre deux partenaires évolutivement éloignés implique une série de modifications morpho-anatomiques (Genre *et al.*, 2008) qui impliquent la formation d'un appareil de prépénétration et d'un appressorium (Bonfante & Genre, 2008) à la surface de la racine, la pénétration de l'hyphe et la colonisation du cortex racinaire où se forment les structures spécialisées, caractéristiques à cette symbiose, appelées arbuscules. Par la suite le CMA va sortir de la racine et croître dans le milieu environnant en formant un réseau de mycélium extra-racinaire (MER) coenocytique interconnecté et multinucléé (Leake *et al.*, 2004). Étant donné sa nature sous-terraine intrinsèque, le MER est continuellement remodelé en réponse aux stress environnementaux, par l'entremise d'extensions et d'interconnections à l'aide de fusions hyphales. Ces dernières constituent un processus appelé anastomoses, qui implique la dégradation des parois cellulaires pour créer un cytoplasme commun entre les hyphes (Glass & Fleissner, 2006). Ce processus représente un important mécanisme de communication intra-hyphal pour le maintient de l'homéostasie dans la colonie de CMA durant la période de croissance et de reproduction (Rayner *et al.*, 1995).

Les champignons mycorhiziens à arbuscules ne possèdent pas de septa dans leurs hyphes, ce qui permet au protoplasme cellulaire de circuler librement, incluant les organelles, et permet de définir la colonie comme étant un syncytium (Friese & Allen, 1991; Bago *et al.*, 1998; Giovannetti *et al.*, 1999). Le réseau mycélien formé par les CMA, interconnecté par l'entremise d'anastomoses, peut lier entre elles des plantes de différentes espèces et/ou appartenant à différentes familles (Giovannetti *et al.*, 2004), créant ainsi un réseau mycorhizien commun (RMC). De plus, la coexistence de nombreux CMA (différents isolats et/ou espèces) dans les communautés du sol est la règle plutôt que l'exception, favorisant les

contacts entre des individus génétiquement distincts. Dans ce contexte, les anastomoses jouent un rôle crucial dans les échanges génétiques au sein d'un même isolat et entre des isolats de CMA évolutivement rapprochés, mais génétiquement divergents (Giovannetti *et al.*, 1999; Croll *et al.*, 2009). Ce mécanisme a été suggéré comme un facteur déterminant au maintient de l'intégrité génétique des membres du phylum des Glomeromycètes (Bever & Wang, 2005; Corradi *et al.*, 2007; Croll *et al.*, 2009; Angelard & Sanders, 2011; Colard *et al.*, 2011).

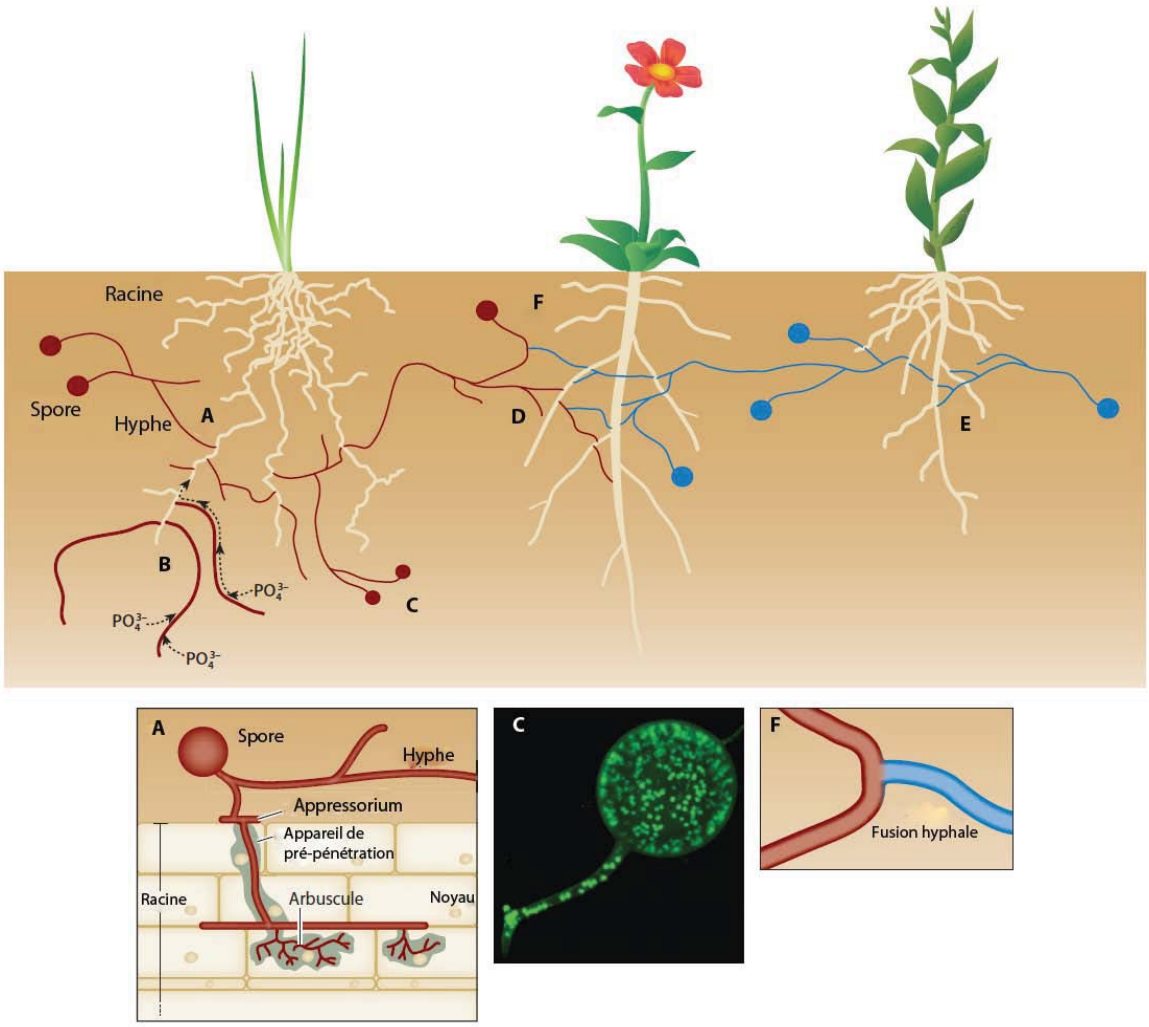


Figure 1.1. Établissement de la symbiose et mode de vie des CMA. Figure modifiée à partir de Sanders & Croll (2010) **(A)** les spores du champignon germent dans le sol. Lorsque les hyphes rencontrent les racines, la symbiose se forme avec la plante. Après avoir formé un appressorium, le champignon pénètre dans la racine, colonise le cortex racinaire et forme des

arbuscules, qui permettent le transfert des nutriments entre les partenaires. **(B)** Le champignon forme alors son MER qui permet l'absorption des nutriments présents dans le sol et leurs transports vers la plante. **(C)** Les hyphes peuvent produire de nouvelles spores multinucléées. **(D)** Les hyphes peuvent continuer à croître dans le sol et coloniser de nouvelles plantes et créer le RMC. **(E)** D'autres CMA génétiquement distincts peuvent coloniser des plantes adjacentes ou la même plante hôte. **(F)** Les hyphes de CMA compatibles peuvent fusionner et former des anastomoses.

1.4 Les défis de la taxonomie

Malgré la distribution ubiquitaire des CMA à l'échelle mondiale, environ 230 espèces sont à ce jour répertoriées. Cela s'explique en partie par la difficulté de les maintenir en culture et de les identifier. Historiquement, le système taxonomique et l'identification des CMA étaient basés sur l'interprétation des caractères morphologiques des spores, plus précisément sur le nombre et la position de parois et membranes internes, ainsi que l'attachement de l'hyphe à la spore. Cependant, cette méthode requiert une grande expertise pour son utilisation adéquate et a été l'objet de débats scientifiques, à savoir si ces critères taxonomiques avaient une origine commune ou étaient indépendants. Des études subséquentes démontrèrent que les parois et membranes des spores ne pouvaient être considérées comme des caractères indépendants puisqu'elles se différentiaient durant la germination de la spore (Franke & Morton, 1994). La structure de la paroi de la spore peut être simple, comme chez *Glomus intraradices* ou complexe avec plusieurs membranes internes chez les *Scutellospora* spp. L'observation d'espèces dimorphiques avec des caractéristiques appartenant au genre *Glomus* et *Acaulospora*, suggère que les spores des Gloméromycètes ne représentent pas une structure homologue (Redecker, 2001). Les risques de mauvaises interprétations reliés à la complexité de l'identification morphologique ont été mis de l'avant lorsqu'il a été démontré que l'espèce modèle de CMA *Glomus intraradices* DAOM-197198 avait fait l'objet d'une mauvaise

classification et faisait en fait parti du clade regroupant les isolats de *Glomus irregulare* (Stockinger *et al.*, 2009).

Les données moléculaires ont grandement changé la systématique et de nouveaux genres et familles ont été mis en place (Walker & Schüßler, 2004; Walker *et al.*, 2007; Krüger *et al.*, 2012). Les espèces sont maintenant décrites avec une combinaison de séquences de gènes nucléaires ribosomaux (ADNr) et de critères morphologiques. Les nouvelles descriptions incluent les séquences d'ADN, ce qui facilite l'identification lorsqu'il y a peu de caractères morphologiques à utiliser (Walker *et al.*, 2007). Cependant, l'identité taxonomique et phylogénétique des CMA continue d'être controversée. Ils ont été initialement assignés aux zygomycètes, un groupe fongique suspecté être paraphylétique (Seif *et al.*, 2006; Liu *et al.*, 2009). Ils ont ensuite été assignés à un phylum distinct, les Glomeromycota (Schussler *et al.*, 2001). Cependant, les phylogénies publiées sont sujettes à questionnement et manquent souvent de support statistique approprié causé par un nombre limité de séquences de gènes nucléaires ribosomaux polymorphiques, un échantillonnage taxonomique insuffisant, ou une combinaison des deux. Dans plusieurs cas, les analyses démontrent de potentiels artéfacts phylogénétiques comme l'attraction des longues branches (Felsenstein, 1978). L'ajout des séquences codantes mitochondriales d'espèces distantes de CMA, *Gigaspora rosea* et *Gigaspora margarita*, a récemment permis de confirmer avec un fort support statistique le positionnement des Glomeromycètes comme groupe sœur des Mortierellales au sein du grand groupe des champignons (Nadimi *et al.*, 2012; Pelin *et al.*, 2012) (voir Article #1 en Annexe).

1.5 Organisation génétique des CMA

L'organisation de la structure génomique nucléaire des CMA a engendré plusieurs interrogations. En effet, ces champignons contiennent plusieurs milliers de noyaux circulant librement dans leurs hyphes coénocytiques et forment des spores multinucléées (Marleau *et al.*, 2011) (Figure 1.2). Comme aucun stade de leur cycle de vie n'est réduit à un seul noyau, ils ne sont donc pas sujets à un goulot d'étranglement génétique. Les CMA ont longtemps été considérés d'anciens organismes asexués. Cependant pratiquement l'ensemble des gènes nécessaires à la méiose a récemment été caractérisé (Halary *et al.*, 2011), ce qui laisse présager la présence d'une forme de reproduction sexuée, voir de la recombinaison cryptique. L'observation d'un niveau élevé de polymorphisme nucléaire intra-isolat a mené la communauté scientifique à se questionner sur la façon dont cette variation était maintenue chez les CMA (voir Boîte 1.1). Il a été suggéré que le polymorphisme observé était réparti à travers la multitude de noyaux retrouvés dans les CMA, cet état est connu sous le nom d'hétérokaryotie (Kuhn *et al.*, 2001; Hijri & Sanders, 2004; Bever & Wang, 2005; Hijri & Sanders, 2005; Corradi *et al.*, 2007) . Toutefois, cette théorie doit être en mesure d'expliquer de quelle façon la diversité est maintenue face au problème de la dérive génétique. L'autre théorie suggérée est que la variation génétique est répartie en différents chromosomes (Pawlowska & Taylor, 2004) ou dans chaque noyau dans des gènes dupliqués, il s'agirait donc d'homokaryotie. Cependant, cette théorie doit être en mesure d'expliquer comment la variation due à l'accumulation de mutations est purgée (Figure 1.3).

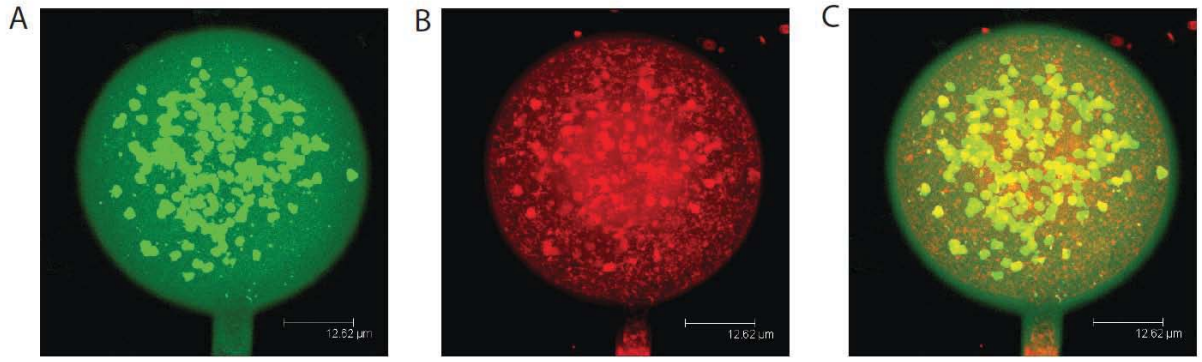


Figure 1.2. Images en microscopie confocale d'une spore vivante de *Rhizophagus sp*. Le panneau **A** montre une spore marquée avec du SytoGreen (les noyaux sont visibles et correspondent aux taches vertes), en utilisant le filtre vert. Le panneau **B**, montre les mitochondries marquées avec du MitoTracker, en utilisant le filtre rouge. Le panneau C présente une combinaison des images **A** et **B**. L'échelle représente 12.62 µm.

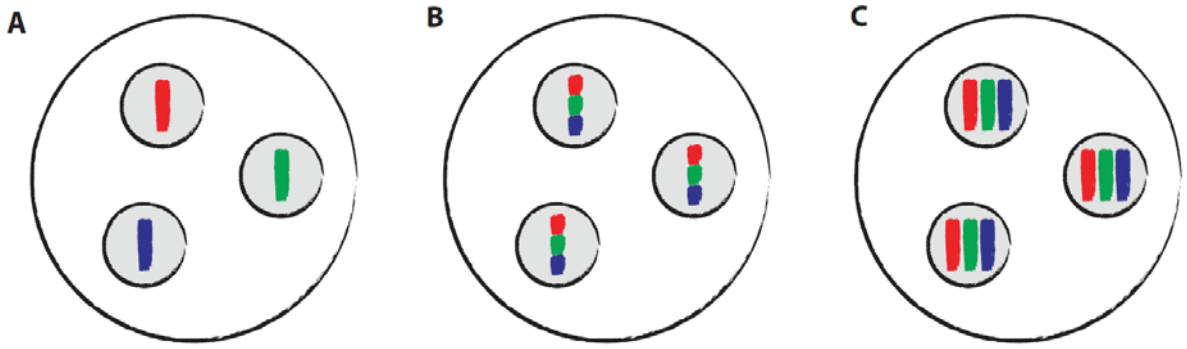


Figure 1.3. Comment est réparti le polymorphisme intra-isolat chez les CMA? (A) Entre différents noyaux possédants des variantes de gènes orthologues **(B)** au sein de mêmes noyaux entre différentes copies paralogues ou **(C)** au sein de mêmes noyaux polyploïdes entre les chromosomes. La dernière option a été écartée, du moins chez l'espèce modèle *G. irregulare*, qui est haploïde.

Boîte 1. Organisation de la variation génétique nucléaire observée chez les CMA

(i) Les premières observations de polymorphisme nucléaire chez les CMA étaient dans la région de l'ADN ribosomal (ADNr), où plusieurs variantes alléliques ont été répertoriées au sein d'un isolat (Sanders *et al.*, 1995; Lloyd-Macgilp *et al.*, 1996; Clapp *et al.*, 2003).
(ii) Deux types de variation intra-isolat ont été décrits: intra-sporale (dans une spore) (Sanders *et al.*, 1995; Clapp *et al.*, 1999; Hijri *et al.*, 1999) et inter-sporale (entre les spores appartenant au même isolat) (Sanders *et al.*, 1995; Clapp *et al.*, 2001; Boon *et al.*, 2010).
(iii) La variation allélique intra-isolat a aussi été observée pour les gènes codant pour des protéines, comme pour l'actine, le facteur d'élongation 1-alpha, la beta tubuline, H+ATPase, la binding protein (BiP) et la P-type II ATpase. (Kuhn *et al.*, 2001; Helgason *et al.*, 2003; Corradi *et al.*, 2004; Corradi & Sanders, 2006; Corradi *et al.*, 2007).
(iv) Le polymorphisme intra-isolat chez les CMA persiste aussi au niveau du transcriptome (Boon *et al.*, 2010; Tisserant *et al.*, 2012).

Preuves supportant l'homokaryotie:

(v) Comme dans un hétérokaryon on s'attend à observer une divergence d'héritabilité des génotypes dans les spores (c.a.d. présence de ségrégation), il a été démontré que les 13 variantes d'un marqueur POL-like (PLS) dans un isolat de *Glomus etunicatum*, étaient présentes dans toutes les spores filles après une génération. Les auteurs conclurent que les variantes coexistaient dans chaque noyau, possiblement sur différents chromosomes (Pawlowska & Taylor, 2004; Pawlowska, 2005).
(vi) Suite à du génotypage multi-locus d'une population de CMA sur deux gènes codants et un marqueur ADNr, il a été suggéré que le champignon forme un mycélium génétiquement uniforme (Stukenbrock & Rosendahl, 2005) et que toutes les variantes observées pouvaient être dérivé d'un seul individu (Rosendahl & Stukenbrock, 2004).

Preuves supportant l'hétérokaryotie:

(vii) En utilisant l'hybridation en fluorescence *in situ* (FISH), deux variantes fortement divergentes de la région ITS de *Scutellospora castanea* étaient ségrégées dans différents noyaux, alors que certains contenaient les deux variantes (Kuhn *et al.*, 2001).
(viii) Le génome de l'espèce de CMA *Glomus irregulare* a été démontré haploïde, ce qui enlève la possibilité d'expliquer la répartition de la variation génétique sur plusieurs chromosomes dans un noyau chez cette espèce (Hijri & Sanders, 2004).
(ix) Le nombre de variantes du gène PLS a été estimé à deux par génomes, ce qui représente un fort contraste aux 13 variantes coexistant dans un isolat de *Glomus etunicatum* (Hijri & Sanders, 2005).
(x) Divergence observée dans le nombre de copies par noyau du gène BiP entre trois isolats de *Glomus irregulare* (Corradi *et al.*, 2007).
(xi) Preuves expérimentales d'échanges génétiques par anastomose entre des isolats génétiquement distincts de *Glomus irregulare*, engendrant un hétérokaryon (Croll *et al.*, 2009).

(xii) Une étude a démontré que les échanges génétiques et la ségrégation engendrent une progéniture ayant différents effets symbiotiques sur la croissance de la plante hôte en comparaison à l'isolat parental (Angelard *et al.*, 2010; Angelard & Sanders, 2011).
(xiii) Le génome de l'espèce modèle de CMA *Glomus irregulare* DAOM-197198 a été très difficile à assembler possiblement due au polymorphisme élevé et possède une taille estimée supérieure à la quantité d'ADN par noyau (Martin *et al.*, 2008).

1.6 Biologie moléculaire face à la problématique du polymorphisme nucléaire

Le polymorphisme nucléaire pose obstacle à plusieurs techniques d'analyse couramment utilisées en biologie moléculaire. Par exemple, lors de l'extraction de l'ADN, on se retrouve avec un mélange de tous les noyaux (c.à.d. une population d'allèles) dans le produit final. Ce problème peut cependant être limité par l'amplification de l'ADN contenu dans une spore unique, où l'on connaît approximativement le nombre de noyaux. Cela permet de limiter dans une certaine mesure le nombre d'allèles présents dans la solution finale. Il est actuellement difficile d'amplifier l'ADN provenant d'un noyau unique en raison de la difficulté d'isolation par microdissection, sans avoir de contamination. Cependant, cette technique a récemment été réussie avec succès dans l'optique du séquençage de quatre noyaux de l'espèce *G. irregulare* (Lin *et al.*, 2014). Une autre limitation auquel les études moléculaires sont confrontées est au niveau de la PCR (polymerase chain reaction). La nature stochastique de cette technique fait en sorte que la diversité génétique des allèles amplifiés ne reflète probablement pas celle retrouvée originellement, surtout dans des organismes potentiellement hautement polymorphiques comme les CMA. L'estimation du nombre de copies d'un gène donné (par qPCR) ou de la ploïdie (par cytométrie en flux, cinétique de réassociation et reconstruction génomique) est aussi problématique puisque l'on n'obtiendra rien de plus qu'une moyenne de la population totale de noyaux présents. De plus, la transformation des

CMA pour l'expression de gènes hybrides marqués avec la green fluorescent protein (GFP) est instable (Helber & Requena, 2008), ce qui limite la possibilité de confirmer la localisation de protéines d'intérêt.

1.7 Nécessité d'outils de diagnostics moléculaires

Les CMA sont fonctionnellement divers et varient énormément dans une vaste gamme de caractéristiques comme leur capacité d'acquisition des nutriments (Jakobsen *et al.*, 1992; Cavagnaro *et al.*, 2005), leur besoin en carbone organique (Koch *et al.*, 2006), leur protection envers les maladies (Maherali & Klironomos, 2007) et dans leur patron de croissance (Hart & Reader, 2002; Koch *et al.*, 2004). Il s'agit d'un constat important puisqu'il a été démontré que l'effet sur le fitness de la plante varie en fonction de l'espèce de CMA et même selon l'isolat de la même espèce (Van Der Heijden *et al.*, 1998). De plus, la ségrégation génétique d'un isolat a des effets significatifs sur la croissance de la plante hôte (Angelard *et al.*, 2010; Angelard & Sanders, 2011). Cela démontre la complexité du maintient d'une communauté de CMA efficace dans les fonctions que l'on veut promouvoir dans les champs.

La caractérisation de ces traits génétiques dépend donc d'outils de diagnostic moléculaires afin d'être en mesure de mesurer la diversité et l'abondance des CMA sur le terrain. Le plus gros défi auquel la recherche sur les CMA est confrontée est l'observation d'une grande variation génétique intra-isolat. De nombreuses études démontrent la variation substantielle dans la séquence de gènes nucléaires (voir Boîte I). Cette situation peut expliquer en partie les difficultés rencontrées par le consortium responsable de l'assemblage du génome nucléaire de *Glomus irregulare* (Martin *et al.*, 2008). La présence de ce polymorphisme nucléaire, combiné

à un manque de données génomiques dans différents groupes phylogénétiques de ces champignons, complique le développement de marqueurs moléculaires pour l'identification et la quantification, ainsi que le positionnement phylogénétique précis des CMA au sein du grand groupe des champignons.

1.8 Marqueur simple copie

Plusieurs études tentent d'amener des solutions à ces problèmes. Par exemple, certains tentent de développer, à l'aide d'une approche bio-informatique, un groupe de marqueurs potentiellement simple copie chez les CMA (Boon, 2012). Ces marqueurs permettraient éventuellement d'avoir une idée de la variation génétique totale retrouvée entre les différents génomes d'un même isolat. Ils permettraient d'adopter une approche d'étude des populations sur les noyaux de ce même isolat. De plus, il serait possible de vérifier comment la diversité génétique de ces marqueurs serait maintenue via les anastomoses ou perdue via la dérive génétique. L'approche des marqueurs simple copie est très intéressante, puisqu'elle permet de diminuer drastiquement le nombre de variantes que l'on s'attend à retrouver et que ces variantes sont nécessairement inter nucléaire. Cette méthode a été employée avec succès pour le locus *sod1* qui code pour la Copper-Zinc Superoxide Dismutase qui est probablement simple copie chez *G. intraradices* (Corradi *et al.*, 2009). Ils ont observé une grande diversité au niveau des nucléotides et des acides aminés encodés entre six espèces de CMA et 14 isolats de *G. intraradices* pour ce locus.

1.9 Marqueurs ciblant les fonctions symbiotiques

Une méthode a été suggérée dans l'optique d'une compréhension plus fonctionnelle à des fins d'application des CMA en agriculture (Gamper *et al.*, 2010). Ils suggèrent une approche utilisée en écologie microbienne utilisant des gènes codant pour des protéines impliquées dans des fonctions clés des interactions symbiotiques des CMA. Pour être en mesure d'être de bons marqueurs, ces gènes devront être liés intimement au processus écosystémique de façon à ce que les chances qu'un changement dans leur expression soit directement lié à un changement de la fonction de l'écosystème soient les plus élevées possible. De plus, le gène marqueur potentiel devrait être conservé entre différentes espèces de CMA, mais distinct des autres organismes présents, ce qui pose un défi considérable. Il a été suggéré d'utiliser le 'High affinity Pi Transporter', impliqué dans l'apport de phosphate, dont la séquence est connue chez trois espèces de CMA appartenant à trois lignées phylogénétiques différentes (Harrison *et al.*, 2002; Karandashov & Bucher, 2005; Javot *et al.*, 2007; Smith *et al.*, 2011). Un autre candidat serait un 'High affinity ammonium transporter' dont la séquence est connue chez *G. intraradices,* qui est impliqué dans l'apport d'ammonium. Des gènes impliqués dans le métabolisme des carbohydrates, des gènes structuraux et de métabolisme général des CMA pourraient aussi être ciblés. Cette approche est intéressante dans la perspective où elle permettrait de cibler l'expression des gènes fonctionnellement importants à l'aide de micropuces à ADNc, ce qui serait très intéressant pour la compréhension globale de l'interface plante-CMA et éventuellement d'accroître leur efficacité plus spécifique sur le terrain. Cependant, cette approche ne contourne pas le problème du polymorphisme et le développement de ces marqueurs s'avérera probablement ardu.

1.10 Marqueurs mitochondriaux et plan de recherche

Une autre approche prometteuse est le développement de marqueurs moléculaires mitochondriaux pour les CMA. Au lieu de viser à développer des marqueurs dans des gènes nucléaires hétérogènes, il semble une bonne alternative d'utiliser un système génétique différent en ciblant le génome mitochondrial, qui a été démontré homogène au sein d'un même isolat (Lee & Young, 2009; Formey *et al.*, 2012). L'ADN mitochondrial est depuis longtemps considéré comme un outil de choix en biologie évolutive en raison de son taux de mutation élevé, ainsi que le fait qu'un individu possède généralement un seul haplotype mitochondrial, hérité maternellement, sans présence de recombinaison. (Birky *et al.*, 1978; Birky, 2001; Sato & Sato, 2012).

Cependant, étant donné le mode de vie particulier des CMA, une meilleure compréhension des processus évolutifs mitochondriaux est nécessaire afin de valoriser l'utilisation de tels marqueurs dans des études de diversité et en génétique des populations. En ce sens, mon projet de doctorat consistait à étudier: **(i)** les vecteurs de divergences inter-isolats et -espèces phylogénétiquement apparentées (**ii)** la plasticité des génomes mitochondriaux (**iii)** l'héritabilité mitochondriale et les mécanismes potentiels de ségrégation, ainsi que **(iv)** la diversité mitochondriale intra-isolat *in situ*.

Mise en contexte (Chapitre 2) - quels sont les vecteurs de divergence mitochondriaux inter-isolats et -espèces ?

Le génome mitochondrial des champignons contient une proportion élevée d'introns, ce qui pourrait donner lieu à du polymorphisme de taille au niveau des gènes entre différentes espèces. De plus, l'absence de variation intra-isolat pour une région mtLSU a été démontré chez les CMA, mais il y avait cependant présence de polymorphisme inter-isolat (Raab *et al.*, 2005). Cela est intéressant dans le sens où le développement de marqueurs mitochondriaux permettrait éventuellement de faire la distinction entre deux isolats d'une même espèce. Le premier génome mitochondrial publié, celui de *Glomus irregulare* FACE#494 (Lee & Young, 2009), confirma l'homogénéité de l'ADN mitochondrial, mais aussi la présence de nombreux éléments mobiles dans les régions inter-géniques. La publication de ce génome était considérée comme une pierre angulaire de la recherche sur les CMA puisque cela ouvrait la porte à la génomique mitochondriale comparative de différentes espèces et isolats de CMA (Lang & Hijri, 2009). Dans cette optique, deux études subséquentes comparèrent la divergence mitochondriale entre cinq isolats de *Glomus irregulare* (Formey *et al.*, 2012) et deux espèces phylogénétiquement apparentées (Beaudet *et al.*, 2013a).

Chapitre 2 - Rapid mitochondrial genome evolution through invasion of mobile elements in two closely related species of arbuscular mycorrhizal fungi

Denis Beaudet, Maryam Nadimi, Bachir Iffis and Mohamed Hijri*

Institut de Recherche en Biologie Végétale, Département de Sciences Biologiques, Université de Montréal, 4101 Rue Sherbrooke Est, Montréal (Québec) H1X 2B2, Canada

Published in: PloS ONE, april 18, 2013. 8(4): e60768. doi:10.1371/journal.pone.0060768

Author contributions: Conceived and designed the experiments: MH. Performed the experiments: DB BI. Analyzed the data: DB MN. Contributed reagents/materials/analysis tools: DB MH. Wrote the paper: DB MH.

2.1 Abstract

Arbuscular mycorrhizal fungi (AMF) are common and important plant symbionts. They have coenocytic hyphae and form multinucleated spores. The nuclear genome of AMF is polymorphic and its organization is not well understood, which makes the development of reliable molecular markers challenging. In stark contrast, their mitochondrial genome (mtDNA) is homogeneous. To assess the intra- and inter-specific mitochondrial variability in closely related *Glomus species*, we performed 454 sequencing on total genomic DNA of *Glomus* sp. isolate DAOM-229456 and we compared its mtDNA with two *G. irregulare* isolates. We found that the mtDNA of *Glomus* sp. is homogeneous, identical in gene order and, with respect to the sequences of coding regions, almost identical t*o G. irregulare*. However, certain genomic regions vary substantially, due to insertions/deletions of elements such as introns, mitochondrial plasmid-like DNA polymerase genes and mobile open reading frames. We found no evidence of mitochondrial or cytoplasmic plasmids in *Glomus* species, and mobile ORFs in *Glomus* are responsible for the formation of four gene hybrids in *atp6, atp9, cox2* and *nad*3, which are most probably the result of horizontal gene transfer and are expressed at the mRNA level. We found evidence for substantial sequence variation in defined regions of mtDNA, even among closely related isolates with otherwise identical coding gene sequences. This variation makes it possible to design reliable intra- and inter-specific markers.

2.2 Keywords

Arbuscular mycorrhizal fungi; Mitochondrial Genome; Comparative Mitochondrial Genomics; Plasmid related *dpo*; mobile endonuclease ORFs; Horizontal Gene Transfer.

2.3 Introduction

Arbuscular mycorrhizal fungi (AMF) are plant root-inhabiting obligate symbionts that form symbiotic associations with approximately 80% of plant species (Wang & Qiu, 2006; Smith & Read, 2008). This symbiosis helps plants to acquire nutrients and protects them from soil-borne pathogens (Azcón-Aguilar & Barea, 1997; St-Arnaud & Vujanovic, 2007) by inducing plant resistance (Datnoff *et al.*, 1995; Cordier *et al.*, 1998; Pozo *et al.*, 2002; Ismail & Hijri, 2012) or inhibiting pathogen growth (Ismail *et al.*, 2011). In return, plants provide carbohydrates, which AMF cannot acquire from extracellular sources. They are an important component of soil microbial communities, as they are able to exchange their genetic material between compatible isolates through a process called anastomosis (Croll *et al.*, 2009). The latter have been hypothesized to be an important factor in maintaining the genetic diversity found in Glomeromycota and to attenuate the effect of genetic drift within a population (Bever & Wang, 2005; Corradi *et al.*, 2007; Croll *et al.*, 2009; Angelard & Sanders, 2011; Colard *et al.*, 2011). AMF are currently thought to reproduce clonally, based on the absence of a recognizable sexual stage (or apparatus). However, this hypothesis has been challenged by the identification of many orthologues of sexually-related genes (Halary *et al.*, 2011; Sanders, 2011; Tisserant *et al.*, 2012), which suggests at least the presence of cryptic recombination. AMF spores and hyphae are multinucleated, but their true genetic organization is currently under debate (Kuhn *et al.*, 2001; Pawlowska & Taylor, 2004; Bever & Wang, 2005; Hijri & Sanders, 2005; Pawlowska & Taylor, 2005). However, evidence strongly suggests that nuclei can be genetically divergent within an AMF individual. Thus, AMF are characterized by considerable within-isolate nuclear genetic diversity even at the expression level (Boon *et al.*, 2010). The presence of such diversity in AMF individuals/populations (Boon *et al.*, 2010;

VanKuren *et al.*, 2013), combined with a lack of molecular data, have hindered the use of nuclear markers to assess questions on community structure, diversity and function. In contrast, AMF mitochondrial (mt) DNA is homogeneous within single isolates (Raab *et al.*, 2005; Lee & Young, 2009), making it a good target for marker development. Following this logic, the mitochondrial large subunit (LSU) rRNA gene has been explored for its usefulness as a marker (Raab *et al.*, 2005; Börstler *et al.*, 2008; Thiéry *et al.*, 2010), although determining its specificity at the isolate level is still challenging for all AMF taxa aside from the model species *G. irregulare*.

Comparative AMF mitochondrial genomics has been proposed as an approach to open up new possibilities for development of strain-specific molecular markers (Lang & Hijri, 2009; Lee & Young, 2009; Formey *et al.*, 2012) given that the type of mitochondrial marker necessary to establish specificity at different divergence levels may vary. This approach has been shown to be a powerful tool for the study of evolutionary relationships among lower fungi (Seif *et al.*, 2005). Unfortunately, only three AMF mitochondrial genomes had been published until recently, including that of *Glomus intraradices* (Lee & Young, 2009; Formey *et al.*, 2012) (renamed to *G. irregulare (Stockinger et al., 2009)* and changed again recently to *Rhizophagus irregularis* based on an exhaustive molecular phylogeny of rRNA genes (Krüger *et al.*, 2012); in the present paper, we will use the older nomenclature) as well as those of two distant AMF species, *Gigaspora rosea* and *Gigaspora margarita* (Nadimi *et al.*, 2012; Pelin *et al.*, 2012). Compared to *G. irregulare*, the Gigasporaceae genomes have an inflated mitochondrial genome size that is mainly the result of extended intergenic regions. These regions are not syntenic and both genomes harbor *cox1* and *rns* genes with exons encoded on different

strands, whose products are joined at the RNA level through either trans-splicing events of group I introns, or base-pairing. The mitochondrial protein sequences in the dataset were sufficient to confirm the phylogenetic relationship of AMF with *Mortierellales* as a sister group. This shows that a broader sampling of AMF mtDNA can answer questions about the evolution of these ecologically important fungi. Formey et al. (2012) have recently sequenced four isolates of *G. irregulare* (Formey *et al.*, 2012) and were able to develop isolate-specific markers using variable regions that were created by the insertion of mobile elements.

Those elements, including linear or circular plasmids and mobile ORF encoding endonucleases (mORFs), are present in a broad range of fungal mitochondrial genomes (For review see (Hausner, 2012)). Plasmids are autonomously-replicating circular or linear extrachromosomal DNA molecules. They are found in three broad types: circular plasmids encoding a DNA polymerase gene (*dpo)* (Griffiths & Yang, 1995), linear plasmids with terminal inverted repeats encoding either a *dpo* or *rpo* (RNA polymerase) gene or both (Klassen & Meinhardt, 2007), and retroplasmids, which usually encode a reverse transcriptase (Kennel & Cohen, 2004). Free linear or circular plasmids encoding *dpo* can be present in the mitochondria of fungi (Griffiths & Yang, 1995) and plants (Brown & Zhang, 1995). Segments have been shown to integrate within the mtDNA of fungi (Bertrand & Griffiths, 1989; Cahan & Kennell, 2005; Ferandon *et al.*, 2008), but plasmid-related *dpo* insertions tend to fragment, shorten (since they are not selected for) and eventually disappear from mitochondrial genomes. Plasmid-related *dpo* insertions have been reported in the AMF *Gigaspora rosea*, but are virtually absent from the closely related paraphyletic zygomycetes. The mobility of mORFs, elements that thrive in *Glomus*, is mediated by the site-specific DNA endonuclease

they encode. This endonuclease cleaves ORF-less alleles by creating a double-strand break in DNA and initiates the insertion and fusion of the mobile element. The same process, called *intron homing*, has been proposed for group I introns (Dujon, 1980). Several lines of phylogenetic evidence support the hypothesis of the evolutionary-independent ancestral origins of mORFs (Bell-Pedersen *et al.*, 1990). These highly mobile elements have the ability to carry group I introns (Dalgaard *et al.*, 1993), intergenic sequences (Sharma *et al.*, 1992), and coding sequences (Eddy, 1992). The first reported case of mitochondrial gene transfer caused by those elements was a mORF-mediated insertion of a foreign *atp6* carboxy-terminal in the blastocladiomycete *Allomyces macrogynus* (Paquin *et al.*, 1994).

The present study compared the mitochondrial genomes of the newly sequenced AMF species *Glomus* sp. DAOM226456 (a *Glomus diaphanum*-like species based on spore morphology) with two isolates of the closely related *G. irregulare*. Along with a highly divergent intron insertion pattern, we found insertions of plasmid-related DNA polymerase and propagation of mobile open reading frame (mORFs) encoding endonucleases in *Glomus* mtDNAs. Our findings have brought to light the first evidence of AMF interspecific exchange of mitochondrial coding sequences entailing formation of gene hybrids in *Glomus* sp. *atp6*, *atp9* (coding for the subunit 6 and 9 of the ATP synthetase complex), *cox2* (cytochrome C oxidase subunit 2) and *nad3* (NADH dehydrogenase subunit 3) genes.

2.4 Materials and Methods

2.4.1 Fungal material

Spores and mycelium of *Glomus* sp. (DAOM-229456*)* and *G. irregulare* (DAOM 197198) were cultivated *in vitro* on a minimal (M) medium with carrot roots transformed with *Agrobacterium rhizogenes*, as described in the literature (Bécard & Fortin, 1988). The medium was liquefied using a 0.82 mM sodium citrate and 0.18 mM citric acid extraction buffer solution. The resulting fungal material was further purified by hand under a binocular microscope, to remove root fragments.

2.4.2 DNA extraction

Spores and mycelium were suspended in 400 µL of the DNeasy Plant Mini Kit AP1 buffer (Qiagen) and crushed with a pestle in 1.5 ml microtubes, and the DNA was purified according to the manufacturer's recommendations. Purified DNA in a final elution volume of 40 µL was stored at -20 ºC until use.

2.4.3 RNA extraction

Fresh *Glomus* sp. fungal material was harvested from *in vitro* cultures. RNA extraction was performed using an E.Z.N.A. Fungal RNA Kit (Omega Biotek) according to manufacturer's recommendations. Total RNA was treated with Turbo DNase (Applied Biosystems) for 30 min at 37 °C to remove residual DNA fragments that could interfere with downstream applications. In order to prevent chemical scission of the RNA during heat inactivation of the DNase at 75 °C for 15 min, EDTA was added at a final concentration of 15 mM. In total, 40 µl of 100 ng/µl RNA was collected and stored at -80 °C until use. The RNA concentration was determined using a Nanophotometer Pearl (Implen).

2.4.4 cDNA synthesis

From the total RNA previously extracted, 500 ng were used for cDNA synthesis with the SuperScript III reverse transcriptase kit (Life Technologies, Canada) according to manufacturer's recommendations, using oligo dT. The only change from these recommendations was the addition of $MgCl_2$ to a final concentration of 15 mM to compensate for the EDTA added in the previous step. In order to remove RNA complementary to the cDNA, 1 µl of RNase ONE ribonuclease (Promega, Canada) was added to the cDNA and incubated at 37°C for 20 min. The resulting cDNA was stored at -20°C until use.

2.4.5 Polymerase chain reaction (PCR)

The proposed intergenic markers to discriminate between *G. irregulare* DAOM197198 and *Glomus* sp. were tested by PCR using the KAPA2G Robust Hotstart ReadyMix PCR kit (KapaBiosystems, Canada). The specific primers used were respectively *rnl-cox2*_197198_spec_F (5'-AAAGGAATTACATCGATTTA-3'), *rnl-cox2*_197198_spec_R (5'-ACAAGAAGGTTTGCATCGCTA-3'), *nad6-cox3*_dia_spec_F (5'-CCACTAGTTAAGCTACCCTCTA-3') and *nad6-cox3*_dia_spec_R (5'-AATCATACCGTGTGAAAGCAAG -3'). The variable length primers were *rnl-cox2*_197198_size_F (5'- TAGGGATCAGTACTTTAGCCAT -3'), *rnl-cox2*_197198_size_R (5'- TCCTTACGGTATGAATGGTAAG -3'), *rnl-cox2*_dia_size_F (5'-AGACTTCTTCAGTTCCACAATCA -3') and *rnl-cox2*_dia_size_R (5'-ATGGCTAAAGTACTGATCCCTAC -3'). For 40 µl of PCR reaction volume, 12 µl of water, 20 µl of 2X PCR buffer, 3.5 µl of (5 µM) forward and reverse primers, and 1 µl of DNA were added. Cycling parameters were 94 °C /3 min, followed by 38 cycles of: 94 °C /30 sec, 54 °C

/25 sec, 72 °C /45 sec and a final elongation at 72 °C. PCR products were separated by electrophoresis in a 1.5% (w/v) agarose gel and visualized with GelRed under UV light.

2.4.6 Reverse transcriptase - polymerase chain reaction (RT-PCR)

Our objective with regard to PCR reactions on cDNA was to assess which regions of the gene hybrid reported in *atp6*, *atp9*, *cox2* and *nad3* were expressed at the mRNA level. For each of the four hybrids, a forward primer designed in the conserved 'core' structure of the gene (*atp6*_core_F: 5'-AGAGCAGTTTGAGATTGTTAAG-3', *atp9*_core_F: 5'-CTGGAGTAGGAGTAGGGATAGT-3', *cox2*_core_F: 5'-CATGGCAATTAGGATTTCAAGA-3' and *nad3*_core_F: 5'-TCGTTCCTTTGTTCGTGCTA-3') was used in combination with three reverse primers designed respectively in the inserted C-terminal (*atp6*_insert_R1: 5'-AGCCTGAATAAGTGCAACAC-3', *atp9*_insert_R1: 5'-GTAAGAAAGCCATCATGAGACA-3', *cox2*_insert_R1: 5'-TGAGAAGAAAGCCATAACAAGT-3' and *nad3*_insert_R1: 5'-AGAAGTATGAAAACCATAGCAATC-3'), the mobile ORF (*atp6*_mORF_R2: 5'-AGTCTTCGAATATACTGGCAG-3', *atp9*_mORF_R2: 5'-TGTCGAGTCTCCAAAGTATGT-3', *cox2*_mORF_R2: 5'-ACTGAATTCCTGTGTTTCGATCT-3' and *nad3*_mORF_R2: 5'-TGACGAATGGTTAGACGATGT-3') and the native C*-terminal portion of the corresponding gene (*atp6*_native_R3: 5'-CGTACCGTCGTAACAAGTAGA-3', *atp9*_native_R3: 5'-CCATCATTAAGGCGAATAGA-3', *cox2*_native_R3: 5'-CTAACAAACTCCCGACTATTACCT-3' and *nad3*_native_R3: 5'-

AGAATGAAGACCATTGCAAC-3'). To verify that there was no residual mitochondrial DNA in the cDNA, the primers Ctrl_positive_*nad5*exon4_689F (5'-ACCATTCTGTTATGTTCTAATGT-3') and Ctrl_positive_*nad5*exon4_689R (5'-GTCTGACTTAGCAGGTTAGTTAAG-3') were designed in *nad5* exon 4 and used as a positive control on cDNA and negative control on RNA. The RT-PCR reactions were carried out using the KAPA2G Robust Hotstart ReadyMix PCR kit (KapaBiosystems, Canada) as described above in the PCR section.

2.4.7 Cloning

Cloning reactions were performed on each successful RT-PCR amplification. The ligation reactions were done using the pGEM-T Easy Vector Systems kit (Promega, Canada) according to manufacturer's recommendations. The transformation was carried out in *E. coli* DH5 alpha competent cells. Bacterial colonies were screened via PCR using T7 and SP6 universal primers as described in the PCR section.

2.4.8 Sequencing, assembly and gene annotation

Glomus sp. total DNA was sequenced using 454 Titanium Flex shotgun technology (one plate) and the respective resulting 1,078,190 reads were assembled with Newbler (Genome Quebec Innovation Center, McGill University, Montreal, Canada). Gene annotation was performed with MFannot (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl), followed by manual inspection and introduction of missing gene features as described in Nadimi et al., (2012). *G. irregulare* isolates 494 and DAOM-197198 mtDNAs (accession numbers FJ648425 and HQ189519 respectively) were used for

comparison. Sequencing of the cloned RT-PCR products was performed on the same sequencing platform, using Sanger technology with T7 and SP6 universal primers.

2.4.9 Phylogenetic analysis

For each gene of interest (*atp6, atp9, cox2* and *nad3*) in 12 AMF species, the dataset contains the corresponding C*-terminal for: *Glomus* sp. DAOM-229456, *G. irregulare* isolate 494, *G. irregulare* DAOM-197198, *G. irregulare* DAOM-240415, *G. irregulare* DAOM-234179, *G. irregulare* DAOM-234328, *G. irregulare* DAOM-213198, *Glomus* sp. DAOM-240422, *G. fasciculatum* DAOM-240159, *G. aggregatum* DAOM-240163, *G. cerebriforme* DAOM-227022, *Gigaspora rosea* DAOM-194757 (accession number JQ693396) and 3 selected fungal representatives: *Mortierella verticillata* (accession number AY863211), *Smittium culisetae* (accession number AY863213) and *Rhizopus oryzae* (accession number AY863212). The sequences were deposited in databases under the accession numbers: JX074786-JX074817. The reference phylogeny was constructed using the concatenated 'core' sequence (without the C*-terminal portion used previously) of the same four genes. The DNA sequence alignments and the inference of maximum likelihood trees using GTR+G (with five distinct gamma categories) were performed using the integrated program MEGA version 5 (Tamura *et al.*, 2011). Bootstrap resampling (1000 replicates) was carried out to quantify the relative support for each branch of the trees. Bayesian analysis were done using MrBayes version 3.2 using the GTR+G model (with five distinct gamma categories), four independant chains, one million cycles, tree sampling every 100 generations and a burn-in value of 25%.

2.5 Results and Discussion

2.5.1 Glomus sp. genome organization and structure

The complete sequence of the *Glomus* sp. 229456 mt genome was a double-stranded circular DNA molecule, exempt of polymorphism, with a size of 87,763 bp. The annotated sequence of *Glomus* sp. was deposited in GenBank under the accession number JX065416. Its mtDNA harbors the typical set of 41 mitochondrial genes found in other AMF (two rRNAs, 14 protein coding genes (PCGs) and 25 tRNAs). The PCGs include three ATP synthetase (*atp*), one cytochrome b (*cob*), three cytochrome C oxydase (*cox*) and seven NADH dehydrogenase (*nad*) genes. Also, 19 ORFs and 31 introns are inserted in this newly sequenced mt genome (Figure 2.1).

2.5.2 Comparative view of three Glomus mtDNAs

The gene content in *Glomus* sp. and *G. irregulare* mitochondrial genomes is similar to that found in zygomycetes, except for *rps3* and *rnpB.* The mtDNAs of both AMF species have the same gene order, and all genes are transcribed from one strand with very similar coding regions except for the insertion of mobile ORF elements (mORFs) in the *atp6, atp9*, *cox2* and *nad3* genes of *Glomus* sp. (Figure 2.2). However, there are many differences in the number of

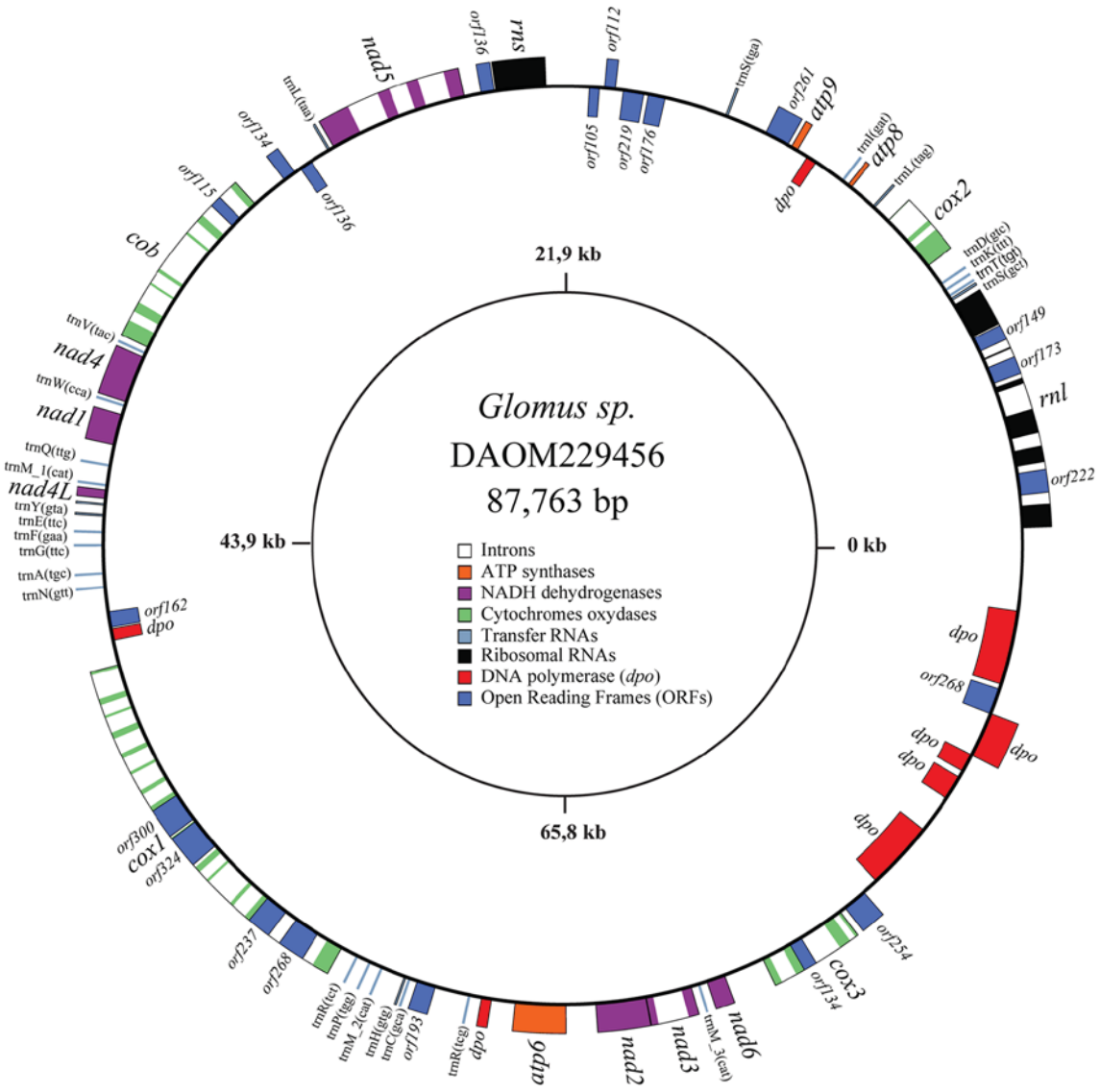


Figure 2.1. The *Glomus* sp. DAOM-229456 mitochondrial genome circular-map, opened upstream of rnl. Genes on the outer and inner circumference are transcribed in a clockwise and counterclockwise direction, respectively. Gene and corresponding product names are*atp6, 8, 9,* ATP synthase subunit 6; *cob*, apocytochrome b; *cox1–3*, cytochrome c oxidase subunits; *nad1–4, 4L, 5–6*, NADH dehydrogenase subunits; *rnl, rns*, large and small subunit rRNAs; A–W, tRNAs, the letter corresponding to the amino acid specified by the particular tRNA followed by their anticodon. Open reading frames smaller than 100 amino acids are not shown.

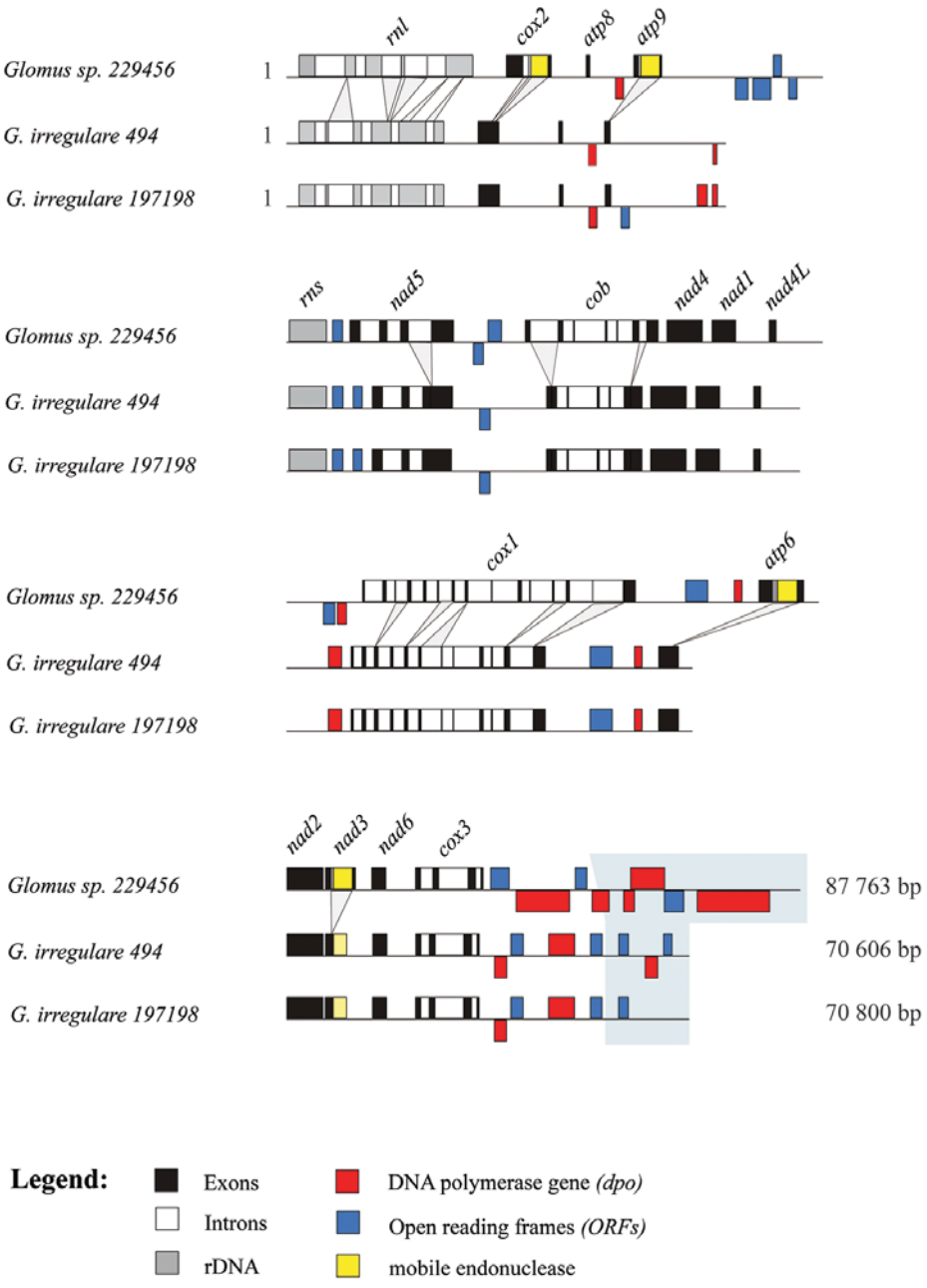


Figure 2.2. Comparative view of the three mitochondrial genomes linear maps. The exons (black), introns (white), rDNA (gray), *dpo* plasmid insertions (red), ORFs (blue) and mobile endonuclease (yellow) are represented. Divergence in intron insertion pattern is indicated by projections. A hyper-variable region in the *cox3-rnl* intergene is boxed in grayscale.

introns, some of which carry more substantial sequence differences than do the coding sequences. The differences in the presence of introns and mORFs explain the inflated genome size of 87,763 bp in *Glomus* sp., as compared to 70,800 bp in *G. irregulare* 197198 (Table 2.1). *Glomus sp*. *cox1* intron 8 is the homolog of an intron inserted at the same position in *Rhizopus oryzae* and angiosperms (with 76 and 79% of sequence identity, respectively) (Lang & Hijri, 2009). *G. irregulare cox1* intron 7 is also inserted at the same position, but has an eroded ORF encoding the homing endonuclease gene, and thus also shares identity with the intron RNA secondary structure of *R. oryzae* and plants. The plant *cox1* intron was thought to have been acquired from a fungal donor, due to the proximity of its clade to that of fungi

Table 2.1. Gene and intron content in AMF and selected fungal mtDNAs.

	Genes										
Species	*rnl, rns*	*atp 6, 8, 9*	*cob*	*cox 1, 2, 3*	*nad 1–6* [a]	*trn A–W*	*rnpB*	*rps3*	*ORFs* [b]	*Intron I* [c]	*Intron II* [c]
Glomus Sp. 229456	2	3	1	3	7	25	0	0	19	31	1
Glomus irregulare 494	2	3	1	3	7	25	0	0	8	26	0
Glomus irregulare 197198	2	3	1	3	7	25	0	0	8	26	0
Gigaspora rosea	2	3	1	3	7	25	0	0	4	13	1
Smittium culisetae	2	3	1	3	7	26	1	1	3	14	0
Mortierella verticillata	2	3	1	3	7	28	1	1	7	4	0
Rhizopus oryzae	2	3	1	3	7	23	1	0	4	9	0
Allomyces macrogynus	2	3	1	3	7	25	0	1	4	26	2
Saccharomyces cerevisiae [d]	2	3	1	3	0	25	1	1	3	9	4

[a]Includes *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* and *nad6*.
[b]Only ORFs greater than 100 amino acids in length are listed, not including intronic ORFs and *dpo* and *rpo* fragments.
[c]Intron I and Intron II denote introns of group I and group II, respectively.
[d]*S. cerevisiae* strain FY 1679 [57].
doi:10.1371/journal.pone.0060768.t001

rather than to the non-vascular plant *Marchantia.* Knowing the extent to which the intron has spread in angiosperms (Cho *et al.*, 1998; Sanchez-Puerta *et al.*, 2008), it would be interesting to see whether such an invasion has also occurred within the Glomeromycota phylum.

Further, intergenic regions differ substantially in sequence: some are identical while others show signs of very fast, substantial changes including point mutations, insertions, deletions

and inversions (Figures 2.2 and 2.3). Most of these differences occur in the *cox3-rnl* intergene, a large hyper-variable region that has been invaded by *dpo* fragments. The variations observed in intergenic regions provide an opportunity to develop species- specific molecular markers as shown in Figure 2.3, and even isolate-specific markers or methods allowing reliable identification and/or quantification of these fungi. Lack of efficient and powerful molecular

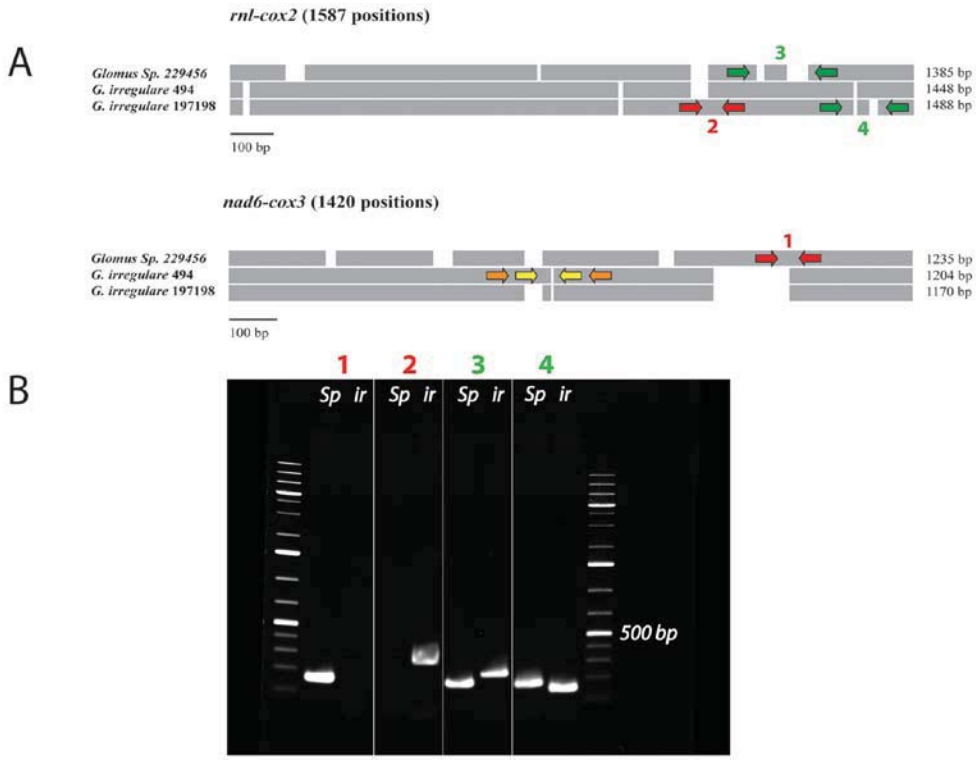


Figure 2.3. Schematic alignment representation of two mitochondrial intergenic regions (*rnl-cox2* and *cox3-nad6*) showing the presence of numerous insertions and deletions (indels). A) The red arrows indicate the approximate position of the PCR primers that yield strain-specific markers, while the green arrows indicate the position of PCR primers that produce a size-specific marker. The yellow and orange arrows indicate potential regions to design, respectively, specific and size-specific markers in *G. irregulare* 494. **B**) Agarose gel electrophoresis figure showing the PCR results of the proposed markers on *Glomus* sp. 229456 (*Gs*) and *G. irregulare*DAOM197198 (*Gi*) respectively for each marker. Marker 1 shows the *Glomus* sp.specific amplification (156 bp), while marker 2 shows the *G. irregulare* 197198 specific marker (263 bp). The size-specific marker 3 yield a length of 160 bp for*Glomus* sp. and 226 bp for *G. irregulare* 197198. Finally, the size-specific marker 4 yield a length of 159 bp for *Glomus* sp. and 131 bp for *G. irregulare* 197198.

markers for AMF identification and quantification constitutes a major problem that limits the analysis of population genetics and field studies in AMF. Mitochondrial DNA is homogeneous within the AMF individuals studied to date, but evidence of genetic polymorphism between *G. irregulare* isolates has been observed in intergenic regions. They harbor highly conserved genes as well as highly variable regions, which promises to facilitate AMF barcoding at different taxonomic levels, an analysis that is currently challenging to carry out using nuclear genes. Hyper-variable intergenic regions with eroded *dpo* insertions and indels in intergenic regions constitute useful mitochondrial areas on which to focus attention in order to develop suitable markers for discriminating isolates of the same species. Intron insertion pattern variations, genome reorganizations (such as gene shuffling) and coding region divergences will make it possible to distinguish between different AMF species, genera and families.

Our 454 pyrosequencing data and direct PCR sequencing showed that *G. irregulare* DAOM197198 and *Glomus* sp. mtDNAs are homogeneous, meaning that all the mitochondrial genomes in a given isolate are essentially identical, in stark contrast to the nuclear genomes. Our results confirm the previous report by Lee *et al.* (2009) suggesting homoplasmy in the first completed Glomeromycota mitochondrial genome of the AMF *G. intraradices* (*G. irregulare* isolate 494). A rapid and effective mitochondrial segregation mechanism was suggested to explain those findings. It was previously demonstrated that isolates of the same species can exchange nuclear material through anastomosis (Croll *et al.*, 2009), but exchange of divergent mitochondrial haplotypes has yet to be shown. This leads us to question whether polymorphism does indeed occur through anastomosis, and for how many generations mitochondrial heteroplasmy is maintained.

2.5.3 Rapid expansion of plasmid-like DNA polymerase sequences in *Glomus*

Plasmid-related DNA polymerase genes are found in mobile mitochondrial plasmids that occur either as free linear or circular DNAs, and have been shown to also insert into mtDNA (for review see (Hausner, 2012)). One striking feature in the comparison of the two closely related *Glomus* species is the presence of numerous *dpo* insertions in the intergenic regions of their mtDNA (Figure 2.2). All three *Glomus* mtDNAs contain a large number of *dpo* fragments, most of which are substantially divergent in sequence and therefore are most likely the result of independent plasmid insertion events. Even the two *G. irregulare* (isolates 494 and DAOM197198), otherwise almost identical in sequence, differ in *dpo* sequence, which supports the interpretation that *dpo* insertion occurs repeatedly and frequently through evolutionary time. A *bona fide* and complete *dpo* gene is present in *Glomus* sp., and its sequence is different from those in *G. irregulare* isolates. Because of its complete length, it most likely results from a recent insertion event. There is no evidence that *dpo* is functional when inserted in mtDNA. As in numerous other cases, *dpo* coding regions are fragmented in *Glomus* and occur on both strands, representing a good indicator of a genomic region experiencing little if any selective evolutionary constraints.

The source of the *dpo* insertions in *Glomus* mtDNA remains elusive. We did not find any free mitochondrial plasmids in our *Glomus* sp. and *G. irregulare* isolate DAOM197198 shotgun data (combining nuclear and mitochondrial DNAs), as we did for *Gigaspora rosea*, where a 3582 bp contig with high sequence coverage was found (Nadimi *et al.*, 2012). However, since the *Glomus* strains used in this study come from aseptic *in vitro* cultures, and even though the *G. rosea* fungal material was extracted from *in vivo* greenhouse pot cultures, we cannot rule

out the possibility that an environmental vector is the source of *dpo* plasmids and is responsible for their propagation in *G. rosea*. Interestingly, *dpo* plasmids have been found to occur in numerous plants, notably in *Daucus carota* (Robison & Wolyn, 2005) which is used as a host plant for AMF cultures *in vitro*. The obligate biotrophic dependence of AMF on plants could be one of the reasons that *dpo* insertions are most abundant in *Glomus* mtDNAs yet virtually absent in mitochondrial sequences of the Blastocladiomycota (except for a single 100 amino acid long fragment occurrence in *Smittium culisetae* mtDNA), which is the closest phylogenetic group to the Glomeromycota.

Mobile element insertions have been shown to trigger genomic rearrangements such as gene shuffling through homologous recombination (Brügger *et al.*, 2004) and even genome linearization (Biessmann *et al.*, 1992; Fricova *et al.*, 2010; Hausner, 2012). Whenever sequence repeats occur, more than one genome conformation may exist, but we have no evidence that this happens in *Glomus* mtDNA. It would be interesting to examine whether numerous recent *dpo* insertions with high sequence similarity might act as genomic repetitions and give rise to genome reorganization in closely related AMF species. Integrated plasmid segments within mitochondrial genomes, even though they are neutral or cryptic, could promote genomic rearrangements.

2.5.4 Mobile ORF elements (mORFs) in *Glomus*

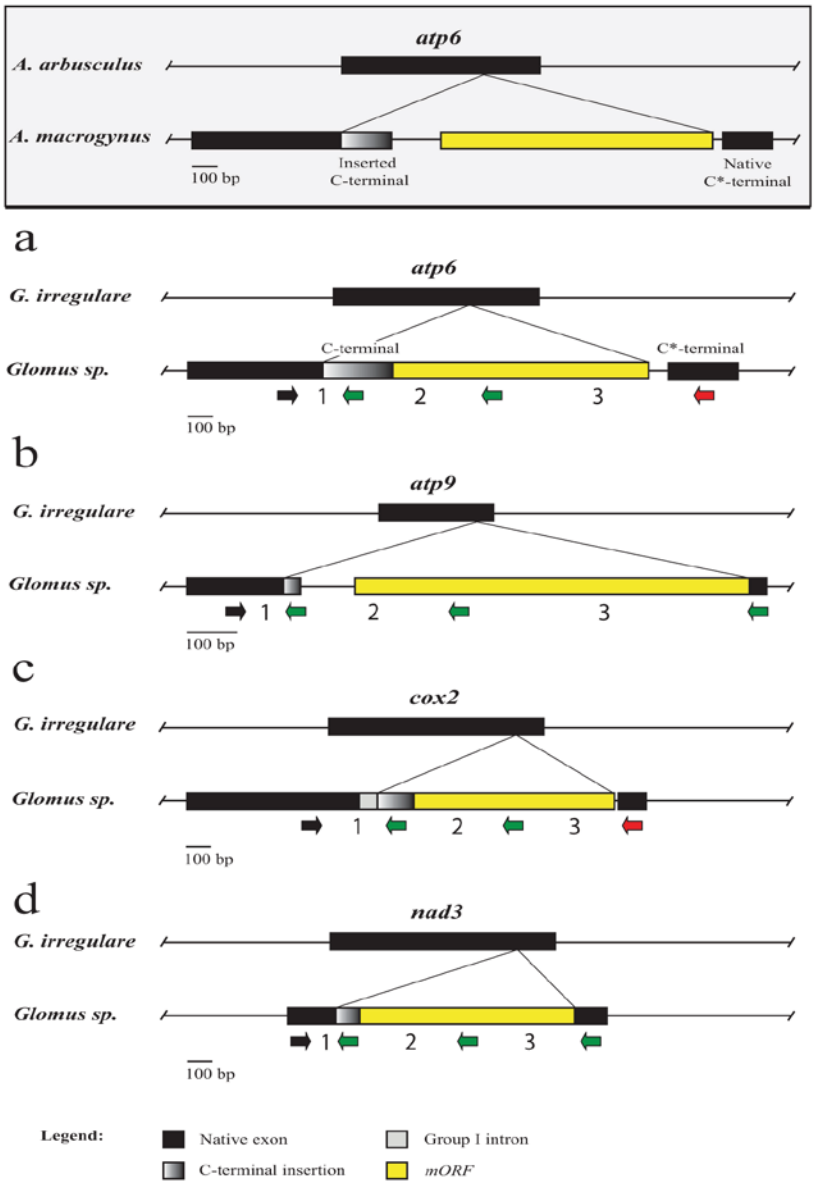
Although most ORF-encoding endonuclease genes are inserted in introns where they have been shown to play a role in propagation, they can also be present in genes in which their evolutionary impact is less obvious. We identified numerous mORFs encoding endonuclease

genes unique to *Glomus* sp. isolate DAOM-229456 mtDNA. When we annotated the sequences of the *atp6*, *atp9*, *cox2* and *nad3* genes, we observed that they all have a peculiar organization. Indeed, these genes harbor a carboxy-terminal duplication (C*-terminal) that was found downstream of a mORF insertion. For example, in the *atp6* gene, the duplicated portion of the C-terminal was found about 1000 bp downstream, following an inserted LAGLIDADG endonuclease ORF. When we compared the DNA sequence of the C*-terminal portion with the corresponding sequence of *G. irregulare* isolate 494, a close relative to *Glomus* sp., we found a 91.2% nucleotide identity. In contrast, the comparison between the *Glomus* sp. *atp6* duplicated carboxy-terminals (C-terminal and C*-terminal) showed a low sequence identity of 63.5%. Interestingly, comparison of the *Glomus sp*. C*-terminal amino acid sequences with the corresponding portion in *G. irregulare* showed 100% identity, indicating that the mutations observed in DNA are all synonymous. However, the comparison of the amino acid sequences of *Glomus sp*. *atp6* carboxy-terminals showed 91% identity.

Surprisingly, when we designed a forward primer in the upstream sequence (5' gene portion) and two reverse primers in the C-terminal and C*-terminal respectively, we found that the C-terminal is transcribed with the upstream sequence resulting in a putative hybrid transcript while the C*-terminal was not expressed into mRNA. Thus we hypothesized that the C-terminal portion could have been acquired from a donor through horizontal gene transfer (HGT). We also observed similar organization in *atp9*, *cox2* and *nad3* genes of *Glomus* sp. where the carboxy-terminal portion (C*-terminal) was replaced partially or completely by one carried by a mORF (C-terminal) encoding a LAGLIDADG endonuclease (except in *atp9*, a GIY-YIG family endonuclease) (Figure 2.4A: a, b, c and d). In *atp6*, the insert lacks a stop

codon and the ORF is in phase with the native gene. In *atp9*, the insert, along with the mORF, completely replaces the native 3' end, while in *cox2* and *nad3* only a portion of the carboxy-terminal is replaced (Table 2.2). The resulting gene hybrids are expressed at the mRNA level in all four cases as shown in Figure 2.4B. After sequencing of the cDNA bands, we found that the mORF and the inserted C-terminal are integral parts of the transcript in all four genes. However, in *atp6* and *cox2*, the native C*-terminal was not expressed into mRNA.

A


atp6
A. arbusculus
A. macrogynus
100 bp
Inserted
C-terminal
Native
C*-terminal
a
atp6
G. irregulare
Glomus sp.
C-terminal
C*-terminal
1
2
3
100 bp
b
atp9
G. irregulare
Glomus sp.
1
2
3
100 bp
c
cox2
G. irregulare
Glomus sp.
1
2
3
100 bp
d
nad3
G. irregulare
Glomus sp.
1
2
3
100 bp
Legend:
Native exon
Group I intron
C-terminal insertion
mORF

B

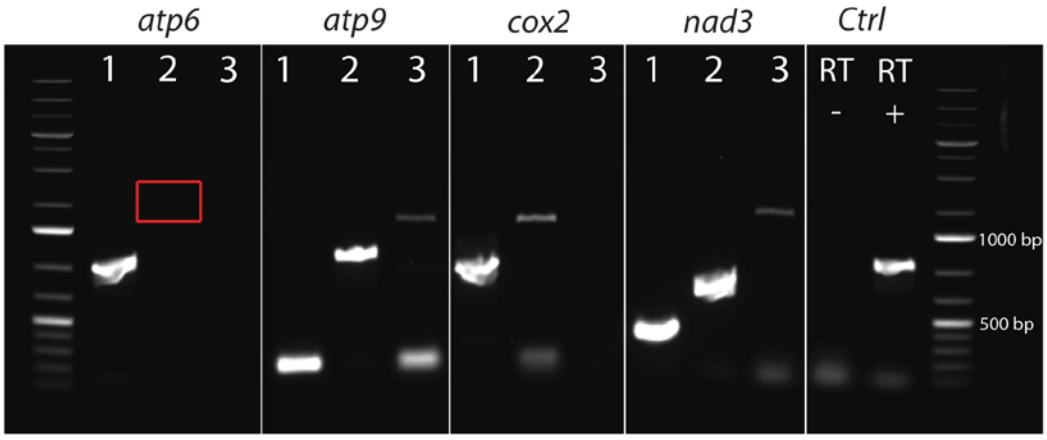

atp6
atp9
cox2
nad3
Ctrl
1 2 3
1 2 3
1 2 3
1 2 3
RT RT
- +
1000 bp
500 bp

Figure 2.4. Comparison of gene hybrids found in *atp6*, *atp9*, *cox2* and *nad3*. A) The *atp6* gene hybrid reported for *Allomyces macrogynus* (grayscale, boxed) is used as a reference in a comparison of the most similar *atp6* (a), *atp9* (b), *cox2*(c) and *nad3* (d) genes in *Glomus* sp. mtDNA. Each occurrence is put in perspective with the gene of a close relative (either *Allomyces arbusculus* or *G. irregulare*) in order to show the insertion point of the foreign element with the projections. Exons are in black, while the inserted foreign C-terminal is shaded in gray. The mobile endonuclease element is in yellow. For each gene, the black arrow indicates the position of the forward primer used in the downstream RT-PCR experiment in combination with three different reverse primers. The green arrows indicate expression at the RNA level of the corresponding portion of the gene, while the red arrows indicate a negative amplification. **B**) Agarose gel electrophoresis figure showing the RT-PCR results. For each gene hybrid, the expression at the RNA level was tested using a forward primer in the conserved gene core and a reverse primer respectively in the inserted C-terminal (1), the mobile endonuclease (2) and the native C*-terminal portion (3). Primers in *nad5* exon 4 were used as a positive control on cDNA (RT +) and negative control on RNA (RT-). The expected size of the amplified fragments was: *atp6* inserted C-terminal (684 bp),*atp9* inserted C-terminal (149 bp), *atp9* mORF (717 bp), *atp9* native C*-terminal (1085 bp), *cox2* inserted C-terminal (938 bp), *cox2* mORF (1291 bp), *nad3*inserted C-terminal (261 bp), *nad3* mORF (597 bp), *nad3* native C*-terminal (1183 bp) and the positive control in *nad5* exon 4 had an expected amplicon size of 689 bp. The red box indicates a faint band that is present on the gel.

Table 2.2. Description of the gene hybrids found in *Glomus* sp. 229456 mtDNA.

	atp6	*atp9*	*cox2*	*nad3*
Total length	1569	1171	1894	1242
CDS length	1569	225	837	1242
Features				
Group I intron	-	-	[684–922]	
Inserted C-terminal	[537–774] [1]	[175–225] [1]	[923–1018] [1]	[194–313] [1]
mORF	[550–1567] [1]	[334–1119] [1]	[1187–1738] [1]	[451–867] [1]
Native C*-terminal	[1582–1860]	[1120–1171] [1]	[1739–1894]	[1116–1238] [1]
Remarks	C-terminal and m*ORF* in phase with native gene	C-terminal in phase with native gene.	Partial inserted C-terminal in phase with native gene.	C-terminal and m*ORF* in phase with native gene.

[1]Gene hybrid features that are expressed at the mRNA level (see Figure 4B).
doi:10.1371/journal.pone.0060768.t002

These gene hybrid structures are similar to that of the *atp6* gene previously described in the *Allomyces macrogynus* (Figure 2.4, grayscale box), a species that belongs to the basal fungal phylum Blastocladiomycota (Paquin *et al.*, 1994). The same scenario has also been observed in the *Rhizopus oryzae atp9* and *Mortierella verticillata cox2* genes (Seif *et al.*, 2005). These hybrids contain a carboxy-terminal duplication as well as a mORF encoding an endonuclease,

which has been biochemically demonstrated to be responsible for the element mobility. In *Allomyces macrogynus*, the inserted C-terminal was shown to have been recently acquired by HGT based on the divergence in sequence it had with the native C*-terminal, while the latter had a perfect sequence identity with the corresponding gene portion of the closely related species *Allomyces arbusculus*.

2.5.5 Evidences of horizontal gene transfer between *Glomus* spp.

The *Glomus* sp. *atp6, atp9*, *cox2* and *nad3* native C*-terminals showed higher nucleotide sequence identity to those of *G. irregulare* 494 (91, 98, 93 and 98%, respectively) than their duplicated C-terminal counterparts (64, 71 and 81 and 73%, respectively) (Supplementary information, Figures S2.1 to S2.4 and Tables S2.1 to S2.4). However at the protein level, the comparison of the C*-terminal amino acid sequences of the *atp6*, *atp9, cox2* and *nad3* genes with the corresponding portion in *G. irregulare* 494 was 100% for *atp6, 94% for atp9, and* 100% for *cox2* and *nad3*. The high sequence identity of the native *Glomus* sp. C*-terminals with *G. irregulare 494,* is in stark contrast to the low similarity observed with the inserted C-terminal portions and points to a recent HGT event, as was described in *Allomyces* spp. *(Paquin et al., 1994)*. However, the HGT hypothesis could likely apply to the *atp6* and *cox2* genes, since their native C*-terminal portion is no longer translated and could undergo rapid divergence. For the *atp9* and *nad3* hybrids, even though it is less parsimonious, the observed sequence divergence between the duplicated portions could have been caused by independent evolution following the mobile element insertion, since both are expressed in the mRNA transcript. It would also be interesting to see if some of the reported gene hybrids can still accomplish their functions at the protein level, given that the mORF and both C*-

terminals are expressed in some cases. They are apparently expressed pseudogenes but post-translational modification mechanisms may be in place to ensure that the resulting protein is functional. We did not find a mORF-less copy of those genes that could have been transferred to the *Glomus* sp. nuclear genome that could explain a pseudogenization in *Glomus* sp. mtDNA.

In regards to the HGT hypothesis, and in order to evaluate whether there is a plausible donor for the duplication, we compared the carboxy-terminal sequence of these genes with those in 11 *Glomus* spp. (to avoid redundancy we didn't add the *G. margarita* sequences since they are identical to *G. rosea*) and three phylogenetically related fungal representatives (Figure 2.5). In all four *Glomus* sp. gene hybrids (*atp6, atp9*, *cox2* and *nad3*), the native C*-terminal sequences cluster within the *Glomus* spp. group as expected given the reference phylogeny (Figure 5, grayscale box), thereby supporting a recent insertion of the foreign element. The *atp6* gene carboxy-terminal comparison (Figure 2.5A) shows that the mORF-derived C-terminal is related to a *Glomus* sp. isolate DAOM213198 with a moderate 60% bootstrap value. Surprisingly, in *atp9* (Figure 2.5B) the inserted C-terminal is even more distantly related to *Glomus* spp. than to *G. rosea*. In *cox2* (Figure 2.5C) the *Glomus* sp. inserted C-terminal and the more divergent AMF species *G. cerebriforme* are in the same cluster. Finally, the *nad3* C-terminal clustered with *Glomus* sp. 213198, as it was the case for *atp6,* with a 79% bootstrap value (Figure 2.5D). Also, the *nad3* gene shows high variability in length in *Glomus* spp., due to the insertion of those elements.

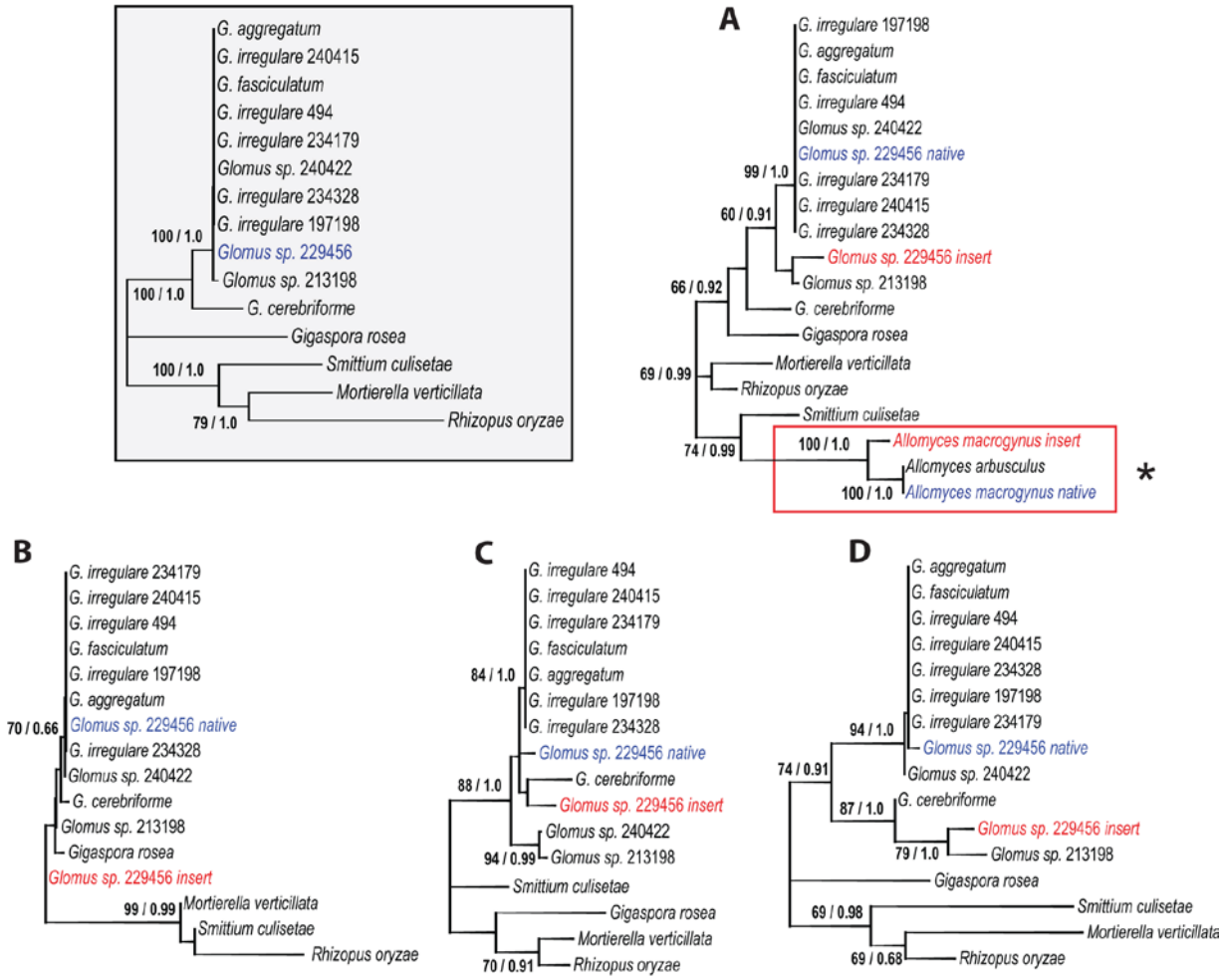


Figure 2.5. Native and inserted C-terminals unrooted maximum likelihood phylogenetic trees. The first number at branches indicates ML bootstrap values with 1000 bootstrap replicates and the second number indicates posterior probability values of a MrBayes analysis with four independant chains. Bayesian inference predict similar trees (not shown). The concatenated tree of the *atp6*, *atp9*, *cox2* and*nad3* 'core' genes (without the duplicated C*-terminal portion) (1489 alignment positions) of selected AMF representatives (grayscale boxed) are compared with those of the *atp6* (298 alignment positions), where the red box with the asterisk point out to the reference *Allomyces* spp. HGT event (Figure 4) (A), *atp9* (51 alignment positions) (B), *cox2* (106 alignment positions) (C) and *nad3* (120 alignment positions) (D) C*-terminals. The *Glomus* sp. native C*-terminals are in blue, while the inserted C-terminals are in red.

In all four cases, the native *Glomus* sp. C*-terminal is nested within the *Glomus* spp. group and the inserted C-terminal is in a different cluster. Although it is difficult to pinpoint the donor of the sequence duplications, due to the possibly complex evolutionary history of those mobile elements with numerous insertion/loss events and 3' end reshufflings, our data suggest HGT from a foreign AMF species, and thus the first reported occurrence in Glomeromycota. The presence of foreign DNA elements could potentially hamper mitochondrial gene phylogeny analysis unless the foreign C-terminals are carefully removed from the native portion of the gene.

2.6 Conclusion

The inclusion of mitochondrial sequences from phylogenetically distant AMF species in the database is essential for developing a better understanding and classification of AMF within fungi. The mitochondrial genome comparison presented here for two closely related AMF species reveals substantial changes in mitochondrial gene sequences, resulting from *dpo* plasmid insertions and mobile ORFs invasions, along with intergenic sequence variation. This illustrates the importance of adding closely related species to the numerous isolates of the same species in the AMF mitochondrial genome collection. Comparative mitochondrial genomics, together with a broader sequencing effort in AMF, opens new avenues for the development of molecular markers at different evolutionary distances. It would be interesting to identify the source of plasmid-related DNA polymerase in AMF mtDNA, which should provide an estimate of the extent to which it is present within the Glomeromycota phylum and an assessment of the consequences on mitochondrial genome organization. Also, the mORF-carried foreign C-terminal described here represents the first reported evidence of HGT in

AMF. The intimate relationship between AMF, the roots of their plant symbiont and soil microorganisms might be a perfect biological context to facilitate such transfers. To what extent the mobilome and HGT may have contributed to AMF evolution is a topic that merits exploration in future studies.

2.7 Acknowledgements

This work is a part of a research project organized and coordinated by Premier Tech. The authors are grateful for financial support from NSERC Cooperative Research and Development (CRD), Premier Tech and CRIBIQ. We would also like to thank Biopterre *centre du développement des bioproduits* and CRBM. Our thanks to Dr. B.F. Lang for bioinformatics assistance and access to an automated organelle genome annotation software, and to Dr. David Morse and Dr. Terrence Bell for comments on the manuscript.

Mise en contexte (Chapitre 3) - quel est l'impact d'insertions d'éléments mobiles sur la plasticité des génomes mitochondriaux des CMA?

Le chapitre précédent démontre une grande divergence dans les régions intergéniques causée par l'insertion et l'érosion d'éléments mobiles comprenant entre autres des 'plasmid-related DNA polymerase genes' (*dpo*) et des cadres de lecture ouverts encodant des endonucléases (mORFs). L'évolution rapide de ces génomes mitochondriaux, due à la présence de ces éléments considérés 'égoïstes', offre l'opportunité de développer des marqueurs permettant de discriminer des isolats et des espèces rapprochés de CMA (Corradi & Bonen, 2012). Cela constituait une avancée importante pour l'identification de souches, mais la présence de ces éléments mobiles soulevait des questions en ce qui a trait à leur impact sur l'organisation de la synténie mitochondriale. L'ADN mitochondrial a longtemps été considéré comme un outil de choix en biologie évolutive en raison de son taux de mutations élevé et la perception conventionnelle qu'un individu ne possède qu'un seul haplotype mitochondrial, hérité maternellement, sans présence de recombinaison (Birky *et al.*, 1978; Birky, 2001; Sato & Sato, 2012). Cependant, des études démontrèrent des cas de recombinaison mitochondriale et d'hétéroplasmie chez les animaux, mais particulièrement chez les plantes et les champignons, ce qui mena à une réévaluation de cette perception (Barr *et al.*, 2005; White *et al.*, 2008). La recombinaison moléculaire est un important mécanisme évolutif qui permet l'élimination de mutations délétères et la création de nouveaux allèles dans une population. L'intégration d'éléments mobiles comme les *dpo* et les SIRs a été démontré être à l'origine de réarrangements génomiques par recombinaison homologue dans d'autres organismes, dont les champignons (Schofield *et al.*, 1992; Bi & Liu, 1996; Paquin *et al.*, 2000; Ferandon *et al.*, 2008; Tanaka *et al.*, 2012). Cependant, l'impact de l'insertion de tels éléments

dans les génomes mitochondriaux des Gloméromycètes reste à déterminer. L'observation de recombinaison mitochondriale chez les CMA modifierait l'interprétation de plusieurs aspects de recherche, principalement pour les analyses phylogénétiques et en génétique des populations (Ballard & Whitlock, 2004).

Chapitre 3 - Mitochondrial genome rearrangements in *Glomus* species triggered by homologous recombination between distinct mtDNA haplotypes

Denis Beaudet, Yves Terrat†, Sébastien Halary†, Ivan Enrique de la Providencia and Mohamed Hijri

Institut de Recherche en Biologie Végétale, Département de Sciences Biologiques, Université de Montréal, 4101 Rue Sherbrooke Est, Montréal (Québec) H1X 2B2, Canada.

† These authors contributed equally to this work.

Published in: Genome Biology and Evolution, August 6, 2013. 5(9): 1628-1643, doi: 10.1093/gbe/evt120.

Author contributions: Conceived and designed the experiments: DB, YT, SH, MH. Performed the experiments: DB. Analyzed the data: DB, YT, SH . Contributed reagents/materials/analysis tools: MH SH YT. Wrote the paper: DB SH YT IdlP MH.

3.1 Abstract

Comparative mitochondrial genomics of arbuscular mycorrhizal fungi (AMF) provide new avenues to overcome long-lasting obstacles that have hampered studies aimed at understanding the community structure, diversity, and evolution of these multinucleated and genetically polymorphic organisms. AMF mitochondrial (mt) genomes are homogeneous within isolates, and their intergenic regions harbor numerous mobile elements that have rapidly diverged, including homing endonuclease genes, small inverted repeats, and plasmid-related DNA polymerase gene (*dpo*), making them suitable targets for the development of reliable strain-specific markers. However, these elements may also lead to genome rearrangements through homologous recombination, although this has never previously been reported in this group of obligate symbiotic fungi. In order to investigate whether such rearrangements are present and caused by mobile elements in AMF, the mitochondrial genomes from two *Glomeraceae* members (i.e. *Glomus cerebriforme* and *Glomus* sp.) with substantial mtDNA synteny divergence, were sequenced and compared with available glomeromycotan mitochondrial genomes. We used an extensive nucleotide/protein similarity network-based approach to investigate *dpo* diversity in AMF as well as in other organisms for which sequences are publicly available. We provide strong evidence of *dpo*-induced inter-haplotype recombination, leading to a reshuffled mitochondrial genome in *Glomus* sp. These findings raise questions as to whether AMF single spore cultivations artificially underestimate mtDNA genetic diversity. We assessed potential *dpo* dispersal mechanisms in AMF and inferred a robust phylogenetic relationship with plant mitochondrial plasmids. Along with other indirect evidence, our analyses indicate that members of the Glomeromycota phylum are potential donors of mitochondrial plasmids to plants.

3.2 Key words

Arbuscular mycorrhizal fungi, mitochondrial genome rearrangements, homologous recombination, plasmid-related DNA polymerase, short inverted repeats, nucleotide/protein similarity network.

3.3 Introduction

The association between arbuscular mycorrhizal fungi (AMF) and plant roots is one of the most widespread symbioses involving plants, and thus has an important role in terrestrial ecosystems (Smith & Read, 2008). In exchange for carbohydrates, AMF improve plant fitness by enhancing mineral nutrient and water uptake (Ruiz-Lozano *et al.*, 1995). AMF have also been shown to alleviate salt stress (Evelin *et al.*, 2009) and to provide protection against root pathogens (St-Arnaud & Vujanovic, 2007; Ismail & Hijri, 2012). Despite the fact that these symbioses contribute to important services in natural, agricultural and reclaimed ecosystems (Gianinazzi *et al.*, 2010), the species richness, community structure and functional diversity of AMF is not well understood (Wehner *et al.*, 2010) due to a lack of reliable molecular tools. The intra-isolate genetic diversity of nuclear DNA and RNA sequences observed in AMF (Sanders *et al.*, 1995; Kuhn *et al.*, 2001; Pawlowska & Taylor, 2004; Stockinger *et al.*, 2009; Boon *et al.*, 2010), has made it difficult to determine evolutionary relatedness at high levels of resolution (i.e. between genetically-similar species and/or isolates).

In contrast, the first mitochondrial (mt) genome sequenced from *Glomus irregulare* (synonym *Rhizophagus irregularis;* formerly *G. intraradices*) was shown to be homogeneous (Lee & Young, 2009), a state often referred to as homoplasmic. Since then, the mitochondrial large

subunit (LSU) rRNA gene has been explored for its usefulness as a marker (Raab *et al.*, 2005; Börstler *et al.*, 2008; Thiéry *et al.*, 2010), but determination of its specificity at the isolate level is challenging due to a lack of genetic variability. Recently, the mt genomes of six *G. irregulare* isolates (Formey *et al.*, 2012) and a closely related species (Beaudet *et al.*, 2013a) were compared. Striking features of these datasets were the presence and variability among isolates of numerous mobile elements such as plasmid-related DNA polymerase genes (*dpo*), short inverted repeats (SIRs) and homing endonuclease genes (HEG), which have given rise to mitochondrial gene hybrids through horizontal gene transfer (HGT). Intra-specific divergence in the eroded portions of these elements revealed the high value of designing much-needed isolate-specific molecular markers (Corradi & Bonen, 2012).

Mitochondrial DNA has long been considered a valuable tool in evolutionary biology because of its high mutation rate and the assumption that individuals only possessed one mtDNA haplotype, directly maternally inherited, without recombination (Birky *et al.*, 1978; Birky, 2001; Sato & Sato, 2012). However, reports of mitochondrial recombination events and heteroplasmy in animals, but especially in plants and fungi, led to re-evaluate this assertion (Barr *et al.*, 2005; White *et al.*, 2008). Molecular recombination is an important evolutionary mechanism that allows deleterious mutations removal and generates allelic diversity in a population. The main causes of heteroplasmy are bi-parental inheritance and/or mutations, which result in the coexistence of mitochondrial genomes that differ in either nucleotide composition (site heteroplasmy) or length (length heteroplasmy). The latter can involve large scale rearrangements or insertion/loss of coding regions (Boursot *et al.*, 1987; Volz-Lingenhöhl *et al.*, 1992). In AMF, transient mitochondrial length heteroplasmy through

anastomosis has recently been demonstrated for geographically distant *G. irregulare in vitro* isolates (de la Providencia *et al.*, 2013). The divergence in length between the two mtDNA haplotypes was caused by a variation in plasmid-related *dpo* insertions. The integration of elements such as *dpo* and SIRs has been shown to produce genomic rearrangements through recombination in other organisms (Schofield *et al.*, 1992; Bi & Liu, 1996; Paquin *et al.*, 2000; Ferandon *et al.*, 2008; Tanaka *et al.*, 2012).

Integrated plasmid-related *dpo* genes have been reported in all AMF mitochondrial genomes that have been sequenced thus far. Plasmids are self-replicating extra-chromosomal DNA molecules that were originally identified in bacteria. Analogous molecules have now been reported in eukaryotes, notably plants and numerous fungal species. The majority of plasmids described in filamentous fungi are strictly mitochondrial, but cases of nuclear localization have been reported in yeasts (reviewed in Griffiths, (1995)) and a *Mucor*-like fungus (Hänfler *et al.*, 1992). Plasmid structure varies widely, from circular molecules to linear invertrons with terminal inverted repeats (TIRs). Both types harbor one or two ORFs encoding DNA and/or RNA polymerase genes (Nargang *et al.*, 1984; Akins *et al.*, 1988; Pande *et al.*, 1989; Sakaguchi, 1990; Shiu-Shing Chan *et al.*, 1991; Court & Bertrand, 1992; Li & Nargang, 1993). Plasmids have also been shown to integrate into fungal mtDNA through recombination, and can either be cryptic or elicit changes in phenotypic expression (Akins *et al.*, 1986; Robison *et al.*, 1991; Griffiths, 1992; Hänfler *et al.*, 1992; Oeser *et al.*, 1993; Hermanns *et al.*, 1994). Such insertions have been reported to be acquired vertically from a common ancestor in plants (Robison & Wolyn, 2005), and also in fungi (Robison *et al.*, 1991). However, numerous cases of horizontal gene transfer of those elements have been reported in fungi,

offering an explanation for their ubiquitous distribution in this group (Collins & Saville, 1990; Griffiths *et al.*, 1990; Kempken *et al.*, 1992; Arganoza *et al.*, 1994; Debets *et al.*, 1994). Furthermore, plant mitochondrial plasmids are suspected to have been acquired from a fungal donor based on their similar structure and low GC content, which is a common feature of fungal mitochondrial genomes (reviewed in Rosewitch & Kistler, (2000)).

Another type of mobile element found in the mtDNA of Glomeromycota, SIRs (Formey *et al.*, 2012), were observed in the mt genome of the green algae *Chlamydomonas reinhardtii* (Boer & Gray, 1991), but their structure is reminiscent of GC clusters in yeast (de Zamaroczy & Bernardi, 1986) and the *Pst* I palindromes of *Neurospora crassa* (Yin *et al.*, 1981). SIRs can be folded into hairpin secondary structures, and have been suspected to have created genome rearrangements in *Chlamydomonas* spp. by intra-molecular recombination (Denovan-Wright & Lee, 1994; Nedelcu & Lee, 1998) through a model that is analogous to what has been observed in fungi (Almasan & Mishra, 1991; Weiller *et al.*, 1991), plants (André *et al.*, 1992) and bacteria (Schofield *et al.*, 1992; Bi & Liu, 1996). They are putatively mobile, although their transposition mechanism is not well understood (Grindley & Reed, 1985; Nakazono *et al.*, 1994). Nevertheless, mobility has been demonstrated in similar repeats that fold into double hairpin structures in *Allomyces* spp. mtDNA (Paquin *et al.*, 2000).

No information about the underlying mechanisms and consequences of such integrations into the mtDNA of Glomeromycota is currently available. Understanding whether or not mitochondrial recombination occurs in AMF is important since it can modify the interpretation of several aspects of research, including phylogenetic and population genetic analyses (Ballard

& Whitlock, 2004). In order to investigate this issue, we compared two novel mt genomes from the Glomeraceae family with a *G. irregulare* isolate. Since all AMF mt genomes from the Glomeraceae family that have been sequenced so far share the same mitochondrial synteny, we used this as a comparative basis for our study. We assessed mt genome evolution with regards to two other species of this ubiquitous AMF family, which showed divergence in mitochondrial gene organization. We wanted to know whether the new mtDNA reported here, from the species *Glomus* sp. DAOM-240422 and *G. cerebriforme* DAOM-227022, could shed light on the impact of mobile element insertions with regards to genome organization, and thus assessed the diversity, distribution, and propagation of mitochondrial *dpo* sequences within the Glomeromycota phylum. We also explored their evolutionary relatedness to other taxa, with respect to their global positioning within all known plasmid-related *dpo* sequences in the public databases.

3.4 Materials and Methods

3.4.1 Fungal material and DNA extraction

Spores and mycelium of *G. irregulare* (DAOM-234179), *Glomus* sp. (DAOM-240422*)* and *G. cerebriforme* (DAOM-229022) were cultivated *in vitro* on a minimal (M) medium solidified with 0.4% (w/v) gellan gum (Gelzantm^{TM}, Gelrite) in association with Ri-T transformed carrot roots following the protocol described by (Bécard & Fortin, 1988). Plates were incubated in the dark in an inverted position at 25°C, and after several weeks, abundant mycelia and spores were produced. Spores and mycelia were extracted from M medium by solubilisation of the gellan gum in 10 mM sodium acetate-citrate buffer (pH 6) (Doner &

Bécard, 1991) and washed in sterile water. The resulting fungal material was further purified by hand under a binocular microscope to remove root fragments. It was further suspended in 400 µL of the DNeasy Plant Mini Kit AP1 buffer (Qiagen) and crushed with a pestle in 1.5 ml micro tubes, and the DNA was purified according to the manufacturer's recommendations. Purified DNA in a final elution volume of 40 µL was stored at -20 ºC until use.

3.4.2 Long polymerase chain reactions (PCRs)

The long PCR protocol was performed using Takara LA Taq™ (Takara Bio, Canada) following the manufacturer's recommendations in a volume of 50 µl containing 0.2 mM dNTP, 1.5 mM MgCl2 and 0.5 µM of each primer and 1 µl of template DNA. The primer pairs summarized in Table 3.1 were used in order to validate the synteny in *Glomus sp*. DAOM-240422. Cycling parameters were 94 °C /3 min, followed by 38 cycles of: 94 °C /30 sec, 54 °C /25 sec, 68 °C /12 min and a final elongation at 72 °C for 8 min. PCR products were separated by electrophoresis in a 1 % (w/v) agarose gel and visualized with GelRed under UV light.

Table 3.1. Long PCR primers used to validate the *Glomus* sp. DAOM-240422 mitochondrial synteny.

mtDNA region	Primers	Sequences (5'-3')	Size (bp)
atp9-cox1	F	CTTGGCTCTATTCGCCTTAATGA	5029
	R	ACCAGGAAGAAGATCATAACGA	
nad4L-rnl	F	TGATAGGATTGATGGGTTTCATAG	13094
	R	GAACATACTTAGCTTTGATGATGGT	
cox3-rns	F	TACGGGTGTACAGCTCTATGAGT	13610
	R	TGGACTACGAGGGTATCTAATCCT	
cox1 control	F	TGCTAAAGATATTGGGACTCTCT	8785
	R	GTAATCTGGTATTCTTCGAGGCA	

3.4.3 Sequencing, assembly and gene annotation

G. irregulare DAOM-234179*, Glomus* sp. DAOM-240422 and *G. cerebriforme* total DNA was sequenced using 454 Titanium FLX shotgun technology and the resulting reads were assembled in one mitochondrial contig using Newbler (Genome Quebec Innovation Center, McGill University, Montreal, Canada). Gene annotation was performed with the automated organellar annotation software MFannot (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl) (for more information on MFannot see Burger *et al*., (2013), followed by manual inspection for the presence of frame shifts and the introduction of missing gene features. Short inverted repeats were also manually annotated using NCBI BLAST, and their secondary structure was predicted using the RNAfold web server (Gruber *et al.*, 2008). The annotated sequences of the three new AMF mitochondrial genomes were deposited in GenBank under the accession numbers KC164354, KC164355 and KC164356 for *G. irregulare* DAOM-234179*, Glomus* sp. DAOM-240422 and *G. cerebriforme,* respectively.

Circular and linear maps of the mitochondrial genomes (Figures 3.1, 3.2) were created using OGDraw v. 1.1 (Lohse *et al.*, 2007).

3.4.4 Network analyses

Two similarity network analyses, where each node represents a *dpo* sequence and an edge represents a sequence similarity/identity were performed using EGN (Halary *et al.*, 2013). The first network was constructed from a dataset of 91 Glomeromycotan *dpo* sequences (from our collection and retrieved from GenBank). All of the sequences were compared to each other using BLASTn (Altschul *et al.*, 1990) according to the following parameters: reciprocal best hit with a minimum of $1E^{-10}$ e-value, and 30% minimum identity covering at least 30% of the smallest sequence. The second and larger network includes an extensive sampling of all DNA polymerases from a broad range of taxa. This dataset was built using translated Glomeromycota *dpo* sequences as queries for BLASTp searches on the Genbank nr database. After retrieving hits that showed a minimal e-value of $1E^{-05}$, we performed a second BLASTp search to be confident that we had covered the largest possible diversity of DNA polymerase gene sequences. The protein similarity network was calculated using BLASTp with a minimal e-value threshold of $1E^{-40}$, 20% minimal similarity covering at least 20 % of the smallest sequence. The network layouts were further produced by Cytoscape software, using an edge-weighted force-directed model, meaning that genes sharing more DNA identity / protein similarity appear closer in the display.

3.4.5 Phylogenetic analyses

Maximum likelihood phylogenetic analyses of AMF mitochondrial genes were performed with a dataset of seven mitochondrial gene nucleotide sequences (*atp6, atp9, cox2, nad1, nad2*, *nad4* and *nad6*) that were concatenated using DAMBE software version 5.2.13 (Xia & Xie, 2001). The alignment was done using MUSCLE version 3.8.31. The analyses were performed using the GTR+G model (with five distinct gamma categories), with 1000 bootstrap replicates. Bayesian analysis was performed using MrBayes version 3.2 with the GTR+G model (with five distinct gamma categories), four independent chains, one million cycles, tree sampling every 10 generations and a burn-in value of 40% (Figure S3.1), which was adequate for Markov chain convergence.

The global *dpo* protein phylogeny (Figure 3.4B) was performed with a dataset extracted from the connected component of the protein similarity network containing fungal and plant sequences. The MCODE Cytoscape plugin (Bader & Hogue, 2003) was used to extract the most connected cluster, i.e. the cluster of nodes which connect the most sequences to each other. This process allows a sampling in which sequences will share a maximum similarity, improving the downstream phylogenetic inference. The selected protein sequences were locally aligned with COBALT version 2.01 (Papadopoulos & Agarwala, 2007). The maximum likelihood phylogenetic analysis was performed with RaxML using the rtREV model with the CAT option (PROTRTREVCAT) and 1000 bootstrap replicates. This phylogenetic model is an amino acid substitution matrix that has been shown to be appropriate for the analyses of polymerase genes (Dimmic *et al.*, 2002). Further, bayesian inference was performed with MrBayes using the rtREV+CAT model, four independent chains, one million cycles, tree

sampling every 10 generations, and a burn-in value of 40 %, which was adequate for Markov chain convergence.

The Glomeromycota *dpo* phylogeny (Figure S3.6) was performed on a conserved domain of 75 amino acids that was present in 60 out of the 91 AMF *dpo* sequences. The sequence alignment was done with COBALT version 2.01. The maximum likelihood analysis was conducted with the rtREV+G model (with five distinct gamma categories), with 1000 bootstrap replicates. Bayesian inference was performed with MrBayes using the rtREV+G model, four independent chains, one million cycles, tree sampling every 10 generations, and a burn-in value of 40 %, which was adequate for Markov chain convergence.

All three phylogenetic analyses were done using the integrative software TOPALi version 2.5 (Milne *et al.*, 2004). The best phylogenetic model was determined using TOPALi version 2.5 model selection test based on the AIC criterion. The tree figures were completed using TreeGraph version 2.0.47 (Stover & Muller, 2010).

3.4.6 Recombination analyses

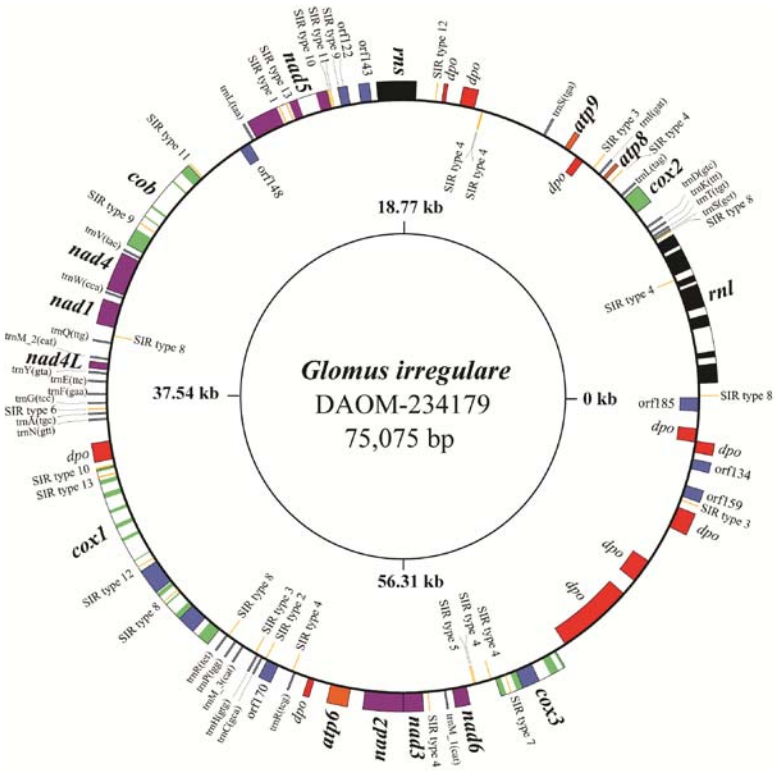
The recombination analyses were performed on the Glomeromycota nucleotide identity network group 1, 2 and 8 (where the *Glomus* sp. 240422 *dpo* inserted in the reshuffled intergenic regions clustered). In order to easily detect the presence of distinct recombination signatures, a distance-based recombination analysis (Figure S3.7) was achieved with the recombination analyses tool (RAT) (Etherington *et al.*, 2005), with the low threshold set to 70% nucleotide identity and the high threshold set to over 90% identity. To further confirm the

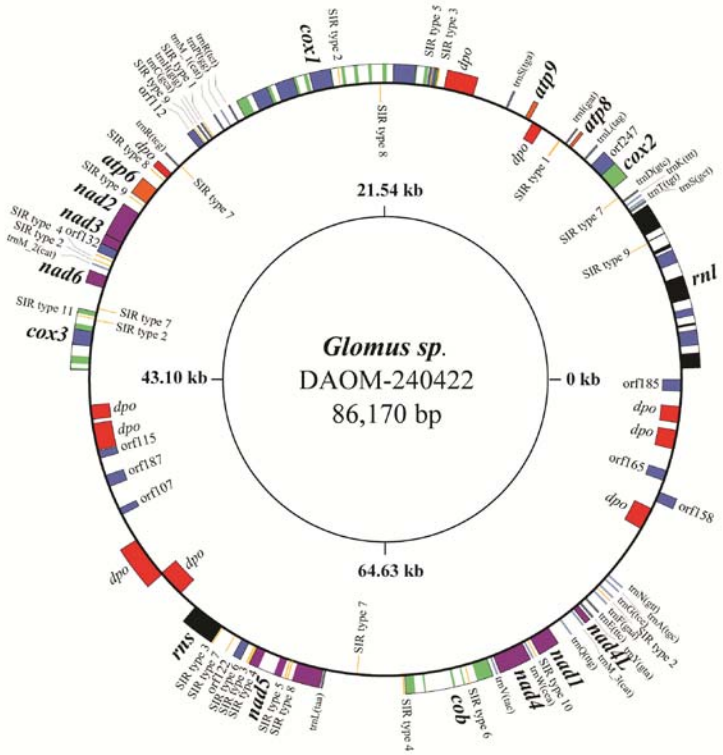
occurrence of the putative recombination events, we tested their statistical significance with a phylogenetic method using the Hidden Markov Model (HMM) with the F84+gaps nucleotide substitution model (Husmeier & Wright, 2001), performed with TOPALi v2.5 (default parameters).

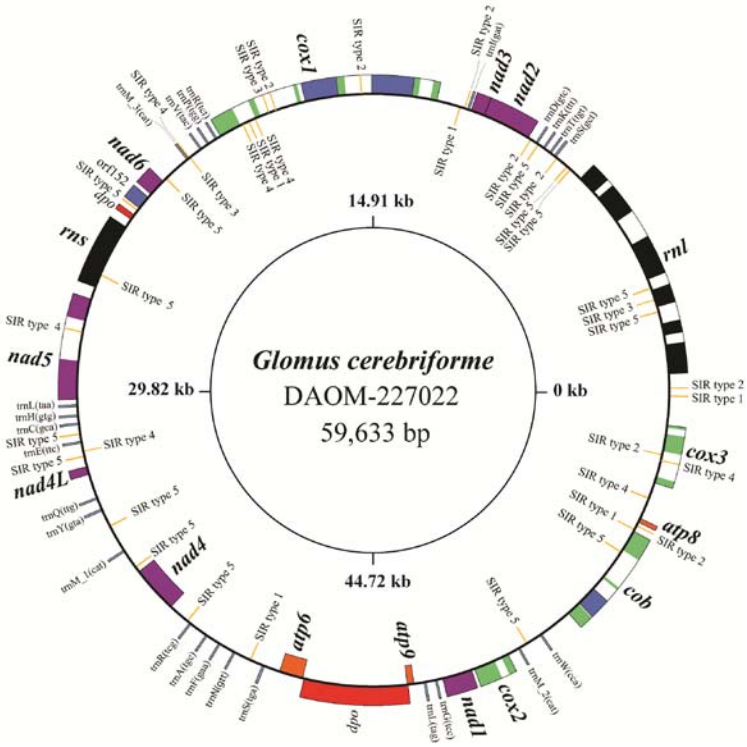
3.5 Results

3.5.1 Mitochondrial genome description and comparison

The complete mtDNA sequences of *G. irregulare* (DAOM-234179), *Glomus* sp. (DAOM-240422) and *G. cerebriforme* (DAOM-229022) are double-stranded circular DNA molecules, are homogeneous within isolate with no DNA sequence polymorphisms, i.e. genetic segregation of mtDNA in these strains appears to be as effective as in other published AMF genome. These mtDNAs harbor the typical set of 41 mitochondrial genes reported so far in AMF (Lee & Young, 2009): two rRNAs, 14 protein coding genes (PCGs) and 25 tRNAs. The PCGs include three ATP synthetase (*atp*), one cytochrome b (*cob*), three cytochrome C oxydase (*cox*) and seven NADH dehydrogenase (*nad*) genes (Figure 3.1). The *G. irregulare* isolate DAOM-234179 mtDNA has a genome size of 75,075 bp with a GC content of 36.7%, which is within range of the sizes of other sequences reported for this species (68,995 to 87,754 bp with a≈ 37% average GC content) (Lee & Young, 2009; Formey *et al.*, 2012; Nadimi *et al.*, 2012). The only true outlying *G. irregulare* isolate with respect to its mitochondrial genome was reported by Formey *et al*., (2012) , (i.e. MUCL43204) and has a size of 87,754 bp, while its conserved mitochondrial genes show only 65.8% identity (including introns and ORFs) with *G. irregulare* isolate 494. The *Glomus* sp.







Legend:

- introns
- ATP synthases
- NADH dehydrogenases
- cytochromes oxydases
- transfer RNAs
- ribosomal RNAs
- small inverted repeats (SIRs)
- DNA polymerase (*dpo*)
- open Reading Frames (ORFs)

Figure 3.1. Comparison of G. irregulare DAOM-234179, Glomus sp. DAOM-240422 and G. cerebriforme mitochondrial genomes. The circular-mapping genomes were opened upstream of *rnl* to allow for easier comparisons. Genes on the outer and inner circumference are transcribed in a clockwise and counterclockwise direction, respectively. Gene and corresponding product names are *atp6, 8, 9,* ATP synthase subunit 6; *cob*, apocytochrome b; *cox1–3*, cytochrome c oxidase subunits; *nad1–4, 4L, 5–6*, NADH dehydrogenase subunits; *rnl, rns*, large and small subunit rRNAs; A-W, tRNAs, the letter corresponding to the amino acid specified by the particular tRNA followed by their anticodon. Open reading frames smaller than 100 amino acids are not shown.

DAOM-240422 and *G. cerebriforme* mtDNAs were 86,170 and 59,633 bp, respectively, with GC contents of 36.7 and 46.7%. The latter is the smallest AMF mitochondrial genome yet reported, and is highly divergent when compared to the available mtDNAs of *Glomus* spp. and *Gigaspora* spp. The inflated GC% observed in *G. cerebriforme* is mostly due to reduced intergenic region size and the lack of mobile elements, which are mostly AT rich regions, compared to the other strains. The difference in the mtDNA sequences observed between these taxa is caused by variation in intron content, *dpo* insertions and the presence of open reading frames (Table 3.2). *G. cerebriforme* genes are encoded on both strands, their order is completely reshuffled compared to the other *Glomus* spp., and they are also phylogenetically distant from those of *G. irregulare* (Figure S3.1). On the other hand, *Glomus* sp. DAOM-240422 showed high similarity to *G. irregulare* with regards to coding sequence identity, but surprisingly, the synteny of these two isolates differed substantially.

3.5.2 Plasmid-related DNA polymerase and sequence diversity of small inverted repeats

We found 10 *dpo* insertions in *G. irregulare* DAOM-234179 and *Glomus* sp. DAOM-240422, making them the most *dpo*-rich AMF isolate reported so far, but only two copies were observed in *G. cerebriforme*. The three isolates harbor a huge *bona fide dpo* with ORF extensions, rendering a structure similar to the mitochondrial plasmid found in other fungi

(Kim *et al.*, 2000). However, we did not find terminal inverted repeats, which are usually a characteristic of recent plasmid integration. We also found numerous SIRs in the three genomes. The 31 SIRs in *G. irregulare* DAOM-234179, 29 in *Glomus* sp. DAOM-240422 and 40 in *G. cerebriforme* were divided into 13, 11 and 5 distinct types, respectively, based on their hairpin secondary structure (Figures S3.3-S3.5). The high number of SIR types, compared to the five previously described by Formey *et al*., result from the characterization of SIRs found in single copy, homologous to those previously described that could not have been detected otherwise. Some types are present in all *G. irregulare* isolates and closely-related species such as *Glomus* sp. DAOM-240422, while others are endogenous to each strain, giving rise to a high SIRs sequence diversity. These elements were always present in intergenic regions, introns or at the edge of endonuclease-encoding ORFs (Tables S3.1-S3.3), and suggest a close relationship with the '*homing*' mechanism for endonuclease integrations, as proposed by Formey *et al*. (2012) . However, the diversity of possible cleaving sites seems to be greater than previously expected with the here-reported novel mtDNA sequences.

3.5.3 Divergent synteny in two novel Glomeraceae species mtDNA

The *G. cerebriforme* mtDNA shows a completely different gene order compared to the other AMF mitochondrial genomes sequenced so far, with many type five SIR insertions in intergenic regions that could potentially be involved in synteny rearrangements. Nevertheless, as mentioned above, only two *dpo* insertions are present, in the *cox3-rnl* and *nad6-rns* intergenic regions. Since *G. cerebriforme* is phylogenetically distant from the publicly available mtDNAs of *Glomus* spp. and *Gigaspora* spp., no reference scaffold is available to

Table 3.2. Arbuscular mycorrhizal fungi mitochondrial genome features

AMF Strain	Genome size (Kb)	Introns [a]	*ORF*s [b]	*dpo*	SIRs [c]	Molecular features
G. irregulare [d] *FACE494*	70,606	26	10	5	26	-
G. irregulare [d] DAOM197198	70,783	26	10	6	27	-
G. irregulare [d] MUCL46239	70,818	26	10	6	28	-
G. irregulare [d] MUCL46240	74,797	26	15	5	32	-
G. irregulare [d] MUCL46241	74,797	26	15	5	32	-
G. irregulare DAOM240415	70,781	26	9	8	?	-
G. irregulare DAOM234328	68,995	26	9	8	?	-
G. irregulare DAOM234179	75,075	26	10	10	31	-
Glomus irregulare [d] MUCL43204	87,754	32	23	4	50	-
Glomus sp. DAOM229456	87,763	32	19	4	?	Gene hybrids/HGT
Glomus sp. DAOM240422	86,170	25	19	10	29	Genome rearrangements
Glomus cerebriforme	59,633	16	4	2	40	Synteny shuffling
Gigaspora margarita	96,998	14	4	0	?	Transplicing/synteny shuffling
Gigaspora rosea	97,349	14	4	0	?	Transplicing/synteny shuffling

[a] All types including group I and II introns
[b] ORFs greater than 100 amino acids encoding unknown protein, LAGLIDADG and GIYYIG endonuclease are listed, including intronic ORFs but not *dpo* fragments.
[c] Short inverted repeats reported so far in AMF mitochondrial genomes
[d] Based on data reported in Formey et al. 2012, but eroded ORFs were excluded.

confirm any gene order reorganization. In contrast, the comparison of *Glomus sp*. DAOM-240422 mt genome to its close relative *G. irregulare* DAOM-234179, revealed a reshuffling of two large gene clusters (Figure 3.2A). One gene cluster encompasses *cox1, atp6, nad2, nad3, nad6 and cox3*, whereas the other contains *rns, nad5, cob, nad4, nad1 and nad4L*. This rearrangement creates three novel intergenic regions in *Glomus* sp. DAOM-240422 which are *atp9-cox1*, *cox3-rns* and *nad4L-rnl*. This result was confirmed by long PCR performed with primers spanning those regions (Figure S3.2). We then assessed the sequence identity between those intergenes and their corresponding regions in *G. irregulare* DAOM-234179 (Figure 3.2B). We compared the *Glomus* sp. DAOM-240422 *atp9-cox1* with *G. irregulare* DAOM-234179 *atp9-rns* and *nad4L-cox1* intergene; the *cox3-rns* with *G. irregulare* DAOM-234179 *cox3-rnl* and *atp9-rns regions*; and finally the *nad4L-rnl* was compared to *nad4L-cox1* and *cox3-rnl*. Each *Glomus* sp. DAOM-240422 intergenic region showed high sequence identity to its putative former counterpart in *G. irregulare* DAOM-234179, but covered only a small fraction of the intergene length, since many more ORFs (mostly *dpo* insertions) are present in *Glomus* sp. DAOM-240422. Local alignment using BLASTn, between the newly formed intergenic regions, did not reveal high similarity matches. We also constructed a sequence identity matrix of all the *dpo* sequences in its mtDNA and the highest match was only 53.1%.

3.5.4 Glomeromycota *dpo* nucleotide identity network and phylogenetic analyses

To test whether the *dpo* sequences in *Glomus* sp. DAOM-240422 were orthologous with the sequences found in other AMF and to assess evolution of *dpo* within the Glomeromycota, we constructed a nucleotide similarity network of all *dpo* genes reported so far in AMF mitochondrial genomes (Figure 3.3A). In total, 12 similarity groups were formed

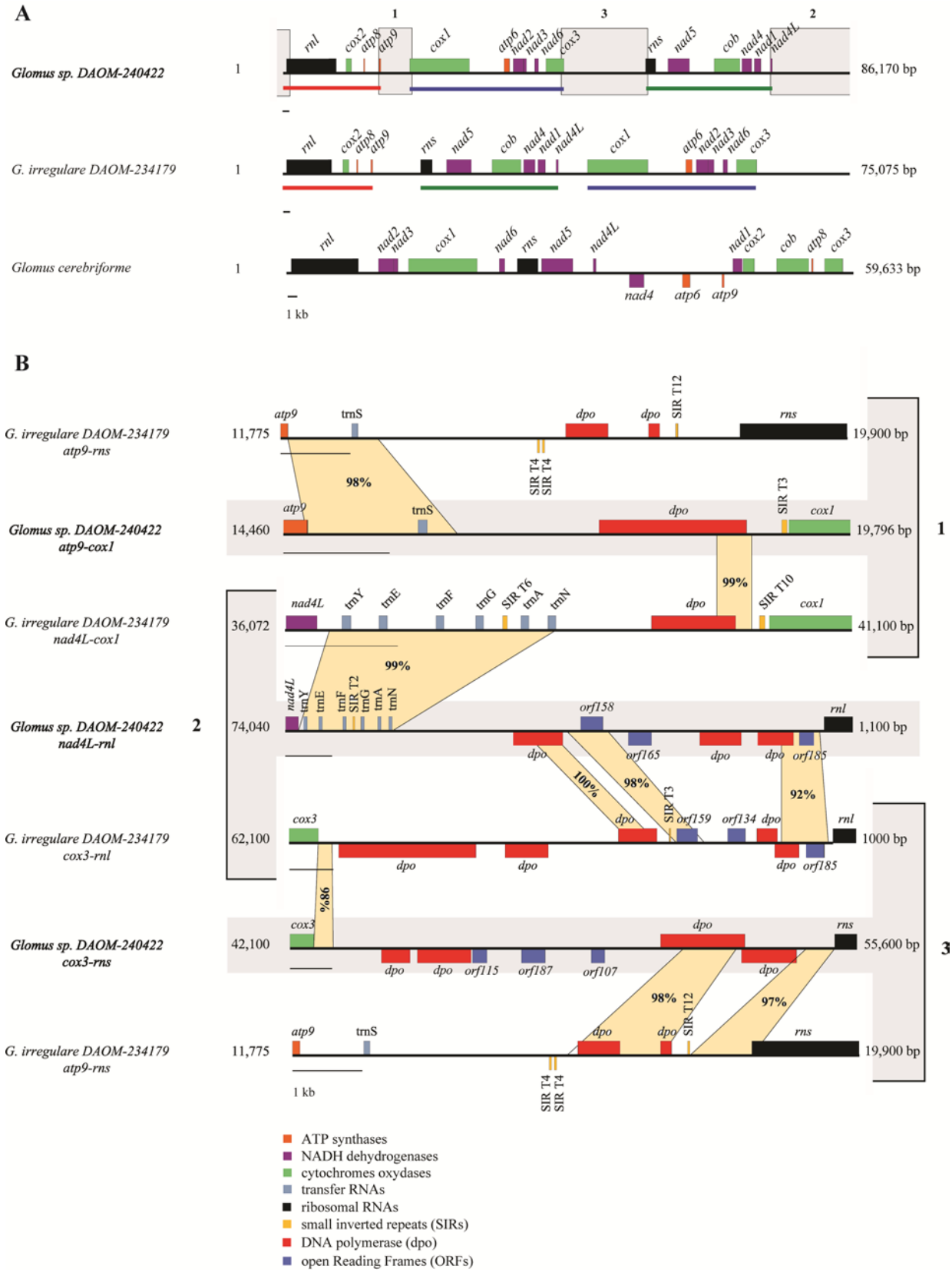
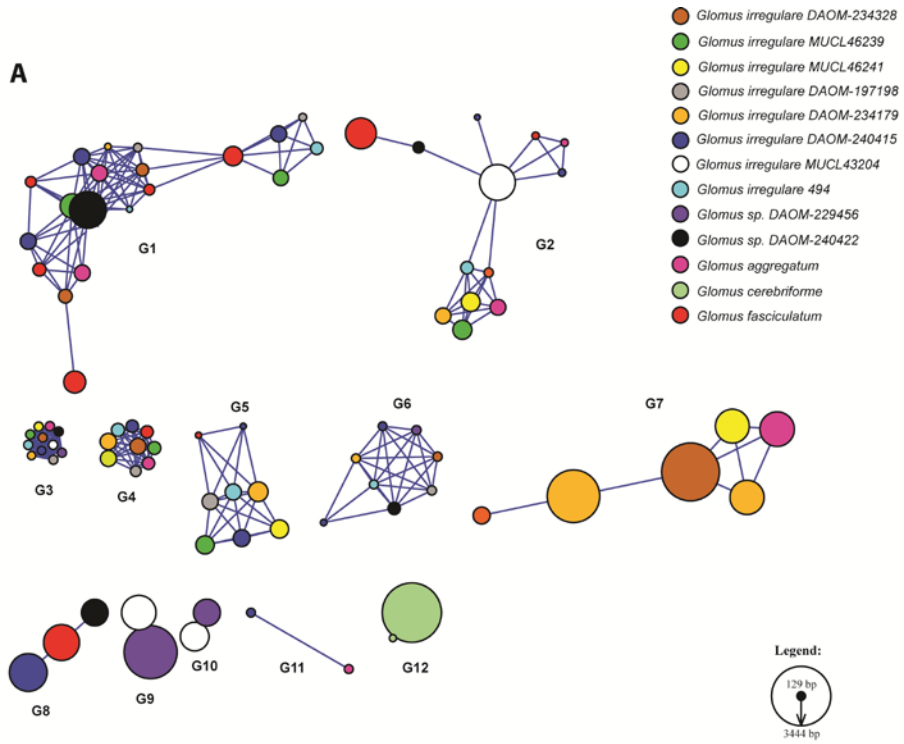

A
Glomus sp. DAOM-240422
G. irregulare DAOM-234179
Glomus cerebriforme
86,170 bp
75,075 bp
59,633 bp
1 kb
B
G. irregulare DAOM-234179 atp9-rns
Glomus sp. DAOM-240422 atp9-cox1
G. irregulare DAOM-234179 nad4L-cox1
Glomus sp. DAOM-240422 nad4L-rnl
G. irregulare DAOM-234179 cox3-rnl
Glomus sp. DAOM-240422 cox3-rns
G. irregulare DAOM-234179 atp9-rns
11,775
19,900 bp
14,460
19,796 bp
36,072
41,100 bp
74,040
1,100 bp
62,100
1000 bp
42,100
55,600 bp
11,775
19,900 bp
98%
99%
100%
98%
92%
%86
98%
97%
1 kb
ATP synthases
NADH dehydrogenases
cytochromes oxydases
transfer RNAs
ribosomal RNAs
small inverted repeats (SIRs)
DNA polymerase (dpo)
open Reading Frames (ORFs)

Figure 3.2. Linear genome representation to compare the mitochondrial synteny between *G. irregulare* DAOM-234179, *Glomus* sp. DAOM-240422 and *G. cerebriforme*. The linear-mapping genomes were opened upstream of *rnl* to allow for easier comparisons. **(A)** Corresponding gene clusters between *G. irregulare* DAOM-234179 and *Glomus* sp. DAOM-240422 are underlined in red, green and blue. The newly formed intergenic regions of *Glomus* sp. DAOM-240422 are boxed in gray and tagged with a number. (**B)** Nucleotide identity comparison with BLASTn between the reshuffled *Glomus* sp. DAOM-240422 intergenic regions (*atp9-cox1* (box 1)*, nad4L-rnl* (box 2) and *cox3-rns* (box 3)) with their putative homologous sequence in *G. irregulare* DAOM-234179. Homologous regions are indicated by the projections with their corresponding percent identity.

and seven singletons were left aside. We observed that five *Glomus* sp. DAOM-240422 *dpo* sequences (black circles) clustered in distinct identity groups (G1, G2, G3, G6 and G8), while the rest were singletons inserted in reshuffled intergenic regions (*cox3-rnl* and *nad4L-rnl*), with low sequence similarity to each other or to known AMF *dpo* sequences. The *Gi. rosea* putative mitochondrial plasmid that was previously described (Nadimi *et al.*, 2012) was also one of the singletons. The two *G. cerebriforme dpo* copies (light green circles) were the sole representatives of similarity group 12. Also, the mtDNA-integrated *dpo* showed a clustering tendency toward their intergenic region localization (Figure 3.3B). Interestingly, three *Glomus* sp. DAOM-240422 *dpo* sequences which are inserted in the reshuffled intergenic regions *atp9-cox1, cox3-rns* and *nad4L-rnl* (Figure 3.2) clustered respectively with the AMF *dpo* group 8 (*cox3-rnl*), group 1 (*atp9-rns*) and group 2 (*cox3-rnl* and the *Glomus fasciculatum atp9-rns*_3 sequence). In order to compare the identity network and the phylogenetic approach, and to assess the relationship of the *dpo* present within each respective mtDNA (i.e. paralogy vs orthology), we constructed a phylogeny based on a conserved protein domain. The alignment of all 91 Glomeromycota *dpo* protein sequences pointed out to a 75 amino acid region that was common in 60 of the sequences (Figure S3.6A). Interestingly, most of the *dpo* were truncated either upstream and/or downstream of that conserved domain.


A
Glomus irregulare DAOM-234328
Glomus irregulare MUCL46239
Glomus irregulare MUCL46241
Glomus irregulare DAOM-197198
Glomus irregulare DAOM-234179
Glomus irregulare DAOM-240415
Glomus irregulare MUCL43204
Glomus irregulare 494
Glomus sp. DAOM-229456
Glomus sp. DAOM-240422
Glomus aggregatum
Glomus cerebriforme
Glomus fasciculatum
G1
G2
G3
G4
G5
G6
G7
G8
G9
G10
G11
G12
Legend:
129 bp
3444 bp

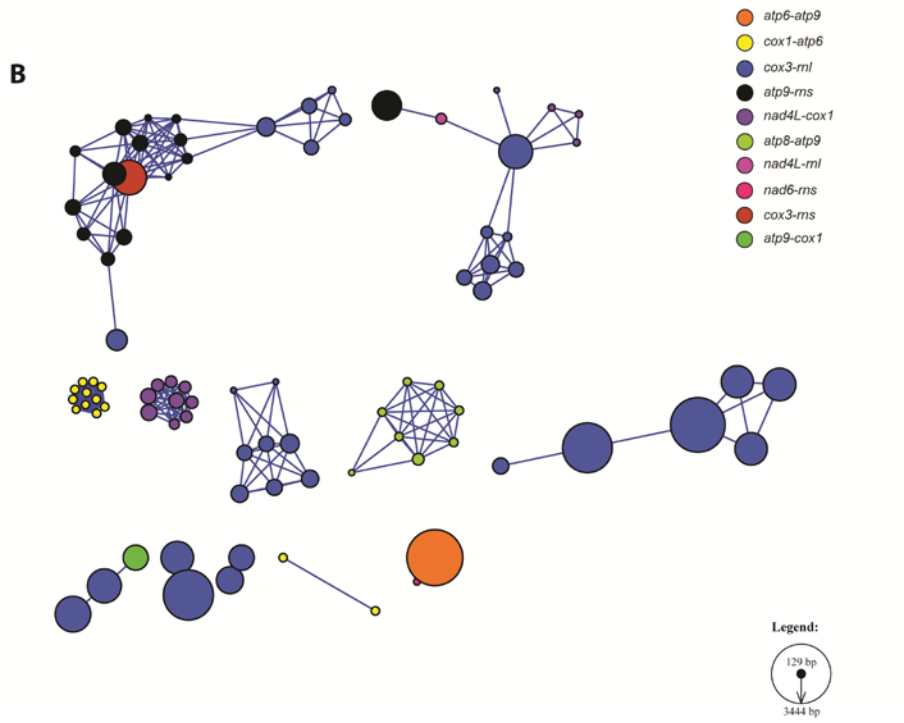

B
atp6-atp9
cox1-atp6
cox3-rnl
atp9-rns
nad4L-cox1
atp8-atp9
nad4L-rnl
nad6-rns
cox3-rns
atp9-cox1
Legend:
129 bp
3444 bp

Figure 3.3. Similarity network of dpo sequences inserted in AMF mitochondrial genomes. (A) Each node represents a *dpo* insert colored by mt genome. The node size is proportional to the sequence length. Two nodes are connected by an edge when they share significant similarity (reciprocal best BLAST hit with a minimum of $1e^{-10}$ score, and 30% minimum identity covering at least 30% of the smallest sequence). The layout was produced by Cytoscape, using an edge-weighted force-directed model, which brings closer sequences sharing more similarity. Twelve homology groups were formed (labeled G1 to G12). There are 97 nodes on that network with 336 edges. **(B)** Same network but nodes are colored by their mitochondrial intergenic region localization. The legend shows a linear size relationship between the smallest *dpo* sequence (*G. irregulare* DAOM-240415_*cox3-rnl*_4, 129 bp) and the largest one (*G. cerebriforme atp6-atp9*, 3444 bp).

The resulting maximum likelihood phylogenetic tree showed robust bootstrap values (> 75%) in the distal branches that support clades corresponding to the similarity groups observed in the gene network, whereas the basal branches displayed were weakly supported (Figure S3.6B). All similarity network groups were represented in the tree with strong statistical support in bootstrap and bayesian inference values, except for the identity group 11 that was not represented, because the sequences didn't harbor the conserved gene core. Surprisingly, this tree revealed only one occurrence of paralogy, in the group 12, for the two *dpo* sequences inserted in *G. cerebriforme* mtDNA (like it was also observed in the nucleotide network). Every other *dpo* phylogenetic clusters represent orthology groups. Lastly, only one of the three *Glomus* sp. DAOM-240422 *dpo* sequences found in reshuffled intergenic regions was present in the phylogenetic analyses, in the group 8 cluster.

3.5.5 *Glomus* sp. DAOM-240422 *dpo* recombination analyses

Three *Glomus* sp. *dpo* sequences present in the reshuffled intergenic regions clustered with *dpo* sequences of other AMF taxa inserted in different mtDNA localization, rather than with paralogous mtDNA sequences (within DAOM-240422 isolate) in the nucleotide network. In order to easily screen for homologous recombination, we performed a distance-based

recombination analysis in which these *Glomus* sp. *dpo* sequences were compared with their three closest relatives in the nucleotide identity network. For the *Glomus* sp. *atp9-cox1* and *cox3-rns* (Figure S3.7A and B) comparison, we observed a disruption in the sequence identity of the *dpo* sequences going from 70-80% identity and spiking abruptly to a near perfect 100% nucleotide identity. When compared together, *Glomus* sp. *nad4L-rnl_1 dpo* sequences started off with < 80% identity to the *G. fasciculatum atp9-rns*_3 *dpo* and then rose sharply to almost 95% identity. Conversely, the *G. irregulare* MUCL43204_*cox3-rnl*_1 *dpo* started at more than 99% similar to the *Glomus* sp. *nad4L-rnl_1 dpo* sequence, but then dropped to < 80% identity (Figure S3.7C). These *dpo* sequences showed typical signatures of homologous recombinants regarding their sequence identity, highly reminiscent to what was found for a recombinant HIV-1 strain (Liitsola *et al.*, 2000).

To further confirm these findings, we performed a statistical phylogenetic recombination analysis using the Hidden Markov Model (HMM). The *Glomus sp*. DAOM-240422 *atp9-cox1* showed a perfect bayesian inference value of 1.0 for each three topologies in a given portion of the alignment (Figure S3.7A), which allowed the finding of putative recombination break points (between two given topologies where one reach a bayesian inference value > 0.95 and the other < 0.05) at nucleotide position 357 and 613 bp of the sequence, respectively. The *Glomus* sp. DAOM-240422 *nad4L-rnl* also showed a perfect bayesian inference value of 1.0, but for only two topologies (I and II), thus allowing for the determination of one putative break point at position 161 bp (Figure S3.7C). Conversely, the *Glomus* sp. DAOM-240422 *cox3-rns* analyses did not provide statistical support for any given topologies tested with the HMM phylogenetic method (Figure S3.7B).

3.5.6 Global protein similarity network with all known DNA polymerase genes in the public database

In order to assess a possible origin and transmission mode of AMF mitochondrial *dpo* sequences, we constructed a protein similarity network using a broad spectrum of *dpo* sequences. Interestingly, most of the glomeromycotan plasmid-related DNA polymerase proteins, 60 in total, were found clustered with 35 other fungi, 28 plants (including two red algae of the genus *Porphyra*), five other eukaryotes (*Physarum polycephalum*, *Ochromonas danica*, two *Placozoan* sp. and a *Cnidaria* species), four bacteria (*Bacillus thuringiensis*, two *Bacillus cereus*, and *Brevibacillus* sp.), and three bacteriophages (Figure 3.4A). All *dpo* sequences are present within their respective mtDNA or are located in a mitochondrial plasmid. The patchy distribution observed among phylogenetically distant taxa such as plants and fungi is similar to that of a *cox1* intron that invaded plant mitochondria and was thought to have been acquired horizontally from a fungal donor (Cho *et al.*, 1998). To assess the evolutionary relationship of AMF DPO proteins with the other taxa present in the cluster, we selected the group gathering the most connected nodes, i.e the sequences sharing the most

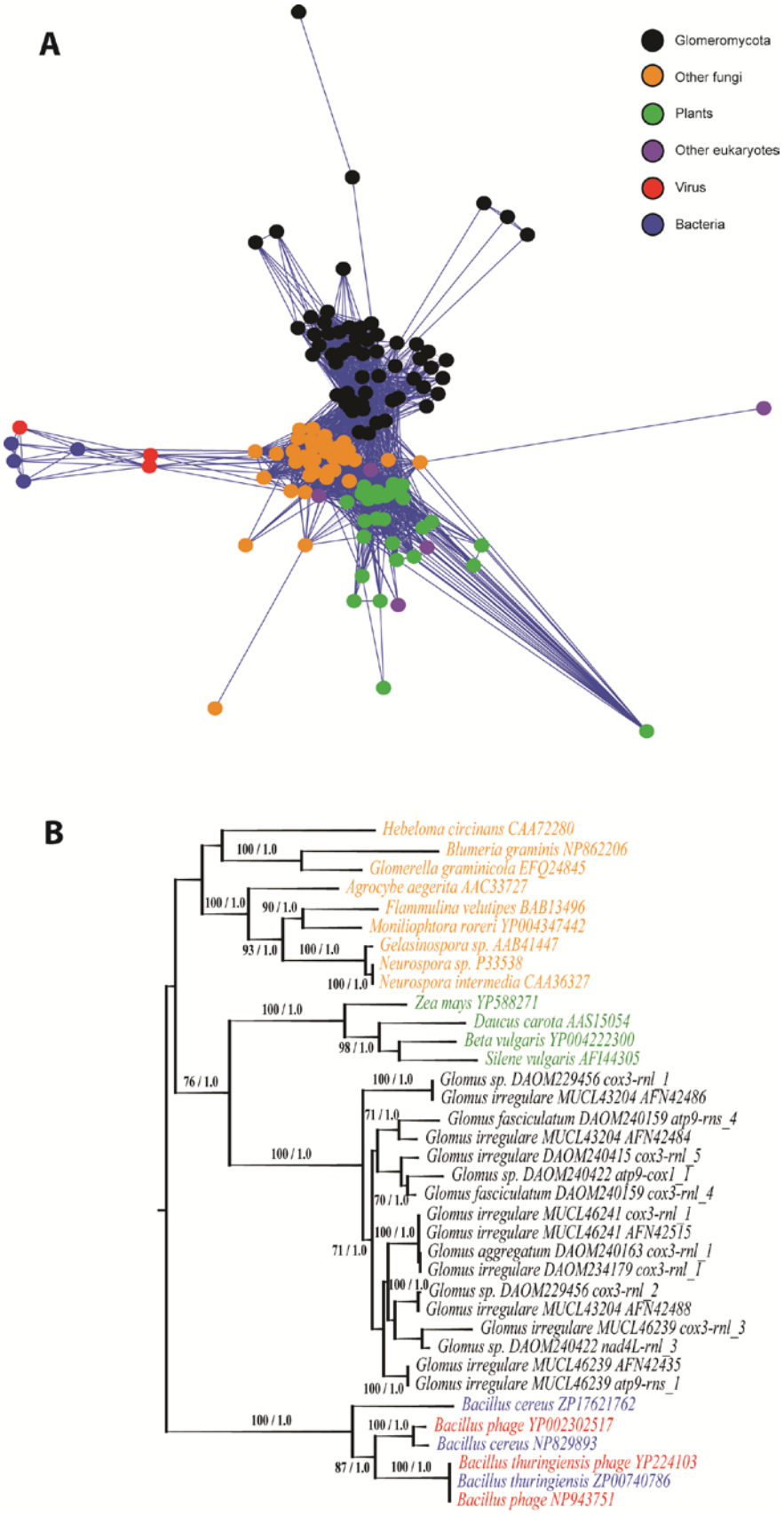

A
Glomeromycota
Other fungi
Plants
Other eukaryotes
Virus
Bacteria
B
Hebeloma circinans CAA72280
Blumeria graminis NP862206
Glomerella graminicola EFQ24845
Agrocybe aegerita AAC33727
Flammulina velutipes BAB13496
Moniliophtora roreri YP004347442
Gelasinospora sp. AAB41447
Neurospora sp. P33538
Neurospora intermedia CAA36327
Zea mays YP588271
Daucus carota AAS15054
Beta vulgaris YP004222300
Silene vulgaris AFI44305
Glomus sp. DAOM229456 cox3-rnl_1
Glomus irregulare MUCL43204 AFN42486
Glomus fasciculatum DAOM240159 atp9-rns_4
Glomus irregulare MUCL43204 AFN42484
Glomus irregulare DAOM240415 cox3-rnl_5
Glomus sp. DAOM240422 atp9-cox1_1
Glomus fasciculatum DAOM240159 cox3-rnl_4
Glomus irregulare MUCL46241 cox3-rnl_1
Glomus irregulare MUCL46241 AFN42515
Glomus aggregatum DAOM240163 cox3-rnl_1
Glomus irregulare DAOM234179 cox3-rnl_1
Glomus sp. DAOM229456 cox3-rnl_2
Glomus irregulare MUCL43204 AFN42488
Glomus irregulare MUCL46239 cox3-rnl_3
Glomus sp. DAOM240422 nad4L-rnl_3
Glomus irregulare MUCL46239 AFN42435
Glomus irregulare MUCL46239 atp9-rns_1
Bacillus cereus ZP17621762
Bacillus phage YP002302517
Bacillus cereus NP829893
Bacillus thuringiensis phage YP224103
Bacillus thuringiensis ZP00740786
Bacillus phage NP943751

Figure 3.4. Sub-cluster of the global network of shared amino acid similarities between the Glomeromycota DPOs and all homologous sequences found on GenBank. (A) Building this dataset was performed using translated Glomeromycota sequences as queries for a BLASTp search on the GenBank nr database (minimal e-value threshold of $1e^{-40}$, 20 % minimal similarity covering at least 20 % of the smallest sequence). The network layouts were further produced by Cytoscape software, using an edge-weighted force-directed model, meaning that genes sharing more protein similarity appear closer in the display. There are 135 nodes in that network with 2520 edges. (**B)** DPO protein maximum likelihood tree obtained with the rtREV+CAT phylogenetic model. The bacterial/viral cluster was used to root the tree. Number on branches indicates bootstrap support values (< 60% cut-off) and bayesian inference values, respectively. The bayesian analyses gave a similar topology (data not shown). The tree includes AMF (black), other fungi (orange), plants (green), virus (red) and bacteria (blue).

similarity all together. The corresponding sequences (including plant, Glomeromycotan and other fungal sequences), were used to perform a maximum likelihood phylogenetic analysis (Figure 3.4B). We used the viral and bacterial proteins present in the sub-network to root the tree, because they present a distal position in the network. The resulting tree showed robust support (100% bootstrap and 1.0 bayesian inference value) for the viral-bacterial group, and surprisingly, the plant DPOs clustered with those of the Glomeromycota as a sister group, supported by a 76% bootstrap and 1.0 bayesian inference value. Except for the other fungi group clade, the phylogeny was robustly supported either for the basal and the distal branches.

3.6 Discussion

3.6.1 Inter-haplotype homologous recombination in *Glomus* sp. DAOM-240422

The numerous type 5 SIRs present in *G. cerebriforme* intergenic regions may have had the potential to trigger intra-molecular gene reshuffling. The two *G. cerebriforme dpo* sequences were shown to be paralogs, with a nucleotide identity of 81%. However, we cannot

be sure that intra-molecular recombination occurred in *G. cerebriforme,* since we do not have closely related taxa as comparators. In the *Glomus* sp. DAOM-240422 mt genome, the two most similar *dpo* sequences share a low nucleotide identity percentage of 53.1%. However, we observe a near perfect sequence identity for PCGs and some intergenic regions with close relative *G. irregulare*. Taken together, these observations suggest that the rearrangements observed in *Glomus* sp. DAOM-240422 are recent and do not result from intra-molecular recombination from an ancient *dpo* duplication. Further, no recent *dpo* duplications were observed in any given AMF strains surveyed. The *Glomus* sp. DAOM-240422 *dpo* sequences that are inserted in the reshuffled intergenic regions clustered in the nucleotide network with *dpo* sequences of other AMF species that were inserted in the putatively former intergenic arrangement. Comparison of the *Glomus* sp. DAOM-240422 *atp9-cox1* and *nad4L-rnl_1 dpo*, with their three closest relatives showed a typical signature of homologous recombination (Figure S3.7A and C), with genes sharing a high sequence identity with different portions of the alignment, where intersections represent putative recombination break points. These recombination events were strongly statistically supported with the HMM method. Even if the *Glomus* sp. 240422 *cox3-rns dpo* recombination was not supported by HMM, it still clustered with *dpo* present in other intergenic regions, and the lack of statistical support might be explained by the absence of its recombination partner in the cluster. These observations strongly suggest that the mitochondrial gene cluster reorganization in *Glomus sp*. DAOM-240422 is the result of *dpo*-mediated homologous recombination between different mitochondrial haplotypes harboring highly similar *dpo* orthologues present in distinct genome localization.

Two types of mobile elements reported in AMF mitochondrial genomes have been previously shown to be involved in genome rearrangements through homologous recombination in other organisms: the plasmid-related *dpo* genes in Basidiomycota (Ferandon *et al.*, 2008) and SIRs in bacteria and plants (Schofield *et al.*, 1992; Bi & Liu, 1996; Tanaka *et al.*, 2012). Furthermore, the fact that the number of *dpo* sequences varies in different intergenic regions among isolates of the same species and closely related ones (Formey *et al.*, 2012; Beaudet *et al.*, 2013a), combined with our findings suggest that: i) there is no evidence of intra-molecular recombination in *Glomus sp*. DAOM-240422, ii) recombination between distinct mtDNA haplotypes harboring different *dpo* insertions could occur, iii) since there is no intra-isolate mtDNA polymorphism, the reshuffled mtDNA haplotype in *Glomus* sp. DAOM-240422 is fixed in the population , iv) those rearrangements are triggered by *dpo* insertions through homologous recombination.

Such recombination events could take place following hyphal fusion (i.e. anastomosis). This last feature not only facilitates the distribution of nutrients and signalling molecules through the entire AMF mycelium, but also plays a key role in genetic exchange (Giovannetti *et al.*, 1999) and segregation (Angelard & Sanders, 2011). Those anastomoses can occur either between genetically different AMF isolates originating from the same experimental field (Croll *et al.*, 2009; Angelard *et al.*, 2010) or between strains isolated from distant locations (e.g from the same (Purin & Morton, 2012), or even different continents (de la Providencia *et al.*, 2013)). Both of these studies suggested that mitochondria might also be exchanged via anastomosis, thus creating a heteroplasmic state.

3.6.2 Evidence of heteroplasmy challenges the concept of an AMF individual

All published AMF mt genomes are homoplasmic, i.e. homogeneous within isolates. In other fungi, mechanisms contributing to homoplasmy are often related to the recombination of parental mtDNAs or by the selection of one of the parental haplotypes due to the presence of segregation mechanisms (reviewed by Birky (2001)). In the absence of external factors, heteroplasmy should be the default state for mtDNA under a simple mutation-drift scenario (White *et al.*, 2008). In yeasts, the presence of a heteroplasmic state was shown to be transient due to the presence of segregation mechanisms, such as nucleoid formation, and mtDNA molecular repair pathways which maintain mitochondrial genome integrity (Zimmer *et al.*, 1991; Hu *et al.*, 1995; Yasuhira & Yasui, 2000). The presence of nucleoids, similar mtDNAs linked by Holliday junctions, induces a genetic bottleneck responsible for a fast segregation (Lockshon *et al.*, 1995; White & Lilley, 1997; MacAlpine *et al.*, 1998). The mt genome homogeneity observed for *in vitro* and *in vivo* AMF cultures could be the result of such effective mtDNA segregation and repair mechanisms that would take place during the sub-cultivation processes.

The AMF colonies are characterized by the interconnectedness of their different parts by means of hyphal fusion (i.e. anastomosis) (Giovannetti et al., 1999; de la Providencia et al., 2005; Voets et al., 2006; Purin and Morton, 2012), which can lead to genetic exchange, even between genetically distinct individuals. This hyphal fusion dynamic in natural populations can also prevent the lost of genetic diversity (Bever & Wang, 2005). Recently, anastomoses between geographically distant *G. irregulare* isolates cultivated *in vitro* were shown to induce transient mitochondrial length heteroplasmy (with variation in *dpo* insertions between the two

mt haplotypes) (de la Providencia *et al.*, 2013). Although *in situ* heteroplasmy have never been demonstrated, these observations, in addition to the inter-haplotype *dpo* homologous recombination evidences presented here, suggest that transient heteroplasmy could occur in the field.

The coexistence of numerous AMF mt haplotypes in the same cytoplasm and the occurrence of homologous mitochondrial recombination might be commonplace in nature, as it was shown in natural populations of the basidiomycete fungus *Armillaria gallica* (Saville *et al.*, 1998). The heteroplasmy stability in natural populations, may provide useful additional information for defining haplotypes, and resolving further the relationships among individuals at a population level (White *et al.*, 2008). Given the increasing use of mitochondrial markers in AMF research, the demonstration that genetically distinct mtDNAs can be exchanged through anastomosis, coexist in a common cytoplasm and give rise to a recombinant haplotype, should lead to a thorough discussion on the AMF individuality concept. If heteroplasmy is confirmed in natural populations, the use of mtDNA as a criterion to define a reliable AMF taxonomic unit would certainly facilitate our understanding of the processes occurring in natural populations, although caution should still be applied since it would be artificial, and would not reflect the genotypic diversity of AMF nuclear DNA, since mitochondrial and nuclear DNA segregation are mostly stochastic mechanisms.

3.6.3 Vertical versus horizontal inheritance of the AMF mitochondrial *dpo*

Given the high genetic diversity of *dpo* and their variable truncation patterns, a classical phylogenetic approach requiring multiple alignment of highly similar sequences, is

not suitable to analyse their overall evolutionary relationships, especially for recombinant ones (Bapteste *et al.*, 2012). Indeed, the classical approach only allowed us to analyze 60 of 91 Glomeromycota *dpo* sequences, on a limited 75 amino acids conserved domain, leading to a tree with poorly supported basal branches (Figure S3.6B). As an alternative, gene similarity network analyses have already been proven to be useful in the study of highly diverse gene families and introgressive descent events, such as recombination and HGT (Bapteste *et al.*, 2012).

The AMF *dpo* nucleotide network highlights the existence of a clustering tendency for *dpo* sequences located in the same intergenic regions (Figure 3.3B). The high similarity (short edges) observed between the sequences of each group and the phylogenetic topology (Figure S3.6B) suggest that they are orthologs that could have recently been acquired vertically from a common ancestor. This is not surprising for the *G. irregulare* isolates, but implies that the other species that shared those sequences have also diverged recently (i.e. *G. fasciculatum* and *G. aggregatum*) given the high within-group sequence similarity. However, the seven singletons left aside by the network nucleotide similarity clustering (five in the closely related species *Glomus* sp. DAOM-240422) suggest that numerous independent plasmid-mediated *dpo* insertions, from unknown origin, and/or followed by a fast divergence might also have occurred. The putative *Gi. rosea* plasmid had low sequence similarity with the other *dpo* insertions in AMF mtDNA and no extra chromosomal *dpo* were found within any of the surveyed AMF genomes. This *dpo* sequence diversity could indicate that AMF mitochondria harbor a broad range of plasmid types, as described in *Neurospora* species, where seven circular and four linear homology groups of mitochondrial plasmids have been reported

(Schulte & Lambowitz, 1991; Yang & Griffiths, 1993; Arganoza *et al.*, 1994; Marcinko-Kuehn *et al.*, 1994).

It has been found that mitochondrial plasmids in plants could be horizontally acquired through pollen (Handa, 2007). Horizontal gene transfer (HGT) has also been hypothesized to be an important factor in the dispersal of mitochondrial plasmids between fungal strains and species. That assumption is based on i) discordant phylogenetic relationships between plasmids and host genomes (Kempken *et al.*, 1992), ii) the geographic distribution of plasmids within and among host species (Arganoza *et al.*, 1994), and iii) direct experimental demonstration of HGT (Collins & Saville, 1990; Griffiths *et al.*, 1990; Debets *et al.*, 1994). Until now, no direct experimental evidence of HGT of *dpo* sequences was provided, and no mitochondrial plasmid has yet been found in AMF (except a putative linear plasmid in *Gi. rosea*). However, the distribution of the plasmid-related *dpo* sequences is uneven between isolates of the same AMF species (Formey *et al.*, 2012) and we showed that there is a large diversity of *dpo* sequences within AMF (12 orthology groups, along with seven singletons). Another intriguing observation is that the *dpo* sequences are also present in all AMF family surveyed including the Glomeraceae, Gigasporaceae, Ambisporaceae, Diversisporaceae and Archaeosporaceae (data not shown, based on a preliminary 454 assembly). A mechanism known as post-fusion incompatibility might be the most plausible explanation for the ubiquitous distribution of AMF mtDNA *dpo*. This mechanism implies a brief hyphal fusion between incompatible AMF species and the formation of septa. This process has already been shown to allow exchange of nuclear genetic material (Croll *et al.*, 2009), and also offers an explanation for the mobile endonuclease-mediated HGT, leading to the formation of gene hybrids in the mitochondrial

genome of *Glomus diaphanum*-like species (Beaudet *et al.*, 2013a). Keeping in perspective that divergent AMF taxa can simultaneously colonize roots of the same plant species, it is likely that a mitochondrial plasmid could spread rapidly through the Glomeromycota phylum even between phylogenetically distant taxa.

3.6.4 Glomeromycota *dpo* are closely related to plant mitochondrial plasmids

More than 50 linear plasmids have been reported in the mitochondria of 20 fungal species, in contrast to the 14 found in eight plant species. These have only been observed twice in animal mtDNA, in the placozoan *Trichoplax adhaerens* and in the moon jelly *Aurelia aurita* (Shao *et al.*, 2006; Signorovitch *et al.*, 2007). This uneven distribution of mitochondrial plasmids within the Eukaryota raises questions about their origin and their mode of transmission. The structure of the plant mitochondrial plasmids consists of an invertron structure, long inverted repeats with proteins covalently bound to their 5' ends. This structure is remarkably similar to fungal plasmids and to some DNA viruses (Rosewich & Kistler, 2000). Additional evidence, such as low GC content in plant plasmids (30% in the *Brassica* 11.6 kb plasmid, 38.9% in the 10.4 kb sugar beet plasmid and 37-39% in the maize S plasmids) compared to their respective mtDNA (*Brassica napus* 45.2% (Handa, 2003), *Beta vulgaris* 43.9% (Kubo *et al.*, 2000) and the maize 44.0% (GenBank accession #DQ490951), supports HGT from a fungal donor as the source of linear mitochondrial plasmids in plants, since low GC content is typically found in fungal mtDNA (Handa, 2008) (with an average of 37% in *G. irregulare* isolates (Handa, 2008; Formey *et al.*, 2012)). Moreover, our phylogenetic analyses robustly support the DPO proteins of plants and Glomeromycota as a sister group. Taken together, this evidence suggests that AMF could be potential donors of

plasmid-related *dpo* to plants. However, species of Glomeromycota colonize the same ecological niche and often co-occur with many pathogenic or endophytic fungi in plant roots, and one of these co-occurring fungi could be the common donor to both plants and AMF, also plant plasmids could have different fungal origins. If this is the case, the organism in question has yet to be sequenced.

To further study the AMF mitochondrial plasmid-related *dpo*s, the mt genome of *Geosiphon pyriformis*, an ancestral member of Glomeromycota that lives in endocytobiotic association with the cyanobacterium *Nostoc punctiforme* (Gehrig *et al.*, 1996), could provide insights into the origin and evolution of those elements in that phylum. Also, some DNA polymerase genes were present in the mtDNA of the ascomycete fungal partner of the lichen *Peltigera malacea (Xavier et al., 2012)*. Those sequences clustered within the *other fungi* group in the global *dpo* network. Another interesting avenue would be to assess the horizontal transmission of mitochondrial plasmids in different fungi-plant symbiotic systems.

3.7 Conclusion and outlook

The mitochondrial genome comparison of the three *Glomus* species showed the extent of mtDNA plasticity with regards to their synteny. It also offers a basis to develop molecular tool kits aimed specifically at identifying and quantifying taxa of the ecologically and agriculturally important Glomeraceae family. However, finding evidence of inter-haplotype homologous recombination suggests the occurrence of mtDNA heteroplasmy in natural populations. Although this has yet to be demonstrated, it raises questions about the artificiality of single spore cultivations of these fungi, which probably induces an underestimation of the

mtDNA diversity *in situ*. Approaches like genotyping-by-sequencing (GBS) and ecotilling could be useful to assess this particular hypothesis. Mitochondrial genome comparisons can also provide insights into molecular processes related to the biology and community structure of AMF. The close phylogenetic relationship between the plasmid-related *dpo* of plants and Glomeromycota suggests that horizontal gene transfer events could have occurred between plants and their fungal symbionts. The intimate relationship between plants and AMF in nature may unravel other transfer events that have taken place through evolution. Also, the role of viruses and bacteria and their interactions with fungi and plants represents an interesting research avenue to generate knowledge about the evolution and origin of mitochondrial plasmids in those groups.

3.8 Acknowledgements

This work is a part of a research project organized and coordinated by Premier Tech. The authors are grateful for financial support from NSERC Cooperative Research and Development (grant number: RDCPJ 395241-09), Premier Tech and CRIBIQ. We would also like to thank Biopterre *centre du développement des bioproduits* and CRBM. Our thanks to Dr. B.F. Lang for bioinformatics assistance and access to an automated organelle genome annotation software, to Dr. Simon Joly for recommendations on the phylogenetic analyses, to Stéphanie Berthiaume for her assistance in long PCRs, to Dr. Terrence Bell and Karen Fisher Favret for English editing and comments on the manuscript. We also think two anonymous reviewers for their helpful comments..

Mise en contexte (Chapitre 4) - homogénéité mitochondriale *in vitro* expliqué par la présence de mécanismes de ségrégation?

Les causes principales expliquant la présence d'hétéroplasmie sont l'héritabilité biparentale et/ou l'accumulation de mutations, qui résultent en la coexistence de différents haplotypes mitochondriaux. Chez les CMA, la présence d'un état hétéroplasmique transitoire a récemment été démontrée, suite au croisement entre deux isolats génétiquement distincts de *Glomus irregulare* (de la Providencia *et al.*, 2013). En absence de facteur externe, l'hétéroplasmie devrait être l'état par défaut en raison de l'accumulation de mutations (White *et al.*, 2008). Chez les levures, il a été démontré que la présence de cet état est temporaire en raison de la présence de mécanismes moléculaires de ségrégation, comme la formation de nucléoides et de processus de réparation de l'ADN mitochondrial, qui contribuent à maintenir l'intégrité du génome (Zimmer *et al.*, 1991; Hu *et al.*, 1995; Yasuhira & Yasui, 2000). La présence de nucléoides, qui sont un regroupement de génomes mitochondriaux maintenu ensemble par des jonctions de Holliday, engendre un goulot d'étranglement génétique responsable d'une ségrégation mitochondriale rapide et efficace observée dans ce groupe fongique (Lockshon *et al.*, 1995; White & Lilley, 1997; MacAlpine *et al.*, 1998). L'homogénéité du génome mitochondrial des CMA observée dans les précédentes études (Lee & Young, 2009; Formey *et al.*, 2012), pourrait possiblement être le résultat de la présence de mécanismes moléculaires orthologues.

Chapitre 4 - Homoplasmy in monosporal cultures arising from crossed-isolates supports the presence of a putative mitochondrial segregation apparatus in *Rhizophagus irregularis*

Laurence Daubois*, Denis Beaudet*, Mohamed Hijri and Ivan de la Providencia

Institut de Recherche en Biologie Végétale, Université de Montréal and Jardin botanique de Montréal, 4101 Rue Sherbrooke Est, Montréal (Québec) H1X 2B2, Canada.

* These authors contributed equally to this work

Submitted to: PloSONE, April, 2014.

Author contributions: Conceived and designed the experiments: IdlP, LD, DB, MH. Performed the experiments: LD, IdlP. Analyzed the data: LD, DB, IdlP. Contributed reagents/materials/analysis tools: MH. Wrote the paper: IdlP, DB, LD, MH.

4.1 Abstract

The occurrence of transient length-heteroplasmy through anastomosis between geographically distant *Rhizophagus irregularis* isolates was previously demonstrated in the progeny spores resulting from crossing experiments. However, the persistence of the heteroplasmy in monosporal culture lines from crossed parental isolates and the mechanisms responsible for mitochondrial segregation are still unknown. Using the arbuscular mycorrhizal fungus model *R. irregularis*, we tested the hypothesis that the previously observed heteroplasmic state occurs and persists in monosporal *in vitro* culture lines, using three different isolate combinations and newly designed TaqMan isolate-specific mitochondrial markers. We found that homoplasmy was the dominant state in all monosporal cultures, with an apparent bias towards one of the parental haplotypes. These results strongly support the presence of a putative mitochondrial segregation proteic machinery in *R. irregularis,* whose complete set of genes were orthologous with those found in other fungi. Our findings suggest that segregation takes place either during spore formation or colony development. Here we present the basic building blocks to explain mtDNA homoplasmy in monosporal cultures arising from crossed isolates in Glomeromycota, an early divergent fungal lineage.

Keywords

Arbuscular mycorrhizal fungi, mitochondrial segregation, anastomosis, mitochondrial nucleoids, protein orthology, monosporal cultures.

4.2 Introduction

It has been widely documented that major food crops associate naturally with beneficial soil microbes, including the arbuscular mycorrhizal fungi (AMF), an ancient group of root-inhabiting fungi belonging to the phylum Glomeromycota (Schussler *et al.*, 2001). These ecologically crucial symbionts enhance the uptake of water and nutrients of the plants they colonize, especially phosphate (Smith & Read, 1997), as well as protect them against pathogens (Ismail *et al.*, 2011) and play a key role in soil structure (Rillig, 2004). The benefits to plants provided by this association with AMF could be enhanced by the manipulation of their fungal partner genetics, as demonstrated by Angelard et al. (2010) and Colard et al. (2011), using the model organism *Rhizophagus irregularis*. In both studies, through the exchange of genetic information via hyphal fusion (i.e. anastomosis), *in vitro* crossed culture lines and segregated culture lines were generated. This process influenced differentially the transcription of symbiosis-specific genes in rice, resulting of an increase in rice growth by a factor of five.

During the last decade numerous studies have been devoted to decipher the nuclear (Tisserant *et al.*, 2013) and mitochondrial genome organization patterns (Beaudet *et al.*, 2013a; Beaudet *et al.*, 2013b), the sexual/asexual paradigm (Halary *et al.*, 2011; Riley *et al.*, 2014) and nuclear segregation process in AMF (Angelard & Sanders, 2011). These data allow new insights into the nuclear inheritance processes and segregation mechanisms that could occur in these fungi. Despite this progress, little is known about the mitochondrial inheritance process and segregation mechanisms. Since the mitochondrial genome encodes essential components of the cellular energy-producing apparatus, understanding mitochondrial DNA (mtDNA)

organization and its inheritance processes in AMF is of paramount importance, in order to manage more efficiently mycorrhizal associations at large scale (Ceballos *et al.*, 2013). A previous study showed that AMF mtDNAs migrate massively in spores during their formation (Marleau *et al.*, 2011), but the faith of each parental haplotype following crosses needs to be investigated. In filamentous fungi, sexual crosses leads to uniparental transmission of mitochondria (Mannella *et al.*, 1979; Lee & Taylor, 1993), whereas mitochondria are biparentally inherited in budding yeast (Okamoto *et al.*, 1998; Berger & Yaffe, 2000). In yeast, microfilaments, such as actin, plays an important role in the positioning and motility of mitochondria, whereas microtubules are the principal mitochondrial transporter in many other fungi. (Westermann & Prokisch, 2002).

The publication of 14 complete AMF mtDNA (Lee & Young, 2009; Formey *et al.*, 2012; Nadimi *et al.*, 2012; Pelin *et al.*, 2012; Beaudet *et al.*, 2013b; de la Providencia *et al.*, 2013) in the last few years, allowed demonstrating that these sequences were all homogeneous within isolates (Raab *et al.*, 2005; Börstler *et al.*, 2008; Lee & Young, 2009; Formey *et al.*, 2012). These studies also shown that genomes are variable in defined intergenic regions between isolates, thus offering an incomparable opportunity to design isolate-specific markers (Corradi & Bonen, 2012; Formey *et al.*, 2012; Beaudet *et al.*, 2013a; de la Providencia *et al.*, 2013). Recently, de la Providencia et al. (2013) (de la Providencia *et al.*, 2013) demonstrated length-heteroplasmy in spores formed near anastomosis regions between geographically distant *R. irregularis in vitro* isolates, using isolate-specific mitochondrial markers. However, no information is available regarding the persistence of this length-heteroplasmic state in *in vitro* monosporal culture lines established from crossed parental isolates. Such a study is needed not

only to understand mtDNA inheritance processes and the persistence of mitochondrial haplotypes in stable culture lines, but also to use mtDNA as a criterion to define a reliable AMF taxonomic unit (Beaudet *et al.*, 2013a; Beaudet *et al.*, 2013b).

Further, no study has yet investigated potential mitochondrial segregation mechanisms occurring in AMF. In *Saccharomyces cerevisiae*, mitochondrial segregation is controlled by the mitochondrial segregation apparatus (MSA), which ensures a coordinated and reliable transmission of mitochondrial organelles and their genomes to the progeny. The MSA in *S. cerevisiae* is a trans-membrane proteic complex consisting of three mitochondrial membrane proteins (i.e. Mmm1, Mdm10 and Mdm12), forming the core component of the apparatus, along with Mmm2, Mdm31 and Mdm32, which interact with the core component (for review see (Chen & Butow, 2005)). This proteic complex links the mitochondrial outer membrane to the actin cytoskeleton and the mitochondrial inner membrane to mitochondrial nucleoids, which are clusters of similar mtDNAs packaged by the proteins Abf2, Aco1 and Ilv5 and maintained together by Holliday junctions (Lockshon *et al.*, 1995; White & Lilley, 1997; MacAlpine *et al.*, 1998). However, the molecular machinery of mtDNA inheritance remains largely unknown in many organisms (Birky, 2001; Sato & Sato, 2012).

In this study, we tested the hypothesis that monosporal culture lines support the occurrence and persistence of the observed length-heteroplasmy (de la Providencia *et al.*, 2013) in *R. irregularis in vitro* spores from crossed cultures of genetically divergent isolates. We also searched for the existence of a putative MSA in *Glomeromycota,* by finding the best reciprocal BLAST of yeast MSA proteins in the published genome of *R. irregularis* (Tisserant *et al.*,

2013) and its transcriptome (Tisserant *et al.*, 2012), and further confirm their orthology with all available sequences on the database. These first building blocks provide the foundation for explaining the effective mitochondrial segregation mechanisms occurring in AMF.

4.3 Materials and Methods

4.3.1 Growth conditions and maintenance of fungal cultures and roots

Monoxenically produced spores of *Rhizophagus irregularis* isolates DAOM-197198 (Pont-Rouge, Quebec, Canada), DAOM-234328 (Finland) and DAOM-240415 (Dufrost, Manitoba, Canada) were provided by the DAOM collection (Ottawa, Ontario, Canada). These three isolates were selected because of their different geographical origins and because their mitochondrial genomes have been fully sequenced (Nadimi *et al.*, 2012; de la Providencia *et al.*, 2013). Spores were subcultured in association with Ri T-DNA transformed chicory (*Cichorium intybus*) roots on a modified minimal (MM) medium (Bécard & Fortin, 1988) solidified with 0.4% (w/v) gellan gum (Sigma). Plates were incubated in the dark in an inverted position at 25° C. Several thousand spores and extraradical mycelia were obtained over a period of 12 weeks. Ri T-DNA transformed chicory roots were routinely propagated by placing actively growing root apexes on M medium with subsequent incubation at 25° C in the dark.

4.3.2 Crossed cultures and monosporal culture lines

Twenty crossed cultures from each combination (DAOM-197198/DAOM-234328;

DAOM-197198/DAOM-240415; DAOM-234328/DAOM-240415) were performed by inoculating a hundred spores in closed vicinity of Ri T-DNA transformed chicory roots, opposing each other at the extreme side of a Petri plate (Figure 4.1). Both colonies were checked weekly and their growth was traced in order to identify interaction zones between mycelia from different isolates, which were characterized by the formation of hyphal contacts. Subsequently, these contacts were checked under a Discovery V12 stereomicroscope (Carl Zeiss, Canada) at magnifications of 6.7-40x. Bright-field microscopy (Axio Imager M1, Carl Zeiss) was also used to observe details of the hyphal interactions at higher magnifications.

After 15 weeks of growth, randomly chosen spores (i.e. progenies) were harvested from the interaction zone of each combination, individually cut out from the mycelium and placed in a new mono-compartment Petri dish (90 mm) containing M medium in the close vicinity of a Ri-T transformed chicory root. For each combination, 50 replicates, consisting of one single spore associated with a chicory root were prepared. Each plate was checked weekly for germination, root colonization and colony development over the next 11 weeks.

4.3.3 DNA extraction

Spores and hyphae were harvested by dissolving the gellan gum matrix in which cultures were grown in a solution containing 0.0083 N sodium citrate and 0.0017 N citric acid. Extracted fungal material was observed under a binocular microscope in order to detect and remove any root contaminants. Spores and hyphae were gently crushed in a 1.5 ml microtube using a sterilized pestle. DNA was extracted using the DNeasy Plant Mini kit (Qiagen, Toronto, ON), according to the manufacturer's instructions.

4.3.4 Marker development, and genotyping by real-time PCR and sequencing of progeny spores

In order to efficiently detect low copy numbers of a given mitochondrial haplotype in each monosporal culture line, TaqMan isolate-specific markers were developed for each parental isolate and were used to genotype monosporal crossed culture lines (Table 4.1 and Supplementary Table S4.1). Genotyping of monosporal cultures lines resulting from each crossing experiment were performed in 3 replicates with approximately 2.5 ng of DNA per replicate. Reactions were carried in 20 µl, using iTaq™ Universal Probes Supermix (Bio-Rad, Canada) with final primers concentration at 0.5 µM and final probes concentrations at 0.1 µM. Multiplex reactions were performed using 5'FAM, 5'VIC and 5'NED dyes and their corresponding quencher. However, due to limited availability of calibrated fluorophores and filter limitations of the instrument, we could only perform multiplex qPCR for combinations DAOM-197198/DAOM-240415 and DAOM-197198/DAOM-234328 and perform singleplex qPCR reactions for the combination DAOM-240415/DAOM234328. Real time PCR assays were performed on a ViiA™ 7 Real-Time PCR System (LifeTechnologies, Canada), with both programs genotyping for multiplex reactions and comparative Ct for singlepex reactions, at 35 cycles. PCR amplicons were visualized on a 2.5% agarose gel stained with GelRed (Invitrogen, Canada). Successful PCR amplicons were sequenced according to the conventional Sanger technique at the Genome Quebec Innovation Center (Montreal, QC). As a control for marker specificity, the new TaqMan markers developed were challenged against *R. irregularis* (DAOM 242422 and DAOM 234179) and *R. clarus* (MUCL 46238). In addition, these markers were also tested in the spore progeny issued from the previous study of de la Providencia *et al*. (2013), in order to corroborate and confirm these earlier results.

4.3.5 Protein orthology and phylogenetic analysis

Each amino acid sequence of the *Saccharomyces cerevisiae* MSA and/or protein involved in nucleoid formation (i.e. MMM1, MDM10, MDM12, MMM2, MDM31, MDM32, ABF2, ACO1 and ILV5) was searched across the *R. irregularis* DAOM-197198 genome assembly (Tisserant *et al.*, 2013) and transcriptome (Tisserant *et al.*, 2012), using TBLASTN. Orthologous candidates in *R. irregularis* genome were recovered using the best-reciprocal BLAST hit to these proteins. Seven putative orthologous protein candidates were retrieved in *R. irregularis;* their GenBank accession numbers are ESA09626, ESA18874, ESA14336, ESA09628, ESA14201, ESA20109 and ESA20644, respectively. Furthermore, clusters of orthologous genes (COGs), gathering sequences from numerous organisms, were determined using STRING version 9.05 (Von Mering *et al.*, 2005) for each protein candidate. The resulting COGs were aligned using COBALT version 2.01 (Papadopoulos & Agarwala, 2007). Finding the best phylogenetic model and performing the maximum likelihood phylogenies was done using the integrative software TOPALI version 2.5 (Milne *et al.*, 2004). Tree figures were completed using TreeGraph version 2.0.47 (Stover & Muller, 2010).

4.4 Results

4.4.1 Germination and fungal development of monosporal culture lines

Cultures from each combination (DAOM-197198/DAOM-234328; DAOM-197198/DAOM-240415; DAOM-234328/DAOM-240415) were performed. After identifying interaction zones between mycelia from different isolates, which were characterized by the formation of hyphal contacts, randomly chosen spores (i.e. progenies) were harvested from the

interaction zone of each combination, individually cut out from the mycelium and placed in a new cultures (Figure 4.1). For each combination, approximately half of the inoculated spores germinated. The observed germination rate corresponds to rates previously observed in four *in vitro* cultured *Glomeraceae* isolates (Marleau *et al.*, 2011), and therefore is likely not influence by an heteroplasmic state. Following root contact and colonisation, some cultures stopped their growth, without producing any spores, and therefore could not be used in this study. The rate of mycelium development and spore production greatly varied among cultures, the combination of isolates DAOM-197198 and DAOM-240415 being the most successful, resulting in 18 colonized dishes containing between 200 and 1000 newly formed spores. The combination of isolates DAOM-197198 and DAOM-234328 produced seven colonized Petri dishes with often fewer than 40 newly formed spores. We obtained 11 poorly developed cultures from isolates combination DAOM-234328/DAOM-240415, and only nine produced sufficient amount of fungal material required for further analysis. In total, 34 monosporal cultures lines were genotyped.

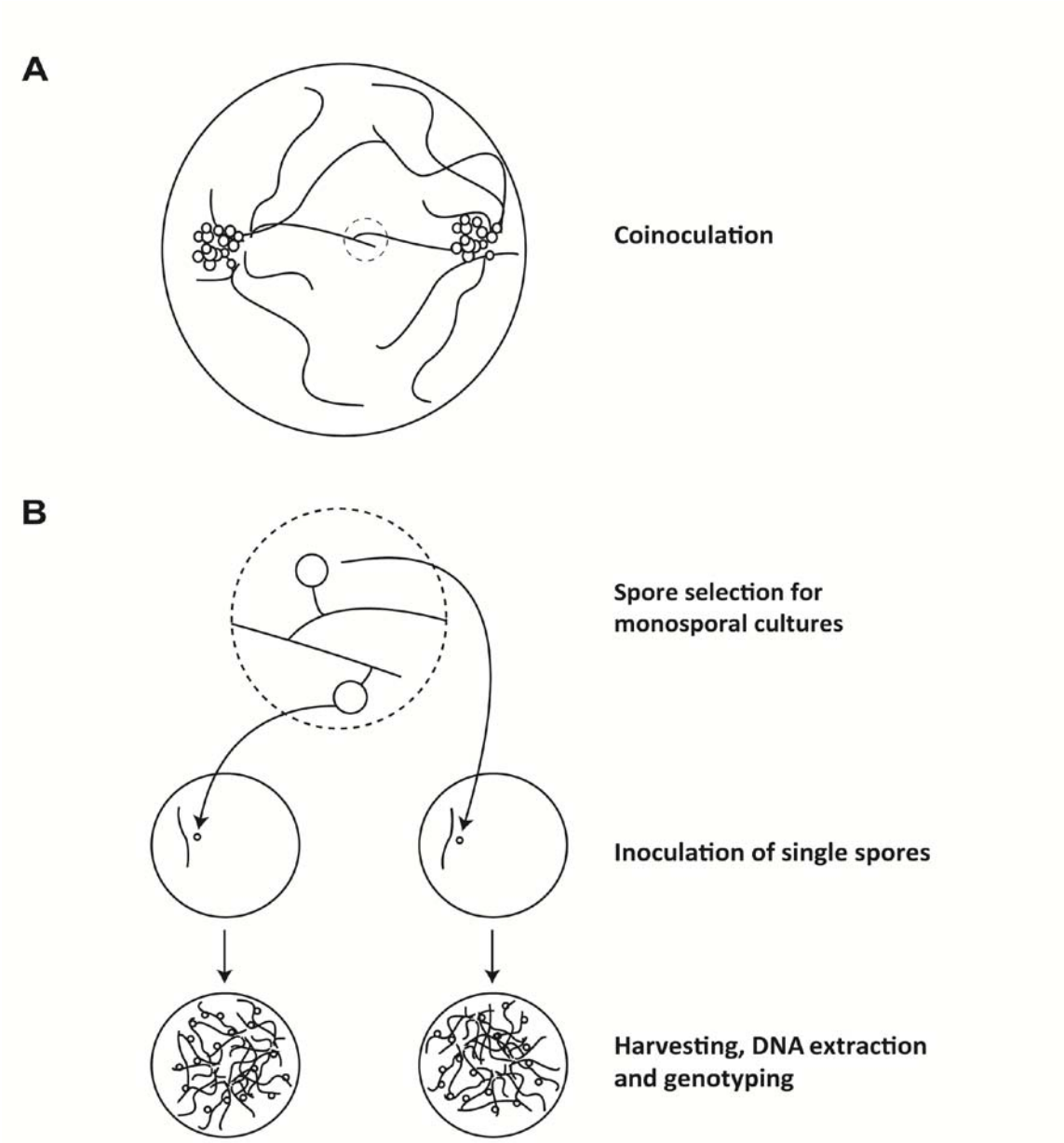


Figure 4.1. Schematic drawing of the experimental design. (A) Observation of hyphal contacts between two clusters of spores from different isolates and **(B)** Identification of spores produced in anastomosis regions and inoculation of a single spore on a new petri dish with Ri-T transformed chicory roots *i.e.* monosporal cultures. After germination and colonisation of the petri dish, spores are extracted and their DNA is used for genotyping of the monosporal culture.

4.4.2 Genotyping analysis

All 34 monosporal culture lines genotyped by qPCR approach presented only one parental mtDNA haplotype. In all nine progenies of the combination DAOM-197198/DAOM-234328, we only detected the DAOM-197198 haplotype, while in the combination DAOM-240415/DAOM-234328, all seven cultures exhibited the DAOM-240415 haplotype. However, among the 18 monosporal cultures of the combination DAOM-197198/DAOM-240415, two cultures showed the DAOM-197198 haplotype and the other 16 cultures were generated with the DAOM-240415 haplotype. Genotyping results are summarized in Table 4.2. qPCR data are not shown in this study but are available upon request. Interestingly, our results showed an apparent selection and/or segregative bias towards a given haplotype in each combination. Indeed, in combinations where DAOM-240415 haplotype was present, it seemed to be preferentially selected over the two other haplotypes. Also, the DAOM-197198 haplotype was largely dominant over the DAOM-234328 haplotype, which was not recovered in any monosporal culture lines. Using the same genetic material from single spores, we also corroborated the heteroplasmy detected in an earlier study by de la Providencia *et al*. (2013) (de la Providencia *et al.*, 2013), although at a lower rate, thus confirming non-self fusion between genetically-close isolates as an important mechanism shaping genetic exchange (Table 4.2). All the markers showed to be isolate-specific when challenged against other *Rhizophagus* isolates (Table 4.1).

Table 4.1. Specificity test of qPCR markers. The marker for each isolate is, at least, specific to the other two isolates used in this study. Adequate controls were carried, in triplicate, both in singleplex and multiplex. Multiplex reaction could not be performed with the combination of 240415 and 234328, because of limitations of the calibrated fluorophores and filter-limitations of the instrument. For each positive assay, in parenthesis is the mean ct value from triplicates with 2 ng of DNA per essay.

		R. irregularis DAOM								*G. cerebriforme* DAOM227022	Water
	Marker\DNA	197198	240415	234328	197198/240415	240415/234328	197198/234328	242422	*234179*		
SINGLEPLEX	197198	X (19)	-	-	X (19)	-	X (19)	X (19)	-	-	-
	240415	-	X (21)	-	X (21)	X (21)	-	-	-	-	-
	234328	-		X (26)	-	X (26)	X (27)	-	X (27)	-	-
MULTIPLEX	197198	X (19)	-	-	X (19)	-	X (19)	-	-	-	-
	240415	-	X (21)	-	X (21)	X (21)	-	-	-	-	-
	197198	X (19)	-	-	X (19)	-	X (19)	X (19)	-	-	-
	234328	-	-	X (26)	-	X (26)	X (27)	-	X (27)	-	-

Table 4.2. Characterization of monosporal cultures from crossing experiments with three different *Rhizophagus irregularis* isolates and single spores from de la Providencia et al. (2013).

Combinations	Inoculated monosporal cultures	Colonized monosporal cultures	Spores per culture	*R. irregularis* haplotype		
				DAOM197198	DAOM240415	DAOM234328
***R. irregularis* 197198/240415**	50	18	100-1000	2	16	NA
***R. irregularis* 197198/234328**	50	9	50-300	9	NA	0
***R. irregularis* 240415/234328**	50	7	30-100	NA	7	0

de la Providencia et al (2013)*

Combinations	Heteroplasmic spores		Homoplasmic spores	
	Previous markers	New markers	Previous markers	New markers
***R. irregularis* 197198/240415**	9	3	1	7
***R. irregularis* 197198/234328**	5	0	5	10
***R. irregularis* 240415/234328**	3	1	7	9

4.4.3 Mitochondrial segregation machinery and nucleoid genes orthology in *R. irregularis*

The genome of *R. irregularis* DAOM-197198 (Tisserant *et al.*, 2013) possesses orthologous genes coding for the core proteic structure of the MSA and involved in the formation of nucleoids in *S. cerevisiae* and other fungi (Figure 4.2). Orthologous candidates in *R. irregularis* corresponded to the best reciprocal blast hit with the corresponding *S. cerevisiae* MSA protein (Table 4.3). To confirm orthology, phylogenetic trees were constructed using all known orthologous genes in fungi and other organisms. All seven phylogenies we performed with clusters of orthologous genes (COGs) supported the idea that *R. irregularis* proteins are orthologs of fungal segregation apparatus and nucleoid proteins (Supplementary Figure S4.1). The presence of one *S. cerevisiae* paralog was observed in the Aco1 phylogeny, as well as two yeast homologs and one paralog in Abf2 COG phylogeny. However, in both cases, the *R. irregularis* protein grouped with the expected *S. cerevisiae* protein, implicated in the mitochondrial segregative processes. We did not found any paralogs in *R. irregularis*. Only two genes were not found in AMF, i.e. mdm31 and mdm32. For these, their role in yeast remains unclear. They are speculated to play an interactive role with the segregation apparatus core components (Dimmer *et al.*, 2005). Further, for these two proteins, the search for COGs regrouped only a few yeast sequences, which leads us to believe they might not be necessary to the segregation process in other fungi. All orthologs found in *R. irregularis* are putatively functional because they are also found in expressed sequence tags (ESTs) (Tisserant *et al.*, 2012).

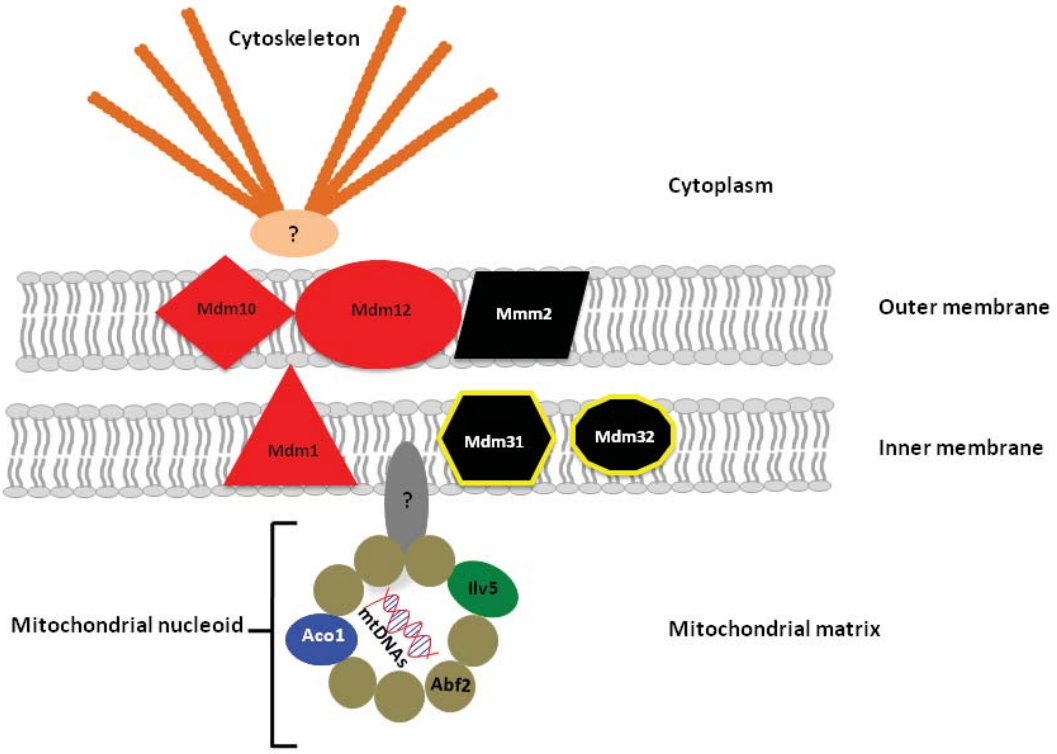


Figure 4.2. Representation of the putative mitochondrial segregation apparatus and nucleoid structure in Rhizophagus irregularis, based on protein orthology with S. cerevisiae. The core components of this machinery consist of a complex of three mitochondrial membrane proteins, which are Mmm1, Mdm10 and Mdm12 (colored in black). Three proteins are thought to interact with the main complex, Mmm2, Mdm31 and Mdm32 (colored in red). The proteins responsible for mtDNA packaging and nucleoid formation are Abf2, Aco1 and Ilv5 (colored in brown, blue and green, respectively). Unknown proteins are thought to link the outer membrane complex to the cytoskeleton (colored in orange) and the nucleoids to the inner membrane by interacting with Mmm1. The yellow outlining shows proteins for which no orthologs were found in *R. irregularis* (Table 4.3).

Table 4.3. Evidence for *Rhizophagus irregularis* protein orthology with the *Saccharomyces cerevisiae* mt segregation and nucleoid proteic machinery.

Protein[a]	Primary function[a]	Best reciprocal BLAST	E-value	Similarity %	Phylogenetic orthologs[b]	mRNA expression[c]
Mdm10 ESA18874	Core components of the mt segregation apparatus	●	1e-32	44	●	●
Mdm12 ESA14336		●	2e-15	52	●	●
Mmm1 ESA09626		●	3e-53	55	●	●
Mmm2 ESA09628	Putative interaction with the mt segregation apparatus	●	5e-11	53	●	●
Mdm31		○	○	○	○	○
Mdm32		○	○	○	○	○
Ilv5 ESA20644	Biosynthesis of Val, Ile and Leu	●	1e-158	76	●	●
Aco1 ESA20109	Citric acid cycle	●	0.0	84	●	●
Abf2 ESA14201	mtDNA packaging	±[d]	1e-14	48	●	●

[a] Described proteins and known functions in *Saccharomyces cerevisiae* (Chen *et al.*, 2005), along with the accession number of the *R. irregularis* ortholog protein candidates.

[b] As shown in supplementary information Figure 4.2.

[c] Based on the *R. irregularis* transcriptome data (Tisserant *et al.*, 2012).

[d] Two proteins in *R. irregularis* and *S. cerevisiae* showed close similarity to each other because they share the same high mobility HMG-BOX proteic domains.

4.5 Discussion

4.5.1 Homoplasmy rather than length-heteroplasmy in monosporal cultures lines from crossed-cultures

Previous results have shown that biparental mtDNA inheritance leading to an heteroplasmic state occurred in the spore progeny from crossed-cultures of divergent *R. irregularis* isolates, however heteroplasmy was not detected in germinated spores (de la Providencia *et al.*, 2013). Using TaqMan markers developed in the present study, we did not detect heteroplasmy in monosporal cultures from crossed parental isolates but confirmed the heteroplasmy status of the crossed cultures spores observed in de la Providencia et al. 2013 (de la Providencia *et al.*, 2013), although at a lower rate. Several factors might explain why we obtained dissimilar results with markers designed in different intergenic regions. The isolate-specific mitochondrial markers are designed in highly variable mobile elements rich regions, which have been shown to be recombination prone (Beaudet *et al.*, 2013b) and might compromise their specificity. Also, the use of the whole genome amplification (WGA) technique (de la Providencia *et al.*, 2013) can introduce significant bias regarding the insertions of SNPs or the formation of chimeras (Pinard *et al.*, 2006; Lasken & Stockwell, 2007). The latter could induce an overestimation of the presence of a given haplotype in a sample, especially since the occurrence of allelic drop-out and preferential amplification is well documented in single cell analysis using a PCR based approach (Findlay *et al.*, 1995). For these reasons, TaqMan qPCR assay is more reliable to assess mtDNA inheritance, given the nature of the highly variable/dynamic regions in which AMF isolate specific markers are designed.

Earlier studies have demonstrated that genetic exchanges occur via vegetative hyphal fusion (i.e. anastomosis), resulting in nuclei coexistence in a common cytoplasm altering both the plant and fungal phenotypes (Croll *et al.*, 2009; Angelard *et al.*, 2010; Angelard & Sanders, 2011; Colard *et al.*, 2011; de la Providencia *et al.*, 2013). Based on the formation of anastomosis between genetically-close *R. irregularis* isolates, several studies have shown that heterogeneous populations of nuclei (Angelard *et al.*, 2010; Colard *et al.*, 2011) and mtDNA (de la Providencia *et al.*, 2013) are randomly inherited at different frequencies (i.e. segregation) (Boon *et al.*, 2013) into the progeny (Marleau *et al.*, 2011). These findings support the paradigm that based on the coenocytic nature of the AMF fungal mycelium, nuclei, mitochondria and other organelles can flow "freely" and migrate between close or distant regions of genetically-close fungal colonies and form the so-called common mycorrhizal networks (CMN) (Selosse *et al.*, 2006).

However, studies on nuclear dynamics along the symbiotic extraradical mycelium challenged this point of view and brought a somewhat discordant note to the supposed continuous stream of mycelial cytoplasm/protoplasm (Bago *et al.*, 1999). Using *in vivo* two-photon microscopy techniques, these authors revealed a patchy distribution of nucleotypes throughout the mycelia and also demonstrated that nuclear flow occurs in pulses, being independent from the cytoplasmic streaming. These studies suggest that mitochondria, like nuclei, might not be equally distributed along hyphae. Therefore, through anastomosis originating from different individuals, the inheritance and segregation processes of both organelles into the spore progeny could be strongly related to the number, type and frequency of each organelle at the interaction zone (i.e. zone of the mycelium where genetically-close individuals fuse and

exchange nutrients and genetic information). The latter could explain the lack of detection of a heteroplasmic state in the progeny spore used to start the monosporal culture. In the event that one haplotype would be crucially underrepresented and therefore the heteroplasmy undectectable by a qPCR assay, such a small amount of a mitochondrial haplotype could easily be lost through stochastic drift in subsequent subcultures, and consequently homoplasmy would persist at the vegetative phase in these fungi. The dominance of some haplotypes we observed would still imply a selection and/or segregative bias of one haplotype over the other.

Together with our observations of a dominant homoplasmic state in the monosporal crossed culture lines, these studies strongly support a plausible mechanism of mtDNA segregation. This in turn offers an explanation for the low frequency, revised in this study (Table 4.2), of heteroplasmic spores resulting from crossed divergent isolates (de la Providencia *et al.*, 2013) and the non-detection in this study of heteroplasmic monosporal cultures lines arising from these spores.

4.5.2 Evidence for a mitochondrial segregation mechanism in *Rhizophagus irregularis*

The identification of orthologs to the MSA and nucleoid proteins in *R. irregularis* genome and transcriptome offers an explanation for the mitochondrial segregation we observed. The *S. cerevisiae* Mmm1, Mdm10 and Mdm12 proteins are an integral part of the mitochondrial membranes and constitute the core components of the MSA. This proteic complex connects the mitochondrial outer membrane to the actin cytoskeleton in yeast (Boldogh *et al.*, 1998; Boldogh *et al.*, 2003), while Mmm1 is thought to link mitochondrial

nucleoids to the inner membrane of the mitochondria (Hobbs *et al.*, 2001). Proteins involved in the MSA have not been reported to play a role in other molecular pathways in *S. cerevisiae*. The fact that we found orthologs of these proteins in AMF gives new insights into the structural intricacies that could potentially hamper the heterogeneous distribution of parental haplotypes in the spore progeny, following hyphal fusion. If indeed mitochondria were linked to the cytoskeleton (i.e. either actin or microtubules), along with their mtDNA bound to the inner membrane, they certainly would not be flowing easily in the cytoplasm. This hypothesis could potentially explain why the fungal protoplasm moves slowly after anastomosis between two divergent isolates as observed earlier (de la Providencia *et al.*, 2013). Also, three proteins of the mitochondrial membrane are also proposed to interact with the core components of the MSA, which are Mdm31, Mdm32 and Mmm2. They are thought to play a role in nucleoid stability, mitochondrial morphology and distribution in yeast (Youngman *et al.*, 2004; Dimmer *et al.*, 2005). However, we only found an ortholog for the Mmm2 protein, which might suggest that other proteins could play the same function as Mdm31 and Mdm32 or that they are not essential in the AMF mitochondrial segregation pathway.

Further, we identified orthologs for the three proteic constituents of mt-nucleoids in AMF. The Abf2 protein is the core packaging element of mt-nucleoids in yeast. It is a non-histone DNA binding high mobility group (HMG) protein, which shows homology to nuclear chromatin proteins (Landsman & Bustin, 1993). This homology explains the difficulty of obtaining the best reciprocal blast (Table 4.3). The COG used to perform the Abf2 phylogeny (Supplementary Figure S4.1) included other proteins that were homologs, and also paralogs, since *S. cerevisiae* underwent an ancient genome duplication (Kellis *et al.*, 2004). The two

other essential components of mt-nucleoid are the bi-functional proteins Aco1 and Ilv5 (Bateman *et al.*, 2002; Chen *et al.*, 2005). They are both highly conserved in AMF, probably due to their metabolic role in the citric acid cycle and amino acid biosynthesis, respectively (Petersen & Holmberg, 1986; Chen & Butow, 2005; Chen *et al.*, 2005). Their expression is metabolically regulated and they play a role in packaging mtDNA into favourable conformations, while protecting DNA against oxydative stress. The detection in *R. irregularis* genome and transcriptome of well conserved orthologous proteins involved exclusively in the MSA, suggest that they are functional and that they might also be involved in a similar process in AMF. The existence of mt-nucleoids in AMF is important because it accelerates the rate at which mtDNA would segregate, since it is directly correlated to the effective population size (White *et al.*, 2008). Nucleoids were also shown to create a genetic bottleneck and to be responsible for the rapid mitochondrial segregation observed in yeast (Birky *et al.*, 1978). Interspecific differences in the morphology, size and distribution of mitochondrial nucleoids have been visualized using different microscopic methods in yeast (Williamson & Fennell, 1976; Miyakawa *et al.*, 1987). To further investigate these findings in *Glomeromycota*, it would be interesting to perform immunofluorescent experiments targeting AMF putative nucleoid proteins, since green fluorescent protein (GFP) tagging has been shown to be transient and render unstable transformants in these multinucleated coenocytic organisms (Helber & Requena, 2008).

4.6 Conclusion

Although mitochondrial homoplasmy seems to be the rule rather than the exception in monosporal progeny originating from *in vitro* crossed-cultures, the coexistence of numerous

mtDNA haplotypes in the same cytoplasm and the occurrence of homologous mitochondrial recombination might be a common status in natural populations. This assumption is based on the fact that in the field, AMF constantly interact with other individuals, potentially giving rise to highly dynamic and frequent hyphal fusions. A detailed study of mitochondrial genetic diversity in natural populations and/or perturbed environments of AMF is needed in order to unravel the underlying fundamentals for further application of mitochondrial markers in population genetic studies. Further, the amenability of the yeast genetic system allowed us to provide genetic evidence supporting the existence of MSA. This provides the building blocks for a better understanding of the mitochondrial inheritance process and segregation in these fungi. Much remains to be learned about how the proteins potentially implicated in these structures interact with other elements and influence mtDNA organization and inheritance in AMF.

4.7 Acknowledgements

This work was supported by The Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grant to MH (grant number: 328098-2012), which is greatly acknowledged. We would also like to thank Dr. S. Halary for recommendations on the phylogenetic analyses and comments on the manuscript, to David Denis for assistance in the *in vitro* culture maintenance, set up of the crossed-cultures and the generation and maintenance of the monosporal cultural lines and to Dr. Karen Fisher Favret for comments on the manuscript.

Mise en contexte (Chapitre 5) - quelle est la diversité mitochondriale intra-isolat *in situ* ?

Les colonies de CMA sont caractérisées par l'inter-connection de différentes parties de leur réseau hyphal par l'entremise d'anastomoses (Giovannetti et al., 1999; de la Providencia et al., 2005; Voets et al., 2006; Purin and Morton, 2012), ce qui peut mener à des échanges génétiques, entre des individus génétiquement distincts. Cette dynamique de fusions hyphales dans les populations naturelles peut aussi prévenir la perte de diversité engendrée par la ségrégation et la dérive génétique (Bever & Wang, 2005). Récemment, des croisements *in vitro* entre des isolats génétiquement distincts de *G. irregulare* ont démontré l'induction d'un statut hétéroplasmique dans la progéniture sporale (de la Providencia *et al.*, 2013). Même si l'hétéroplasmie des CMA n'a jamais été observée dans les populations naturelles (*in situ*), ces résultats expérimentaux, combinés à l'observation de recombinaison homologue entre différents haplotypes mitochondriaux (Beaudet *et al.*, 2013b), suggèrent la présence d'hétéroplasmie *in situ*. La coexistence chez les CMA de différents haplotypes mitochondriaux au sein d'un cytoplasme commun et la recombinaison homologue entre ces haplotypes pourrait constituer des évènements rencontrés fréquemment dans les populations naturelles, comme il a été démontré dans des populations du Basidiomycète *Armillaria gallica* (Saville *et al.*, 1998).

Chapitre 5 - Mitochondrial genetic polymorphism in Glomeromycota leads to the coexpression of protein coding gene variants

Denis Beaudet*†, Ivan de la Providencia*†, Manuel Labridy, Alice Roy-Bolduc, Laurence Daubois and Mohamed Hijri

Institut de Recherche en Biologie Végétale, Département de Sciences Biologiques, Université de Montréal, 4101 Rue Sherbrooke Est, Montréal (Québec) H1X 2B2, Canada.

† These authors contributed equally to this work.

Accepted with revisions in NewPhytologist

Author contributions: Conceived and designed the experiments: IdlP, DB, MH. Performed the experiments: ML, LD. Analyzed the data: DB, IdlP, ARB. Contributed reagents/materials/analysis tools: MH. Wrote the paper: DB, IdlP, ARB, ML.

5.1 Abstract

Arbuscular mycorrhizal fungi (AMF) are multi-nucleated and coenocytic organisms, and the extent of their intra-isolate nuclear genetic variation has been a source of debate. Conversely, their mitochondrial genomes (mtDNAs) have appeared to be homogeneous within isolates. Although several lines of evidence have challenged mtDNA homogeneity in AMF, extensive survey of investigating intra-isolate allelic diversity has not previously been undertaken. In this study, we used a conventional PCR-based approach on selected mitochondrial regions with a high fidelity Taq polymerase, followed by cloning and Sanger sequencing. Two isolates of *Rhizophagus irregularis* were used, one cultivated *in vitro* for several generations (DAOM-197198) and the other recently isolated from the field (DAOM-242422). We also investigated the persistence of mtDNA variability at the transcriptional level. We found substantial intra-isolate allelic variation within the mtDNA that persists in the transcriptome. This variation was also shaped by the presence of mitochondrial DNA copies within nuclear genomes (Numts). Our study suggests that genetic variation in Glomeromycota is higher than had been previously assumed, and might be grossly underestimated in most NGS-based AMF studies in both mitochondrial and nuclear genomes.

5.2 Keywords

Arbuscular mycorrhizal fungi, mitochondria, heteroplasmy, NGS and Sanger sequencing, *Rhizophagus irregularis*, protein co-expression.

5.3 Introduction

The success of land plant colonization around 450 million years ago and its further evolution was facilitated by their interaction with the arbuscular mycorrhizal fungi (AMF) (Brundrett, 2002), an ancient group of root-inhabiting fungi belonging to the phylum Glomeromycota (Schussler *et al.*, 2001). They are considered to be obligate biotrophs and during their interaction with plants, these fungi are rewarded with fixed carbon (Kiers *et al.*, 2011) in exchange for highly specialized services (eg. increase of nutrients and water uptake and resistance against pathogens) that enhance plant health and fitness (reviewed in (Smith & Read, 2008)). The welfare of these services could be enhanced by modifying the shape of their genomic organization (Angelard & Sanders, 2011) through segregated lines and crossed-cultures.

However, the coenocytic nature of the AMF mycelium, the absence of a developmental stage reduced to a single nuclei combined with the ability of genetically divergent isolates to fuse and exchange genetic information (Croll *et al.*, 2009; Angelard & Sanders, 2011; de la Providencia *et al.*, 2013) have challenged the interpretation on how genetic variation is organized and maintained within these fungi over multiples generations (Sanders & Croll, 2010). Two possible scenarios have been proposed, homokaryosis and heterokaryosis. The homokaryotic state suggests that the nuclear polymorphism is the result from orthologous allelic variants partitioned between chromosomes (polyploidy) or paralogous copies within a chromosome (Pawlowska & Taylor, 2004) while in the heterokaryotic state, different allelic variants could be evenly partitioned among distinct nuclei or be presented in a consortium of complementary nuclei (Kuhn *et al.*, 2001; Hijri & Sanders, 2005). These hypothesis might not

exclude each other since the genetic variation among and within nuclei is likely to be a continuum between these two scenarios, being shaped by modest rates of hyphal fusion and segregation (Bever & Wang, 2005). Recent next generation sequencing (NGS) based-studies, investigating nuclear polymorphism (Tisserant *et al.*, 2013) by comparing single nucleus genomes in the model organism *Rhizophagus irregularis* (Lin *et al.*, 2014), revealed that genetic variation among nuclei is low, thus supporting the homokaryotic organization and haploidy in this fungus.

In contrast with the controversial debate about the nuclear genome organization, the publication of 14 mitochondrial genomes (mtDNA) (Lee & Young, 2009; Formey *et al.*, 2012; Nadimi *et al.*, 2012; Pelin *et al.*, 2012; Beaudet *et al.*, 2013b; de la Providencia *et al.*, 2013) has revealed high intra-isolate homogeneity, with no apparent polymorphism, even while thoroughly investigated (Lee & Young, 2009; Formey *et al.*, 2012). However, the detection of transient length-heteroplasmy caused by non-self hyphal fusion between divergent isolates (de la Providencia *et al.*, 2013), evidences of mitochondrial horizontal gene transfer (Beaudet *et al.*, 2013a), the occurrence of homologous recombination events between distinct mtDNA haplotypes (Beaudet *et al.*, 2013b) and the existence of a high rate of short indels detected in NGS studies (Formey *et al.*, 2012), suggested that coexistence of numerous mtDNA haplotypes in the same cytoplasm might be a common status in natural AMF populations. This assumption is based on the fact that in the environment AMF constantly interact with each other through hyphal fusions, potentially giving rise to fungal cells harbouring multiple genomes. Further, in the absence of external factors, such as selection pressure, heteroplasmy should be the default state for mtDNA under a combination of mutation accumulation and

genetic drift (White *et al.*, 2008). For these reasons, we investigated the occurrence of intra-isolate heteroplasmy in Glomeromycota, and whether mtDNA variation persists at the transcriptome level, as already demonstrated for nuclear genes (Boon *et al.*, 2010).

In order to test our hypothesis and investigate the presence of intra-isolate allelic variations within AMF mtDNA, we returned to the basics, using a reliable PCR-based approach with a high fidelity Taq polymerase, followed by cloning and Sanger sequencing of selected mtDNA regions. We used two *Rhizophagus irregularis* isolates for this study, DAOM-197198 and DAOM-242422, recently isolated from petroleum-polluted soil. We expected to find higher intra-isolate allelic diversity of mtDNA in the isolate DAOM-242422, which has recently been isolated from a soil that is highly contaminated with petroleum, than in DAOM-197198, which has been cultured for more than 40 generations *in vitro*, and was previously shown to have homogeneous mtDNA(Formey *et al.*, 2012; Nadimi *et al.*, 2012). Surprisingly, we observed substantial intra-isolate allelic variation within the mtDNA of both strains. Most interestingly, we observed mtDNA variability in the model isolate *R. irregularis* DAOM-197198 that had been overlooked by previous next generation sequencing (NGS) based studies. This variation persisted at the transcriptome level, resulting in the co-expression of several distinct transcripts.

5.4 Material and methods

5.4.1 Fungal growth conditions

Culture of *R. irregularis* DAOM-197198 was routinely maintained in our domestic collection in association with Ri-T-DNA transformed chicory (*Cichorium intybus* L.) roots

following the protocol described in Cranenbrouck et al. (Cranenbrouck *et al.*, 2005). This isolate has been distributed to several laboratories worldwide and maintained under *in vitro* conditions since 1992 (Cardenas-Flores *et al.*, 2010). Cultures from DAOM-197198 maintained by the Eastern Cereal and Oilseed Research Centre, Ottawa, ON (GINCO-Canada) and Premier Tech Biotechnology, Rivière-du-loup, QC, Canada (www.premiertech.com) were also used in this study. Pure cultures of *R. irregularis* DAOM-242422 were obtained by trapping AMF propagules from highly petroleum polluted-soils in *Allium porrum* L. (leek) and further association to Ri-T-DNA transformed chicory roots as described in (Cranenbrouck *et al.*, 2005). Only the first generation (G1) of pure (i.e. monoxenic) cultures were considered for this study (Figure 5.1).

5.4.2 DNA extraction

Following spores and mycelium extraction from the gellan gel (Doner & Bécard, 1991), samples were pulverized in liquid nitrogen and total genomic DNA for both cultures was extracted using the DNeasy Plant Mini kit (Qiagen, Rockville, MD) according to the manufacturer's instructions. DNA was stored at -20 °C until use.

5.4.3 Molecular marker development

Multiple alignments were performed on the variable intergenic regions of the complete sequenced *R. irregularis* mitochondrial genomes. Further, three intergenic regions were selected since they are non-coding and prone to accumulation of mutations. These regions were chosen based on their length, *cob-nad4* (400 bp), *cox2-atp8* (962 bp) and *nad4-nad1* (427 bp), to facilitate the PCR amplification and downstream cloning and sequencing steps.

The mitochondrial NADH dehydrogenase subunit I (*nad1*) protein coding gene (642 bp) and the single copy nuclear 40S ribosomal protein S2 (*rps2*) (670 bp) were used as controls, since the first has been shown to be homogenous in *R. irregularis* DAOM-197198 mtDNA (Formey *et al.*, 2012; Nadimi *et al.*, 2012) and the latter was previously shown polymorphic within an isolate of *R. irregularis* (Boon, 2012) (Table 5.1).

Table 5.1. PCR primers used in the mitochondrial diversity screening and in the RT-PCR experiments on the nad4 C-terminal structural variants.

mtDNA region	Primer	Sequences (5'-3')	Size (bp)
nad1	F	ATGCTATTCTTCTTACTGGAATC	642
	R	CTGCTACGAGCTCTTGTTCTG	
Nuclear *rps2*	F	GTTCCTGTTACGAAACTTGGG	670
	R	CCAGTAACCTTGCAAGGAACAG	
cob-nad4	F	GATCTTCTGCTTTCCGACCAT	783
	R	TGGACAAACTGGAAGTGGCT	
cox2-atp8	F	TGGTGTCCTTCATTATGGTA	962
	R	ACGTTGAAGCAGTTGAGGA	
nad4-nad1	F	GGAGTCCTAGCCGTTACCTT	747
	R	AAGGTTGAAGAATCCCGTAA	
RT-PCR	**Primer**	**Sequences (5'-3')**	
nad4 cDNA-specific	F	TGGCAGTACCACTAACGGCTAAC	
	S-R1	AGCCTTCTACCGGCATAGGT	
	S-R2	TTAGCCCGTTCTAACTACGG	
	S-R3	TTTCTAACTACGGGAGATTTTGGA	
	S-R4	GATGATACAGAAAAATCTAGGATGACG	
	Ctrl-R	CTAGATTAGCCCGTTCTAACTACG	

5.4.4 Polymerase chain reaction (PCR)

The pre-cloning PCR mixture was made up of 1X PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM of each deoxynucleotide triphosphate (dNTP), 0.5 mM of each primer, and 1 ul unit of Phusion Taq (Agilent Technologies), and 1 ul of DNA template in a volume of 20 ul. Thermal cycling parameters were as follow: Initial denaturation at 94°C for 3 min; 35 cycles at 94°C for 30 sec, 54°C for 25 sec, 72°C for 72 sec and the final elongation at 72°C for 10 min. PCRs were done on a Eppendorf Mastercycle ProS. PCR products were separated by electrophoresis using 1% (w/v) agarose gel using Gelred, visualized under ultraviolet light and images were recorded by Gel-Doc system (Bio-Rad, Canada).

5.4.5 Cloning, post-cloning PCR and sequencing

PCR products were purified with the Qiaquick PCR purification Kit and quantified using Qubit according to the manufacturer's instructions. PCR products were cloned using the strata clone ultra Blunt PCR cloning Kit (Agilent technologies) following the manufacturer's recommendations. 2 ul of ligation product were incubated at room temperature during 5 min with 3 µl of Strata clone Cloning Buffer and 1 ul of Strataclone vector Mix. 1 µl of ligation product were used to transform one tube of competent cells. After a 45 sec heat shock at 42°C, 250 ul of SOC medium were added to the transformed cells and incubated at 37°C with agitation for one hour. 100 ul of transformed cells were spread on LB agar plates containing X-Gal (40 mg/ml), IPTG (100 mM) and ampicilin (100 mg/ml). After an overnight incubation at 37°C, white bacterial colonies were spiked and transferred into 20 µl of PCR master mix for amplification. Screening PCR protocol was performed using KAPA2G Taq as follows: Initial denaturation at 94°C for 3min then 35 cycles at 94°C for 30 sec, 54°C for 25 sec, 72°C for 72

sec and the final elongation at 72°C for 10 min. Sanger sequencing was commercially performed on an Applied Biosystems 3730xl DNA analyzer at the Genome Quebec Innovation Centre (McGill University, Montreal, Canada).

5.4.6 RNA extraction and cDNA synthesis

RNA extraction was only performed on eight-weeks old DAOM-197198 cultures using an E.Z.N.A. fungal RNA Kit (Omega Biotek) according to manufacturer's recommendations. Residual DNA fragments were removed as described in Beaudet *et al*., (2013a). In total, 50 ul of 100 ng/ul of RNA was collected and stored at -80 °C until use. The SuperScript III reverse transcriptase kit (Life Technologies, Canada) was used for cDNA synthesis modifying manufacturer's recommendations according to Beaudet et al. Beaudet *et al.*, (2013a), with oligo dt (12-18) and gene specific primers. Resulting cDNA were stored at -20 °C until use.

5.4.7 mRNA expression experiment

In order to perform PCR confirming the mRNA expression of the four *nad4* C-terminal structural variants found, five pairs of variant specific primers were designed, along with an intergenic control region, (Table 1). Amplifications with cDNA as template were carried for all variants, and multiple controls were included to ensure the specificity of reactions. PCR amplification were carried using 1X PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM of each deoxynucleotide triphosphate (dNTP), 0.5 mM of each primer, and 1 ul unit of Taq (Kapa biosystems), and 2 ul of cDNA template in a total volume of 20 ul. Thermal cycling parameters were as follow: initial denaturation at 95°C for 2 min; 35 cycles at 95°C for 30 sec, 52 to 55°C for 30 sec, 72°C for 40 sec and final elongation at 72°C for 12 min. PCRs were

performed on a Eppendorf Mastercycle ProS. PCR products were separated by electrophoresis using 2% (w/v) agarose gel using Gelred, visualized under ultraviolet light and images were recorded by Gel-Doc system (Bio-Rad, Canada).

5.4.8 Sequence and diversity analysis

Sequences were edited, cleaned and assembled in Geneious Pro v6.1.2. (Biomatters Ltd, Auckland, New Zealand). Alignments were done using MUSCLE v.3.5 (Edgar, 2004). In the presence of a putative SNP site, we checked the sharpness of the nucleotide sequences and further confirmed its occurrence if it was present on both sequencing strands. If only one sequence identified the SNP, the anomaly was considered a sequencing error. The presence of chimeras was inspected on the Bellerophon software (Huber *et al.*, 2004) using the Huber-Hugenholtz correction parameter with a 300 bp window. Nucleotide consensus sequences were deposited on GenBank database and are registered under the accession numbers KJ775870-.KJ776340.

For each selected region, the consensus sequences retrieved were imported in the program Mothur version 1.29.2 (Schloss *et al.*, 2009) in order to compute allelic frequencies. Two sequences with a single nucleotide difference were considered distinct alleles. To compare allelic diversity among isolates and loci sampled at different depth (different number of clones), we performed a coverage-based rarefaction and extrapolation analysis using the iNEXT package version 1.0 (Hsieh, 2013) in R version 3.0.2 (Team, 2013) with 1000 bootstrap re-sampling for the construction of 95% confidence intervals, the details of this method are described in (Chao *et al.*, 2013) and (Chao & Jost, 2012). The estimated sample coverage,

which is a measure of sample completeness (Good, 1953), and the Chao-1 index, which is an estimator of asymptotic species richness (Chao, 1984), were also computed with the iNEXT package.

5.4.9 Searching for mitochondrial DNA copies (numts) in nuclear genomes

The BLASTn analysis were performed on the nuclear assemblies of *R. irregularis* DAOM-197198 (Tisserant *et al.*, 2013; Lin *et al.*, 2014) using all the mitochondrial loci investigated in this study as query. We considered the four single nuclei (N6, N31, N33 and N36) and two DNA samples (DNA1 and DNA2) of Lin et al. (2014) as independent assemblies, along with the one published by Tisserant et al. 2013. The presence of potential numts was considered if they met the following criteria: more than 100 nucleotides in length, with an e-value of 1E-20 or higher and flanked by non-mitochondrial sequences on the contig.

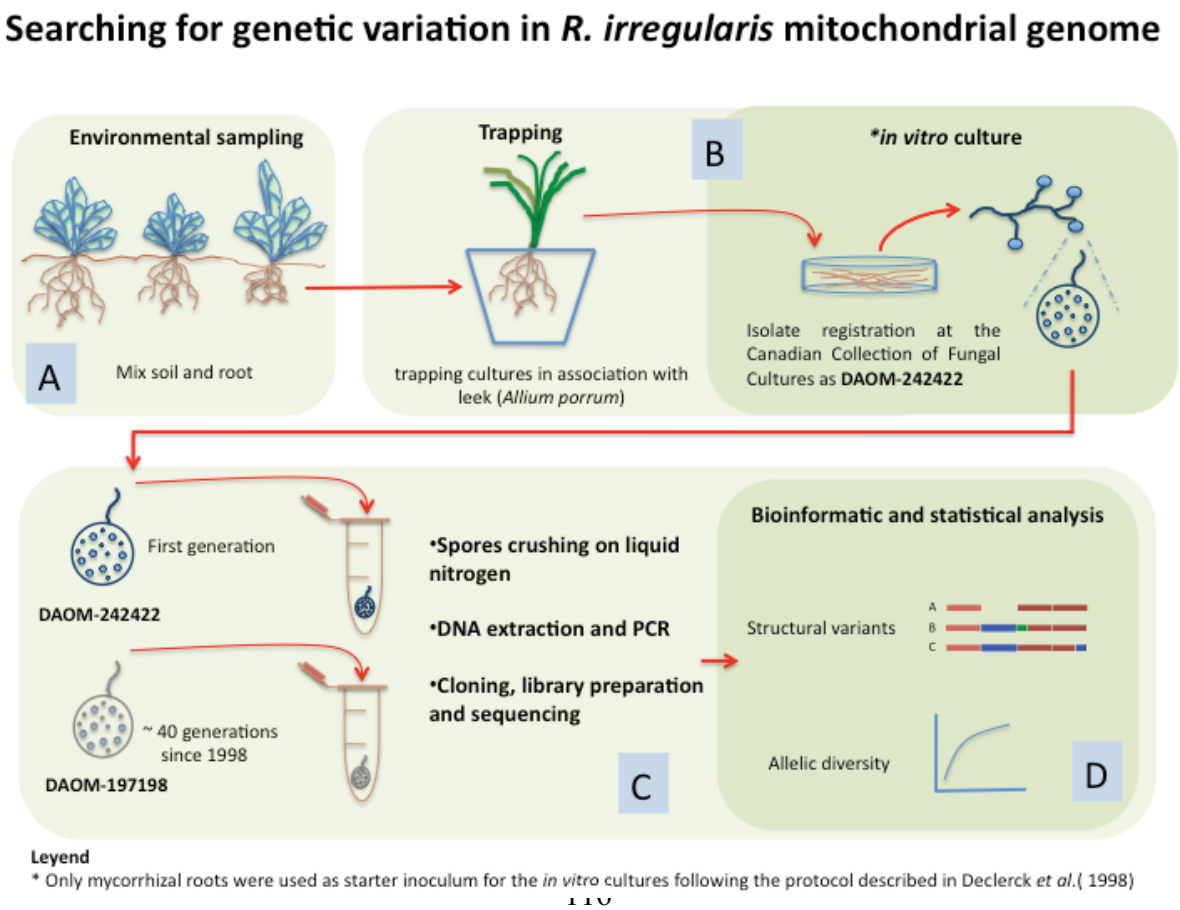


Figure 5. 1. Experimental design and detection of the genetic variation in *R. irregularis* isolates mitochondrial genome using a PCR-cloning and Sanger sequencing approach.

5.5 Results

5.5.1 Allelic diversity comparison between two *R. irregularis* isolates

A total of 471 consensus clone sequences were obtained by Sanger sequencing and further analyzed for the presence of chimeras and allelic variations. In total, seven putative chimeras were detected in the *R. irregularis* DAOM-197198 *nad4-nad1* intergenic region, but were kept in the downstream analysis after manual inspection, since their preference scores were close to 1.0. A summary presenting the results for each of the five regions is shown in Table 5.2 and the rarefaction and extrapolation curves for the different loci and isolates are displayed in Figure 5.2. We also investigated the recently published nuclear genome assemblies (Tisserant *et al.*, 2013; Lin *et al.*, 2014) for the presence of potential nuclear mitochondrial copies (numts) of the regions we investigated in this study. We found the presence of partial *nad4* and *nad1* CDS, along with partial *cox2-atp8* sequences on nuclear contigs. No potential numts were found for the *cob-nad4* intergenic region. Surprisingly, a complete *nad4-nad1* mitochondrial intergenic region, with almost perfect identity to its mitochondrial counterpart, was observed on a nuclear contig, but it was only present in one (DNA2) of the seven assemblies (Table 5.3).

As expected, we could not compute the rarefaction and extrapolation statistics for the *nad1* gene in the DAOM-197198 isolate since it did not show any variation. Indeed, the 47 clones

sequenced for the DAOM-197198 strain displayed a unique allele. For the same locus, eight distinct variants were identified for the DAOM-242422 isolate and the Chao-1 estimator indicates it could reach a diversity of 29 alleles. These eight variants harboured 12 polymorphic sites, with interestingly 11 non-synonymous mutations. The nuclear *rps2* gene was represented by five different alleles and the estimated Chao-1 diversity was close to 10 for both isolates. All mitochondrial intergenic regions exhibited intra-isolate genetic variation, generally between two and five different alleles. In almost all cases, the number of polymorphic sites in the sequences alignment was close to the number of alleles, which means that the divergence between the allelic variants was only one or two SNPs. Further, we noted the presence of a dominant allele in each region (Table 5. 2). Surprisingly, the DAOM-197198 isolate exhibited high diversity for the *nad4-nad1* locus, with 20 observed variants and over 56 estimated alleles. The estimated coverage (Table 5. 2) and the shape of the rarefraction curves (Figure 5. 2) indicate that most regions reached saturation for the two isolates, with the exception of the DAOM-197198 *nad4-nad1* region. In the latter, the estimated coverage is only 63% whereas all the other values are close or above 90%. This shows that, in general, the sequencing depth was sufficient to capture most of the allelic diversity present within each isolate.

5.5.2 Structural allelic variants and their mRNA expression

As described above, the *nad4-nad1* mitochondrial intergenic region of the model *R. irregularis* isolate DAOM-197198 exhibited high diversity (Figure 5.3A), with seven structural variants showing the presence of indels (Figure 5.3B). Four of these variants showing the presence of indels were variable in the C-terminal region of the NADH

dehydrogenase subunit 4 (*nad4*) mitochondrial protein-coding gene. The observed variation in the *nad4* C-terminal region did not cause any frameshift that could potentially hamper downstream transcription, but rather gave four alternative stop codons. These variants could

Table 5.2. Number of clones, allelic diversity, polymorphic sites and estimated coverage for the different loci and isolates.

Locus	Strain DAOM	Number of clones	Number of alleles	Dominant allele frequency	# of unique alleles	Polymorphic sites [a]	Chao-1 estimator (lci-hci) [b]	Good's coverage (%) [c]
nad1	197198	47	1	47	0	0	1,00 (1,00-1,00)	100
	242422	56	8	49	7	12 [d]	28,62 (6,54-50,71)	88
Nuclear *rps2*	197198	40	5	36	4	4	10,85 (4,21-17,49)	90
	242422	51	5	46	3	5	9,41 (0,92-17,89)	94
cob-nad4	197198	51	3	49	2	2	3,98 (2,91-5,04)	96
	242422	54	5	50	4	4	10,89 (4,19-17,59)	93
cox2-atp8	197198	46	3	43	1	2	3,49 (0,83-6,15)	98
	242422	37	3	35	2	2	3,97 (2,82-5,12)	95
nad4-nad1	197198	40	20	16	15	22	56,56 (13,76-126,89)	63
	242422	49	2	48	1	1	2,00 (1,06-2,94)	98

[a] Indels ≥ 2 nucleotides were counted as a single polymorphic site.
[b] The lower confidence interval (lci) and higher confidence interval (hci) values are the bounds on the upper and lower 95% confidence intervals for the average Chao-1 values. In other words, the observed richness was between those two numbers in 950 of the 1,000 bootstrap iterations (Chao, 1984).
[c] The estimated sample coverage, which is a measure of sample completeness (Good, 1953).
[d] Out of the 12 SNPs found, 11 were non-synonymous mutations in the *nad1* CDS.

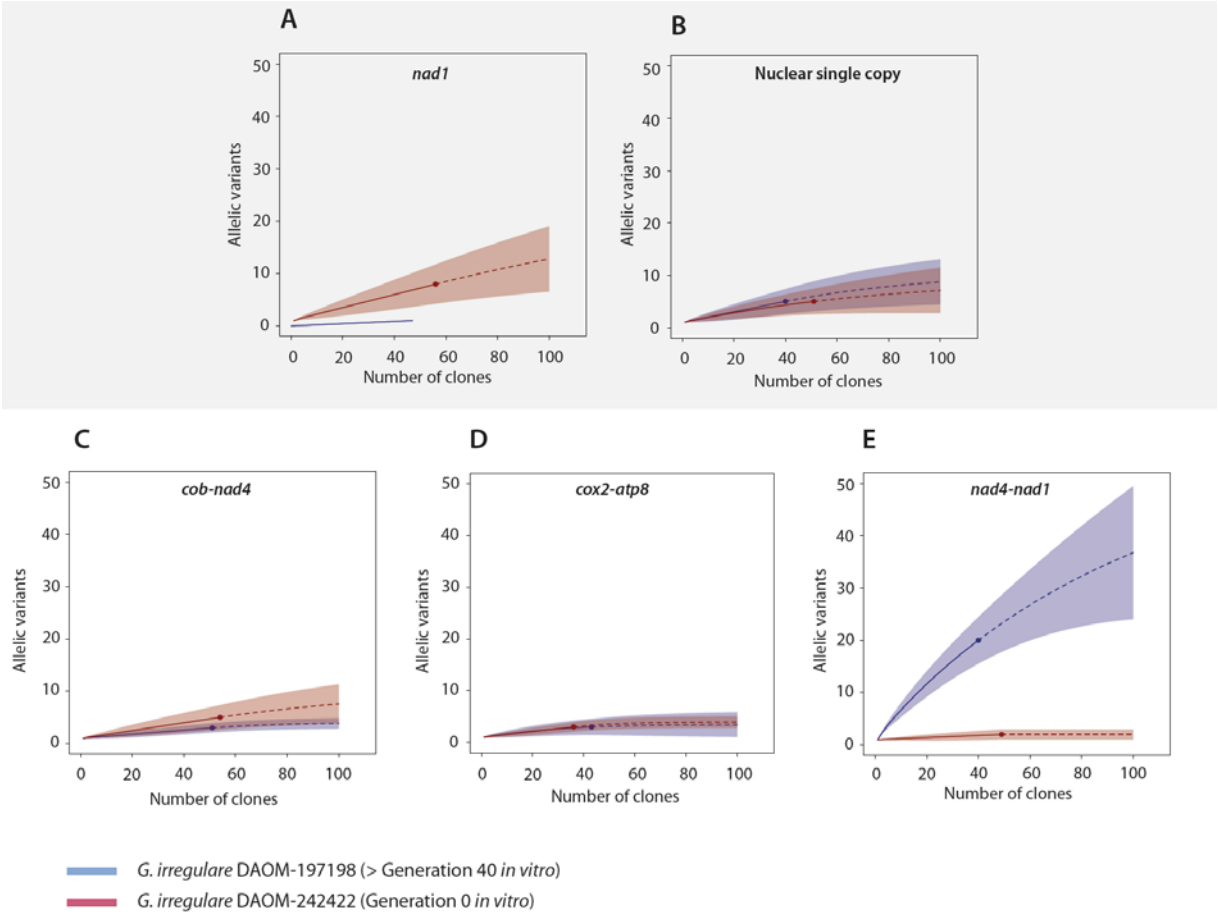


Figure 5. 2. Rarefaction curves inferred with the allelic diversity present at each investigated mitochondrial loci. The *nad1* gene (A) and the nuclear single copy gene 40S rDNA protein S2 (B) were used as controls (shaded in gray) and compared to three mitochondrial intergenic regions. The *cob-nad4* (C), *cox2-atp8* (D) and *nad4-nad1* (E) intergenic regions were investigated. The coverage based rarefaction (solid lines) and an extrapolation up to 100 clones (dashed lines) are shown. A confidence interval of 95% (shaded area, based on a bootstrap method with 1000 replicates) of the allelic diversity from cloning sequencing (filled circles) is represented. The comparison was done with *R. irregularis* model isolate DAOM-197198 (blue line) and the first generation *R. irregularis* strain DAOM-242422 (red line).

potentially give rise to four distinct *nad4* proteins ranging from 485 to 499 amino acids in length (Figure 5.3C). This unexpected structural variation was not found in DAOM-242422 isolate. Interestingly, the 454 sequencing reads of the DAOM-197198 isolate showed the

presence of these indels (data not shown), which were previously overlooked. Also, to rule out the hypothesis of in house contamination, we confirmed the presence of the allelic variation in the two other DAOM-197198 isolates originating from different laboratories (i.e. GINCO-Canada and Premier Tech) (Supplementary information Figure S5.2). Further, we assessed the mRNA expression of the four potential *nad4* proteic variants. We conducted a RT-PCR experiment on the DAOM-197198 cDNA using specific markers for each of the variants. The results showed that all four variants are expressed (Supplementary information Figure S5.3) and thus give rise to the co-expression of at least four distinct *nad4* proteins within a common cytoplasm.

5.6 Discussion

5.6.1 Mitochondrial genetic diversity is higher than meets the eye

Although, there is no unambiguous concluding evidence of homoplasmy in Glomeromycota, it has been consensually agreed that low levels of mtDNA variation is ubiquitous and intrinsic feature of AMF mtDNA (Raab *et al.*, 2005; Börstler *et al.*, 2008; Lee & Young, 2009; Formey *et al.*, 2012). Furthermore, low levels of intra-isolate variation have been mainly attributed to sequencing errors generated by the NGS platforms and was ignored by the downstream assembly process which often purge variation within a polymorphic sample, hampering the detection of SNPs and indels (Miller *et al.*, 2010). In this study we examined the inter- and intra mtDNA variation in two *R. irregularis* isolates using cloning and Sanger sequencing approaches in selected mitochondrial regions and we have found that

indeed, there is a substantial amount of mtDNA variation that even persist in the transcriptome.

The two *R. irregularis* isolates showed the presence of polymorphism either in the mitochondrial intergenic loci or in the two control loci we selected for this study, the *nad1* gene and the single copy nuclear *rps2* gene. The surprise came mostly from the *nad4-nad1* intergenic region, where the model *R. irregularis* isolate DAOM-197198, known for its mtDNA homogeneity (Formey *et al.*, 2012; Nadimi *et al.*, 2012), harboured a stunning number of 20 distinct alleles, that could reach up to a richness of 56.56 based on the Chao-1 estimator. Our study covered only 63% suggesting that not all the diversity could have been recovered (Table 5.2). In this particular region, seven out of these 20 allelic variants show the presence of indels which is reminiscent of the divergence usually observed between isolates of the same species (Formey *et al.*, 2012; Beaudet *et al.*, 2013a), rather than within an individual. In order to rule out the possibility of in-house contamination, we corroborated the presence of these indels in the same isolate originating from two other laboratories (Supplementary information Figure S5.2).

Also, we revisited the 454 sequencing reads (i.e raw data) of the DAOM-197198 strain (Nadimi *et al.*, 2012) and we found indels in the set of reads of the *nad4-nad1* region, however the downstream assembly purged that variation in the consensus contig. Formey and colleagues (Formey *et al.*, 2012) observed a similar scenario. The authors identified high rate of short indels in reads when mapping them on the nuclear and mitochondrial genome, but were further discarded as 454 sequencing errors. Our main concern is mostly in the filtering

algorithms (Miller *et al.*, 2010) commonly used to analyse the SNPs or indels diversity present in NGS reads. In fact, the usual filtering parameters consider a variant only if it is present more than two times in a read set (Lee & Young, 2009; Formey *et al.*, 2012), which is already challenged by a wide range of genomic phenomena (i.e. homopolymeric repeats) (Reviewed in (Miller *et al.*, 2010). This can be problematic in studies aiming to investigate for low level mutations (LLMs), like in population genetics, heteroplasmic mtDNAs or in multigenomic organisms such as AMF, where a substantial sequencing depth is required in order to confirm such rare variants (Miller *et al.*, 2010).

Table 5. 3. Blastn survey of all the putative mtDNA copies of the investigated mitochondrial loci in this study, present in the nuclear genome assemblies of R. irregularis DAOM-197198.

Nuclear assembly		mtDNA copies in the nuclear genome (numts) [a]				
		nad1 [b]	***nad4*** [b]	***cox2-atp8*** [b]	***cob-nad4***	***nad4-nad1***
Lin et al. 2014 [c]	DNA1	○	JARA01006962	JARA01006115	○	○
	DNA2	○	JARB01006781	JARB01003323	○	JARB01002805 [d]
	N31	○	JAQW01003701	JAQW01006974	○	○
	N33	○	JAQX01002384 JAQX01002367	○	○	○
	N36	JAQY01000696	JAQY01000696 JAQY01007470	○	○	○
	N6	○	JAQZ01003472	JAQZ01005991	○	○
Tisserant et al. 2013		○	AUPC01008815	○	○	○

[a] Numts > 100 nucleotides were taken into account.
[b] Only partial coding gene sequences were found in the nuclear genome.
[c] The reference genome available on GenBank, assembled based on the reads used for genome assemblies in JAQW00000000, JAQX00000000, JAQY00000000, JAQZ00000000, JARA00000000 and JARB00000000, was not used in the analysis.
[d] The *bona fide* complete *nad4-nad1* intergenic region found on a nuclear contig. Its putative presence was investigated by syntenic PCR amplifications.

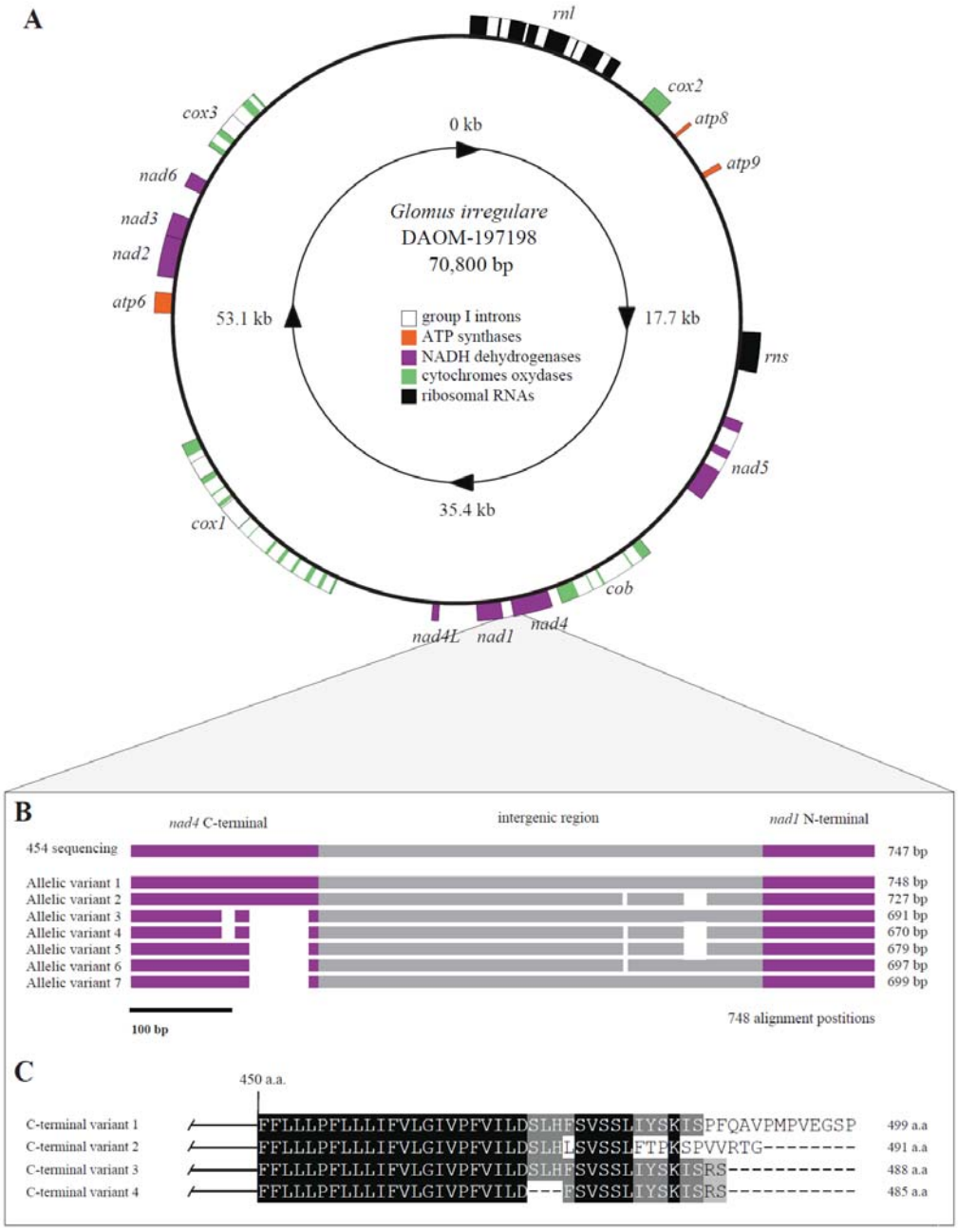


Figure 5. 3. Schematic representation of the intra-isolate mitochondrial allelic diversity found in the model AMF isolate R. irregularis DAOM-197198. The simplified representation of *R. irregularis* mtDNA, where the open reading frames, DNA polymerase genes and small inverted repeats are not shown **(A)**, serves as template to locate the investigated loci in the present study. The seven structural variant found in the *nad4-nad1* intergenic region are shown in an alignment **(B)**. The four *nad4* C-terminal protein variants are shown in an amino acid alignment and are expressed at the mRNA level (Supplementary Information Figure S5.3) **(C)**.

5.6.2 Mitochondrial protein variants co-expression in *R. irregularis* DAOM-197198

The *R. irregularis* DAOM-197198 isolate harboured at least 20 distinct alleles in the *nad4-nad1* intergenic region (Figure 5.3A), seven of them were structural variants (Figure 5.3B). Four of these variants had indels located in the C-terminal portion of the mitochondrial *nad4* protein coding gene. The divergence observed in the C-terminal region of the *R. irregularis* DAOM-197198 was reminiscent of the endonuclease-mediated partial gene duplication previously reported in AMF mtDNA and other fungi (Paquin *et al.*, 1994; Beaudet *et al.*, 2013a), but we have not found an endonuclease ORF nor any eroded remnants downstream of the sequence. The presence of these indels in the *nad4* CDS was rather surprising since their occurrence did not cause any frame-shifts that could hamper their rightful translation, but rather could potentially give rise to the expression of at least four protein variants with length ranging between 485 and 499 amino acids (Figure 5.3C). Their expression was confirmed by RT-PCR on the cDNA, thus testified for the co-expression of mitochondrial protein variants within a common cytoplasm. Further, the *R. irregularis* DAOM-242422 isolate had eight distinct alleles for the *nad1* locus with 11 out of 12 non-synonymous mutations in the polymorphic sites. Although we did not test their mRNA expression, this could also lead to the co-expression of at least eight different *nad1* protein variants in that particular strain. The differential selection pressures between the originating environment of the two *R. irregularis* isolates (i.e. one from an oil-contaminated soil and the other cultivated *in vitro* for more than 40 generations) might have played a role in shaping the allelic diversity we observed in this study. However, in light of these results, protein variants co-localization within a fungal cell might be intrinsic in AMF and probably constitutes an important evolutionary mechanism. It would be interesting to see whether the segregation or differential expression of mitochondrial variants could have an effect on AMF

fitness but also on the host plant, like was previously shown for the nuclear genome (Angelard & Sanders, 2011).

5.6.3 Mitochondrial DNA copies in nuclear genomes (numts)

The presence of 16 single-base substitutions and three regions of indels that were each supported by at least two reads in the *nad4-nad1* mitochondrial intergenic regions has already been noted (Lee & Young, 2009). However, these variants were attributed to the presence of numts. With the recently assemble nuclear genome or *R. irregularis* DAOM-197198, it is worthy to mention the detection of mitochondrial pseudogenes in the nuclear genome. We considered the four single nuclei (N6, N31, N33 and N36) and two DNA samples (DNA1 and DNA2) of Lin et al. 2014 as independent assemblies, along with the one published by Tisserant et al. 2013. When we searched for possible numts of the mitochondrial region we investigated in this study, we found partial coding sequences (CDS) of the *nad4* and *nad1* gene. Taking into account these observations, we hypothesize that the variation we found in this particular region might have been shaped by the ongoing evolutionary process of natural DNA transfer from mitochondria to the nucleus (Hazkani-Covo *et al.*, 2010). Numts are widespread in eukaryotes (Hazkani-Covo *et al.*, 2010), and are thought to be usually non-functional (Bensasson *et al.*, 2001). Their presence does not compromise the validity and novelty of our results since we have demonstrated the unambiguous presence of allelic variants in the mtDNA through syntenic amplifications and confirmed their mRNA expression, thus supporting intrinsic heteroplasmy.

Surprisingly, we found a *bona fide* complete *nad4-nad1* intergenic region (with again partial CDS) on a nuclear contig. However, that particular contig was present in only one of the seven investigated nuclear genome assembly. It could therefore be the result of a false assembly,

especially since the intergenic region was perfectly conserved, which is unexpected since a rapid erosion is usually observed for these transfers (Hazkani-Covo *et al.*, 2010). Interestingly, if the presence of this particular numt is confirmed, it would mean that the high allelic variation we observe in the *nad4-nad1* region might be shared not only between mt-haplotypes, but also among distinct nuclei.

5.6.4 Intra-isolate mtDNA genome heterogeneity in Glomeromycota

It is difficult to disentangle the genetics of AMF, mainly because there is no stage in their life cycle where a cell harbours a single nucleus. The absence of such a genetic bottleneck allows multiple nuclei and mitochondria to be inherited through generations, possibly resulting in individuals containing a variable number of diverse genomes (Boon *et al.*, 2013). Indeed, some studies have corroborated this issue by detecting genotype inheritance at different frequencies (i.e. segregation) of known or anonymous alleles into the progeny (Angelard & Sanders, 2011; Boon *et al.*, 2013). Others also pointed out that genetic variation has never been exhaustively sampled for these fungi (Boon *et al.*, 2013), indicating that AMF could not be described as organisms harbouring a single genome sequence. The detection of numerous variants of mtDNA in both *R. irregularis* isolates but also the co-expression of at least four protein variants in the model organism DAOM-197198, suggests that an AMF isolate harbour a population of different mitochondrial genomes. The existence of multiple genomes both in nuclei and mitochondria might have played a key role to the evolutionary success of these putative asexual fungi.

5.7 Conclusions and outlook

Our findings suggest that the allelic diversity in *R. irregularis* might be critically underestimated and raise the question of whether the polymorphism detected in this study was overlooked in previous NGS based studies (Formey *et al.*, 2012; Nadimi *et al.*, 2012) and probably in other published mitochondrial and nuclear genomes. We have detected a substantial amount of variation in two isolates of *R. irregularis* and as already found for nuclear markers (Boon *et al.*, 2010) it persisted at transcriptional level, thus rejecting the hypothesis of homoplasmy. Further, the presence of this intra-isolate mitochondrial variation was shown difficult to delimitate from NGS sequencing errors. Future research might be addressed to study whether co-expression of mtDNA variants and their segregation might influence the AM symbiosis. This aspect might be of relevant importance in agro-ecosystems since the manipulation of the AMF genetics was shown to shape plant fitness.

5.8 Acknowledgements

This work was supported by The Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grant to MH (grant number: 328098-2012) and by funds of Genome Canada and Genome Québec, which are greatly acknowledged. We thank Petromont Inc. (ConocoPhillips Canada) for allowing us to access to the Varennes field site. We also thank David Denis for technical support on the isolation and maintenance of *in vitro* cultures, Drs Terrence Bell, Royce Steeves and Karen Fisher Favret for comments on the manuscript.

Chapitre 6 - Discussion générale

6.1 Résumé des résultats et faits saillants de la thèse

La présente thèse de doctorat a démontré la présence de variation génétique substantielle intra- et inter-spécifique, engendrées par l'invasion d'éléments mobiles (comme les *dpo*, les *mORF* et les SIR) dans les génomes mitochondriaux des CMA, donnant lieu à une évolution moléculaire rapide des régions intergéniques et des introns. Cette variation permettait pour la première fois de développer des marqueurs spécifiques à des isolats de la même espèce. De plus, l'observation de duplications partielles de la région C-terminale de plusieurs gènes mitochondriaux codant pour des protéines, a permis d'inférer par analyses phylogénétiques, la possibilité de transfert latéral de gènes entre souches de CMA (Chapitre 2). Ces indices d'échanges génétiques entre souches laissaient présager que les anastomoses (c.a.d. fusions hyphales) pourraient être un mécanisme permettant le mélange d'haplotypes mitochondriaux, à l'instar des génomes nucléaires (Croll *et al.*, 2009).

L'insertion d'éléments mobiles, comme ceux énumérés précédemment, dans les régions intergéniques soulevait des questions intéressantes, puisqu'ils avaient été démontrés être à l'origine de réarrangements génomiques chez d'autres organismes. À l'aide d'une approche évolutive par réseaux de similarité de gènes sur toutes les séquences de *dpo* présentes dans les bases de données publiques, ainsi que dans les génomes mitochondriaux des CMA, on a été en mesure de démontrer des évènements de recombinaisons homologues entre des haplotypes mitochondriaux distincts, menant à des réarrangements génomiques, ainsi qu'un cas possible de transfert latéral de gènes entre plantes et CMA (Chapitre 3). Cette approche analytique par réseau

sur les éléments mobiles constituait une première. De plus, l'observation des signatures moléculaires de recombinaison homologue entre différents haplotypes mitochondriaux permettait encore une fois d'amener l'hypothèse de la colocalisation de plusieurs haplotypes mitochondriaux au sein d'un même isolat de CMA.

Cependant, tous les isolats séquencés et les analyses de polymorphisme mitochondrial intra-isolat (Lee & Young, 2009; Formey *et al.*, 2012), démontraient la grande homogénéité de ce génome. Cette apparente contradiction, avec les indices démontrant des échanges génétiques et de la recombinaison entre différents haplotypes, a amené à l'étude les échanges génétiques mitochondriaux, ainsi que la ségrégation des haplotypes parentaux dans les spores filles résultantes de croisements. Pour ce faire, à l'aide de marqueurs spécifiques développés dans cette thèse, une première étude démontra que les croisements entre des isolats génétiquement distincts de *G. irregulare* engendraient des spores filles pouvant posséder les deux haplotypes parentaux (de la Providencia *et al.*, 2013) (voir article #2 Annexe). Cependant, il restait à savoir si cet état hétéroplasmique était maintenu lors de la germination et du développement de la colonie. Cet aspect fût abordé (Chapitre 4) et démontra que l'état hétéroplasmique était absent et/ou perdu dans toutes les cultures initialisées avec des spores filles issues de croisements inter-isolats. Il y avait cependant la présence fortement majoritaire d'un haplotype parental donné dans chacun des croisements. Cela suggérait la présence de mécanismes moléculaires de ségrégation lors de la formation et/ou de la germination des spores. C'est dans cette optique que nous avons démontré pour la première fois la présence de tous les gènes orthologues aux champignons filamenteux et aux levures codant pour un appareil de ségrégation mitochondrial, ce qui permettrait d'expliquer en partie la ségrégation efficace et l'homogénéité intra-isolat des génomes mitochondriaux, même suite à des croisements entre isolats génétiquement distincts.

Malgré le fait que les mécanismes de ségrégation moléculaire pouvaient fournir une piste d'explication concernant l'homogénéité de l'ADN mitochondrial pour les souches *in vitro*, même suite à des croisements entre isolats distincts, une incertitude subsistait pour les souches de CMA évoluant en milieu naturel. En effet, la coexistence de nombreux CMA (différents isolats et/ou espèces) dans les communautés du sol est la règle plutôt que l'exception, ce qui favorise les contacts entre des individus génétiquement distincts. Dans ce contexte, une dynamique permanente de fusions hyphales pourrait jouer un rôle crucial dans les échanges génétiques au sein d'un même isolat et entre des isolats de CMA évolutivement rapprochés, mais génétiquement divergents. Cela donnerait potentiellement lieu à des isolats intrinsèquement hétéroplasmiques. En utilisant une souche de CMA de génération I provenant d'un sol pollué, comparé à un isolat ayant été cultivé *in vitro* plus de 40 générations, on a démontré la présence de polymorphisme génétique mitochondrial dans les deux isolats investigués. De plus, cette variation était exprimée dans l'ARN messager dans un des isolats (Chapitre 5). Ces résultats étaient plus qu'intéressants puisqu'ils suggéraient que d'autres études basées sur le séquençage nouvelle génération aurait potentiellement sous-estimé cette variation. Ce chapitre amenait plusieurs nouveaux arguments afin de considérer les CMA comme des organismes possédant une population de génomes mitochondriaux et nucléaires distincts.

6.2 Séquençage des génomes mitochondriaux des CMA: potentiel et limitations

Dans le cadre du projet dans lequel s'intégrait ma thèse de doctorat, 12 souches de CMA ont été séquencées jusqu'à maintenant, dont cinq isolats de l'espèce modèle *Glomus irregulare* (Tableau 6.1) . Cela nous a permis de faire une analyse comparative exhaustive de la diversité intra- et inter-spécifique d'espèces de la famille des Glomeraceae. De plus, une espèce de CMA

distante a été complétée, soit *Gigaspora rosea* de la famille des Gigasporaceae (voir article #1 Annexe). Une tentative avait été effectuée sur trois autres espèces distantes, soit *Entrophospora colombiana*, *Diversispora spurca* et *Ambispora leptoticha.* Cependant, en raison de la quantité limitée de matériel biologique et de la présence de contaminants externes (bactéries, levures, etc.), due à la méthode de maintient des cultures *in vivo*, très peu de contig mitochondriaux appartenant à ces souches ont été récupéré dans les données de séquençage. Pour ces espèces, la grande majorité des données de lecture du séquençage 454 obtenues étaient d'origine bactérienne (Figure 6.1). Cette problématique récurrente dans le cadre de ma thèse a amené à réfléchir sur une méthode plus efficace pour le maintient des cultures de CMA qui permettrait de séquencer proprement un plus grand éventail d'espèces.

Il y a deux types de cultures de CMA, i) la culture *in vitro,* qui permet de cultiver de façon stérile des spores de CMA. La culture se fait dans des pétris avec un milieu de culture en gélose. Des plantes modifiées génétiquement pour ne produire que des racines (généralement la carotte), sont placées en co-culture avec les CMA sur ce milieu (Bécard & Fortin, 1988). La récolte des spores se fait par liquéfaction du milieu de culture et filtration des spores. ii) la culture *in vivo,* de son côté, est généralement réalisée dans des pots conventionnels avec des plants de poireaux inoculés avec des spores de CMA. Ces pots sont placés dans des sacs en plastique transparent qui permettent de fermer le système et de limiter l'évaporation de l'eau et les contaminations croisées. La récolte se fait en prélevant de la terre et des racines de la plante pour ensuite extraire les spores par centrifugation dans une solution à gradient de sucrose. Il y a plusieurs inconvénients liés à ces méthodes de culture actuellement utilisées.

Tableau 6. 1. Souches dont les génomes mitochondriaux sont disponibles ou en cours d'analyse.

Souches	Isolats	Culture	Statut	Nu. Acc.	Référence
Glomus fasciculatum	DAOM240159	In vitro	En cours		
Glomus cerebriforme	DAOM227022	In vitro	Public	KC164356	Beaudet et al. 2013b
Glomus aggregatum	DAOM240163	In vitro	En cours	-	-
Glomus sp.	DAOM229456	In vitro	Public	JX065416	Beaudet et al. 2013a
Glomus sp.	DAOM240422	In vitro	Public	KC164355	Beaudet et al. 2013b
G. irregulare	DAOM197198	In vitro	Public	HQ189519	Nadimi et al. 2012
G. irregulare	DAOM234179	In vitro	Public	KC164354	Beaudet et al. 2013b
G. irregulare	DAOM240415	In vitro	Public	JX993113	de la Providencia et al. 2013
G. irregulare	DAOM213198	In vitro	Public	KF591215/6	Namidi & Hijri Submitted
G. irregulare	DAOM234328	In vitro	Public	JX993114	de la Providencia et al. 2013
G. irregulare	494	In vitro	Public	FJ648425	Lee & Young 2009
G. irregulare	MULC46239	In vitro	Public	JQ514224	Formey et al. 2012
G. irregulare	MULC46240	In vitro	Public	JQ514225	Formey et al. 2012
G. irregulare	MULC43204	In vitro	Public	JQ514224	Formey et al. 2012
Gigaspora rosea	DAOM194757	In vivo	Public	JQ693396	Nadimi et al. 2012
Gigaspora margarita	BEG34	In vivo	Public	JQ041882	Pelin et al. 2012
Ambispora leptoticha	GA401A	In vitro	En cours	-	-
Ent. colombiana	CL148	In vivo	En cours	-	-
Diversispora spurca	SC151	In vivo	En cours	-	-
Glomus mosseae	FL156	In vivo	En cours	-	-

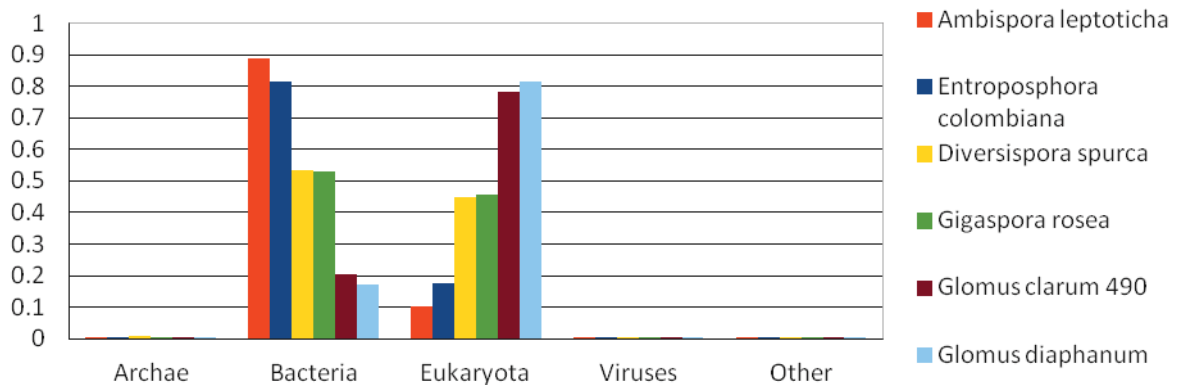


Figure 6.1. Histogramme représentant le pourcentage global de séquences obtenues en fonction de leur origine évolutive dans les données brutes de pyroséquençage 454.

Parmi les inconvénients liés à la méthode de culture *in vitro*, on note particulièrement que le nombre d'espèces de CMA pouvant être cultivées est très limité, la récolte des spores est laborieuse et demande beaucoup de temps, la période de récolte des spores est limitée (car le milieu a une réserve finie en nutriments), on doit constamment faire des sous-cultures pour maintenir la collection et il y a seulement quelques possibilités de système racinaire (ceux de la chicorée et la carotte sont les plus couramment utilisés). De plus, les désavantages liés à la méthode de culture *in vivo* sont: problématique de contamination croisée lors de l'arrosage, maintient de la culture qui demande beaucoup de temps (arrosage, fertilisation, etc.), difficulté à maintenir la culture stérile ce qui amène des problèmes comme le pourrissement des racines, la prolifération d'algues ou les maladies causées par des pathogènes, humidité excessive dans le système créant beaucoup de condensation, récolte des spores de CMA qui est laborieuses, demande beaucoup de temps et est destructive. Toutes ces problématiques ont mené au développement d'un nouveau système de maintient des cultures de CMA (voir Annexe 5).

6.3 Développement et utilisation de marqueurs moléculaires mitochondriaux

L'utilisation du génome mitochondrial comme marqueur dépend de deux aspects, i) le maintient de l'intégrité du génome à l'intérieur d'une espèce et ii) l'accumulation des mutations entre deux espèces. Jusqu'à maintenant les alignements de 16 intergènes mitochondriaux pour 14 souches de CMA ont permis de constater la variabilité inter-isolat et inter-spécifique en ce qui a trait à la présence d'éléments mobiles qui divergent rapidement, créant de nombreuses insertions et délétions (indels). Cette divergence offre la possibilité de développer une multitude de marqueurs (spécifiques, taille spécifique, ainsi que généraliste à certain genre). Certaines régions intergéniques sont même beaucoup trop variables pour les présenter en une figure en raison d'insertions d'éléments mobiles dégradés qui encodent une ADN polymérase (*dpo*) provenant de plasmides mitochondriaux. Ce sont ces régions qui offrent la possibilité de développer des marqueurs spécifiques entre des isolats de la même espèce (Figure 6.3). Nous en avons développé pour la plupart des souches d'intérêt et certains d'entre eux sont présentement utilisés dans des projets expérimentaux et par une entreprise privée pour l'élaboration d'un processus de validation de procédé industriel. Les intergènes démontrent une grande variabilité, mais les régions codantes sont très conservées. Il est possible de se servir de ces régions pour développer des marqueurs généralistes, mais spécifiques à certains groupes de CMA (Figure 6.4).

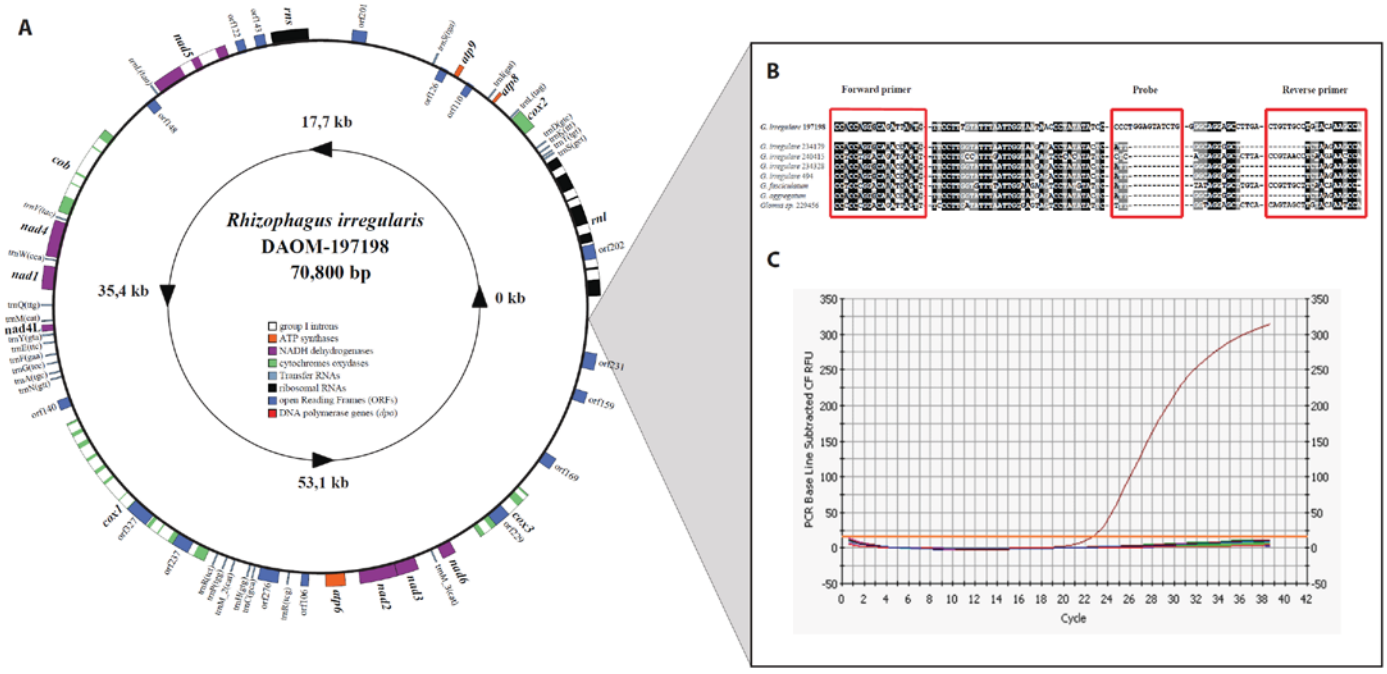


Figure 6.3. Développement et validation d'un marqueur spécifique à un isolat de CMA.
A) Représentation graphique circulaire du génome mitochondrial de *Glomus irregulare* DAOM-197198. **B)** Alignement d'une portion de la région intergénique *cox3-rnl* où le marqueur spécifique a été déterminé. L'alignement a été effectué en utilisant quatre isolats distincts de *G. irregulare*, avec trois espèces génétiquement rapprochées. Les encadrés rouges correspondent à la position des amorces et de la sonde TaqMan. **C)** La spécificité du marqueur a aussi été testée par amplifications qPCR sur l'ADN d'autres isolats de G. irregulare: DAOM-234179, DAOM-240415, DAOM-234328, DAOM-2207225 and DAOM-46328, et sur l'ADN d'autres espèces proches: *Glomus* sp. DAOM-229456, *G. fasciculatum* DAOM-240159, *G. clarum* DAOM-240429 and DAOM-234281, *Glomus* sp. DAOM-240422**,** *G. aggregatum* DAOM-240163. Seulement G. irregulare DAOM-197198 a été amplifié avec succès avec le marqueur spécifique.

L'utilisation de ces marqueurs moléculaires mitochondriaux est très prometteuse, tout particulièrement en raison de leur capacité de faire la distinction entre deux isolats de la même espèce, mais plusieurs problématiques possibles doivent être prises en considération. En effet, ma thèse de doctorat a permis de situer dans un contexte évolutif et biologique l'utilisation de marqueurs moléculaires mitochondriaux dans une optique d'identification et de quantification des CMA, ce qui soulève plusieurs interrogations en ce qui a trait à l'utilisation judicieuse de tels marqueurs. En premier lieu, le premier aspect pouvant poser problème est au niveau de

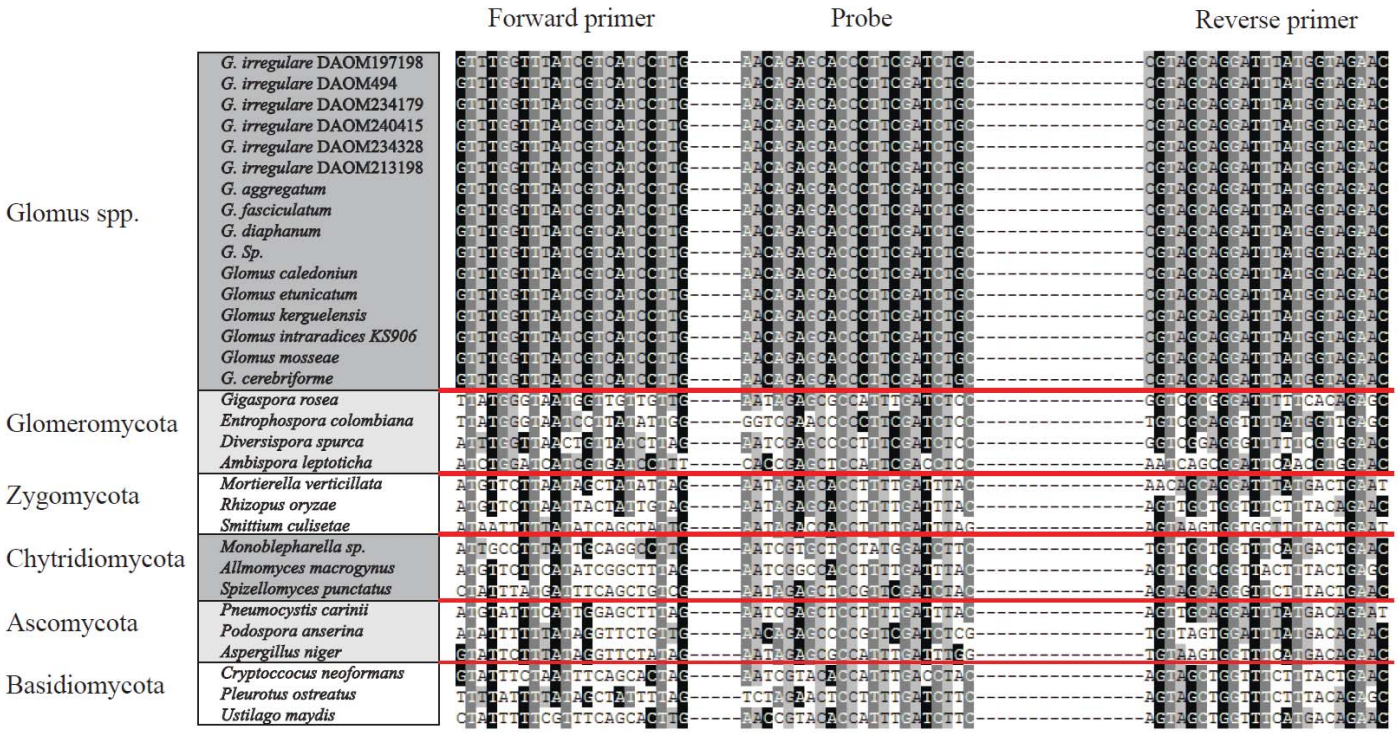


Figure 6.4. Développement et d'un marqueur généralistique, spécifique au genre *Glomus*. Alignement d'une portion du gène *nad1* où le marqueur a été déterminé. L'alignement a été effectué en utilisant six isolats distincts de *G. irregulare*, avec 10 espèces du genre *Glomus*, en plus d'autres familles de Gloméromycètes, ainsi que de différents phylums du règne des Champignons. La spécificité du marqueur a aussi été testée par BLASTn sur la base de données GenBank.

l'identification puisque les régions intergéniques où ces marqueurs sont développés contiennent plusieurs insertions de *dpo*. Ces éléments semblent se propager rapidement dans tous le phylum des Gloméromycètes, puisque l'on en trouve dans tous les genres de CMA séquencés jusqu'à maintenant (Nadimi *et al.*, 2012; Pelin *et al.*, 2012), mais aussi chez d'autres champignons. Non seulement leur distribution semble ubiquitaire, mais des copies similaires sont dupliquées dans différentes régions mitochondriales intra- et/ou inter-isolats. Cela peut engendrer facilement une non-spécificité des marqueurs, soit en raison de la présence d'une copie paralogue, orthologue, ou de recombinaison *in vitro*/séquence chimérique. Pour ces raisons, les marqueurs potentiels doivent être testés rigoureusement sur un grand nombre d'espèces génétiquement proches et

isolats de la même espèce, avec un nombre de cycles PCR approprié (généralement plus de 35 cycles) ou par qPCR. Même avec une procédure rigoureuse en laboratoire, il sera difficile d'évaluer la spécificité d'un tel marqueur dans l'environnement.

Une autre problématique potentielle de l'utilisation des marqueurs mitochondriaux est au niveau de la quantification par qPCR. Cette méthodologie est basée sur le nombre de copies de génome mitochondrial contenu dans une spore d'une souche de CMA donnée. Cette méthode peut théoriquement être efficace dans un milieu contrôlé où la souche en question n'est pas soumise à des stress extérieurs variables. Cependant, il a été démontré que la simple présence de strigolactone, une phytohormone sécrétée par les plantes, a un puissant effet sur l'activité mitochondriale des CMA (Besserer *et al.*, 2006). Dans un milieu naturel, on pourrait donc s'attendre à ce que le nombre de copies d'ADN mitochondrial par spore soit très variable en fonction des conditions environnantes. De plus, on a démontré la présence de plusieurs copies d'ADN mitochondrial dans le génome nucléaire (numts) de l'espèce *G. irregulare* (Chapitre 5). La présence de ces numts ajoute une incertitude quant à la quantification du nombre de copies mitochondriales, puisque le résultat global va être influencé par les duplications présentes du marqueur utilisé. Puisque le processus de transfert d'ADN mitochondrial vers le génome nucléaire est très dynamique (Hazkani-Covo *et al.*, 2010), il serait possible que les régions/gènes transférées varient en fonction de l'espèce, voire même de l'isolat. Cela ferait en sorte qu'un marqueur généraliste à un groupe de CMA ne pourrait être utilisé pour leur quantification individuelle au niveau de l'espèce, puisque les données auraient la possibilité d'être biaisées par la présence de ces copies nucléaires.

6.4 Les frontières de l'individualité chez les CMA: impact sur l'utilisation de marqueurs moléculaires

L'interprétation des données obtenues dans des études de génétiques de population et/ou de diversité des CMA peut être difficile en raison du mode de vie et de l'organisation génétique de ces champignons énigmatiques. Même si la controverse demeure concernant la présence d'un état homokaryotique ou hétérokaryotique intrinsèque au sein d'un isolat de CMA, il a été démontré que des isolats génétiquement distincts peuvent échanger du matériel génétique nucléaire et/ou mitochondrial (Croll *et al.*, 2009; de la Providencia *et al.*, 2013). Cela permet d'engendrer une progéniture arborant un génotype mosaïque. Il a même été démontré que les échanges génétiques entre isolats ainsi que la ségrégation génétique engendraient des modifications phénotypiques sur les isolats de CMA résultants, puisqu'ils avaient un effet variable sur le fitness de leur plante hôte (Angelard *et al.*, 2010; Angelard & Sanders, 2011). Ces données expérimentales, combinés aux connaissances apportées par cette thèse de doctorat en ce qui a trait à la dynamique mitochondriale et la connaissance du mode de vie promiscuitaire des CMA (formant un réseau de mycélium commun à l'aide d'anastomoses), suggèrent que l'hétérokaryotie et l'hétéroplasmie sont des états intrinsèques dans les populations naturelles. Cependant, il semble y avoir présence de mécanismes de ségrégation et de dérive génétique lorsqu'un isolat est sous-cultivé *in vitro* de nombreuses générations. Cela a le potentiel d'engendrer des situations conflictuelles lors de l'utilisation de marqueurs moléculaires (Figure 6.5).

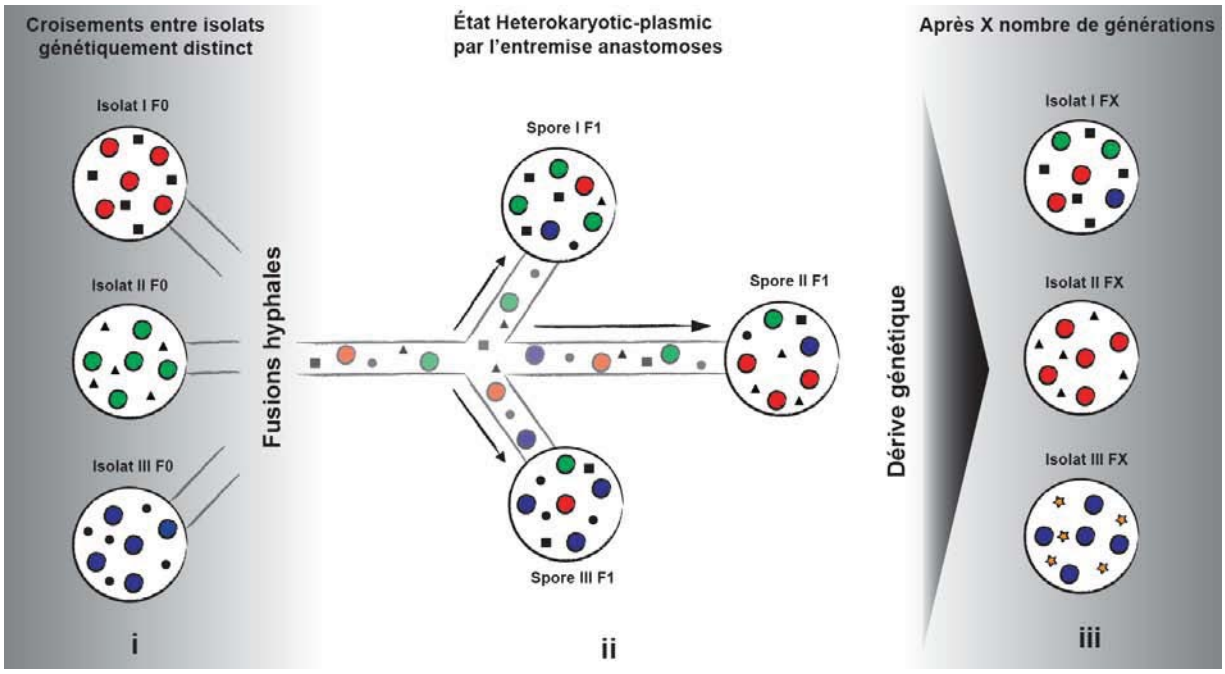


Figure 6.5. Représentation schématique de croisements entre des isolats de CMA génétiquement divergents (isolat I, II and III F0). Chaque isolat de départ possède un haplotype mitochondrial (symboles noirs) et un génotype (cercles colorés) unique et homogène.

À titre d'exemple, suite à des fusions hyphales entre des isolats génétiquement distincts, mais homogènes au niveau de leur génome mitochondrial et nucléaire (Figure 6.5 i), le réseau de mycélium commun engendre des spores filles hetero-karyotiques et -plasmiques (spore I, II and III F1) (Figure 6.5 ii). Ces spores sont alors sous-cultivées pour un nombre X de générations et le génotype hypothétique résultant d'une dérive et ségrégation génétique est représenté (isolate I, II and III FX) (Figure 6.5 iii). Le symbole en forme d'étoile représente un haplotype mitochondrial recombinant qui pourrait éventuellement être créé tel que démontré précédemment (Beaudet *et al.*, 2013b). Les génotypes obtenus provenant des croisements entre isolats génétiquement distincts, suivis par de la dérive et ségrégation génétique, peuvent engendrer des états difficiles à

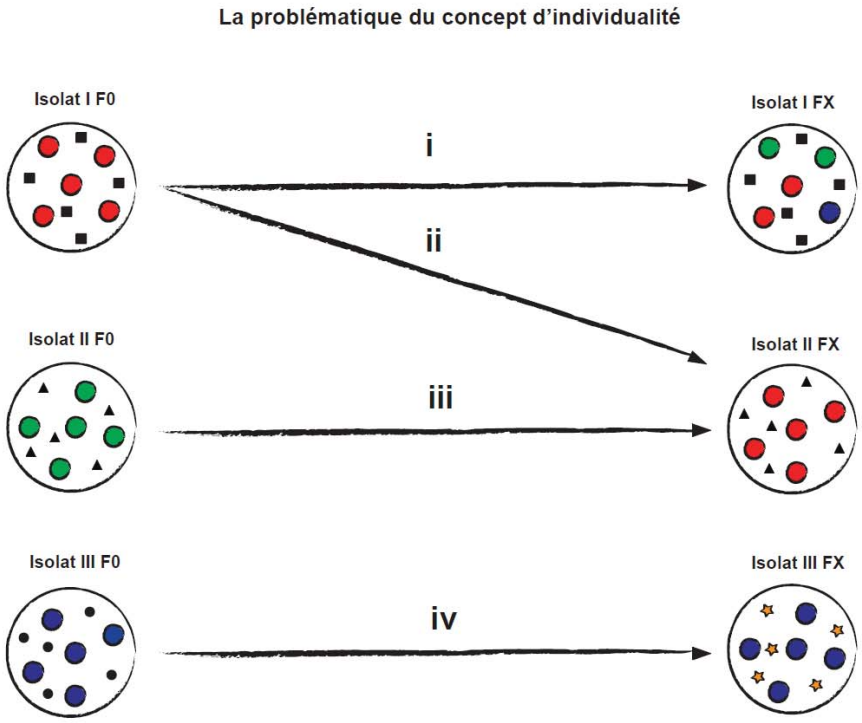


Figure 6.6. Exemples de cas où la résolution du concept d'individu est problématique chez les CMA, tant pour les marqueurs nucléaires que mitochondriaux.

résoudre avec des marqueurs moléculaires. Deux isolats pourraient avoir le même haplotype mitochondrial homogène, avec un génotype nucléaire mosaïque (I F0 and I FX) (Figure 6.6 i). Un isolat pourrait partager le même génotype nucléaire avec un autre (I F0 and II FX) (Figure 6.6 ii), alors que ce dernier partagerait son haplotype mitochondrial avec un autre isolat (II F0 and II FX) (Figure 6.6 iii). Finalement, un isolat pourrait avoir le même génotype nucléaire qu'un autre, mais posséder un ADN mitochondrial recombinant unique (Figure 6.6 iv).

Le concept de population est grandement utilisé en écologie, en génétique des populations et en biologie évolutive. De nombreuses définitions peuvent être trouvées, mais le thème récurrent est qu'une population sert à décrire un ensemble d'individus d'une certaine espèce. On possède très peu de connaissances sur ce qui constitue un individu chez les CMA (Rosendahl, 2008). Chez les

autres groupes de champignons, il n'existe pas non plus un consensus dans la communauté scientifique sur ce qui constitue un individu. Cela est dû en partie au fait qu'un individu peut être défini génétiquement ou physiologiquement. Les situations conflictuelles hypothétiques présentées plus hautes démontrent la complexité de définir l'individualité génétique de ces organismes lorsqu'ils sont dans leur milieu naturel, tant par l'entremise de marqueurs mitochondriaux que nucléaires. Cette individualité est dynamique et constamment modifiée par les interactions fusionnelles avec des isolats compatibles. Dans ce contexte, l'utilisation de marqueurs moléculaires mitochondriaux pourrait être utiles pour distinguer des groupes de CMA constituant des entités de compatibilité d'anastomoses distinctes (groupes de compatibilité par anastomoses) comme c'est le cas chez *Rhizoctonia solani* (Bolkan, 1985; Guillemaut *et al.*, 2003), par exemple au niveau de l'espèce. Pour ce qui est d'identifier un isolat, cela devrait demeurer un concept artificiel en laboratoire et se limiter à des fins pratiques et conceptuelles.

Chapitre 7 - Conclusion générale et perspectives

La génomique évolutive mitochondriale des CMA a permis de révéler les vecteurs de divergences intra- et inter-spécifiques, en plus de cibler les régions propices au développement de marqueurs moléculaires mitochondriaux. L'observation de signatures moléculaires typiques aux échanges génétiques inter-espèces et de recombinaison mitochondriale entre haplotypes génétiquement distincts, a permis d'ouvrir les perspectives sur la dynamique mitochondriale et l'hétéroplasmie dans un même isolat. De plus, cette apparente contradiction avec l'homogénéité mitochondriale intra-isolat généralement observée a amené à investiguer les échanges génétiques à l'aide de croisements d'isolats génétiquement distincts et de décrire des mécanismes de

ségrégation moléculaire orthologues à ceux retrouvés chez les levures et les champignons filamenteux. Finalement, l'observation d'hétéroplasmie *in situ*, ainsi que la co-expression de variantes de protéines dans une souche *in vitro* a amené le questionnement du concept d'individualité chez ce groupe particulier de champignons.

En raison de la disponibilité limitée de matériel biologique et de problème de contamination, il a été difficile d'obtenir des génomes provenant d'espèces distantes de CMA. Il serait très informatif d'être en mesure de séquencer le génome mitochondrial d'espèces de CMA représentant les différentes grandes familles de Gloméromycètes. L'obtention de ces séquences permettrait entre autres un positionnement phylogénétique plus précis au sein du grand groupe des champignons. Il serait aussi intéressant de déterminer la corrélation entre le nombre de copies de génomes mitochondriaux, l'expression des gènes mitochondriaux et l'activité fonctionnelle de gènes impliqués dans les fonctions symbiotiques, comme le transport du phosphate, puisqu'il s'agit de mécanismes de transport actif. De plus, il sera important de déterminer l'étendue de la présence de numts dans les génomes nucléaires des CMA, mais surtout de voir si leur présence varie en fonction de l'espèce ou de l'isolat étudier. Cette dernière éventualité aurait un impact majeur sur l'utilisation de marqueurs mitochondriaux pour l'identification et la quantification d'inoculum endomycorhizien. Finalement, cette thèse de doctorat aura permis d'établir les bases, tant fondamentales qu'appliquées, pour l'élaboration et l'utilisation d'un kit destiné au suivi d'application des CMA en agriculture. D'autres études devront être effectuées afin de déterminer les meilleures stratégies possible d'applications, en tenant compte des différents facteurs liés à l'environement, la physiologie et la génétique de ces organismes particuliers et fascinants.

Bibliographie

Akins RA, Grant DM, Stohl LL, Bottorff DA, Nargang FE, Lambowitz AM. 1988. Nucleotide sequence of the Varkud mitochondrial plasmid of Neurospora and synthesis of a hybrid transcript with a leader derived from mitochondrial RNA. *Journal of Molecular Biology* **204**(1): 1-25.

Akins RA, Kelley RL, Lambowitz AM. 1986. Mitochondrial plasmids of neurospora: Integration into mitochondrial DNA and evidence for reverse transcription in mitochondria. *Cell* **47**(4): 505-516.

Almasan A, Mishra NC. 1991. Recombination by sequence repeats with formation of suppressive or residual mitochondrial DNA in Neurospora. *Proceedings of the National Academy of Sciences* **88**(17): 7684-7688.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**(3): 403-410.

André C, Levy A, Walbot V. 1992. Small repeated sequences and the structure of plant mitochondrial genomes. *Trends in Genetics* **8**(4): 128-132.

Angelard C, Colard A, Niculita-Hirzel H, Croll D, Sanders IR. 2010. Segregation in a Mycorrhizal Fungus Alters Rice Growth and Symbiosis-Specific Gene Transcription. *Current Biology* **20**(13): 1216-1221.

Angelard C, Sanders IR. 2011. Effect of segregation and genetic exchange on arbuscular mycorrhizal fungi in colonization of roots. *New Phytologist* **189**(3): 652-657.

Arganoza MT, Min J, Hu Z, Akins RA. 1994. Distribution of seven homology groups of mitochondrial plasmids in Neurospora: evidence for widespread mobility between species in nature. *Current Genetics* **26**(1): 62-73.

Azcón-Aguilar C, Barea JM. 1997. Arbuscular mycorrhizas and biological control of soil-borne plant pathogens – an overview of the mechanisms involved. *Mycorrhiza* **6**(6): 457-464.

Bader G, Hogue C. 2003. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4**(1): 2.

Bago B, Azcón-Aguilar C, Piché Y. 1998. Architecture and Developmental Dynamics of the External Mycelium of the Arbuscular Mycorrhizal Fungus Glomus intraradices Grown under Monoxenic Conditions. *Mycologia* **90**(1): 52-62.

Bago B, Zipfel W, Williams R, Piché Y. 1999. Nuclei of symbiotic arbuscular mycorrhizal fungi as revealed by in vivo two-photon microscopy. *Protoplasma* **209**(1-2): 77-89.

Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. *Molecular Ecology* **13**(4): 729-744.

Bapteste E, Lopez P, Bouchard F, Baquero F, McInerney JO, Burian RM. 2012. Evolutionary analyses of non-genealogical bonds produced by introgressive descent. *Proceedings of the National Academy of Sciences* **109**(45): 18266-18272.

Barr CM, Neiman M, Taylor DR. 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytologist* **168**(1): 39-50.

Bateman JM, Perlman PS, Butow RA. 2002. Mutational bisection of the mitochondrial DNA stability and amino acid biosynthetic functions of ilv5p of budding yeast. *Genetics* **161**(3): 1043-1052.

Beaudet D, Nadimi M, Iffis B, Hijri M. 2013a. Rapid Mitochondrial Genome Evolution through Invasion of Mobile Elements in Two Closely Related Species of Arbuscular Mycorrhizal Fungi. *PLoS ONE* **8**(4): e60768.

Beaudet D, Terrat Y, Halary S, de la Providencia IE, Hijri M. 2013b. Mitochondrial Genome Rearrangements in Glomus Species Triggered by Homologous Recombination between Distinct mtDNA Haplotypes. *Genome Biology and Evolution* **5**(9): 1628-1643.

Bécard G, Fortin JA. 1988. Early events of vesicular–arbuscular mycorrhiza formation on Ri T-DNA transformed roots. *New Phytologist* **108**(2): 211-218.

Bécard G, Kosuta S, Tamasloukht M, Séjalon-Delmas N, Roux C. 2004. Partner communication in the arbuscular mycorrhizal interaction. *Canadian Journal of Botany* **82**(8): 1186-1197.

Bell-Pedersen D, Quirk S, Clyman J, Belfort M. 1990. Intron mobility in phage T4 is dependent upon a distinctive class of endonucleases and independent of DNA sequences encoding the intron core: mechanistic and evolutionary implications. *Nucleic Acids Research* **18**(13): 3763-3770.

Bensasson D, Zhang D, Hartl DL, Hewitt GM. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology and Evolution* **16**(6): 314-321.

Berger KH, Yaffe MP. 2000. Mitochondrial DNA inheritance in Saccharomyces cerevisiae. *Trends in microbiology* **8**(11): 508-513.

Bertrand H, Griffiths AJF. 1989. Linear plasmids that integrate into mitochondrial DNA in Neurospora. *Genome* **31**: 155-159.

Besserer A, Puech-Pagès V, Kiefer P, Gomez-Roldan V, Jauneau A, Roy S, Portais J-C, Roux C, Bécard G, Séjalon-Delmas N. 2006. Strigolactones Stimulate Arbuscular Mycorrhizal Fungi by Activating Mitochondria. *PLoS Biology* **4**(7): e226.

Bever JD, Wang M. 2005. Arbuscular mycorrhizal fungi: Hyphal fusion and multigenomic structure. *Nature* **433**(7022): E3-E4.

Bi X, Liu LF. 1996. DNA rearrangement mediated by inverted repeats. *Proceedings of the National Academy of Sciences* **93**(2): 819-823.

Biessmann H, Valgeirsdottir K, Lofsky A, Chin C, Ginther B, Levis RW, Pardue ML. 1992. HeT-A, a transposable element specifically involved in "healing" broken chromosome ends in Drosophila melanogaster. *Molecular and Cellular Biology* **12**(9): 3910-3918.

Birky CW. 2001. The inheritance of genes in mitochondria and chloroplasts: Laws, Mechanisms, and Models. *Annual Review of Genetics* **35**(1): 125-148.

Birky CW, Strausberg RL, Forster JL, Perlman PS. 1978. Vegetative segregation of mitochondria in yeast: Estimating parameters using a random model. *Molecular and General Genetics MGG* **158**(3): 251-261.

Boer PH, Gray MW. 1991. Short dispersed repeats localized in spacer regions of Chlamydomonas reinhardtii mitochondrial DNA. *Current Genetics* **19**(4): 309-312.

Boldogh I, Vojtov N, Karmon S, Pon LA. 1998. Interaction between mitochondria and the actin cytoskeleton in budding yeast requires two integral mitochondrial outer membrane proteins, Mmm1p and Mdm10p. *The Journal of Cell Biology* **141**(6): 1371-1381.

Boldogh IR, Nowakowski DW, Yang H-C, Chung H, Karmon S, Royes P, Pon LA. 2003. A protein complex containing Mdm10p, Mdm12p, and Mmm1p links mitochondrial membranes and DNA to the cytoskeleton-based segregation machinery. *Molecular biology of the cell* **14**(11): 4618-4627.

Bolkan HA. 1985. Anastomosis groups and pathogenicity of Rhizoctonia solani isolates from Brazil. *Plant Disease* **69**(7): 599-601.

Bonfante P, Genre A. 2008. Plants and arbuscular mycorrhizal fungi: an evolutionary-developmental perspective. *Trends in Plant Science* **13**(9): 492-498.

Boon E. 2012. *The evolution of inter-genomic variation in arbuscular mycorrhizal fungi.* Article thesis, University of Montreal Montreal.

Boon E, Zimmerman E, Lang BF, Hijri M. 2010. Intra-isolate genome variation in arbuscular mycorrhizal fungi persists in the transcriptome. *Journal of Evolutionary Biology* **23**(7): 1519-1527.

Boon E, Zimmerman E, St-Arnaud M, Hijri M. 2013. Allelic Differences within and among Sister Spores of the Arbuscular Mycorrhizal Fungus Glomus etunicatum Suggest Segregation at Sporulation. *PLoS ONE* **8**(12): e83301.

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**(2): 452-465.

Boursot P, Yonekawa H, Bonhomme F. 1987. Heteroplasmy in mice with deletion of a large coding region of mitochondrial DNA. *Molecular Biology and Evolution* **4**(1): 46-55.

Brown GG, Zhang M 1995. Mitochondrial plasmids: DNA and RNA. In: Levings CS III VIe ed. *The molecular biology of plant mitochondria*. Dordrecht: Kluwer, 61-91.

Brügger K, Torarinsson E, Redder P, Chen L, Garrett R. 2004. Shuffling of Sulfolobus genomes by autonomous and non-autonomous mobile elements. *Biochemical Society Transactions* **32**: 179-183.

Brundrett MC. 2002. Coevolution of roots and mycorrhizas of land plants. *New Phytologist* **154**(2): 275-304.

Bucher M, Wegmüller S, Drissner D. 2009. Chasing the structures of small molecules in arbuscular mycorrhizal signaling. *Current Opinion in Plant Biology* **12**(4): 500-507.

Burger G, Gray MW, Forget L, Lang BF. 2013. Strikingly Bacteria-Like and Gene-Rich Mitochondrial Genomes throughout Jakobid Protists. *Genome Biology and Evolution* **5**(2): 418-438.

Cahan P, Kennell J. 2005. Identification and distribution of sequences having similarity to mitochondrial plasmids in mitochondrial genomes of filamentous fungi. *Molecular Genetics and Genomics* **273**(6): 462-473.

Cardenas-Flores A, Draye X, Bivort C, Cranenbrouck S, Declerck S. 2010. Impact of multispores in vitro subcultivation of Glomus sp. MUCL 43194 (DAOM 197198) on vegetative compatibility and genetic diversity detected by AFLP. *Mycorrhiza* **20**(6): 415-425.

Cavagnaro TR, Smith FA, Smith SE, Jakobsen I. 2005. Functional diversity in arbuscular mycorrhizas: exploitation of soil patches with different phosphate enrichment differs among fungal species. *Plant, Cell & Environment* **28**(5): 642-650.

Ceballos I, Ruiz M, Fernández C, Peña R, Rodríguez A, Sanders IR. 2013. The in vitro mass-produced model mycorrhizal fungus, Rhizophagus irregularis, significantly increases yields of the globally important food security crop cassava. *PLoS ONE* **8**(8): e70633.

Chao A. 1984. Nonparametric Estimation of the Number of Classes in a Population. *Scandinavian Journal of Statistics* **11**(4): 265-270.

Chao A, Gotelli N, Hsieh TC, Sander E, Ma KH, Colwell RK, Ellison AM. 2013. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs*.

Chao A, Jost L. 2012. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* **93**(12): 2533-2547.

Chen XJ, Butow RA. 2005. The organization and inheritance of the mitochondrial genome. *Nature Reviews Genetics* **6**(11): 815-825.

Chen XJ, Wang X, Kaufman BA, Butow RA. 2005. Aconitase couples metabolic regulation to mitochondrial DNA maintenance. *Science* **307**(5710): 714-717.

Cho Y, Qiu Y-L, Kuhlman P, Palmer JD. 1998. Explosive invasion of plant mitochondria by a group I intron. *Proceedings of the National Academy of Sciences* **95**(24): 14244-14249.

Clapp J, Helgason T, Daniell T, Peter J, Young W 2003. Genetic Studies of the Structure and Diversity of Arbuscular Mycorrhizal Fungal Communities. In: Heijden MA, Sanders I eds. *Mycorrhizal Ecology*: Springer Berlin Heidelberg, 201-224.

Clapp JP, Fitter AH, Young JPW. 1999. Ribosomal small subunit sequence variation within spores of an arbuscular mycorrhizal fungus, Scutellospora sp. *Molecular Ecology* **8**(6): 915-921.

Clapp JP, Rodriguez A, Dodd JC. 2001. Inter- and intra-isolate rRNA large subunit variation in Glomus coronatum spores. *New Phytologist* **149**(3): 539-554.

Colard A, Angelard C, Sanders IR. 2011. Genetic Exchange in an Arbuscular Mycorrhizal Fungus Results in Increased Rice Growth and Altered Mycorrhiza-Specific Gene Transcription. *Applied and Environmental Microbiology* **77**(18): 6510-6515.

Collins RA, Saville BJ. 1990. Independent transfer of mitochondrial chromosomes and plasmids during unstable vegetative fusion in Neurospora. *Nature* **345**(6271): 177-179.

Cordier C, Pozo MJ, Barea JM, Gianinazzi S, Gianinazzi-Pearson V. 1998. Cell Defense Responses Associated with Localized and Systemic Resistance to Phytophthora parasitica Induced in Tomato by an Arbuscular Mycorrhizal Fungus. *Molecular Plant-Microbe Interactions* **11**(10): 1017-1028.

Corradi N, Bonen L. 2012. Mitochondrial genome invaders: an unselfish role as molecular markers. *New Phytologist* **196**(4): 963-965.

Corradi N, Croll D, Colard A, Kuhn G, Ehinger M, Sanders IR. 2007. Gene Copy Number Polymorphisms in an Arbuscular Mycorrhizal Fungal Population. *Applied and Environmental Microbiology* **73**(1): 366-369.

Corradi N, Kuhn G, Sanders IR. 2004. Monophyly of β-tubulin and H+-ATPase gene variants in Glomus intraradices: consequences for molecular evolutionary studies of AM fungal genes. *Fungal Genetics and Biology* **41**(2): 262-273.

Corradi N, Ruffner B, Croll D, Colard A, Horak A, Sanders IR. 2009. High-Level Molecular Diversity of Copper-Zinc Superoxide Dismutase Genes among and within Species of Arbuscular Mycorrhizal Fungi. *Appl. Environ. Microbiol.* **75**(7): 1970-1978.

Corradi N, Sanders I. 2006. Evolution of the P-type II ATPase gene family in the fungi and presence of structural genomic changes among isolates of Glomus intraradices. *BMC Evolutionary Biology* **6**(1): 21.

Court DA, Bertrand H. 1992. Genetic organization and structural features of maranhar; a senescence-inducing linear mitochondrial plasmid of Neurospora crassa. *Current Genetics* **22**(5): 385-397.

Cranenbrouck S, Voest L, Bivort C, Renard L, Strullu D, Declerck S. 2005. Methodologies for in vitro cultivation of arbuscular mycorrhizal fungi with root organs.: In: Declerck S, Strullu DG, Fortin JA (eds). In vitro culture of mycorrhizas. Springer-Verlag: pp 341-375.

Croll D, Giovannetti M, Koch AM, Sbrana C, Ehinger M, Lammers PJ, Sanders IR. 2009. Nonself vegetative fusion and genetic exchange in the arbuscular mycorrhizal fungus Glomus intraradices. *New Phytologist* **181**(4): 924-937.

Dalgaard JZ, Garrett RA, Belfort M. 1993. A site-specific endonuclease encoded by a typical archaeal intron. *Proceedings of the National Academy of Sciences* **90**(12): 5414-5417.

Datnoff LE, Nemec S, Pernezny K. 1995. Biological Control of Fusarium Crown and Root Rot of Tomato in Florida Using Trichoderma harzianum and Glomus intraradices. *Biological Control* **5**(3): 427-431.

de la Providencia IE, Nadimi M, Beaudet D, Rodriguez Morales G, Hijri M. 2013. Detection of a transient mitochondrial DNA heteroplasmy in the progeny of crossed genetically divergent isolates of arbuscular mycorrhizal fungi. *New Phytologist* **200**(1): 211-221.

de Zamaroczy M, Bernardi G. 1986. The GC clusters of the mitochondrial genome of yeast and their evolutionary origin. *Gene* **41**(1): 1-22.

Debets F, Yang X, Griffiths AF. 1994. Vegetative incompatibility in Neurospora: its effect on horizontal transfer of mitochondrial plasmids and senescence in natural populations. *Current Genetics* **26**(2): 113-119.

Denovan-Wright EM, Lee RW. 1994. Comparative Structure and Genomic Organization of the Discontinuous Mitochondrial Ribosomal RNA Genes of Chlamydomonas eugametos and Chlamydomonas reinhardtii. *Journal of Molecular Biology* **241**(2): 298-311.

Dimmer KS, Jakobs S, Vogel F, Altmann K, Westermann B. 2005. Mdm31 and Mdm32 are inner membrane proteins required for maintenance of mitochondrial shape and stability of mitochondrial DNA nucleoids in yeast. *The Journal of Cell Biology* **168**(1): 103-115.

Dimmic MW, Rest JS, Mindell DP, Goldstein RA. 2002. rtREV: An Amino Acid Substitution Matrix for Inference of Retrovirus and Reverse Transcriptase Phylogeny. *Journal of Molecular Evolution* **55**(1): 65-73.

Doner L, Bécard G. 1991. Solubilization of gellan gels by chelation of cations. *Biotechnology Techniques* **5**: 25-28.

Dujon B. 1980. Sequence of the intron and flanking exons of the mitochondrial 21S rRNA gene of yeast strains having different alleles at the co and rib-1 loci. *Cell* **20**(1): 185-197.

Eddy S. 1992. *Introns in the T-seven bacteriophages.* Dissertation thesis, University of Colorado Boulder.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**(5): 1792-1797.

Etherington GJ, Dicks J, Roberts IN. 2005. Recombination Analysis Tool (RAT): a program for the high-throughput detection of recombination. *Bioinformatics* **21**(3): 278-281.

Evelin H, Kapoor R, Giri B. 2009. Arbuscular mycorrhizal fungi in alleviation of salt stress: a review. *Annals of Botany* **104**(7): 1263-1280.

Felsenstein J. 1978. Cases in which Parsimony or Compatibility Methods Will be Positively Misleading. *Systematic Zoology* **27**(4): 401-410.

Ferandon C, Chatel SEK, Castandet B, Castroviejo M, Barroso G. 2008. The Agrocybe aegerita mitochondrial genome contains two inverted repeats of the nad4 gene arisen by duplication on both sides of a linear plasmid integration site. *Fungal Genetics and Biology* **45**(3): 292-301.

Findlay I, Urquhart A, Quirke P, Sullivan K, Rutherford AJ, Litford RJ. 1995. Simultaneous DNA 'fingerprinting', diagnosis of sex and single-gene defect status from single cells. *Molecular Human Reproduction* **1**(2): 85-93.

Formey D, Molès M, Haouy A, Savelli B, Bouchez O, Bécard G, Roux C. 2012. Comparative analysis of mitochondrial genomes of Rhizophagus irregularis – syn. Glomus irregulare – reveals a polymorphism induced by variability generating elements. *New Phytologist* **196**(4): 1217-1227.

Franke M, Morton J. 1994. Ontogenetic comparisons of arbuscular mycorrhizal fungi Scutellospora heterogama and Scutellospora pellucida: revision of taxonomic character concepts, species descriptions, and phylogenetic hypotheses. *Canadian Journal of Botany* **72**(1): 122-134.

Fricova D, Valach M, Farkas Z, Pfeiffer I, Kucsera J, Tomaska L, Nosek J. 2010. The mitochondrial genome of the pathogenic yeast Candida subhashii: GC-rich linear DNA with a protein covalently attached to the 5′ termini. *Microbiology* **156**(7): 2153-2163.

Friese CF, Allen MF. 1991. The spread of VA mycorrhizal fungal hyphae in the soil: inoculum types and external hyphal architecture. *Mycologia* **83**(4): 409-418.

Gamper HA, van der Heijden MGA, Kowalchuk GA. 2010. Molecular trait indicators: moving beyond phylogeny in arbuscular mycorrhizal ecology. *New Phytologist* **185**(1): 67-82.

Gehrig H, Schüßler A, Kluge M. 1996. Geosiphon pyriforme, a fungus forming endocytobiosis withNostoc (Cyanobacteria), is an ancestral member of the glomales: Evidence by SSU rRNA Analysis. *Journal of Molecular Evolution* **43**(1): 71-81.

Genre A, Chabaud M, Faccio A, Barker DG, Bonfante P. 2008. Prepenetration apparatus assembly precedes and predicts the colonization patterns of arbuscular mycorrhizal fungi within the root cortex of both Medicago truncatula and Daucus carota. *The Plant Cell Online* **20**(5): 1407-1420.

Gianinazzi S, Gollotte A, Binet M-N, Tuinen D, Redecker D, Wipf D. 2010. Agroecology: the key role of arbuscular mycorrhizas in ecosystem services. *Mycorrhiza* **20**(8): 519-530.

Gilbert N. 2009. The disappearing nutrient. *Nature* **461**(8): 716-718.

Giovannetti M, Azzolini D, Citernesi AS. 1999. Anastomosis Formation and Nuclear and Protoplasmic Exchange in Arbuscular Mycorrhizal Fungi. *Applied and Environmental Microbiology* **65**(12): 5571-5575.

Giovannetti M, Sbrana C, Avio L, Strani P. 2004. Patterns of below-ground plant interconnections established by means of arbuscular mycorrhizal networks. *New Phytologist* **164**(1): 175-181.

Glass NL, Fleissner A 2006. Re-wiring the network: understanding the mechanism and function of anastomosis in filamentous ascomycete fungi. *Growth, Differentiation and Sexuality*: Springer, 123-139.

Good IJ. 1953. The population frequencies of species and the estimation of population parameters. *Biometrika* **40**(3-4): 237-264.

Griffiths AF, Kraus S, Barton R, Court D, Myers C, Bertrand H. 1990. Heterokaryotic transmission of senescence plasmid DNA in Neurospora. *Current Genetics* **17**(2): 139-145.

Griffiths AJ. 1995. Natural plasmids of filamentous fungi. *Microbiological Reviews* **59**(4): 673-685.

Griffiths AJF. 1992. Fungal Senescence. *Annual Review of Genetics* **26**(1): 351-372.

Griffiths AJF, Yang X. 1995. Recombination between heterologous linear and circular mitochondrial plasmids in the fungus *Neurospora*. *Molecular and General Genetics* **249**(1): 25-36.

Grindley NDF, Reed RR. 1985. Transpositional Recombination in Prokaryotes. *Annual Review of Biochemistry* **54**(1): 863-896.

Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. 2008. The Vienna RNA Websuite. *Nucleic Acids Research* **36**(suppl 2): W70-W74.

Guillemaut C, Edel-Hermann V, Camporota P, Alabouvette C, Richard-Molard M, Steinberg C. 2003. Typing of anastomosis groups of Rhizoctonia solani by restriction analysis of ribosomal DNA. *Canadian journal of microbiology* **49**(9): 556-568.

Halary S, Malik S-B, Lildhar L, Slamovits CH, Hijri M, Corradi N. 2011. Conserved Meiotic Machinery in Glomus spp., a Putatively Ancient Asexual Fungal Lineage. *Genome Biology and Evolution* **3**: 950-958.

Halary S, Mclnerney JO, Lopez P, Bapteste E. 2013. EGN: a wizard for construction of gene and genome similarity networks. *BMC Evolutionary Biology* **13**(1): 146.

Handa H. 2003. The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (Brassica napus L.): comparative analysis of the mitochondrial genomes of rapeseed and Arabidopsis thaliana. *Nucleic Acids Research* **31**(20): 5907-5916.

Handa H. 2007. Investigation of the origin and transmission of linear mitochondrial plasmid based on phylogenetic analysis in Japanese rapeseed varieties. *Genome* **50**(2): 234-240.

Handa H. 2008. Linear plasmids in plant mitochondria: Peaceful coexistences or malicious invasions? *Mitochondrion* **8**(1): 15-25.

Hänfler J, Teepe H, Weigel C, Kruft V, Lurz R, Wöstemeyer J. 1992. Circular extrachromosomal DNA codes for a surface protein in the (+) mating type of the zygomycete Absidia glauca. *Current Genetics* **22**(4): 319-325.

Harrison MJ, Dewbre GR, Liu J. 2002. A phosphate transporter from Medicago truncatula involved in the acquisition of phosphate released by arbuscular mycorrhizal fungi. *The Plant Cell Online* **14**(10): 2413-2429.

Hart M, Reader R. 2002. Host plant benefit from association with arbuscular mycorrhizal fungi: variation due to differences in size of mycelium. *Biology and Fertility of Soils* **36**(5): 357-366.

Hausner G 2012. Introns, Mobile Elements, and Plasmids Organelle Genetics. In: Bullerwell CE ed.: Springer Berlin Heidelberg, 329-357.

Hazkani-Covo E, Zeller RM, Martin W. 2010. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS genetics* **6**(2): e1000834.

Helber N, Requena N. 2008. Expression of the fluorescence markers DsRed and GFP fused to a nuclear localization signal in the arbuscular mycorrhizal fungus Glomus intraradices. *New Phytologist* **177**(2): 537-548.

Helgason T, Watson IJ, Young JPW. 2003. Phylogeny of the Glomerales and Diversisporales (Fungi: Glomeromycota) from actin and elongation factor 1-alpha sequences. *FEMS Microbiology Letters* **229**(1): 127-132.

Hermanns J, Asseburg A, Osiewacz HD. 1994. Evidence for a life span-prolonging effect of a linear plasmid in a longevity mutant of Podospora anserina. *Molecular and General Genetics MGG* **243**(3): 297-307.

Hijri M, Hosny M, van Tuinen D, Dulieu H. 1999. Intraspecific ITS Polymorphism inScutellospora castanea(Glomales, Zygomycota) Is Structured within Multinucleate Spores. *Fungal Genetics and Biology* **26**(2): 141-151.

Hijri M, Sanders IR. 2004. The arbuscular mycorrhizal fungus Glomus intraradices is haploid and has a small genome size in the lower limit of eukaryotes. *Fungal Genetics and Biology* **41**(2): 253-261.

Hijri M, Sanders IR. 2005. Low gene copy number shows that arbuscular mycorrhizal fungi inherit genetically different nuclei. *Nature* **433**(7022): 160-163.

Hobbs AEA, Srinivasan M, McCaffery JM, Jensen RE. 2001. Mmm1p, a mitochondrial outer membrane protein, is connected to mitochondrial DNA (mtDNA) nucleoids and required for mtDNA stability. *The Journal of Cell Biology* **152**(2): 401-410.

Hsieh TC. 2013. *iNEXT: iNterpolation and EXTrapolation for species diversity*.

Hu J, Vanderstraeten S, Foury F. 1995. Isolation and characterization of ten mutator alleles of the mitochondrial DNA polymerase-encoding MIP1 gene from Saccharomyces cerevisiae. *Gene* **160**(1): 105-110.

Huber T, Faulkner G, Hugenholtz P. 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20**(14): 2317-2319.

Husmeier D, Wright F. 2001. Detection of Recombination in DNA Multiple Alignments with Hidden Markov Models. *Journal of Computational Biology* **8**(4): 401-427.

Ismail Y, Hijri M. 2012. Arbuscular mycorrhisation with Glomus irregulare induces expression of potato PR homologues genes in response to infection by Fusarium sambucinum. *Functional Plant Biology* **39**(3): 236-245.

Ismail Y, McCormick S, Hijri M. 2011. A Fungal Symbiont of Plant-Roots Modulates Mycotoxin Gene Expression in the Pathogen Fusarium sambucinum. *PLoS ONE* **6**(3): e17990.

Jakobsen I, Abbott LK, Robson AD. 1992. External hyphae of vesicular-arbuscular mycorrhizal fungi associated with Trifolium subterraneum L. *New Phytologist* **120**(3): 371-380.

James JD. 1998. Mycorrhizal Symbiosis. S.E. Smith and D.J. Read. *Plant Growth Regulation* **25**(1): 71-71.

Javot H, Pumplin N, Harrison MJ. 2007. Phosphate in the arbuscular mycorrhizal symbiosis: transport properties and regulatory roles. *Plant, Cell & Environment* **30**(3): 310-322.

Karandashov V, Bucher M. 2005. Symbiotic phosphate transport in arbuscular mycorrhizas. *Trends in Plant Science* **10**(1): 22-29.

Kellis M, Birren BW, Lander ES. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. *Nature* **428**(6983): 617-624.

Kempken F, Hermanns J, Osiewacz HD. 1992. Evolution of linear plasmids. *Journal of Molecular Evolution* **35**(6): 502-513.

Kennel JC, Cohen SM 2004. Fungal mitochondria: genomes, genetic elements and gene expression. In: Arora DK ed. *The handbook of fungal biotechnology*. New York: Marcel Dekker Inc. , 131-143.

Kiers ET, Duhamel M, Beesetty Y, Mensah JA, Franken O, Verbruggen E, Fellbaum CR, Kowalchuk GA, Hart MM, Bago A, Palmer TM, West SA, Vandenkoornhuyse P, Jansa J, Bucking H. 2011. Reciprocal Rewards Stabilize Cooperation in the Mycorrhizal Symbiosis. *Science* **333**(6044): 880-882.

Kim E-K, Jeong J-H, Youn HS, Koo YB, Roe J-H. 2000. The terminal protein of a linear mitochondrial plasmid is encoded in the N-terminus of the DNA polymerase gene in white-rot fungus Pleurotus ostreatus. *Current Genetics* **38**(5): 283-290.

Klassen R, Meinhardt F 2007. Linear protein primed replicating plasmids in eukaryotic microbes. *Microbial linear plasmids*. Berlin, Germany: Springer, 188-226.

Koch AM, Croll D, Sanders IR. 2006. Genetic variability in a population of arbuscular mycorrhizal fungi causes variation in plant growth. *Ecology Letters* **9**(2): 103-110.

Koch AM, Kuhn G, Fontanillas P, Fumagalli L, Goudet J, Sanders IR. 2004. High genetic variability and low local diversity in a population of arbuscular mycorrhizal fungi. *Proceedings of the National Academy of Sciences of the United States of America* **101**(8): 2369-2374.

Krüger M, Krüger C, Walker C, Stockinger H, Schüßler A. 2012. Phylogenetic reference data for systematics and phylotaxonomy of arbuscular mycorrhizal fungi from phylum to species level. *New Phytologist* **193**(4): 970-984.

Kubo T, Nishizawa S, Sugawara A, Itchoda N, Estiati A, Mikami T. 2000. The complete nucleotide sequence of the mitochondrial genome of sugar beet (Beta vulgaris L.) reveals a novel gene for tRNACys(GCA). *Nucleic Acids Research* **28**(13): 2571-2576.

Kuhn G, Hijri M, Sanders IR. 2001. Evidence for the evolution of multiple genomes in arbuscular mycorrhizal fungi. *Nature* **414**: 745-748.

Landsman D, Bustin M. 1993. A signature for the HMG-1 box DNA-binding proteins. *BioEssays* **15**(8): 539-546.

Lang B, Franz, Hijri M. 2009. The complete Glomus intraradices mitochondrial genome sequence – a milestone in mycorrhizal research. *New Phytologist* **183**(1): 3-6.

Lasken RS, Stockwell TB. 2007. Mechanism of chimera formation during the Multiple Displacement Amplification reaction. *BMC Biotechnology* **7**(1): 19.

Leake J, Johnson D, Donnelly D, Muckle G, Boddy L, Read D. 2004. Networks of power and influence: the role of mycorrhizal mycelium in controlling plant communities and agroecosystem functioning. *Canadian Journal of Botany* **82**(8): 1016-1045.

Lee J, Young JPW. 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**(1): 200-211.

Lee SB, Taylor J. 1993. Uniparental inheritance and replacement of mitochondrial DNA in Neurospora tetrasperma. *Genetics* **134**(4): 1063-1075.

Li Q, Nargang FE. 1993. Two Neurospora mitochondrial plasmids encode DNA polymerases containing motifs characteristic of family B DNA polymerases but lack the sequence Asp-Thr-Asp. *Proceedings of the National Academy of Sciences* **90**(9): 4299-4303.

Liitsola K, Holm K, Bobkov A, Pokrovsky V, Smolskaya T, Leinikki P, Osmanov S, Salminen M. 2000. An AB Recombinant and Its Parental HIV Type 1 Strains in the Area of the Former Soviet Union: Low Requirements for Sequence Identity in Recombination. *AIDS Research and Human Retroviruses* **16**(11): 1047-1053.

Lin K, Limpens E, Zhang Z, Ivanov S, Saunders DGO, Mu D, Pang E, Cao H, Cha H, Lin T, Zhou Q, Shang Y, Li Y, Sharma T, van Velzen R, de Ruijter N, Aanen DK, Win J, Kamoun S, Bisseling T, Geurts R, Huang S. 2014. Single Nucleus Genome Sequencing Reveals High Similarity among Nuclei of an Endomycorrhizal Fungus. *PLoS Genet* **10**(1): e1004078.

Liu Y, Leigh JW, Brinkmann H, Cushion MT, Rodriguez-Ezpeleta N, Philippe H, Lang BF. 2009. Phylogenomic Analyses Support the Monophyly of Taphrinomycotina, including Schizosaccharomyces Fission Yeasts. *Molecular Biology and Evolution* **26**(1): 27-34.

Lloyd-Macgilp SA, Chambers SM, Dodd JC, Fitter AH, Walker C, Young JPW. 1996. Diversity of the ribosomal internal transcribed spacers within and among isolates of Glomus mosseae and related mycorrhizal fungi. *New Phytologist* **133**(1): 103-111.

Lockshon D, Zweifel SG, Freeman-Cook LL, Lorimer HE, Brewer BJ, Fangman WL. 1995. A role for recombination junctions in the segregation of mitochondrial DNA in yeast. *Cell* **81**(6): 947-955.

Lohse M, Drechsel O, Bock R. 2007. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current Genetics* **52**(5): 267-274.

MacAlpine DM, Perlman PS, Butow RA. 1998. The high mobility group protein Abf2p influences the level of yeast mitochondrial DNA recombination intermediates in vivo. *Proceedings of the National Academy of Sciences* **95**(12): 6739-6743.

Maherali H, Klironomos JN. 2007. Influence of Phylogeny on Fungal Community Assembly and Ecosystem Functioning. *Science* **316**(5832): 1746-1748.

Mannella C, Pittenger T, Lambowitz A. 1979. Transmission of mitochondrial deoxyribonucleic acid in Neurospora crassa sexual crosses. *Journal of Bacteriology* **137**(3): 1449.

Marcinko-Kuehn M, Yang X, Debets F, Jacobson DJ, Griffiths AJF. 1994. A kalilo-like linear plasmid in Louisiana field isolates of the pseudohomothallic fungus Neurospora tetrasperma. *Current Genetics* **26**(4): 336-343.

Marleau J, Dalpe Y, St-Arnaud M, Hijri M. 2011. Spore development and nuclear inheritance in arbuscular mycorrhizal fungi. *BMC Evolutionary Biology* **11**(1): 51.

Martin F, Gianinazzi-Pearson V, Hijri M, Lammers P, Requena N, Sanders IR, Shachar-Hill Y, Shapiro H, Tuskan GA, Young JPW. 2008. The long hard road to a completed Glomus intraradices genome. *New Phytologist* **180**(4): 747-750.

Miller JR, Koren S, Sutton G. 2010. Assembly algorithms for next-generation sequencing data. *Genomics* **95**(6): 315-327.

Milne I, Wright F, Rowe G, Marshall DF, Husmeier D, McGuire G. 2004. TOPALi: software for automatic identification of recombinant sequences within DNA multiple alignments. *Bioinformatics* **20**(11): 1806-1807.

Miyakawa I, Sando N, Kawano S, Nakamura SI, Kuroiwa T. 1987. Isolation of morphologically intact mitochondrial nucleoids from the yeast, Saccharomyces cerevisiae. *Journal of Cell Science* **88**(4): 431-439.

Nadimi M, Beaudet D, Forget L, Hijri M, Lang BF. 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**(9): 2199-2210.

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**(2): 273-281.

Nargang FE, Bell JB, Stohl LL, Lambowitz AM. 1984. The DNA sequence and genetic organization of a Neurospora mitochondrial plasmid suggest a relationship to introns and mobile elements. *Cell* **38**(2): 441-453.

Nedelcu AM, Lee RW. 1998. Short repetitive sequences in green algal mitochondrial genomes: potential roles in mitochondrial genome evolution. *Molecular Biology and Evolution* **15**(6): 690-701.

Oeser B, Rogmann-Backwinkel P, Tudzynski P. 1993. Interaction between mitochondrial DNA and mitochondrial plasmids in Claviceps purpurea: analysis of plasmid-homologous sequences upstream of the IrRNA-gene. *Current Genetics* **23**(4): 315-322.

Okamoto K, Perlman PS, Butow RA. 1998. The sorting of mitochondrial DNA and mitochondrial proteins in zygotes: preferential transmission of mitochondrial DNA to the medial bud. *The Journal of Cell Biology* **142**(3): 613-623.

Pande S, Lemire EG, Nargang FE. 1989. The mitochondrial plasmid from Neurospora intermedia strain Labelle-1b contains a long open reading frame with blocks of amino acids characteristic of reverse transcriptases and related proteins. *Nucleic Acids Research* **17**(5): 2023-2042.

Papadopoulos JS, Agarwala R. 2007. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **23**(9): 1073-1079.

Paquin B, Laforest M-J, Lang BF. 2000. Double-Hairpin Elements in the Mitochondrial DNA of Allomyces: Evidence for Mobility. *Molecular Biology and Evolution* **17**(11): 1760-1768.

Paquin B, Laforest MJ, Lang BF. 1994. Interspecific transfer of mitochondrial genes in fungi and creation of a homologous hybrid gene. *Proceedings of the National Academy of Sciences* **91**(25): 11807-11810.

Parniske M. 2008. Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nat Rev Micro* **6**(10): 763-775.

Pawlowska TE. 2005. Genetic processes in arbuscular mycorrhizal fungi. *FEMS Microbiology Letters* **251**(2): 185-192.

Pawlowska TE, Taylor JW. 2004. Organization of genetic variation in individuals of arbuscular mycorrhizal fungi. *Nature* **427**: 733-737.

Pawlowska TE, Taylor JW. 2005. Arbuscular mycorrhizal fungi: Hyphal fusion and multigenomic structure (reply). *Nature* **433**(7022): E4-E4.

Pelin A, Pombert JF, Salvioli A, Bonen L, Bonfante P, Corradi N. 2012. The mitochondrial genome of the arbuscular mycorrhizal fungus Gigaspora margarita reveals two unsuspected trans-splicing events of group I introns. *New Phytologist* **194**(3): 836-845.

Petersen JGL, Holmberg S. 1986. The ILVS gene of Saccharomyces cerevisiae is highly expressed. *Nucleic Acids Research* **14**(24): 9631-9651.

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**(1): 216.

Pozo MJ, Cordier C, Dumas-Gaudot E, Gianinazzi S, Barea JM, Azcón-Aguilar C. 2002. Localized versus systemic effect of arbuscular mycorrhizal fungi on defence responses to Phytophthora infection in tomato plants. *Journal of Experimental Botany* **53**(368): 525-534.

Purin S, Morton J. 2012. Anastomosis behavior differs between asymbiotic and symbiotic hyphae of Rhizophagus clarus. *Mycologia* **In press**: DOI: 10.3852/3812-3135.

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**(12): 1315-1322.

Rayner ADM, Griffith GS, Ainsworth AM 1995. Mycelial interconnectedness. *The growing fungus*: Springer, 21-40.

Redecker D. 2001. Two new families of Glomales, Archaeosporaceae and Paraglomaceae, with two new genera Archaeospora and Paraglomus, based on concordant molecular and morphological characters.

Redecker D, Kodner R, Graham LE. 2000. Glomalean Fungi from the Ordovician. *Science* **289**(5486): 1920-1921.

Riley R, Charron P, Idnurm A, Farinelli L, Dalpé Y, Martin F, Corradi N. 2014. Extreme diversification of the mating type–high-mobility group (MATA-HMG) gene family in a plant-associated arbuscular mycorrhizal fungus. *New Phytologist* **201**(1): 254-268.

Rillig M, Wright S, Eviner V. 2002. The role of arbuscular mycorrhizal fungi and glomalin in soil aggregation: comparing effects of five plant species. *Plant and Soil* **238**(2): 325-333.

Rillig MC. 2004. Arbuscular mycorrhizae, glomalin, and soil aggregation. *Canadian Journal of Soil Science* **84**(4): 355-363.

Rillig MC, Mardatin NF, Leifheit EF, Antunes PM. 2010. Mycelium of arbuscular mycorrhizal fungi increases soil water repellency and is sufficient to maintain water-stable soil aggregates. *Soil Biology and Biochemistry* **42**(7): 1189-1191.

Robison MM, Royer JC, Horgen PA. 1991. Homology between mitochondrial DNA of Agaricus bisporus and an internal portion of a linear mitochondrial plasmid of Agaricus bitorquis. *Current Genetics* **19**(6): 495-502.

Robison MM, Wolyn DJ. 2005. A mitochondrial plasmid and plasmid-like RNA and DNA polymerases encoded within the mitochondrial genome of carrot *Daucus carota*. *Current Genetics* **47**(1): 57-66.

Rosendahl S. 2008. Communities, populations and individuals of arbuscular mycorrhizal fungi. *New Phytologist* **178**(2): 253-266.

Rosendahl S, Stukenbrock EH. 2004. Community structure of arbuscular mycorrhizal fungi in undisturbed vegetation revealed by analyses of LSU rDNA sequences. *Molecular Ecology* **13**(10): 3179-3186.

Rosewich UL, Kistler HC. 2000. Role of horizontal gene transfer in the evolution of fungi. *Annual Review of Phytopathology* **38**(1): 325-363.

Roy-Bolduc A, Hijri M. 2011. The Use of Mycorrhizae to Enhance Phosphorus Uptake: A Way Out the Phosphorus Crisis. *Journal of Biofertilizers & Biopesticides* **02**(01).

Ruiz-Lozano JM, Azcon R, Gomez M. 1995. Effects of arbuscular-mycorrhizal glomus species on drought tolerance: physiological and nutritional plant responses. *Applied and Environmental Microbiology* **61**(2): 456-460.

Sakaguchi K. 1990. Invertrons, a class of structurally and functionally related genetic elements that includes linear DNA plasmids, transposable elements, and genomes of adeno-type viruses. *Microbiological Reviews* **54**(1): 66-74.

Sanchez-Puerta MV, Cho Y, Mower JP, Alverson AJ, Palmer JD. 2008. Frequent, Phylogenetically Local Horizontal Transfer of the cox1 Group I Intron in Flowering Plant Mitochondria. *Molecular Biology and Evolution* **25**(8): 1762-1777.

Sanders Ian R. 2011. Fungal Sex: Meiosis Machinery in Ancient Symbiotic Fungi. *Current biology : CB* **21**(21): R896-R897.

Sanders IR, Alt M, Groppe K, Boller T, Wiemken A. 1995. Identification of ribosomal DNA polymorphisms among and within spores of the Glomales: application to studies on the genetic diversity of arbuscular mycorrhizal fungal communities. *New Phytologist* **130**(3): 419-427.

Sanders IR, Croll D. 2010. Arbuscular Mycorrhiza: The Challenge to Understand the Genetics of the Fungal Partner. *Annual Review of Genetics* **44**(1): 271-292.

Sato M, Sato K. 2012. Maternal inheritance of mitochondrial DNA: Degradation of paternal mitochondria by allogeneic organelle autophagy, allophagy. *Autophagy* **8**(3): 424-425.

Saville BJ, Kohli Y, Anderson JB. 1998. mtDNA recombination in a natural population. *Proceedings of the National Academy of Sciences* **95**(3): 1331-1335.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Horn DJV, Weber CF. 2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology* **75**(23): 7537-7541.

Schofield MA, Agbunag R, Miller JH. 1992. DNA inversions between short inverted repeats in Escherichia coli. *Genetics* **132**(2): 295-302.

Schulte U, Lambowitz AM. 1991. The LaBelle mitochondrial plasmid of Neurospora intermedia encodes a novel DNA polymerase that may be derived from a reverse transcriptase. *Molecular and Cellular Biology* **11**(3): 1696-1706.

Schussler A, Schwarzott D, Walker C. 2001. A new fungal phylum, the Glomeromycota: phylogeny and evolution. *Mycological Research* **105**: 1413-1421.

Seif E, Cadieux A, Lang BF. 2006. Hybrid E. coli—Mitochondrial ribonuclease P RNAs are catalytically active. *RNA* **12**(9): 1661-1670.

Seif E, Leigh J, Liu Y, Roewer I, Forget L, Lang BF. 2005. Comparative mitochondrial genomics in zygomycetes: bacteria-like RNase P RNAs, mobile elements and a close source of the group I intron invasion in angiosperms. *Nucleic Acids Research* **33**(2): 734-744.

Selosse M-A, Richard F, He X, Simard SW. 2006. Mycorrhizal networks: des liaisons dangereuses? *Trends in ecology & evolution (Personal edition)* **21**(11): 621-628.

Shao Z, Graf S, Chaga OY, Lavrov DV. 2006. Mitochondrial genome of the moon jelly Aurelia aurita (Cnidaria, Scyphozoa): A linear DNA molecule encoding a putative DNA-dependent DNA polymerase. *Gene* **381**(0): 92-101.

Sharma M, Ellis RL, Hinton DM. 1992. Identification of a family of bacteriophage T4 genes encoding proteins similar to those present in group I introns of fungi and phage. *Proceedings of the National Academy of Sciences* **89**(14): 6658-6662.

Shiu-Shing Chan B, Court DA, Vierula PJ, Bertrand H. 1991. The kalilo linear senescence-inducing plasmid of Neurospora is an invertron and encodes DNA and RNA polymerases. *Current Genetics* **20**(3): 225-237.

Signorovitch AY, Buss LW, Dellaporta SL. 2007. Comparative Genomics of Large Mitochondria in Placozoans. *PLoS Genet* **3**(1): e13.

Smith S, Read D, eds. 2008. *Mycorrhizal Symbiosis*. Cambridge: UK: Academic press.

Smith SE, Jakobsen I, Grønlund M, Smith FA. 2011. Roles of arbuscular mycorrhizas in plant phosphorus nutrition: interactions between pathways of phosphorus uptake in arbuscular mycorrhizal roots have important implications for understanding and manipulating plant phosphorus acquisition. *Plant Physiology* **156**(3): 1050-1057.

Smith SE, Read DJ. 1997. *Mycorrhizal symbiosis*: Academic Pr.

St-Arnaud M, Vujanovic V. 2007. Effect of the arbuscular mycorrhizal symbiosis on plant diseases and pests. *In: Hamel C, Plenchette C (eds) Mycorrhizae in Crop Production. Haworth Food & Agricultural Products Press, Binghampton NY*: 67-122.

Stockinger H, Walker C, Schüßler A. 2009. 'Glomus intraradices DAOM197198', a model fungus in arbuscular mycorrhiza research, is not Glomus intraradices. *New Phytologist* **183**(4): 1176-1187.

Stover B, Muller K. 2010. TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics* **11**(1): 7.

Strack D, Fester T, Hause B, Schliemann W, Walter MH. 2003. Review Paper: Arbuscular Mycorrhiza: Biological, Chemical, and Molecular Aspects. *Journal of Chemical Ecology* **29**(9): 1955-1979.

Stukenbrock EH, 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**(3): 743-752.

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*.

Tanaka Y, Tsuda M, Yasumoto K, Yamagishi H, Terachi T. 2012. A complete mitochondrial genome sequence of Ogura-type male-sterile cytoplasm and its comparative analysis with that of normal cytoplasm in radish (Raphanus sativus L.). *BMC Genomics* **13**(1): 352.

Team RC 2013. R: A language and environment for statistical computing.In: R Foundation for Statistical Computing, Vienna, Austria.

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 Phylogenetics and Evolution* **55**(2): 599-610.

Tisserant E, Kohler A, Dozolme-Seddas P, Balestrini R, Benabdellah K, Colard A, Croll D, Da Silva C, Gomez SK, Koul R, Ferrol N, Fiorilli V, Formey D, Franken P, Helber

N, Hijri M, Lanfranco L, Lindquist E, Liu Y, Malbreil M, Morin E, Poulain J, Shapiro H, van Tuinen D, Waschke A, Azcón-Aguilar C, Bécard G, Bonfante P, Harrison MJ, Küster H, Lammers P, Paszkowski U, Requena N, Rensing SA, Roux C, Sanders IR, Shachar-Hill Y, Tuskan G, Young JPW, Gianinazzi-Pearson V, Martin F. 2012. The transcriptome of the arbuscular mycorrhizal fungus Glomus intraradices (DAOM 197198) reveals functional tradeoffs in an obligate symbiont. *New Phytologist* **193**(3): 755-769.

Tisserant E, Malbreil M, Kuo A, Kohler A, Symeonidi A, Balestrini R, Charron P, Duensing N, dit Frey NF, Gianinazzi-Pearson V. 2013. Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proceedings of the National Academy of Sciences* **110**(50): 20117-20122.

Van Der Heijden MGA, Boller T, Wiemken A, Sanders IR. 1998. Different arbuscular mycorrhizal fungal species are potential determinants of plant community structure. *Ecology* **79**(6): 2082-2091.

VanKuren NW, den Bakker HC, Morton JB, Pawlowska TE. 2013. Ribosomal RNA gene diversity, effective population size, and evolutionary longevity in asexual Glomeromycota. *Evolution* **67**(1): 207-224.

Volz-Lingenhöhl A, Solignac M, Sperlich D. 1992. Stable heteroplasmy for a large-scale deletion in the coding region of Drosophila subobscura mitochondrial DNA. *Proceedings of the National Academy of Sciences* **89**(23): 11528-11532.

Von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, Jouffre N, Huynen MA, Bork P. 2005. STRING: known and predicted protein–protein associations, integrated and transferred across organisms. *Nucleic Acids Research* **33**(suppl 1): D433-D437.

Walker C, Schüßler A. 2004. Nomenclatural Clarifications and New Taxa in the Glomeromycota Pacispora. *Mycological Research* **108**(9): 981-982.

Walker C, Vestberg M, Demircik F, Stockinger H, Saito M, Sawaki H, Nishmura I, Schüßler A. 2007. Molecular phylogeny and new taxa in the Archaeosporales (Glomeromycota): Ambispora fennica gen. sp. nov., Ambisporaceae fam. nov., and emendation of Archaeospora and Archaeosporaceae. *Mycological Research* **111**(2): 137-153.

Wang B, Qiu YL. 2006. Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* **16**(5): 299-363.

Wehner J, Antunes PM, Powell JR, Mazukatow J, Rillig MC. 2010. Plant pathogen protection by arbuscular mycorrhizas: A role for fungal diversity? *Pedobiologia* **53**(3): 197-201.

Weiller GF, Bruckner H, Kim SH, Pratje E, Schweyen RJ. 1991. A GC cluster repeat is a hotspot for mit− macro -deletions in yeast mitochondrial DNA. *Molecular and General Genetics MGG* **226**(1): 233-240.

Westermann B, Prokisch H. 2002. Mitochondrial dynamics in filamentous fungi. *Fungal Genetics and Biology* **36**(2): 91-97.

White DJ, Wolff JN, Pierson M, Gemmell NJ. 2008. Revealing the hidden complexities of mtDNA inheritance. *Molecular Ecology* **17**(23): 4925-4942.

White MF, Lilley D. 1997. Characterization of a Holliday junction-resolving enzyme from Schizosaccharomyces pombe. *Molecular and Cellular Biology* **17**(11): 6465-6471.

Williamson D, Fennell D. 1976. The use of fluorescent DNA-binding agent for detecting and separating yeast mitochondrial DNA. *Methods in Cell Biology* **12**: 335-351.

Xavier BB, Miao VPW, Jónsson ZO, Andrésson ÓS. 2012. Mitochondrial genomes from the lichenized fungi Peltigera membranacea and Peltigera malacea: Features and phylogeny. *Fungal Biology* **116**(7): 802-814.

Xia X, Xie Z. 2001. DAMBE: Software Package for Data Analysis in Molecular Biology and Evolution. *Journal of Heredity* **92**(4): 371-373.

Yang X, Griffiths AJF. 1993. Plasmid diversity in senescent and nonsenescent strains of Neurospora. *Molecular and General Genetics MGG* **237**(1): 177-186.

Yasuhira S, Yasui A. 2000. Alternative Excision Repair Pathway of UV-damaged DNA inSchizosaccharomyces pombe Operates Both in Nucleus and in Mitochondria. *Journal of Biological Chemistry* **275**(16): 11824-11828.

Yin S, Heckman J, RajBhandary UL. 1981. Highly conserved GC-rich palindromic DNA sequences flank tRNA genes in Neurospora crassa mitochondria. *Cell* **26**(3, Part 1): 325-332.

Youngman MJ, Hobbs AEA, Burgess SM, Srinivasan M, Jensen RE. 2004. Mmm2p, a mitochondrial outer membrane protein required for yeast mitochondrial shape and maintenance of mtDNA nucleoids. *The Journal of Cell Biology* **164**(5): 677-688.

Zimmer M, Krabusch M, Wolf K. 1991. Characterization of a novel open reading frame, urf a, in the mitochondrial genome of fission yeast: Correlation of urf a; mutations with a mitochondrial mutator phenotype and a possible role of frameshifting in urf a; expression. *Current Genetics* **19**(2): 95-102.

Zimmerman E, St-Arnaud M, Hijri M 2010. Sustainable Agriculture and the Multigenomic Model: How Advances in the Genetics of Arbuscular Mycorrhizal Fungi will Change Soil Management Practices. In: Bouarab K, Brisson N, Daayf F eds. *Molecular Plant-microbe interactions.*: CAB International, 269-287.

Annexes

1- Supplementary Information (Chapter 2)

Rapid mitochondrial genome evolution through invasion of mobile elements in two closely related species of arbuscular mycorrhizal fungi

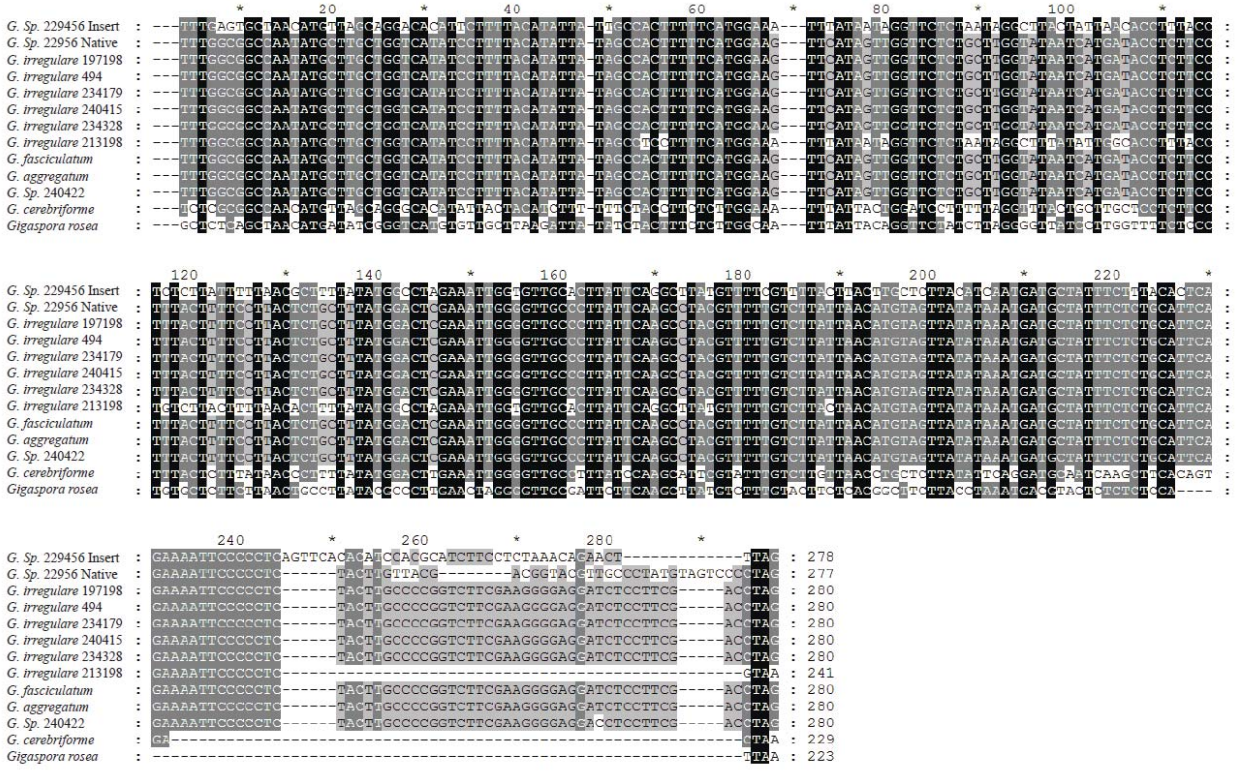


Figure S2.1. Multiple DNA sequence alignment of numerous AMF representatives of the *atp6* native C-terminals along with the *Glomus* sp. *229456* putative foreign inserted C*-terminal.

Table S2.1. Sequence identity matrix of the atp6 native C-terminals along with the *Glomus* sp. 229456 putative foreign inserted C*-terminal.

Seq->	Gsp229456 Insert	Gsp229456 Native	Gi197198	Gi494	Gi234179	Gi240415	Gi234328	Gsp213198	fascicula	aggregatum	Gsp240422	cerebri	G._rosea
Gsp229456 Insert	ID	63.5%	68.1%	68.1%	68.1%	68.1%	68.1%	74.1%	68.1%	68.1%	68.1%	56.1%	50.3%
Gsp229456 Native	63.5%	ID	91.2%	91.2%	91.2%	91.2%	91.2%	72.5%	91.2%	91.2%	90.8%	56.6%	49.0%
Gi197198	68.1%	91.2%	ID	100.0%	100.0%	100.0%	100.0%	71.7%	100.0%	100.0%	99.6%	56.0%	48.5%
Gi494	68.1%	91.2%	100.0%	ID	100.0%	100.0%	100.0%	71.7%	100.0%	100.0%	99.6%	56.0%	48.5%
Gi234179	68.1%	91.2%	100.0%	100.0%	ID	100.0%	100.0%	71.7%	100.0%	100.0%	99.6%	56.0%	48.5%
Gi240415	68.1%	91.2%	100.0%	100.0%	100.0%	ID	100.0%	71.7%	100.0%	100.0%	99.6%	56.0%	48.5%
Gi234328	68.1%	91.2%	100.0%	100.0%	100.0%	100.0%	ID	71.7%	100.0%	100.0%	99.6%	56.0%	48.5%
Gsp213198	74.1%	72.5%	71.7%	71.7%	71.7%	71.7%	71.7%	ID	71.7%	71.7%	71.7%	63.9%	59.3%
fascicula	68.1%	91.2%	100.0%	100.0%	100.0%	100.0%	100.0%	71.7%	ID	100.0%	99.6%	56.0%	48.5%
aggregatum	68.1%	91.2%	100.0%	100.0%	100.0%	100.0%	100.0%	71.7%	100.0%	ID	99.6%	56.0%	48.5%
Gsp240422	68.1%	90.8%	99.6%	99.6%	99.6%	99.6%	99.6%	71.7%	99.6%	99.6%	ID	56.0%	48.5%
cerebri	56.1%	56.6%	56.0%	56.0%	56.0%	56.0%	56.0%	63.9%	56.0%	56.0%	56.0%	ID	63.7%
G._rosea	50.3%	49.0%	48.5%	48.5%	48.5%	48.5%	48.5%	59.3%	48.5%	48.5%	48.5%	63.7%	ID

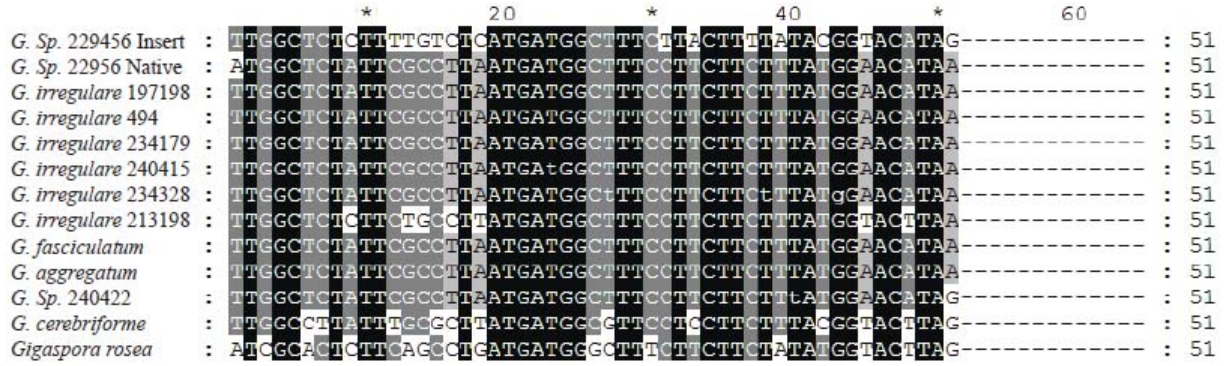


Figure S2.2. Multiple DNA sequence alignment of numerous AMF representatives of the *atp9* native C-terminals along with the *Glomus* sp. *229456* putative foreign inserted C*-terminal.

Table S2.2. Sequence identity matrix of the atp9 native C-terminals along with the *Glomus* sp. 229456 putative foreign inserted C*-terminal.

Seq->	Gsp229456 Insert	Gsp229456 Native	Gi197198	Gi494	Gi234179	Gi240415	Gi234328	Gi213198	fascicula	aggregatum	Gsp240422	cerebri	G._rosea
Gsp229456 Insert	ID	70.5%	72.5%	72.5%	72.5%	72.5%	72.5%	80.3%	72.5%	72.5%	74.5%	74.5%	70.5%
Gsp229456 Native	70.5%	ID	98.0%	98.0%	98.0%	98.0%	98.0%	84.3%	98.0%	98.0%	96.0%	74.5%	72.5%
Gi197198	72.5%	98.0%	ID	100.0%	100.0%	100.0%	100.0%	86.2%	100.0%	100.0%	98.0%	76.4%	70.5%
Gi494	72.5%	98.0%	100.0%	ID	100.0%	100.0%	100.0%	86.2%	100.0%	100.0%	98.0%	76.4%	70.5%
Gi234179	72.5%	98.0%	100.0%	100.0%	ID	100.0%	100.0%	86.2%	100.0%	100.0%	98.0%	76.4%	70.5%
Gi240415	72.5%	98.0%	100.0%	100.0%	100.0%	ID	100.0%	86.2%	100.0%	100.0%	98.0%	76.4%	70.5%
Gi234328	72.5%	98.0%	100.0%	100.0%	100.0%	100.0%	ID	86.2%	100.0%	100.0%	98.0%	76.4%	70.5%
Gsp213198	80.3%	84.3%	86.2%	86.2%	86.2%	86.2%	86.2%	ID	86.2%	86.2%	84.3%	78.4%	80.3%
fascicula	72.5%	98.0%	100.0%	100.0%	100.0%	100.0%	100.0%	86.2%	ID	100.0%	98.0%	76.4%	70.5%
aggregatum	72.5%	98.0%	100.0%	100.0%	100.0%	100.0%	100.0%	86.2%	100.0%	ID	98.0%	76.4%	70.5%
Gsp240422	74.5%	96.0%	98.0%	98.0%	98.0%	98.0%	98.0%	84.3%	98.0%	98.0%	ID	78.4%	72.5%
cerebri	74.5%	74.5%	76.4%	76.4%	76.4%	76.4%	76.4%	78.4%	76.4%	76.4%	78.4%	ID	68.6%
G._rosea	70.5%	72.5%	70.5%	70.5%	70.5%	70.5%	70.5%	80.3%	70.5%	70.5%	72.5%	68.6%	ID

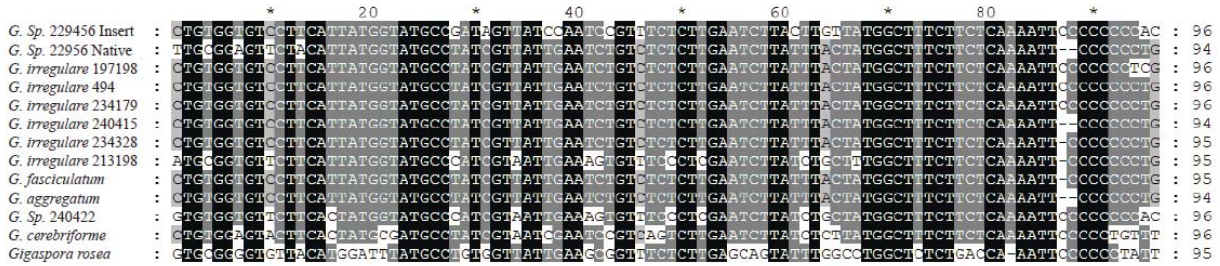


Figure S2.3. Multiple DNA sequence alignment of numerous AMF representatives of the *cox2* native C-terminals along with the G*lomus* sp. *229456* putative foreign inserted C*-terminal.

Table S2.3. Sequence identity matrix of the *cox2* native C-terminals along with the *Glomus* sp. 229456 putative foreign inserted C*-terminal.

Seq->	Gsp229456 Insert	Gsp229456 Native	Gi197198	Gi494	Gi234179	Gi240415	Gi234328	Gsp213198	fascicula	aggregatum	Gsp240422	cerebri	G._rosea
Gsp229456 Insert	ID	81.2%	87.5%	88.5%	88.5%	86.4%	87.5%	79.1%	87.5%	86.4%	83.3%	79.1%	63.5%
Gsp229456 Native	81.2%	ID	90.6%	92.7%	92.7%	94.6%	93.6%	85.2%	93.6%	94.6%	81.2%	78.1%	66.6%
Gi197198	87.5%	90.6%	ID	97.9%	97.9%	95.8%	96.8%	83.3%	96.8%	95.8%	84.3%	82.2%	65.6%
Gi494	88.5%	92.7%	97.9%	ID	100.0%	97.9%	98.9%	85.4%	98.9%	97.9%	85.4%	82.2%	66.6%
Gi234179	88.5%	92.7%	97.9%	100.0%	ID	97.9%	98.9%	85.4%	98.9%	97.9%	85.4%	82.2%	66.6%
Gi240415	86.4%	94.6%	95.8%	97.9%	97.9%	ID	98.9%	85.2%	98.9%	100.0%	83.3%	80.2%	64.5%
Gi234328	87.5%	93.6%	96.8%	98.9%	98.9%	98.9%	ID	86.3%	100.0%	98.9%	84.3%	81.2%	65.6%
Gsp213198	79.1%	85.2%	83.3%	85.4%	85.4%	85.2%	86.3%	ID	86.3%	85.2%	92.7%	73.9%	62.5%
fascicula	87.5%	93.6%	96.8%	98.9%	98.9%	98.9%	100.0%	86.3%	ID	98.9%	84.3%	81.2%	65.6%
aggregatum	86.4%	94.6%	95.8%	97.9%	97.9%	100.0%	98.9%	85.2%	98.9%	ID	83.3%	80.2%	64.5%
Gsp240422	83.3%	81.2%	84.3%	85.4%	85.4%	83.3%	84.3%	92.7%	84.3%	83.3%	ID	77.0%	61.4%
cerebri	79.1%	78.1%	82.2%	82.2%	82.2%	80.2%	81.2%	73.9%	81.2%	80.2%	77.0%	ID	59.3%
G._rosea	63.5%	66.6%	65.6%	66.6%	66.6%	64.5%	65.6%	62.5%	65.6%	64.5%	61.4%	59.3%	ID

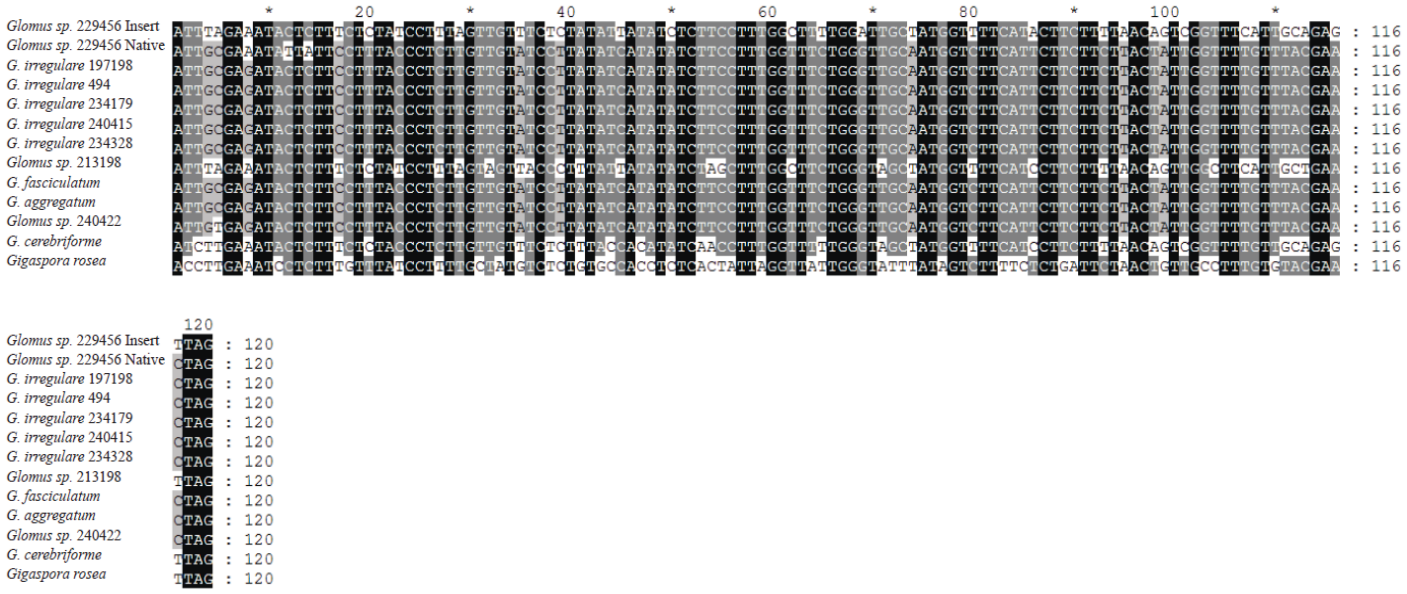


Figure S2.4. Multiple DNA sequence alignment of numerous AMF representatives of the *nad3* native C-terminals along with the *Glomus* sp. *229456* putative foreign inserted C*-terminal

Table S2.4. Sequence identity matrix of the *nad3* native C-terminals along with the *Glomus* sp. 229456 putative foreign inserted C*-terminal.

Seq->	Gsp229456 insert	Gsp229456 native	Gi197198	Gi494	Gi234179	Gi240415	Gi234328	Gsp213198	fascicula	aggregatum	Gsp240422	cerebri	G_rosea
Gsp229456 Insert	ID	73%	74%	74%	74%	74%	74%	88%	74%	74%	74%	85%	61%
Gsp229456 Native	73%	ID	98%	98%	98%	98%	98%	73%	98%	98%	97%	75%	63%
Gi197198	74%	98%	ID	100%	100%	100%	100%	73%	100%	100%	99%	76%	63%
Gi494	74%	98%	100%	ID	100%	100%	100%	73%	100%	100%	99%	76%	63%
Gi234179	74%	98%	100%	100%	ID	100%	100%	73%	100%	100%	99%	76%	63%
Gi240415	74%	98%	100%	100%	100%	ID	100%	73%	100%	100%	99%	76%	63%
Gi234328	74%	98%	100%	100%	100%	100%	ID	73%	100%	100%	99%	76%	63%
Gi213198	88%	73%	73%	73%	73%	73%	73%	ID	73%	73%	73%	83%	61%
Fascicula	74%	98%	100%	100%	100%	100%	100%	73%	ID	100%	99%	76%	63%
aggregatum	74%	98%	100%	100%	100%	100%	100%	73%	100%	ID	99%	76%	63%
Gsp240422	74%	97%	99%	99%	99%	99%	99%	73%	99%	99%	ID	77%	64%
Cerebri	85%	75%	76%	76%	76%	76%	76%	83%	76%	76%	77%	ID	68%
G_rosea	61%	63%	63%	63%	63%	63%	63%	61%	63%	63%	64%	68%	ID

2 - Supplementary Information (Chapter 3)

Mitochondrial genome rearrangements in *Glomus* species triggered by homologous recombination between distinct mtDNA haplotypes

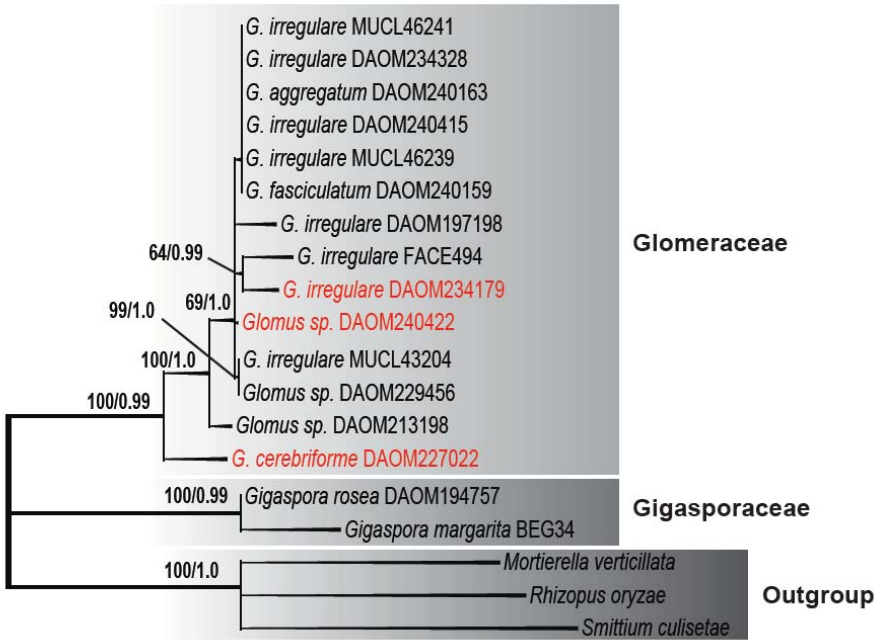


Figure S3.1. Unrooted maximum likelihood tree using concatenated mitochondrial coding gene sequences of all published mt genomes and the ones from our collection. The analysis was performed using the general time reversible (GTR) model with five distinct gamma categories with invariable sites. The first number at branches indicates ML bootstrap values (< 60% cut-off) with 1000 bootstrap replicates and the second number indicates posterior probability values of a MrBayes analysis with four independant chains. Bayesian inference predict similar trees (not shown). The phylogeny was conducted on the concatenated nucleotide sequences of seven conserved mitochondrial protein coding genes (*atp6, atp9, cox2, nad1, nad2, nad4* and *nad6*, for a total of 6632 alignment positions). The strains we compared in this study are in red. We used Blastocladiomycota species as an outgroup.

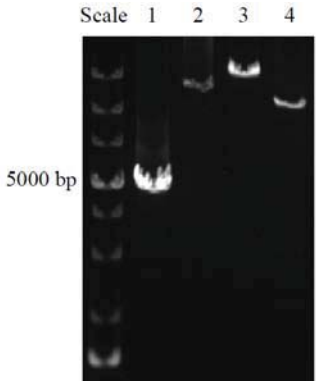


Figure S3.2. Agarose gel electrophoresis figure showing the long PCR amplifications on *Glomus* sp. 240422 reshuffled intergenic regions. Each newly formed intergenic region: *atp9-cox1* (1), *nad4L-rnl* (2), *cox3-rns* (3) and the *cox1* gene as a positive control (4), have been amplified by long PCR with primers spanning the entire mitochondrial region of interest. The expected size of the amplicons was 5029, 13094, 13610 and 8785 bp, respectively.

Table S3.1. Distribution of small inverted repeats (SIRs) found in *G. irregulare* 234179 depending on their subtypes and genome localization.

Coordinates	SIR subtypes	Genome localization
23424-23454	T1	*nad5* intron 2
50561-50532	T2	intergenic
10440-10473	T3	intergenic
50069-50036		intergenic
70802-70835		intergenic
4843-4811	T4	*rnl* intron 5
9544-9576		intergenic
15520-15488		intergenic
15559-15527		intergenic
51757-51725		intergenic
57301-57333		intergenic
59286-59254		intergenic
59929-59897		intergenic
59218-59186	T5	intergenic
38002-38043	T6	intergenic
60364-60433	T7	*cox3* intron 1
14-49	T8	intergenic
6580-6619		intergenic
35087-35058		intergenic
45911-45946		*cox1* intron 9
48788-48749		intergenic
21246-21279	T9	GIYYIG endonuclease
30632-30665		*cob* intron 4
21621-21668	T10	intergenic
40281-40328		intergenic
21539-21577	T11	intergenic
27613-27651		intergenic
17439-17471	T12	intergenic
44203-44235		*cox1* intron 7
23188-23221	T13	*nad5* intron 2
40530-40563		*cox1* intron 1

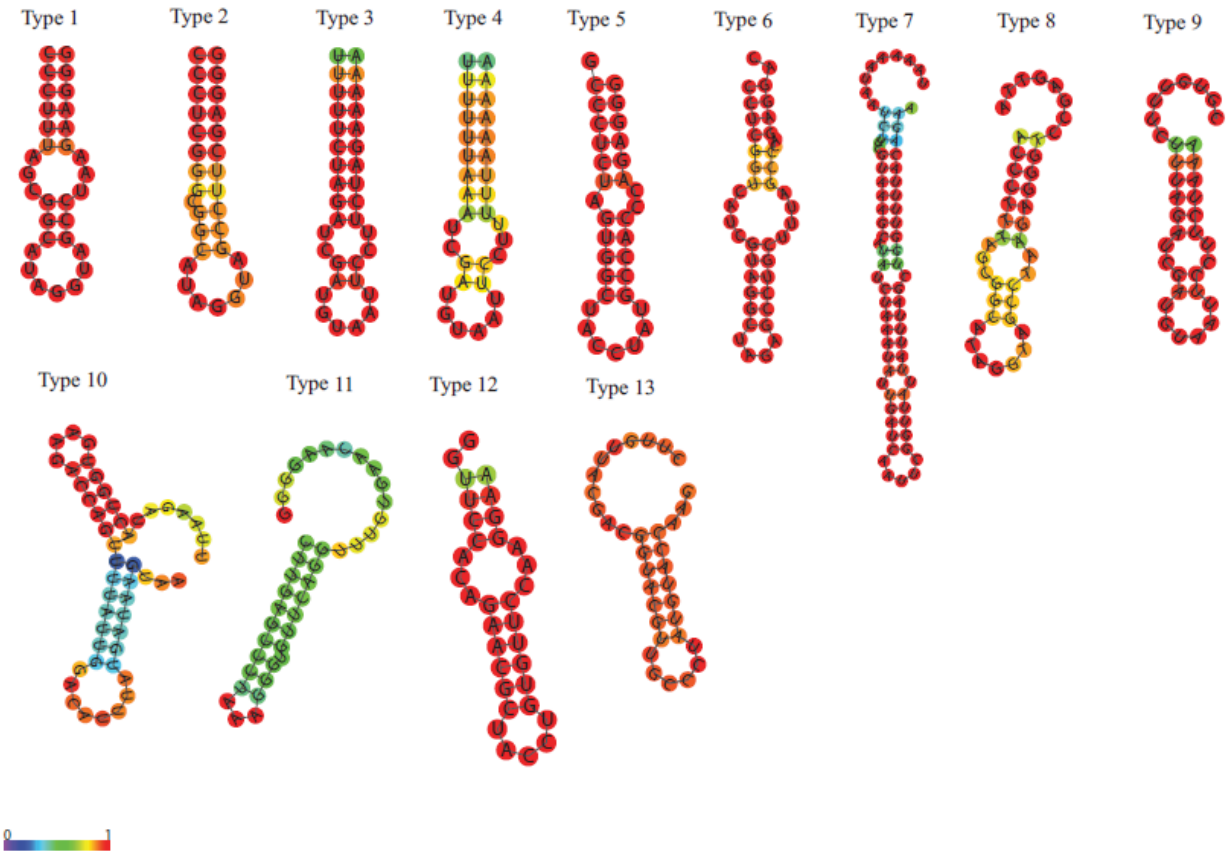


Figure S3.3. Minimum free energy structure (MFE) drawing encoding base-pair probabilities of the 13 small inverted repeats (SIRs) subtypes found in *G. irregulare* 234179. The structures above is colored by base-pairing probabilities. For unpaired regions the color denotes the probability of being unpaired.

Table S3.2. Distribution of small inverted repeats (SIRs) found in *Glomus* sp. 240422 depending on their subtypes and genome localization.

Coordinates	SIR subtypes	Genome localization
12924-12860	T1	intergenic
29969-30001		intergenic
23574-23606	T2	*cox1* intron 7
37717-37751		intergenic
40333-40299		*cox3* intron 1
75489-75523		intergenic
19153-19200	T3	intergenic
56641-56678		intergenic
58428-58475		intergenic
37481-37526	T4	intergenic
58345-58392		intergenic
65429-65475		intergenic
19403-19436	T5	LAGLIDADG endonuclease
59985-60018		*nad5* intron 2
58051-58085	T6	GIYYIG endonuclease
68487-68521		*cob* intron 4
9094-9062	T7	intergenic
32040-32008		intergenic
39795-39763		intergenic
57324-57356		intergenic
63330-63298		intergenic
21809-21771	T8	*cox1* intron 4
33049-33086		intergenic
60219-60252		*nad5* intron 2
6556-6526	T9	*rnl* intron 6
30438-30468		intergenic
34561-34591		intergenic
71457-71490	T10	intergenic
40229-40298	T11	*cox3* intron 1

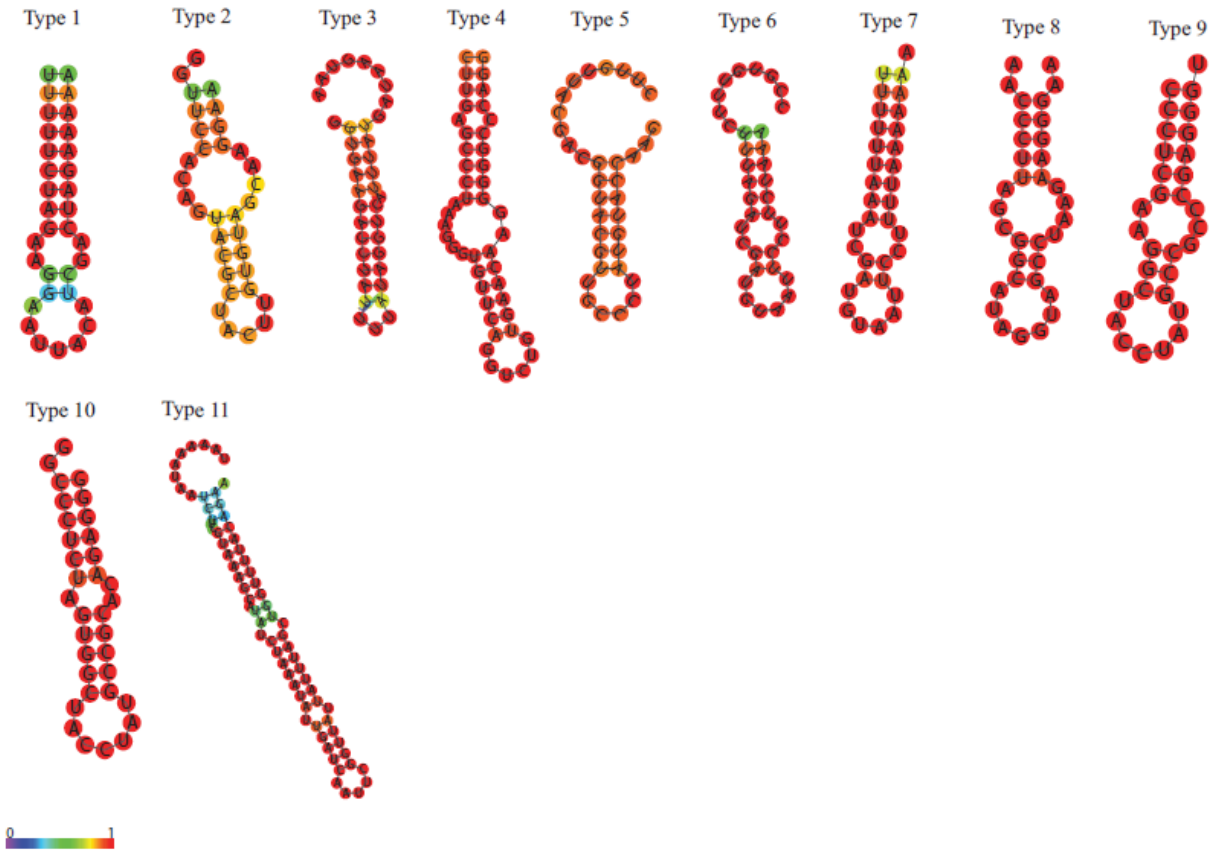


Figure S3.4. Minimum free energy structure (MFE) drawing encoding base-pair probabilities of the 11 small inverted repeats (SIRs) subtypes found in *Glomus* sp. 240422. The structures above is colored by base-pairing probabilities. For unpaired regions the color denotes the probability of being unpaired.

Table S3.3. Distribution of small inverted repeats (SIRs) found in *G. cerebriforme* depending on their subtypes and genome localization.

Coordinates	SIR subtypes	Genome localization
2308-2345		intergenic
12096-12059		intergenic
19133-19096	T1	*cox1* intron 5
40686-40649		intergenic
55004-54967		intergenic
59429-59466		intergenic
49-83		intergenic
8558-8524		intergenic
9588-9554		intergenic
11906-11938	T2	intergenic
15627-15661		*cox1* intron 2
17966-18000		*cox1* intron 4
55119-55151		intergenic
57567-57534		*cox3* intron 2
2968-2935		*rnl* exon 3
15702-15735	T3	*cox1* intron 4
21257-21224		intergenic
18819-18786		*cox1* intron 5
19233-19266		*cox1* intron 5
21247-21280		intergenic
28073-28108	T4	*nad5* intron 1
31801-31768		intergenic
56115-56072		intergenic
57203-57249		*cox3* intron 1
2564-2521		*rnl* intron 2
3361-3318		*rnl* intron 3
6356-6399		*rnl* exon 5
8003-7962		intergenic
8146-8104		intergenic
9312-9269		intergenic
22273-22231		intergenic
23607-23651	T5	intergenic
25943-25987		*rns*
31020-31064		intergenic
31925-31969		intergenic
34361-34317		intergenic
35912-35869		intergenic
38313-38271		intergenic
49930-49889		intergenic
54164-54121		*cob* intron 2

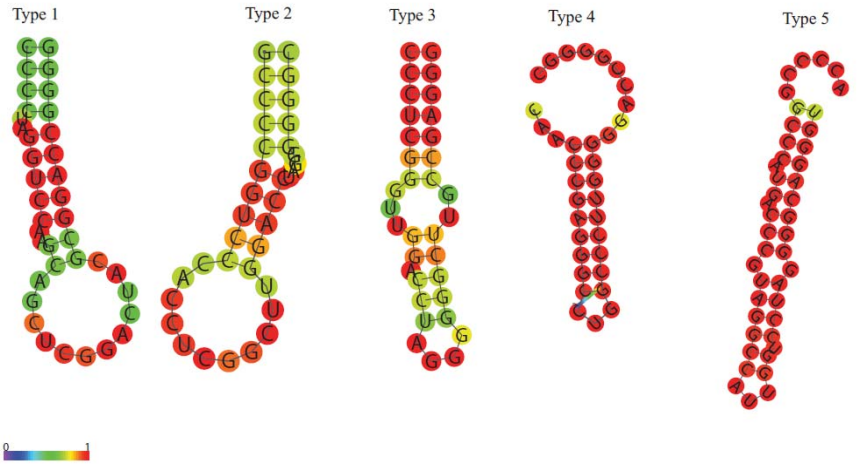


Figure S3.5. Minimum free energy structure (MFE) drawing encoding base-pair probabilities of the five small inverted repeats (SIRs) subtypes found in *G. cerebriforme*. The structures above is colored by base-pairing probabilities. For unpaired regions the color denotes the probability of being unpaired.

Fig. S6-A

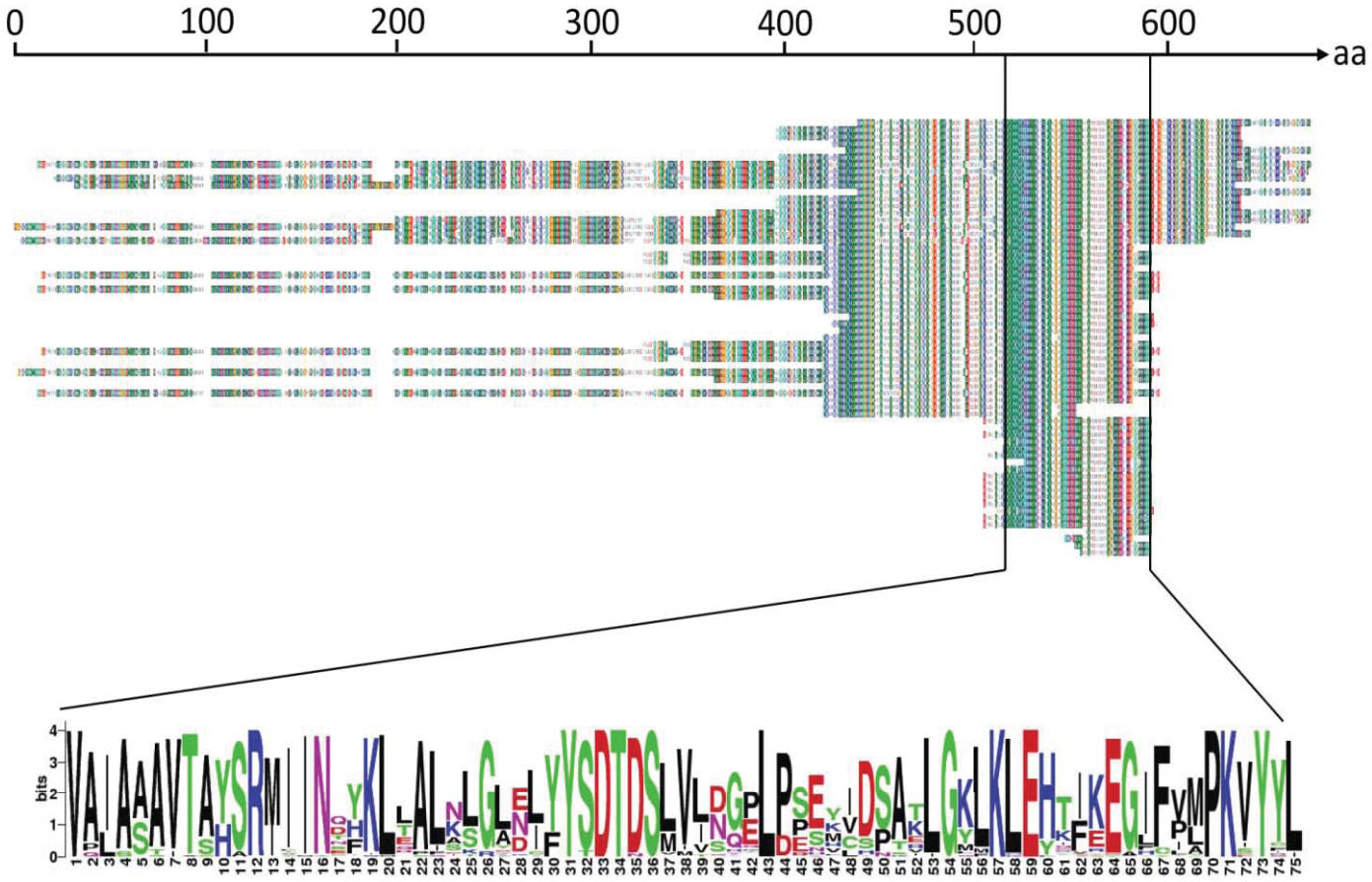


Fig. S6-B

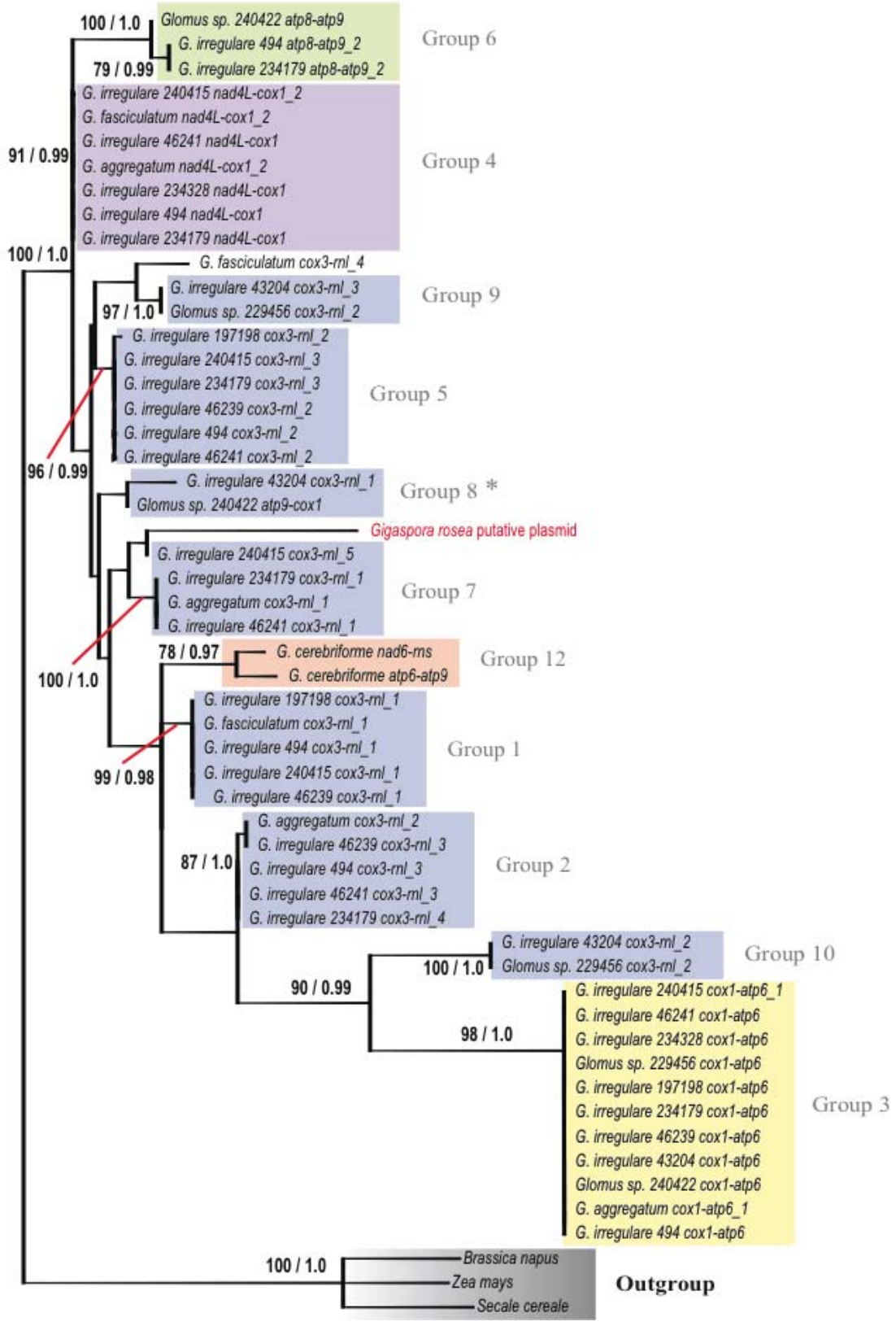

100 / 1.0
Glomus sp. 240422 atp8-atp9
G. irregulare 494 atp8-atp9_2
79 / 0.99
G. irregulare 234179 atp8-atp9_2
Group 6
G. irregulare 240415 nad4L-cox1_2
G. fasciculatum nad4L-cox1_2
G. irregulare 46241 nad4L-cox1
91 / 0.99
G. aggregatum nad4L-cox1_2
G. irregulare 234328 nad4L-cox1
G. irregulare 494 nad4L-cox1
G. irregulare 234179 nad4L-cox1
Group 4
100 / 1.0
G. fasciculatum cox3-rnl_4
G. irregulare 43204 cox3-rnl_3
97 / 1.0
Glomus sp. 229456 cox3-rnl_2
Group 9
G. irregulare 197198 cox3-rnl_2
G. irregulare 240415 cox3-rnl_3
G. irregulare 234179 cox3-rnl_3
G. irregulare 46239 cox3-rnl_2
G. irregulare 494 cox3-rnl_2
G. irregulare 46241 cox3-rnl_2
Group 5
96 / 0.99
G. irregulare 43204 cox3-rnl_1
Glomus sp. 240422 atp9-cox1
Group 8 *
Gigaspora rosea putative plasmid
G. irregulare 240415 cox3-rnl_5
G. irregulare 234179 cox3-rnl_1
G. aggregatum cox3-rnl_1
G. irregulare 46241 cox3-rnl_1
Group 7
78 / 0.97
G. cerebriforme nad6-rns
G. cerebriforme atp6-atp9
Group 12
100 / 1.0
G. irregulare 197198 cox3-rnl_1
G. fasciculatum cox3-rnl_1
G. irregulare 494 cox3-rnl_1
G. irregulare 240415 cox3-rnl_1
G. irregulare 46239 cox3-rnl_1
Group 1
99 / 0.98
G. aggregatum cox3-rnl_2
G. irregulare 46239 cox3-rnl_3
87 / 1.0
G. irregulare 494 cox3-rnl_3
G. irregulare 46241 cox3-rnl_3
G. irregulare 234179 cox3-rnl_4
Group 2
G. irregulare 43204 cox3-rnl_2
100 / 1.0
Glomus sp. 229456 cox3-rnl_2
Group 10
90 / 0.99
G. irregulare 240415 cox1-atp6_1
98 / 1.0
G. irregulare 46241 cox1-atp6
G. irregulare 234328 cox1-atp6
Glomus sp. 229456 cox1-atp6
G. irregulare 197198 cox1-atp6
G. irregulare 234179 cox1-atp6
G. irregulare 46239 cox1-atp6
G. irregulare 43204 cox1-atp6
Glomus sp. 240422 cox1-atp6
G. aggregatum cox1-atp6_1
G. irregulare 494 cox1-atp6
Group 3
100 / 1.0
Brassica napus
Zea mays
Secale cereale
Outgroup

Figure S3.6. Glomeromycota *dpo* maximum likelihood phylogenetic tree based on a 75 amino acids conserved domain. The conserved proteic domain found in 60 out of 91 AMF *dpo* sequences where the consensus sequence is shown **(A)** was used to conduct a phylogenetic analysis where the number on branches correspond to ML bootstrap values (< 60% cut-off) based on a thousand bootstrap replicates and bayesian inference values, respectively **(B)**. The different groups referred to the ones identified in the Glomeromycota *dpo* nucleotide identity network (Figure 3A) and their box color correspond to the ones in Figure 3B based on their mtDNA localization. Three plants *dpo* sequences were used as an outgroup. *The Gigaspora rosea* putative mitochondrial plasmid is colored in red. The group 8 marked with an asterisk correspond to one of the recombination event identified (Figure S7A).

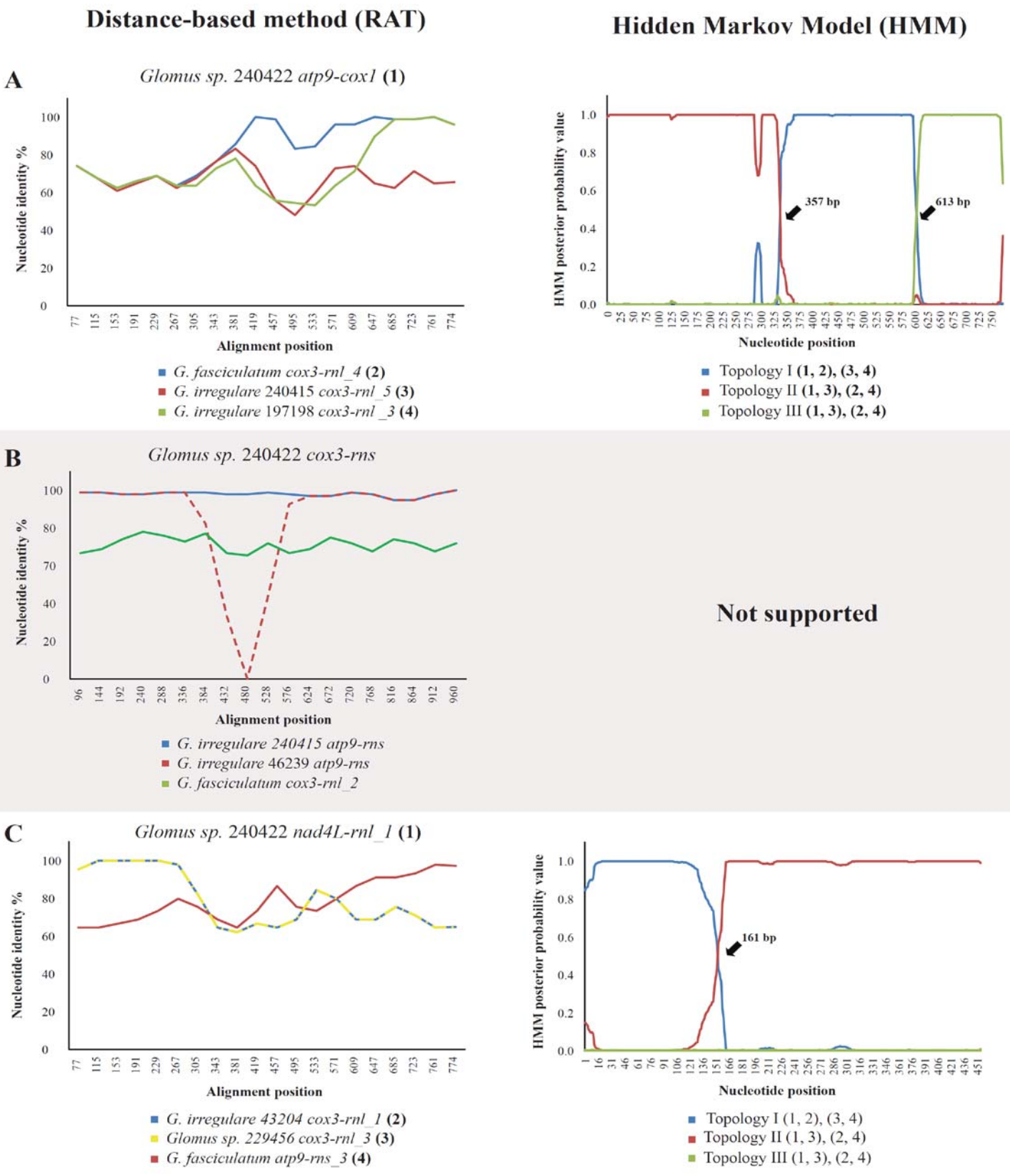
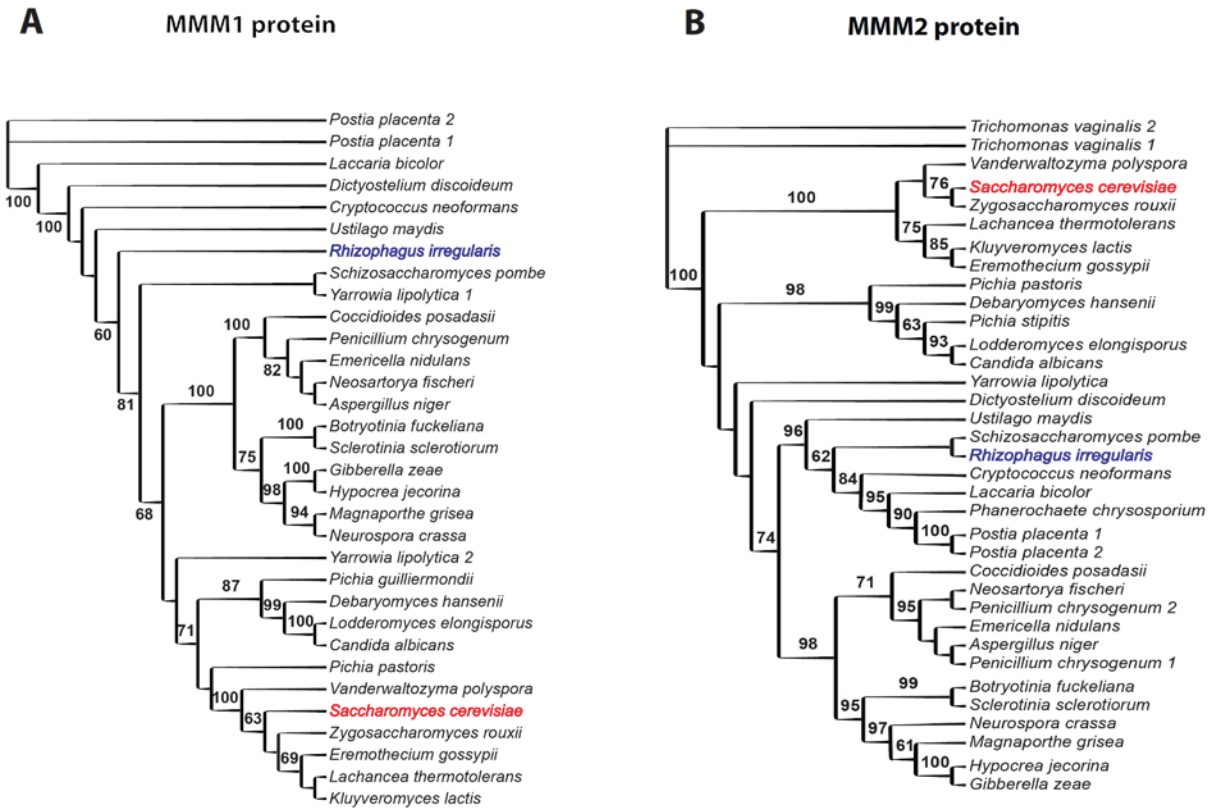
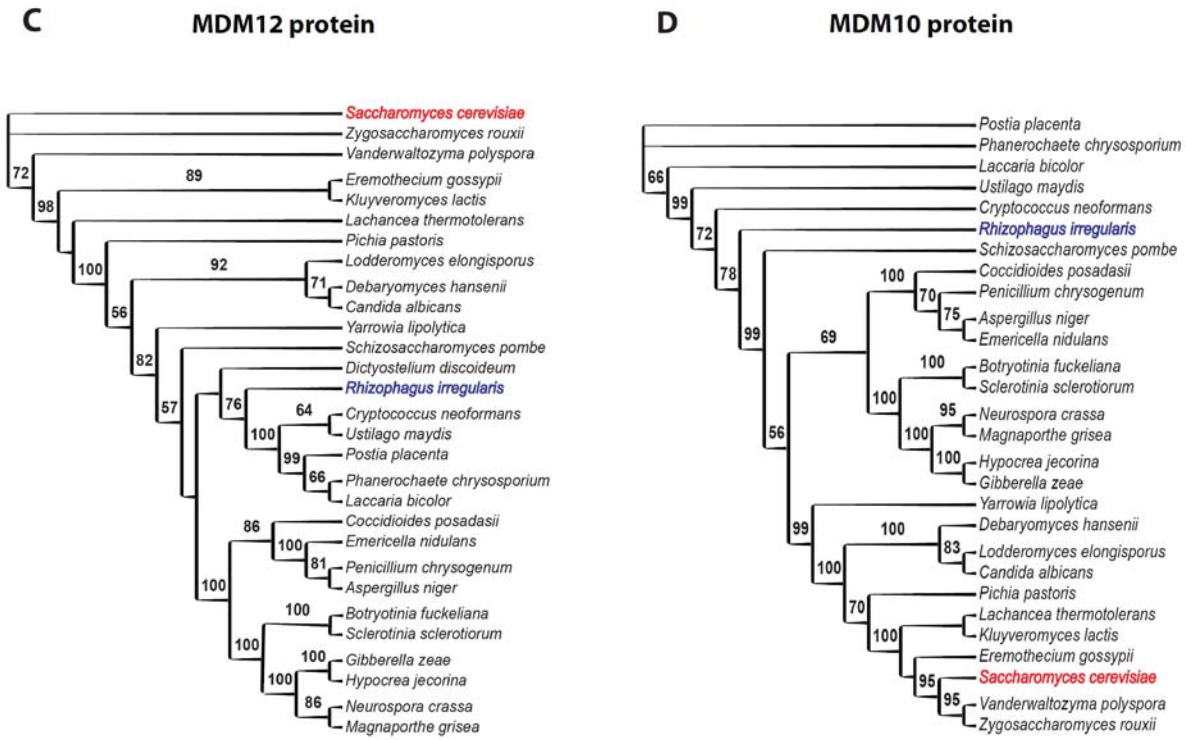

Distance-based method (RAT)
Hidden Markov Model (HMM)
A
Glomus sp. 240422 atp9-cox1 (1)
Nucleotide identity %
Alignment position
G. fasciculatum cox3-rnl_4 (2)
G. irregulare 240415 cox3-rnl_5 (3)
G. irregulare 197198 cox3-rnl_3 (4)
HMM posterior probability value
357 bp
613 bp
Nucleotide position
Topology I (1, 2), (3, 4)
Topology II (1, 3), (2, 4)
Topology III (1, 3), (2, 4)
B
Glomus sp. 240422 cox3-rns
Nucleotide identity %
Alignment position
G. irregulare 240415 atp9-rns
G. irregulare 46239 atp9-rns
G. fasciculatum cox3-rnl_2
Not supported
C
Glomus sp. 240422 nad4L-rnl_1 (1)
Nucleotide identity %
Alignment position
G. irregulare 43204 cox3-rnl_1 (2)
Glomus sp. 229456 cox3-rnl_3 (3)
G. fasciculatum atp9-rns_3 (4)
HMM posterior probability value
161 bp
Nucleotide position
Topology I (1, 2), (3, 4)
Topology II (1, 3), (2, 4)
Topology III (1, 3), (2, 4)

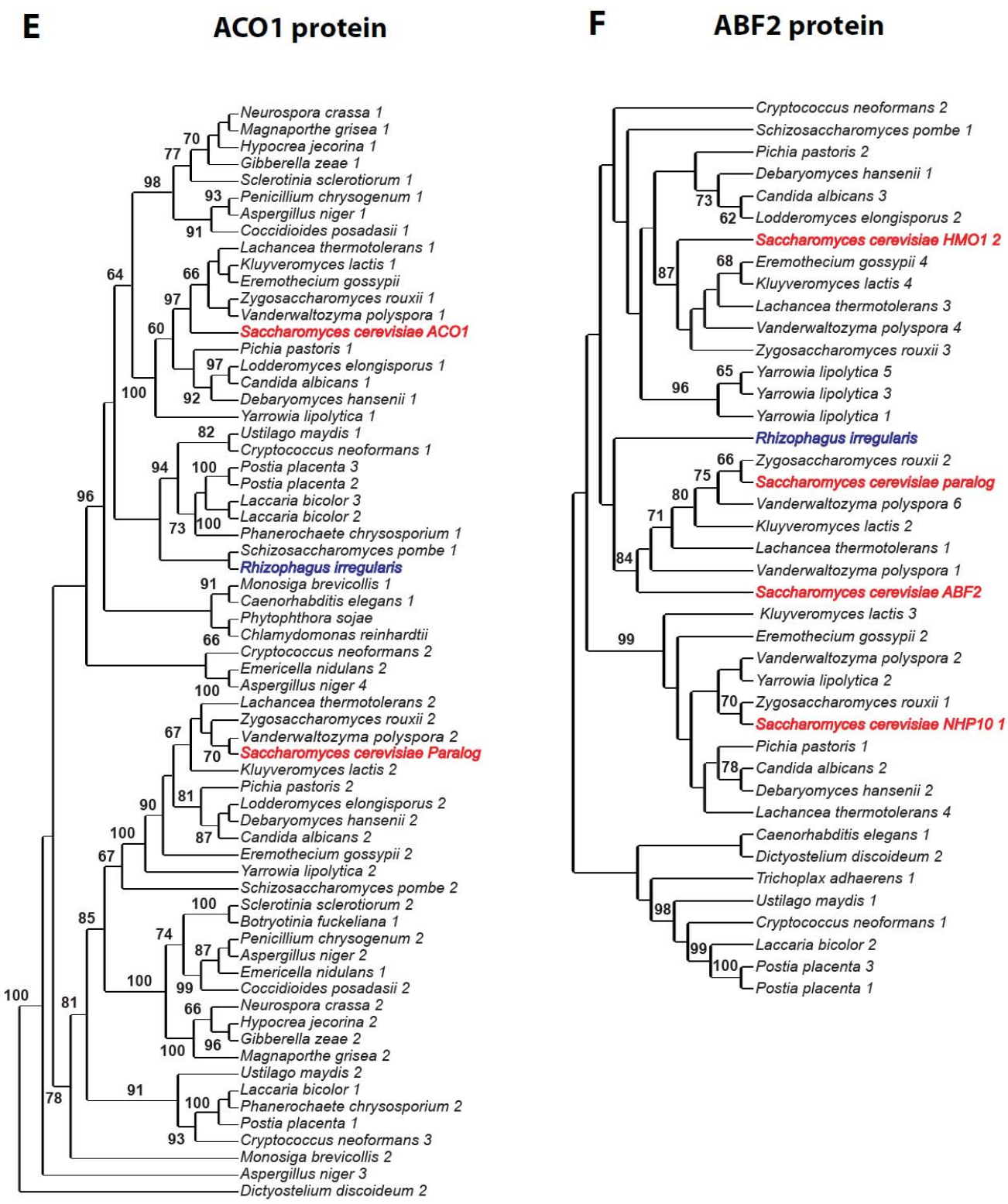
Figure S3.7. Graphical representation of a distance-based and a hidden markov model (HMM) phylogenetic recombination analysis. Each of the three *dpo* sequences present in the *Glomus* sp. 240422 reshuffled intergenic region that clustered in the network in group 1, 2 and 8, respectively (see Figure 3): *Glomus* sp. 240422 *atp9-cox1 dpo* **(A)**, the *cox3-rns* **(B)** and the *nad4L-rnl_1* **(C)**, were aligned with their three closest relatives. The black arrows indicate putative homologous recombination breakpoints based on a HMM 0.95 bayesian inference value. Numbers in parentheses for each *dpo* sequences in the distance-based method refer to the ones tested for each topology in the phylogenetic HMM method.

3 - Supplementary Information (Chapter 4)

Homoplasmy in monosporal cultures arising from crossed-isolates supports the presence of a putative mitochondrial segregation apparatus in *Rhizophagus irregularis*




C
MDM12 protein
Saccharomyces cerevisiae
Zygosaccharomyces rouxii
Vanderwaltozyma polyspora
Eremothecium gossypii
Kluyveromyces lactis
Lachancea thermotolerans
Pichia pastoris
Lodderomyces elongisporus
Debaryomyces hansenii
Candida albicans
Yarrowia lipolytica
Schizosaccharomyces pombe
Dictyostelium discoideum
Rhizophagus irregularis
Cryptococcus neoformans
Ustilago maydis
Postia placenta
Phanerochaete chrysosporium
Laccaria bicolor
Coccidioides posadasii
Emericella nidulans
Penicillium chrysogenum
Aspergillus niger
Botryotinia fuckeliana
Sclerotinia sclerotiorum
Gibberella zeae
Hypocrea jecorina
Neurospora crassa
Magnaporthe grisea
D
MDM10 protein
Postia placenta
Phanerochaete chrysosporium
Laccaria bicolor
Ustilago maydis
Cryptococcus neoformans
Rhizophagus irregularis
Schizosaccharomyces pombe
Coccidioides posadasii
Penicillium chrysogenum
Aspergillus niger
Emericella nidulans
Botryotinia fuckeliana
Sclerotinia sclerotiorum
Neurospora crassa
Magnaporthe grisea
Hypocrea jecorina
Gibberella zeae
Yarrowia lipolytica
Debaryomyces hansenii
Lodderomyces elongisporus
Candida albicans
Pichia pastoris
Lachancea thermotolerans
Kluyveromyces lactis
Eremothecium gossypii
Saccharomyces cerevisiae
Vanderwaltozyma polyspora
Zygosaccharomyces rouxii


E
ACO1 protein
Neurospora crassa 1
Magnaporthe grisea 1
Hypocrea jecorina 1
Gibberella zeae 1
Sclerotinia sclerotiorum 1
Penicillium chrysogenum 1
Aspergillus niger 1
Coccidioides posadasii 1
Lachancea thermotolerans 1
Kluyveromyces lactis 1
Eremothecium gossypii
Zygosaccharomyces rouxii 1
Vanderwaltozyma polyspora 1
Saccharomyces cerevisiae ACO1
Pichia pastoris 1
Lodderomyces elongisporus 1
Candida albicans 1
Debaryomyces hansenii 1
Yarrowia lipolytica 1
Ustilago maydis 1
Cryptococcus neoformans 1
Postia placenta 3
Postia placenta 2
Laccaria bicolor 3
Laccaria bicolor 2
Phanerochaete chrysosporium 1
Schizosaccharomyces pombe 1
Rhizophagus irregularis
Monosiga brevicollis 1
Caenorhabditis elegans 1
Phytophthora sojae
Chlamydomonas reinhardtii
Cryptococcus neoformans 2
Emericella nidulans 2
Aspergillus niger 4
Lachancea thermotolerans 2
Zygosaccharomyces rouxii 2
Vanderwaltozyma polyspora 2
Saccharomyces cerevisiae Paralog
Kluyveromyces lactis 2
Pichia pastoris 2
Lodderomyces elongisporus 2
Debaryomyces hansenii 2
Candida albicans 2
Eremothecium gossypii 2
Yarrowia lipolytica 2
Schizosaccharomyces pombe 2
Sclerotinia sclerotiorum 2
Botryotinia fuckeliana 1
Penicillium chrysogenum 2
Aspergillus niger 2
Emericella nidulans 1
Coccidioides posadasii 2
Neurospora crassa 2
Hypocrea jecorina 2
Gibberella zeae 2
Magnaporthe grisea 2
Ustilago maydis 2
Laccaria bicolor 1
Phanerochaete chrysosporium 2
Postia placenta 1
Cryptococcus neoformans 3
Monosiga brevicollis 2
Aspergillus niger 3
Dictyostelium discoideum 2
F
ABF2 protein
Cryptococcus neoformans 2
Schizosaccharomyces pombe 1
Pichia pastoris 2
Debaryomyces hansenii 1
Candida albicans 3
Lodderomyces elongisporus 2
Saccharomyces cerevisiae HMO1 2
Eremothecium gossypii 4
Kluyveromyces lactis 4
Lachancea thermotolerans 3
Vanderwaltozyma polyspora 4
Zygosaccharomyces rouxii 3
Yarrowia lipolytica 5
Yarrowia lipolytica 3
Yarrowia lipolytica 1
Rhizophagus irregularis
Zygosaccharomyces rouxii 2
Saccharomyces cerevisiae paralog
Vanderwaltozyma polyspora 6
Kluyveromyces lactis 2
Lachancea thermotolerans 1
Vanderwaltozyma polyspora 1
Saccharomyces cerevisiae ABF2
Kluyveromyces lactis 3
Eremothecium gossypii 2
Vanderwaltozyma polyspora 2
Yarrowia lipolytica 2
Zygosaccharomyces rouxii 1
Saccharomyces cerevisiae NHP10 1
Pichia pastoris 1
Candida albicans 2
Debaryomyces hansenii 2
Lachancea thermotolerans 4
Caenorhabditis elegans 1
Dictyostelium discoideum 2
Trichoplax adhaerens 1
Ustilago maydis 1
Cryptococcus neoformans 1
Laccaria bicolor 2
Postia placenta 3
Postia placenta 1

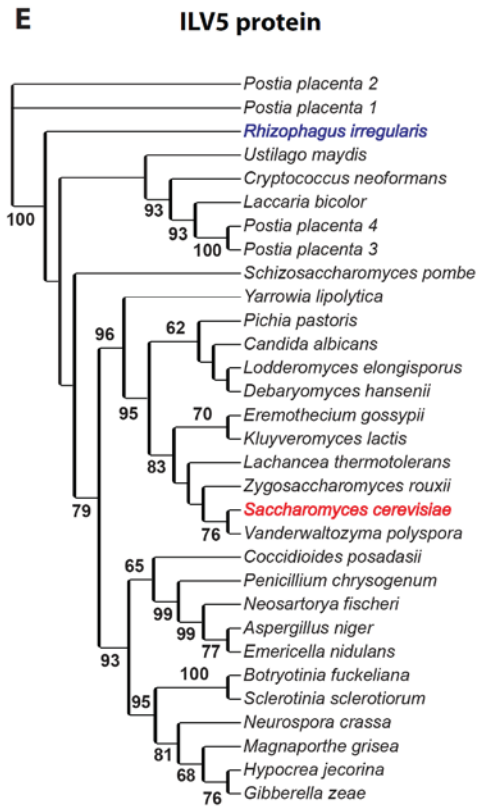


Figure S4.1. Maximum likelihood phylogenetic trees based on seven proteins implicated in the mitochondrial segregation process and nucleoid formation in Saccharomyces cerevisiae, along with their closest orthologs found in fungi and other organisms. Each phylogeny was performed accordingly to its predicted model: the MMM1 protein phylogeny was done using the JTT+I+G model **(A)**, MMM2 with JTT+I+G **(B)**, MDM12 with JTT+G **(C)**, MDM10 with WAG+I+G **(D)**, ACO1 with WAG+I+G **(E),** ABF2 with WAG+I+G **(F)**, and finally the ILV5 phylogeny was performed using the WAG+G model **(G)**. Numbers on branches correspond to bootstrap support values (< 60% cut-off) on 1000 replicates. The *Saccharomyces cerevisiae* orthologs and paralogs are labeled in red, while the *Rhizophagus irregularis* sequences are in blue.

Table S4.1. Isolate-specific primers used to discriminate the three *Rhizophagus irregularis* isolates

Primer	Primer sequence (5' - 3')	Probe sequence (5' – 3')	Amplicon size (bp)	Position of primers in mtDNA		
				R. irregularis DAOM197198	*R. irregularis* DAOM240415	*R. irregularis* DAOM234328
197198F	GCTCATTCTTGAGACTCTGGAAC	CTGCCCCTGCAACTCCCGTA	187	68559-		
197198R	CTCCTTAGCTAGTTCCTCAGTGGT			68746		
240415F	TTCCTAGGGTTGCAGAATCGATC	TCCTTGAATAGGCCGTAACGGCA	199		68960-	
249415R	AGAGATTTGTAGATTCTTTACCAGGA				69160	
234328F	GCTAATTTCCCTTTGTATCACTTGA	AGAAGCAAAGGAGAAAGGGCTT	191			62989-
234328R	GTATCCATTACAATCGTCCATGGTG	GCTT				63180

4 - Supplementary Information (Chapter 5)

Evidences for a population of different mitochondrial genomes and their co-expression in an ancient root-inhabiting fungus

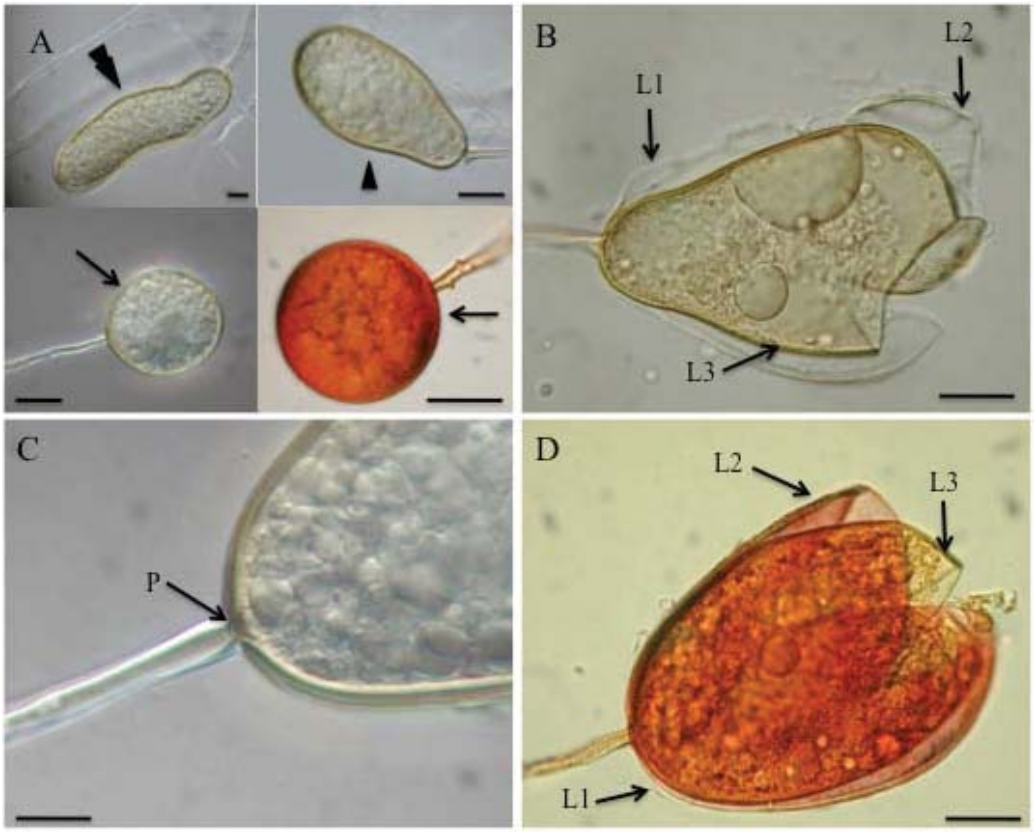


Figure S5.1. Morphological description of *Rhizophaghus irregularis* spores DAOM 242422 under *in vitro* conditions. A) Spores are hyaline to to pale yellow, mostly ovoid (black arrow) 65-70 µm to oblong (arrowhead) to irregular 65-77X96-240 µm (double arrowhead), scale bars=30 µm. **B)** The outermost layer (L1) is mucilaginous and hyaline ~ 1.5 µm thick, the second layer (L2) is rigid, smooth and hyaline ~ 1.5 and L3 is smooth and pale yellow ~ 1.5-3.2 µm. L1 and L2 were closely attached to each other forming a unique shell, easily detachable for the third layer (L3) scale bar=20 µm. **C)** The subtending hyphae (6.5-8 µm) was hyaline and straight, the pore was wide open at the spore-base but sometimes occluded by a funnel-shaped like to curved septum (P), scale bar=15 µm **D)** Spores in Melzer's reagent stained from red-violet in L1-L2 to brown in L3, scale bar=20 µm.

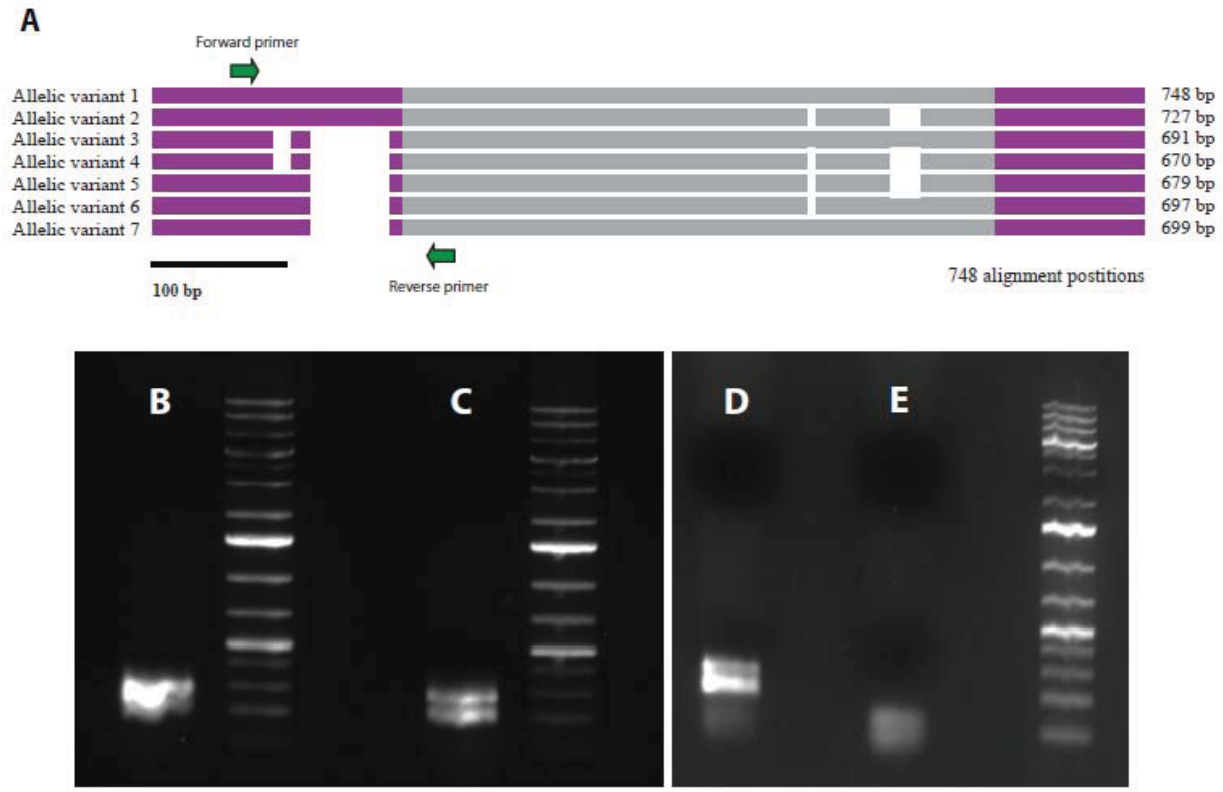


Figure S5.2. Confirmation of the occurrence of length heteroplasmy in the model *R. irregularis* isolate DAOM-197197 originating from three different locations. (A) The seven structural variants found in the *nad4-nad1* intergenic region are shown in an alignment with the primers designed to confirmed heteroplasmy in the different samples (green arrows). **(B-E)** Electrophoresis gel of the PCR reactions performed on *R. irregularis* DAOM-197198 originating from our laboratory located in Montreal-QC-Canada, PremierTech biotechnology located in Rivière-du-loup-QC-Canada and Agrifood Canada located in Ottawa-ON-Canada, respectively, along with the negative PCR control. All DAOM-197198 isolates showed the presence of double bands at the investigated locus, thus confirming the occurrence of heteroplasmy in all strains.

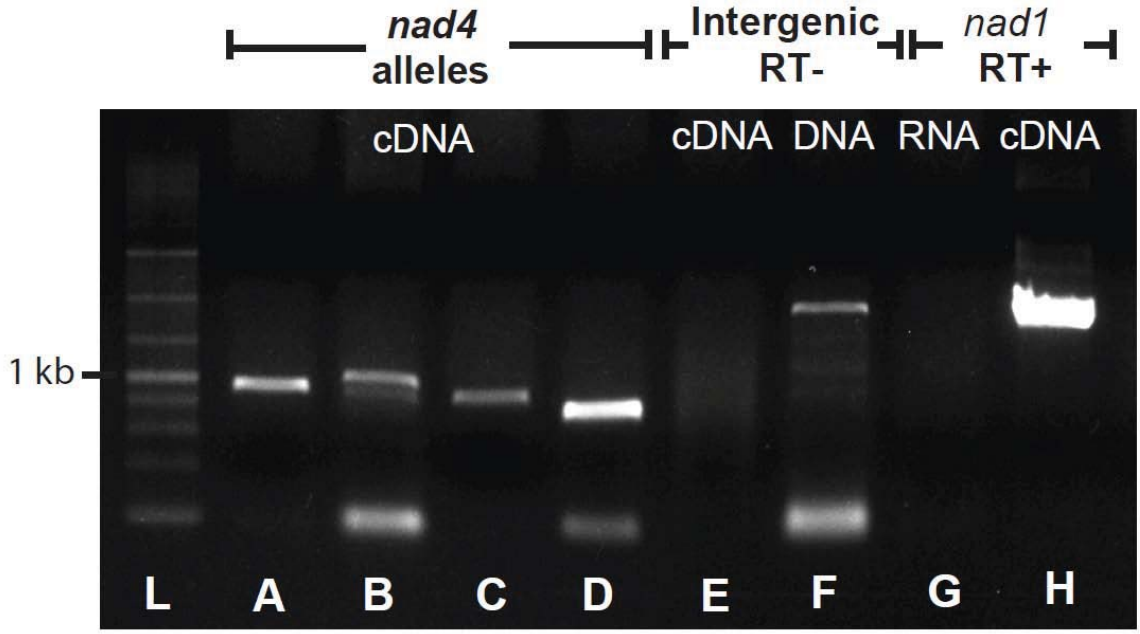


Figure S5.3. Electrophoresis gel of RT-PCR reactions performed on *R. irregularis* DAOM-197198 cDNA showing the expression of the four *nad4* allelic variants. (L) 1 kb ladder, **(A-D)** mRNA expression of all four *nad4* variants, **(E-F)** reverse transcription negative control in a mitochondrial intergenic region on cDNA and DNA templates, respectively, **(G-H)** reverse transcription positive control in the nad1 gene performed on RNA and cDNA templates, respectively.

5 - Bioréacteur *in vivo* stérile pour le maintient des collections de CMA

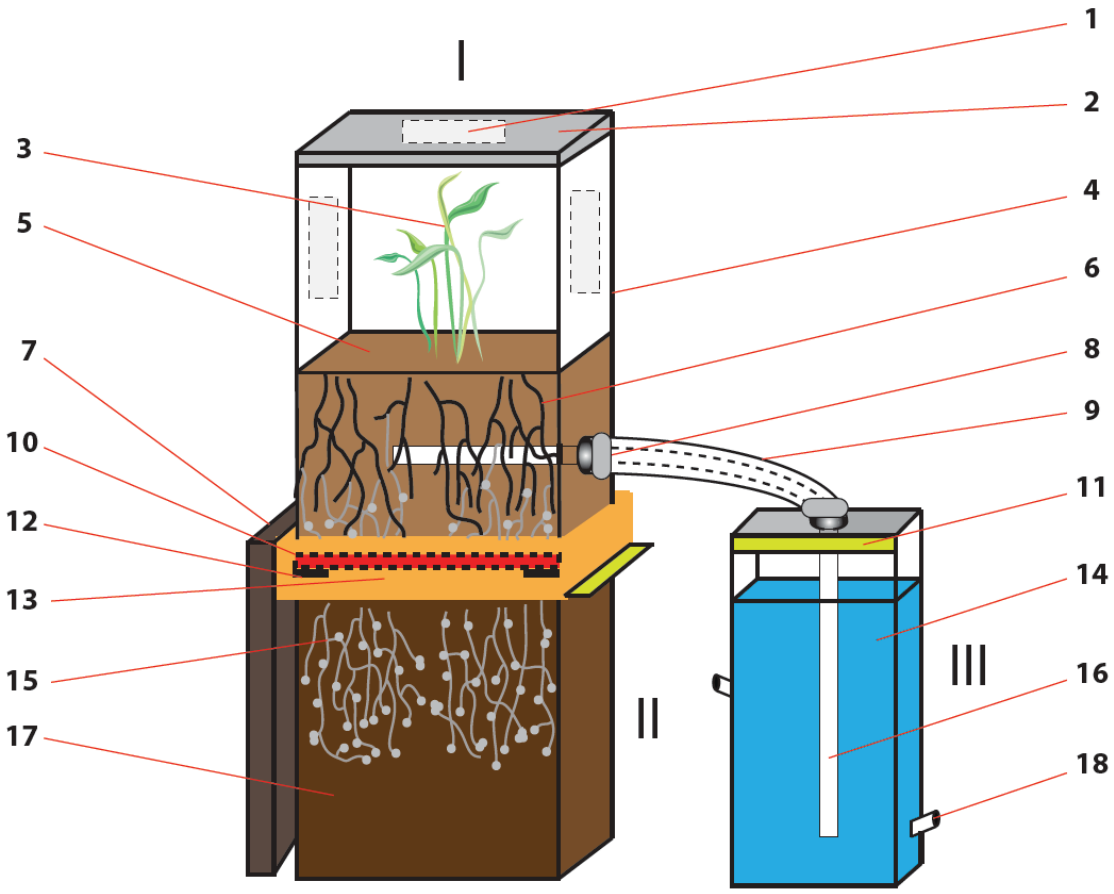


Figure 1. Schématisation du bioréacteur *in vivo* stérile pour cultiver les CMA.

Trois compartiments principaux du système: **I)** Chambre de croissance des plantes et de production de biomasse racinaire. **II)** Chambre de croissance des CMA pour la production de spores. **III)** Chambre d'alimentation en eau et/ou nutriments. Composantes individuelles du système: **1)** membrane permettant les échanges gazeux avec le milieu extérieur, **2)** couvercle autoclavable en polypropylène, **3)** plante au choix de l'utilisateur, **4)** contenant autoclavable en polycarbonate ou polypropylène, **5)** substrat pour les plantes au choix (terre, sable, tourbe, vermiculite, perlite, etc.), **6)** racine de plante, **7)** paroi de support physique au système qui est fixée au coupleur de compartiments (13), **8)** système de vis/écrou faisant le lien entre le tuyau (9) et l'intérieur du compartiment I, **9)** tuyau autoclavable de polypropylène/silicone avec écrou, **10)** membrane de cellophane ou de nylon permettant passage des hyphes de champignons et non des racines de plantes, **11)** système de fixation par glissement ou loquet du compartiment III et du coupleur de compartiments, **12)** excroissance à l'intérieur du coupleur de compartiments (13) permettant d'y déposer la membrane de cellophane/nylon (10), **13)** coupleur de compartiments qui maintient ensemble le compartiment I et II, **14)** eau et/ou solution de nutriments à la

discrétion de l'utilisateur, **15)** spore et hyphes du CMA, **16)** tige en coton, nylon ou polyester traversant le tuyau et pénètre dans le compartiment I permettant l'humidification du système par capillarité, **17)** substrat pour les CMA au choix (terre, sable, tourbe, vermiculite, perlite, apatite, etc.), **18)** becs permettant l'alimentation du compartiment III en liquide et de brancher en série plusieurs réacteurs par l'entremise de tuyaux séparés par des filtres ou le mouvement serait généré par une pompe.

Ce nouveau système de maintient des collections réponds à plusieurs besoins: i) empêche la contamination croisée entre différentes cultures de CMA, puisque l'on ouvre jamais le système, ii) simplifie le maintient d'une collection de CMA par la facilité d'apporter des intrants au système par le bec d'alimentation (18) du compartiment III, iii) possibilité de brancher plusieurs réacteurs en série (18) où l'alimentation serait gérée par un système de pompe, iv) constitue la première possibilité de cultiver les CMA *in vivo* de façon stérile, v) maintient un taux d'humidité constant par apport graduel de l'eau grâce au système de distribution par capillarité (16), vi) offre une solution simple pour la récolte des spores produites par le système symbiotique (par séparation du compartiment II), qui sont exempt de contamination (racines, autres champignons, etc.), vii) permet de cultiver de nombreuses espèces de CMA puisque les conditions du système sont flexibles (plantes utilisées, substrats utilisés, conditions abiotiques, etc.), ce qui n'est pas le cas pour le système de culture *in vitro*, viii) permet une production de spores continue par le système puisque le compartiment II est rechargeable de substrat et que l'apport en nutriment se fait continuellement. De plus, il offre plusieurs avantages: peu de maintient, peu couteux, productivité de spores théoriquement accrues, production en continu, récolte des spores très facile et non destructive, réutilisable, permet de cultiver de nombreuses espèces de CMA, alimentation en intrants peut se faire à la chaîne à l'aide d'un tuyau commun, permet le maintient des cultures stériles. Un brevet provisoire a été déposé pour ce système et un prototype construit. Il est présentement en phase de tests préliminaires, afin de démontrer son efficacité pour remplacer les systèmes actuels de maintient de cultures de CMA.

Group I Intron–Mediated *Trans*-splicing in Mitochondria of *Gigaspora rosea* and a Robust Phylogenetic Affiliation of Arbuscular Mycorrhizal Fungi with Mortierellales

Maryam Nadimi,†,[1] Denis Beaudet,†,[1] Lise Forget,[2] Mohamed Hijri,[1] and B. Franz Lang*,[2]

[1]Institut de Recherche en Biologie Végétale, Département de Sciences Biologiques, Université de Montréal, Montréal, Québec, Canada

[2]Département de Biochimie, Centre Robert-Cedergren, Université de Montréal, Montréal, Québec, Canada

†These authors contributed equally to this work.

*Corresponding author: [redacted]

Associate editor: Andrew Roger

Abstract

Gigaspora rosea is a member of the arbuscular mycorrhizal fungi (AMF; Glomeromycota) and a distant relative of *Glomus* species that are beneficial to plant growth. To allow for a better understanding of Glomeromycota, we have sequenced the mitochondrial DNA of *G. rosea*. A comparison with *Glomus* mitochondrial genomes reveals that Glomeromycota undergo insertion and loss of mitochondrial plasmid-related sequences and exhibit considerable variation in introns. The gene order between the two species is almost completely reshuffled. Furthermore, *Gigaspora* has fragmented *cox1* and *rns* genes, and an unorthodox initiator tRNA that is tailored to decoding frequent UUG initiation codons. For the fragmented *cox1* gene, we provide evidence that its RNA is joined via group I–mediated *trans*-splicing, whereas *rns* RNA remains in pieces. According to our model, the two *cox1* precursor RNA pieces are brought together by flanking *cox1* exon sequences that form a group I intron structure, potentially in conjunction with the *nad5* intron 3 sequence. Finally, we present analyses that address the controversial phylogenetic association of Glomeromycota within fungi. According to our results, Glomeromycota are not a separate group of paraphyletic zygomycetes but branch together with Mortierellales, potentially also Harpellales.

Key words: arbuscular mycorrhizal fungi (AMF), mitochondrial genome, intron evolution, phylogeny, tRNA structure, genetic code.

Introduction

Arbuscular mycorrhizal fungi (AMF) is a group of ubiquitous soil-borne fungi that form symbiotic associations with the majority of vascular plants (Parniske 2008). AMF are obligate biotrophs, that is, they are unable to grow without a host plant that provides them with carbohydrates; in turn, AMF transfer nutrients such as phosphate to the plant (reviewed in Strack et al. 2003). At the cellular level, AMF are characterized by the formation of large, multinucleate hyphae, and asexual spores (e.g., Marleau et al. 2011). Apparently, the genetic segregation of the hundreds of distinct nuclei that are present in these species does not follow canonical but rather population rules, and recent analyses demonstrate substantial sequence variation in certain nuclear genes (Hijri and Sanders 2005; Croll and Sanders 2009; Croll et al. 2009; Boon et al. 2010). It is therefore no surprise that the *Glomus irregulare* nuclear genome project has turned into a sequence assembly nightmare (Martin et al. 2008). Yet in stark contrast, the first complete *Glomus* mitochondrial DNAs (mtDNAs) that have been deciphered recently by 454 sequencing are homogeneous in sequence (Lee and Young 2009 and GenBank #FJ648425; Bullerwell CE, Forget L, Lang BF, unpublished data), that is, genetic segregation of mtDNA is as effective in *Glomus* as in other fungi. In these two cases, long homopolymer stretches that introduce systematic pyrosequencing error are surprisingly absent. In other, more A+T-rich mtDNAs, however, we have observed intolerable levels of 454 sequence error (close to one per 1 kbp sequence on average in a heterolobosean amoeba; Bullerwell CE, Forget L, Lang BF, unpublished data) causing frameshifts in several protein-coding genes. It therefore remains advisable to carefully examine homopolymer-rich sequences for potential error, for instance by resequencing with Sanger technology.

The taxonomic and phylogenetic identity of AMF have been, and continues to be, controversial. Initially assigned to zygomycetes, a fungal taxon that is strongly suspected to be paraphyletic (e.g., Schwarzott et al. 2001; Seif et al. 2006; Hibbett et al. 2007; Liu, Leigh, et al. 2009), AMF have been recently moved into a separate fungal phylum, Glomeromycota (Hibbett et al. 2007). Yet the underlying published phylogenies are controversial and often lack significant statistical support, either due to a limited amount of sequence data (based on one or only few gene sequences), poor taxon sampling, or a combination of both. In many instances, analyses further suffer from potential phylogenetic artifacts such as long-branch attraction (e.g., Felsenstein 1978).

A phylogenetic data set of complete mtDNA sequences is currently restricted to two *G. irregulare* isolates (Lee and

Young 2009 and GenBank #FJ648425) having identical coding sequences. Likewise, a previous phylogenomic analysis with a large number of nuclear genes (Liu, Leigh, et al. 2009) had only limited taxon sampling. Accordingly, phylogenetic analyses with both mitochondrial and nuclear genome data have provided only a tentative answer to the question of where AMF belong within Fungi. In some cases, they both show a weak affinity of Mortierellales with Glomeromycota (Lee and Young 2009; Liu, Leigh, et al. 2009), unsupported however by strict statistical analysis (such as the AU test, Shimodaira 2002). An updated, comprehensive fungal phylogenomic analysis with nuclear sequence data published in 2011 (Ebersberger et al. 2012) nicely summarizes the confusing state of the art, commenting that "at the moment, available data do not allow to confidently attach glomeromycetes to the phylogenetic backbone of the fungi." In the latter phylogenomic analysis, Mortierellales are shown separate from Mucorales, that is, excluding Mortierellales from a monophyletic taxon Mucoromycotina favored by others and in contradiction to conclusions reached in a previous phylogenomic analysis (Liu, Leigh, et al. 2009). Evidently, better taxon sampling of genomic data sets is required to resolve these questions, in particular by adding to both mitochondrial and nuclear gene data sets members of Mortierellales, and AMF lineages that are distant from Glomeraceae.

Our rationale for sequencing the *Gigaspora rosea* mtDNA is that Gigasporaceae are at a large evolutionary distance to *Glomus* species, with clearly distinct morphological characteristics. Gigasporaceae form auxiliary cells in the extraradical mycelium, and giant spores that are usually larger than 200 μm and visible to the naked eye. These atypical spores are formed individually from hyphae and contain funnel-shaped hyphal attachments that extend from a specialized bulbous sporogenous cell. Another motive for sequencing mtDNAs from additional Glomeromycota relates to the origin and distribution of introns, other endonuclease-based mobile elements, and mitochondrial plasmid-like inserts in mtDNAs that are frequent in *Glomus* species (Lee and Young 2009; Lang BF, unpublished data). In fact, one of the group I introns of a glomeromycotan *cox1* gene might have been transmitted to plant mitochondria (Vaughn et al. 1995; Adams et al. 1998; Seif et al. 2005; Lang and Hijri 2009), potentially as part of symbiotic plant–AMF interactions. Group I introns are mobile genetic elements that may insert into intron-less gene copies (i.e., with respect to a given intron insertion point) by a process called intron homing (Colleaux et al. 1986; Lambowitz and Zimmerly 2004). They are further characterized by an RNA that folds into a distinct secondary structure, consisting of up to nine base-paired helical regions (P1–P9) that are involved in bringing respective exons into close proximity and in the splicing reaction itself (Anziano et al. 1982; Michel et al. 1982; Waring et al. 1982). In many but not all cases (Lang et al. 2007), group I introns are self-splicing in vitro, that is, the RNA is a ribozyme capable of catalyzing its own excision from precursor RNA (Cech et al. 1983; Jacquier and Rosbash 1986; Schmelzer and Schweyen 1986; Van der Veen et al. 1986).

Intron splicing usually occurs in *cis*, but a few cases of *trans*-splicing are known, and we provide evidence in this paper that this also applies to a *G. rosea* mitochondrial gene. For organelle genes, *trans*-splicing is usually mediated by group II introns (for a review, see Bonen 2008) but in a few recently discovered instances also by group I introns (Burger et al. 2009; Grewe et al. 2009; Pombert and Keeling 2010; Hecht et al. 2011). Curiously, all known *trans*-splicing group I introns are located in *cox1*, with identical insertion points in the quilwort *Isoetes* and the spikemoss *Selaginella* but also in the evolutionarily distant *Gigaspora* and the metazoan *Trichoplax* (supplementary fig. S1, Supplementary Material online). *Trans*-splicing entails the joining and ligation of discontinuous coding regions that are transcribed separately, that is, located on distinct RNA molecules. For group I and II introns, these separate transcripts may be brought together by interaction of partial (i.e., fragmented) intron RNA sequences that fold into the typical RNA structure that allows splicing to take place. In more complicated, rare instances, the flanking sequences of precursor transcripts alone do not code for the full intron structure, but additional helper RNAs are involved. In the two known instances, three distinct RNA molecules in *trans* are required to form a complete intron RNA structure (Goldschmidt-Clermont et al. 1991; Knoop et al. 1997). In any case, claims for intron-mediated *trans*-splicing should be based on evidence for 1) the presence of a mature RNA that is ligated exactly at predicted exon–intron junctions (e.g., based on multiple sequence alignments) and 2) an inferred intron RNA structure that is complete, matching previously established conservation rules perfectly. Note that not all discontinuous genes require *trans*-splicing. For instance, known fragmented rRNAs remain in separate transcripts, folding into a functional ribosomal structure, but without the requirement for *trans*-splicing (e.g., Boer and Gray 1988; Schnare and Gray 1990).

In this article, we report the complete mtDNA sequence of *G. rosea*, which encodes two fragmented genes, transcripts of one of which undergo group I intron–mediated *trans*-splicing. We further report the results of a phylogenetic analysis of mitochondrial proteins that allows more confident positioning of Glomeromycota within Fungi.

Materials and Methods

Fungal Material

Spores of *G. rosea* (DAOM194757) were extracted from soil of in vivo pot cultures with leek as a host plant, using a wet-sieve cascade (400, 250, and 60 μM). Spores were further purified by centrifugation through a water/50% sucrose step gradient (50-ml tube; 5 min at 5,000 rpm in an Eppendorf 5804 centrifuge; Esch et al. 1994). The resulting spore-containing layer from the upper interface of the 50% sucrose solution was collected, and clean spores were sorted manually under a binocular.

DNA Purification

Spores were suspended in 400 μl of the DNeasy Plant Mini Kit AP1 buffer (Qiagen) and crushed with a pestle in 1.5-ml

microtubes. DNA was purified with the same kit according to the manufacturer's recommendations, with a final elution volume of 40 µl. Purified DNA was stored at −20 °C until use.

RNA Purification

Freshly harvested *G. rosea* spores from in vivo pot cultures were vortexed with 1 ml TRIzol reagent (Invitrogen) and glass beads for 10 min. Other RNA extraction steps followed the manufacturer's protocol. Total RNA was then purified with the RNeasy mini kit (Qiagen), treated with DNase using the Turbo DNA free kit (Applied Biosystem), and stored at −20 °C. RNA quality was checked by electrophoresis on a 1.5% agarose gel, stained with ethidium bromide, and visualized under UV light.

Reverse Transcriptase–Polymerase Chain Reaction

We have used reverse transcriptase–polymerase chain reaction (RT-PCR) to test for intron excision, *trans*-splicing, and the presence of (apparently nonintron) sequence inserts in rRNAs. In all experiments, PCR amplifications were performed either with or without prior conversion of RNA into cDNA (i.e., addition of RT). As an additional control, genomic DNA was PCR amplified.

To test for potential *trans*-splicing of *cox1* and *rns*, primers were designed within the coding regions that flank predicted break points. In case of *cox1*, primers were located in *cox1* exon 4 (5′-CTGTATTGGTCACTGCCGT-3′) and exon 5 (5′-AAAGGCTGAAATAACATGGCT-3′). For *rns*, the two respective primers were 5′-ACCTTGATC-CAGCCAACTAGA-3′ and 5′-CACACTATCGTATCTCAG-CGTC-3′. To test for functionality of the RT-PCR assay (positive control), an internal region of *cox1* exon 4 was amplified (primers 5′-CTCTAGCAGGGACCCAGTC-3′ and 5′-CCGGATCATAGAAGCAGGT-3′).

To verify whether several inserts in *rnl* and *rns* remain at the RNA level (suspected because of an apparent lack of conserved intron RNA secondary structure), primers were designed within the flanking regions of these inserts. For *rns*, the primers were 5′-GTGCATTGTCATCA-CAGGTG-3′, 5′-CGAGTTACAGAGCACAGTTCG-3′, and 5′-GTCTCGTAACAAGCCTCCTTAAC-3′ (the latter is located inside the predicted insert). In the case of *rnl*, five primers were designed to test inserts 3, 4, and 5 (5′-AAG-TAGAGGCTCCAGAAGCAG-3′, 5′-GAAGGGTTTCATGA-GTAAGGTGA-3′, 5′-CAGGTCTGCAGGTTCACG-3′, 5′-CGTTCAGTCTTAACACTTGGC-3′, and 5′-CCCCTTTTT-AGTGCCGC-3′). We further verified that the two predicted *rnl* introns are indeed excised at the RNA level, using primer (5′-GGGCGCGTCTGTTTACTTA-3′) upstream of the first intron (group II), in conjunction with a primer downstream of the second group IB intron (5′-CGAGTACCGGTACCAGAGTAGGT-3′).

For samples undergoing cDNA synthesis, relevant primers (0.1 µl of 5 µM stock solutions), 2 µl RNA solution (~1 to 10 ng), and 8 µl of RNase-free water were mixed, denatured at 75 °C for 2 min, and immediately placed on ice for 2 min. Then, 4 µl RT buffer, 2 µl of each dNTP (10 mM stocks), 4 µl water, and 0.1 µl (1.5 U) AMV RT enzyme (Roche Applied Science) were added. After incubation at 45 °C for 45 min in the reaction mix, the RNA was denatured again at 75 °C for 2 min, placed on ice, fresh AMV RT enzyme was added, and the samples incubated for another 45 min at 45 °C. Reaction products were stored at −20 °C until use.

Subsequent PCR amplifications were carried out with the Expand High Fidelity PCR system (Roche Applied Science) in a total volume of 40 µl, containing either cDNAs (4 µl) or positive and negative controls (0.4 µl RNA or 0.2 µl genomic DNA), primers (4 µl each from 5 µM stocks), 4 µl PCR buffer (10×, without $MgCl_2$), 3.2 µl $MgCl_2$ (25 mM), 0.4 µl (1.5 U) DNA polymerase, 4 µl dNTPs (2 mM), and 1 µl RNase (5 µg/ml). Cycling parameters were 4 min at 95 °C, followed by 33 cycles of 15 s at 95 °C, 20 s at a 53 to 57 °C temperature gradient, 20 s at 72 °C, and a final elongation step of 4 min at 72 °C. PCR products were separated by electrophoresis in a 1.5% (w/v) agarose gel and visualized with ethidium bromide under UV light.

Sequencing, Assembly, and Gene Annotation

Gigaspora rosea total DNA was sequenced by 454 shotgun technology, and the resulting 587,881 reads were assembled with Newbler (Genome Quebec Innovation Center, McGill University, Montreal; Titanium Flex, ½ plate). The resulting 26 mitochondrial contigs were combined into a single circular-mapping DNA by PCR amplification of total DNA and Sanger sequencing of the PCR fragments. Regions with potential 454 sequence error (due to homopolymer motifs) were resequenced in the same way. Gene annotation was performed with MFannot (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl), followed by manual inspection and addition of missing gene features. Mfannot predicts group I and II introns, tRNAs, RNase P RNA, and 5S rRNA with Erpin (Gautheret and Lambert 2001) as a search engine, based on RNA structural profiles developed in house. Exons of protein-coding genes are inferred in a first round with Exonerate (Slater and Birney 2005) and then for less well-conserved genes with HMM profile searches (Eddy 2008; based on models for all known mtDNA-encoded proteins). Miniexons that are not recognized by Exonerate and that may be as short as 3 nucleotides are inferred by the presence of missing conserved protein regions (with reference to multiple protein alignments including other species) and orphan introns. The precise placement of small exons is based on the best fit of HMM protein profiles and on the fit with conserved nucleotide sequence profiles of respective (group I or II) exon–intron boundaries. Genes encoding the small and large rRNA subunits are predicted with HMM profiles that cover the most highly conserved domains, allowing precise predictions of the small subunit rRNA termini but only approximate positioning of large subunit rRNA ends. The latter termini, as well as the precise exon–intron structure of rRNA genes,

Table 1. Gene and Intron Content in Selected Fungal mtDNAs.

	Genes and Introns										
Species	*rnl, rns*	*atp6, 8, 9*	*cob*	*cox1, 2, 3*	*nad1–6*[a]	*trn A-W*	*rnpB*	*rps3*	ORFs[b]	Intron I[c]	Intron II[c]
Gigaspora rosea	2	3	1	3	7	25	0	0	4	13	1
Glomus irregulare 494	2	3	1	3	7	25	0	0	8	24	0
G. irregulare 197198	2	3	1	3	7	25	0	0	8	26	0
Smittium culisetae	2	3	1	3	7	26	1	1	3	14	0
Mortierella verticillata	2	3	1	3	7	28	1	1	7	4	0
Rhizopus oryzae	2	3	1	3	7	23	1	0	4	9	0
Allomyces macrogynus	2	3	1	3	7	25	0	1	4	26	2
Saccharomyces cerevisiae[d]	2	3	1	3	0	25	1	1	3	9	4

[a] Includes *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, and *nad6*.
[b] Only ORFs greater than 100 amino acids in length are listed, not including intronic ORFs and *dpo* and *rpo* fragments.
[c] Intron I and Intron II denote introns of group I and group II, respectively.
[d] *S. cerevisiae* strain FY 1679 (Foury et al. 1998).

are predicted manually (using comparative structure modeling principles). Furthermore, RT-PCR experiments are evaluated to distinguish between true introns and sequence insertions that are not removed by splicing. In any case, automated annotations are curated manually to account for MFannot warnings (e.g., potential *trans*-spliced genes, gene fusions, frameshifts, alternative translation initiation sites, failure to identify miniintrons, etc.), to correct potential errors, and to find features that are not recognized by the automated procedures. The annotated sequences (mtDNA and plasmid) are available at GenBank (JQ693395 and JQ693396).

Phylogenetic Analysis

The data set contains 13 mitochondrion-encoded proteins (*Cox1*, *2*, *3*, *Cob*, *Atp6*, *9* and *Nad1*, *2*, *3*, *4*, *4L*, *5*, *6*) and includes sequences from all zygomycetes for which complete mtDNA sequences are available. Protein collections were managed and automatically aligned, trimmed, and concatenated with Mams (developed in house; Lang BF, Rioux P, unpublished data). Mams uses Muscle (Edgar 2004) for an initial alignment, followed by a refinement step with HMMalign (S. Eddy; http://hmmer.janelia.org). The final data set contains 37 taxa and 3,664 amino acid positions and is available from the authors upon request.

Phylogenetic analyses were performed by Bayesian inference (PhyloBayes; Lartillot and Philippe 2004) using the CAT+Gamma model, six discrete categories, four independent chains, 10,000 cycles (corresponding to ~550, 000 generations), and the −dc parameter to remove constant sites. The site removal reduces the total number of amino acid positions to 3048. The first 1,700 cycles were removed as burn-in. The robustness of internal branches was evaluated based on 100 jackknife (60%) replicates, as modeling of duplicated sequence sites generated by bootstrap analysis is problematic with the Bayesian approach.

Maximum likelihood (ML) analysis was performed with RaxML-HPC v7.2.2 (Stamatakis 2006), using the LG model (PROTGAMMALGF), and the fast bootstrapping option (100 replicates).

Results and Discussion

Comparison of *G. irregulare* and *G. rosea* Mitochondrial Genomes

The mitochondrial genome of *G. rosea* has been completely sequenced by a combination of 454 sequencing from total DNA and filling of the remaining 26 gaps (due to low sequence coverage and 454 sequence error) by PCR amplification and Sanger sequencing. The size of *G. rosea* mtDNA is relatively large (97,349 bp) when compared with those of the two *G. irregulare* isolates (70,606 and 70,799 bp) and contained several homopolymer stretches that required correction of the 454 data by traditional Sanger sequencing. In addition, paired-end sequencing was performed to avoid misassembly of the genome due to the presence of long sequence repeats. Whenever sequence repeats occur in a genome, more than one genome conformation may exist due to recombination events, but we have no evidence that this occurs in *G. rosea*. The large genome size of *G. rosea* is essentially due to extended intergenic regions, not a variation in the number or size of introns (table 1), or the presence of mitochondrial plasmid insertions that are otherwise common in fungi.

The high number of initial sequencing gaps was surprising as according to our experience, the same amount of total 454 sequence information resulted in complete or almost complete coverage of *Glomus* spp. mtDNA. Among possible explanations, a substantial contamination with DNA from foreign species seems unlikely, as spores were carefully cleaned and sorted individually prior to extraction of total DNA. It remains possible either that the copy number of mtDNA is lower in *G. rosea* than in other AMF species, that its nuclear genome is much more complex, or that the spores harbor endosymbionts accounting for a substantial increase in total DNA relative to mtDNA content. Given the available sequence data, we have no reason to favor any of these scenarios.

As in most other fungi, the *G. rosea* mtDNA maps as a circular molecule (fig. 1) but is likely organized as a linear multimeric concatamer in vivo (Bendich 1996). To allow for easier genome comparison, we have opened the circle at a previously introduced standard position, upstream of *rnl*. Conservation of mitochondrial gene order between

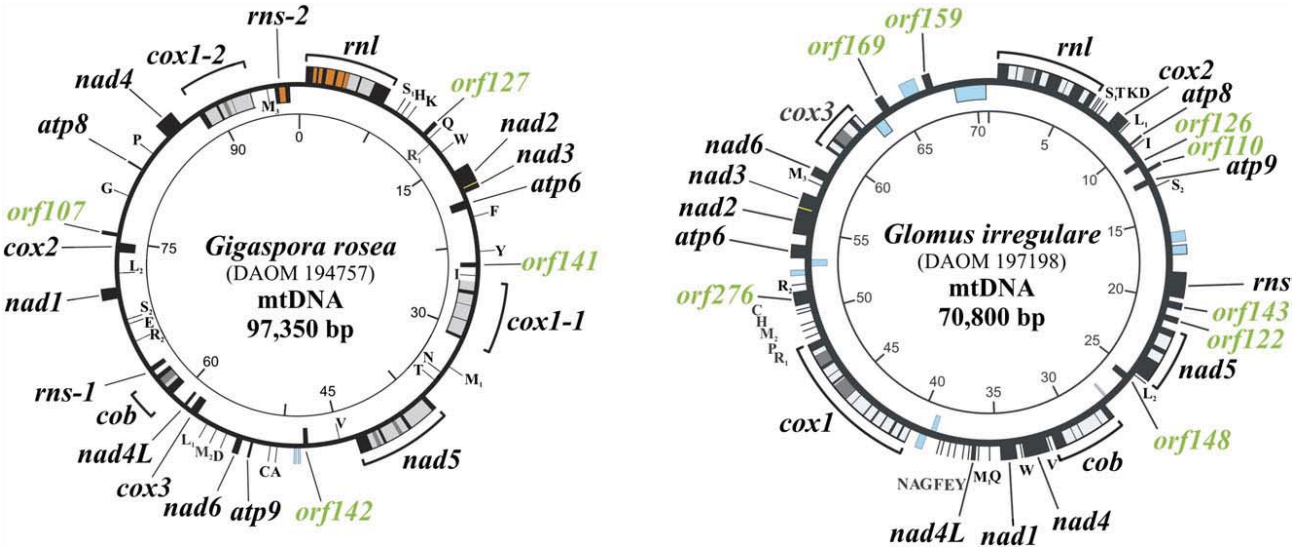


Fig. 1. Comparison of *Gigaspora rosea* and *Glomus irregulare* mitochondrial genomes. The circular-mapping genomes of *G. rosea* and *G. irregulare* were opened upstream of *rnl* to allow for easier comparisons. Genes on the outer and inner circumference are transcribed in clockwise and counterclockwise direction, respectively. Arcs indicate coding regions interrupted by introns. Gene fragments in *G. rosea* are numbered −1 and −2. Boxes of coding regions are filled black, intron ORFs dark gray, and introns light gray. Regions with similarity to mitochondrial plasmid-like DNA polymerase are marked light blue and insertions in *rns* and *rnl* that are not excised from the rRNA (i.e., are not introns) in orange. Gene and corresponding product names are *atp6*, ATP synthase subunit 6; *cob*, apocytochrome b; *cox1–3*, cytochrome c oxidase subunits; *nad1–4*, *4L*, *5–6*, NADH dehydrogenase subunits; *rnl*, *rns*, large and small subunit rRNAs; A-W, tRNAs, the letter corresponding to the amino acid specified by the particular tRNA.

Gigaspora and *Glomus* is virtually nonexistent and, in contrast *to G. irregulare*, genes are encoded on both strands of the *G. rosea* mtDNA. These genes belong to the basic fungal set (table 1), including all required tRNAs. Three tRNAs have a CAU anticodon, one each for initiator and methionine elongator tRNAs and a third one in which a potential lysidine modification of the anticodon C residue would allow decoding of AUA as isoleucine (e.g., Muramatsu et al. 1988; Weber et al. 1990). Notably, the predicted initiator methionine tRNA exhibits structural features in the anticodon loop that never occur in orthodox tRNAs, entailing a most unorthodox modification of decoding preferences (for more details, see below).

As in *G. irregulare* (Lee and Young 2009), genes for the ribosomal protein *rps3* and the RNA component of mitochondrial RNase P (*rnpB*) are absent, although they occur in most zygomycetes (Seif et al. 2005). However, whereas *rps3* may have been transferred to the nuclear genome, this inference remains more contentious for *rnpB* as it would require targeting of a relative large RNA molecule from the cytoplasm back into mitochondria. An alternative solution would be a nucleus-encoded protein-only RNase P activity that replaces the ribozyme, as it occurs in animal and plant mitochondria (Holzmann et al. 2008; Gobert et al. 2010).

Mitochondrial Plasmid Insertions

Mitochondrial autonomously replicating plasmids are widespread. They may carry their own genes for DNA replication (*dpo*) and transcription (*rpo*), and in some cases, plasmids are known to integrate into mtDNA. There is no biochemical or genetic evidence that complete or fragmented versions of *dpo* and *rpo* are functional when inserted into mtDNA; in fact, plasmid insertion into mtDNAs is followed by rapid genome reorganization and loss (e.g., Bertrand et al. 1986; Vierula and Bertrand 1992; Baidyaroy et al. 2000; Hausner 2011), confirming the view that *dpo* and *rpo* are required for plasmid replication and transcription only.

We have used our shotgun data containing both mitochondrial and nuclear sequence reads to search for plasmid-related sequences. *G. irregulare* mtDNA carries many more and larger plasmid-related DNA polymerase (*dpo*) gene fragments than *G. rosea* (filled light blue boxes; two tiny fragments in *G. rosea* are at map position ~50 kbp; fig. 1). In addition, *G. rosea* has an apparently complete, plasmid-like *dpo* (but no *rpo*) in a 3,582-bp contig. Given its relatively high sequence coverage with respect to nuclear genes, this element is likely a freestanding plasmid that only encodes *dpo*. We have not succeeded in demonstrating a circular plasmid version by PCR amplification (including the use of a dedicated long PCR kit; data not shown) and therefore assume that the plasmid is linear. Its cellular location (cytoplasmic or mitochondrial) remains unknown. The *dpo* gene is translated with the standard genetic code, but this would be compatible with both a cytoplasmic and mitochondrial location (for a more detailed discussion of the mitochondrial genetic code, see below). Whatever its cellular location, the *G. rosea* plasmid *dpo* is most closely related to mitochondrial *dpo* fragments, in both *Glomus* and *Gigaspora*, belonging to the family of plasmids that repeatedly invade mitochondrial genomes. A freestanding form of a mitochondrial plasmid has so far not been described for glomeromycotan species.

Evolution of the Mitochondrial Genetic Code in Glomeromycota

Some members of the paraphyletic zygomycete lineages, such as *Rhizopus oryzae*, but also the rapidly evolving *Smittium culisetae*, have retained the standard genetic code for the regular set of mitochondrial protein-coding genes, a trait inherited from their bacterial ancestors (Seif et al. 2005). However, *Mortierella verticillata* reassigns UGA "stop" codons as tryptophan, one each in *nad3* and *nad4*. Likewise, UGA(Trp) codons are also present in the *S. culisetae* group I intronic open reading frame (ORF)283 and ORF248, encoding homing endonucleases of the LAGLI/DADG type (Seif et al. 2005).

In Glomeromycota, deviations from the standard genetic code also vary. A few clear-cut instances of UGA(Trp) codons are found in the two published *G. irregulare* mtDNAs but not in *G. rosea*. In turn, translation initiation with either AUG(Met) or GUG(Met) leads to seven proteins in *G. rosea* (i.e., a surprising 50% of standard protein-coding genes) that are severely truncated at their amino-terminal. In these instances, sequence similarity at the protein level extends further upstream to potential UUG initiation codons (supplementary fig. S2, Supplementary Material online).

UUG initiation is infrequent in fungal mitochondria (e.g., an isolated instance in *Mortierella*; GenBank #AY863211; supplementary fig. S2, Supplementary Material online). It does occur in bacteria where initiation of translation depends on a specific initiator methionine tRNA that recognizes AUG, GUG, and UUG, in descending order of efficiency. In the case of UUG initiation, only the second and third codon positions (U and G) are able to interact directly with the tRNA anticodon (i.e., the C and A of the CAU anticodon), accounting for the very low effectiveness of UUG as an initiation codon, despite stabilization and precise positioning of translation initiation codons by Shine–Dalgarno sequence motifs. These motifs occur at a defined distance upstream of initiation codons, yet in mitochondria, they are only known from the jakobid flagellate *Reclinomonas americana* (Lang et al. 1997). How then are UUG codons recognized in mitochondria of *G. rosea*?

A mechanism for more precise positioning of translation starts is known from mitochondria of monoblepharidalian chytrid fungi and the sea anemone *Metridium senile*, in which almost every protein-coding gene has a guanosine residue upstream of the predicted AUG or GUG start codons (Bullerwell, Forget, et al. 2003). This conserved G residue converts the triplet into a more stable quadruplet initiation codon, correlating with the presence of an unorthodox cytosine residue at position 37 in the anticodon loop of the initiator tRNAs. This feature permits a more stable 4-bp interaction between CAUC anticodons and quartet GAUG/GGUG codons. Inspection of *G. rosea* sequences does not confirm the presence of quadruplet initiation codons, yet its predicted initiator tRNA is highly unorthodox, with a G residue at position 32 that is otherwise always a pyrimidine (for a recent review on mitochondrial tRNA structure, see Lang et al. 2011). In fact, in a 3D structure, position G-32 of the anticodon loop is in close vicinity to both the U-36 of the anticodon and the first U of UUG initiation codons. Interaction of G-32 with both U residues would thus stabilize the U–U pair. In other words, all three nucleotides of the UUG codon are capable of forming stable interactions with this unorthodox initiator tRNA, implying that translation initiation with UUG becomes as effective as with AUG: in accord with the high incidence of inferred UUG translation initiation codons in *G. rosea* mitochondria.

Robust Phylogenetic Association of Glomeromycota with Mortierellales

The Glomeromycota are part of zygomycetes, a phylogenetically heterogenous (paraphyletic) taxon (Seif et al. 2005; James et al. 2006; Liu, Steemkamp, et al. 2009). According to a current proposal, Glomeromycota constitute an independent fungal phylum (Hibbett et al. 2007), yet phylogenomic analyses with nuclear and mitochondrial genes (e.g., Seif et al. 2006; Liu, Leigh, et al. 2009) do not support this view. When the complete mtDNA data of *G. irregulare* are included in the analysis, this fungus associates with *Mortierella* (Lang and Hijri 2009; Lee and Young 2009): an affinity that was also observed with a nuclear data set (Liu, Leigh, et al. 2009), although in both cases based on poor taxon sampling and without significant statistical support. Evidently, additional zygomycete nuclear or mitochondrial genome data are required to address this question and demonstrate beyond reasonable doubt that Glomeromycota merit the status of an independent phylum.

When *G. rosea* data are included in a data set containing all standard mtDNA-encoded proteins, Glomeromycota branch with *Mortierella* as noted above, but with better statistical support (PP 0.99; 66% jackknife value [JV]; fig. 2). In addition, *S. culisetae*, a rapidly evolving species belonging to Harpellales (grouping at a basal position with ML and the applied global model of protein evolution; data not shown), occupies a sistergroup position with Glomeromycota when PhyloBayes and the more realistic CAT model are used for phylogenetic inference (modeling site-wise heterogeneity of amino acid positions, one of the major sources of systematic phylogenetic error in global models) (Lartillot and Philippe 2004; Lartillot et al. 2007). However, the inclusion of *Smittium* in a monophyletic group including *Mortierella* and the two Glomeromycota draws only low statistical support (a posterior probability value of 0.68% and 43% JV), which we attribute to a combination of low phylogenetic signal (use of a single representative for Harpellales) and conflicting long-branch attraction (because rapidly evolving) toward a basal position in the tree. When removing Smittium from the data set, the support for Mortierella forming a monophyletic group with Glomeromycota increases substantially (PP 1.0; 93% JV)

In summary, our results further support the view that Glomeromycota and Mortierellales are sister taxa. This

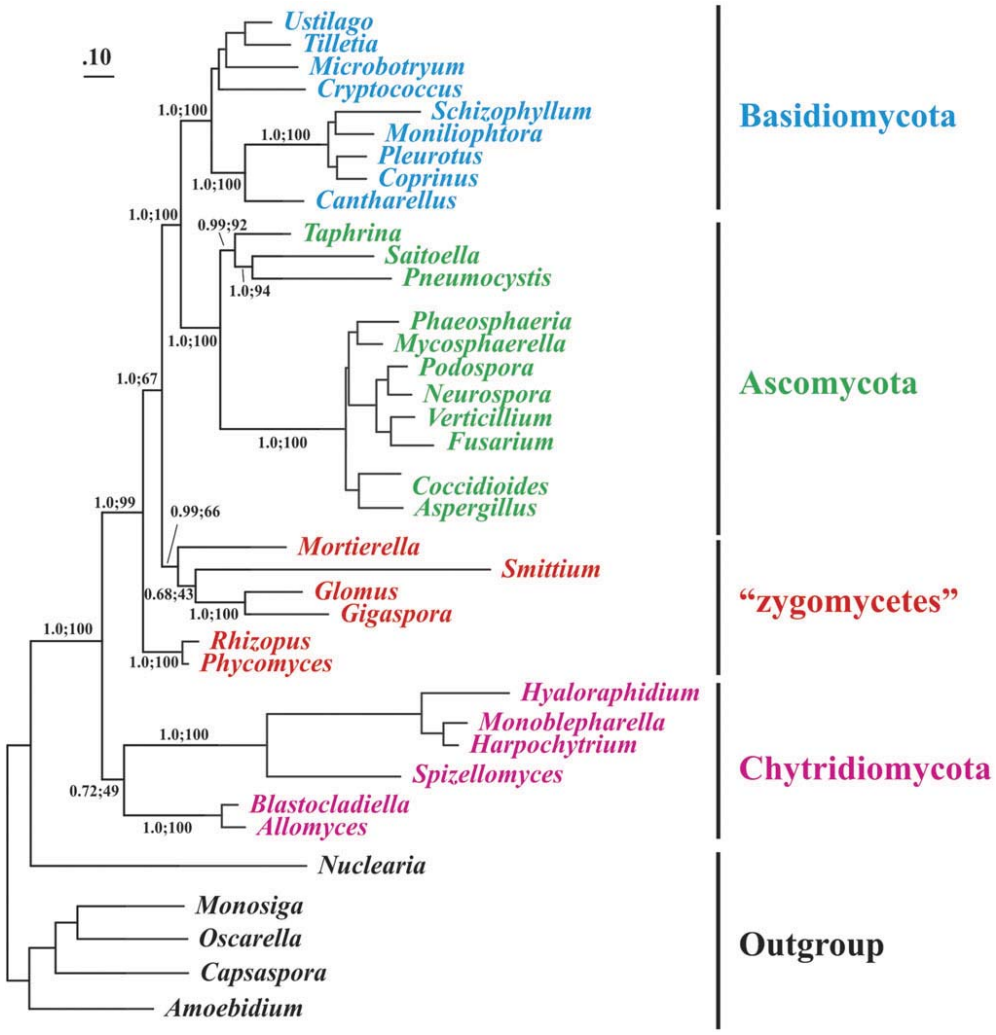


Fig. 2. Phylogenetic positioning of Glomeromycota with mitochondrial protein data. The tree shown was inferred with PhyloBayes, the CAT+Gamma model and six discrete categories (Lartillot and Philippe 2004) based on 13 concatenated proteins. The first number at branches indicates posterior probability values of a PhyloBayes analysis with four independent chains and the second number jackknife supports values. ML inference predicts a similar tree (not shown), except for *Smittium* and Blastocladiales.

inference and the potential inclusion of Harpellales in the same monophyletic group remain to be tested with extended species sampling across "zygomycetes." In particular, representatives of Harpellales with more moderate evolutionary rates need to be identified and data from Mortierellales and Glomeromycota added. Contrary to our previously more pessimistic assessment (Lang and Hijri 2009), broad taxon sampling of mtDNAs alone may turn out to be sufficient to resolve phylogenetic relationships of Glomeromycota with confidence. Whatever the outcome, checking the consistency of mitochondrial phylogenies with phylogenomic analyses based on nuclear gene sequences will be important as a means of informing us about hidden phylogenetic artifacts.

Group I Intron–Mediated *Trans*-splicing of *G. rosea cox1*

The *G. rosea cox1* gene (encoding subunit 1 of respiratory chain complex IV) is broken up into two fragments that are located at a distance of ~30 kbp, encoded on the same strand of the mitochondrial genome. Formally, a *cox1* gene with a 30 kbp intron could be postulated, but this scenario is most unlikely for the following reasons. The two *cox1* fragments are separated by a total of 15 genes, including several tRNAs that are on the same strand. It is known that tRNAs are rapidly and efficiently processed from RNA precursors, so that a 30-kb RNA precursor would be rapidly fragmented within the large intervening region. In addition, the *cox1* break points correspond exactly to a position where *G. irregulare* carries a group I intron, favoring the hypothesis of group I intron–mediated *trans*-splicing.

Examples of group I intron–mediated *trans*-splicing have been reported previously (Burger et al. 2009; Grewe et al. 2009; Pombert and Keeling 2010; Hecht et al. 2011), which in all three cases occur in the most intron-rich mitochondrial gene, *cox1*. The insertion points of *trans*-spliced group I introns vary (supplementary

Table 2. Presence (●) or Absence (○) at Cognate Insertion Site of the Shared Introns between *Gigaspora* and *Glomus* in Representatives of Other Fungi.

Species	Accession	Introns								
Glomeromycota										
Glomus irregulare	FJ648425	*rnli1939*	*nad5i417*	*nad5i717*	○	*cobi429*	*cox1i386*	*cox1i731*	*cox1i867*	*cox1i1108*
Gigaspora rosea		●	●	●	*nad5i954*	●	●	±	●	●
Basidiomycota										
Cryptococcus neoformans	NC004336	○	○	○	○	○	○	○	○	○
Pleurotus ostreatus	EF204913	○	○	○	○	○	●	○	○	●
Ascomycota										
Pneumocystis carinii	GU133622	○	○	○	○	○	○	○	○	○
Podospora anserina	NC001329	○	○	○	○	○	○	○	●	●
Zygomycota										
Mortierella verticillata	AY863211	○	○	○	○	○	○	○	○	○
Rhizopus oryzae	AY863212	○	○	○	○	○	●	○	○	○
Chytridiomycota										
Monoblepharella	AY182007	○	●	●	○	○	○	○	○	●
A. macrogynus	NC001715	○	○	●	○	●	●	○	●	●

NOTE.—The representation ± indicates a partial group I intron secondary structure. The gray columns represent the *cox1* intron that is *trans*-spliced in *G. rosea*, and the *nad5* intron believed to be involved in the *trans*-splicing. Introns were named according to Dombrovska and Qiu (2004). Shared intron positions are based on the coding sequence of the gene in which they are inserted in the first Glomeromycota mitochondrial genome published (Lee and Young 2009). Unique introns to *G. rosea* are named according to their positions in the gene in which they were identified. Of the 13 group I introns in *G. rosea* and the 24 in *G. irregulare* 494, only seven are shared at cognate insertion site. Note that the *G. rosea* nad5i954 intron that may complement a functional secondary structure of the *trans*-spliced *cox1* intron is not present in *G. irregulare* and in the other represented fungal species.

fig. S1, Supplementary Material online; table 2), with exceptions in the related species *Isoetes* and *Selaginella* and in *G. rosea* and *Trichoplax* where introns occupy the same (otherwise frequent) insertion point. In the published examples, a complete canonical group I intron RNA secondary structure can be inferred by pairing the flanking portion of the discontinuous exons and is confirmed by use of group I intron prediction algorithms (Lang et al. 2007; Beck and Lang 2009). Yet in *G. rosea*, certain structural core elements appear to be either missing or to deviate significantly from established rules. As other group I introns in *G. rosea*, protein-coding genes do not display many unusual features, and as a catalytically conserved intron RNA core is essential for intron excision, we predict the use of an additional third helper RNA fragment in *trans*, similar to the situation of fragmented introns in *Chlamydomonas* (Goldschmidt-Clermont et al. 1991) and *Oenothera* (Knoop et al. 1997). According to our assessment, the third intron in the *nad5* gene matches perfectly the two partial discontinuous intron sequences, completing a comprehensive intron RNA secondary structure (fig. 3*A* and *B*).

To provide evidence that the mitochondrial *cox1* is not a pseudogene in *G. rosea* but indeed *trans*-spliced in vivo, we have conducted RT-PCR experiments, using primers located in the respective flanking exons. Sequencing of the resulting PCR product confirms that the *cox1* exons are accurately ligated in vivo (fig. 3*C*). Sequencing of the PCR product confirmed sequence identity with the respective *cox1* exons, without any sequence modification. In the placozoan animal example as mentioned above, *trans*-splicing combines with an additional RNA editing step (Burger et al. 2009). No PCR product was obtained with total genomic DNA, ruling out the possibility that a contiguous *cox1* gene exists in the nuclear genome.

Gene for the Small Subunit rRNA in Two Pieces

The other fragmented *Gigaspora* gene is *rns*, coding for the small subunit rRNA. In similar cases (including both the large and the small subunit rRNAs), *trans*-splicing has never been observed, and break points usually occur in variable loops of the rRNA secondary structure rather than in highly conserved regions as in protein-coding genes (Lang 1984). Secondary and tertiary interactions between the unligated rRNA fragments are sufficient for folding into a complete rRNA structure and formation of functional ribosomes. According to our rRNA secondary structure model, the breakpoint of *G. rosea* is in a variable loop (fig. 4). RT-PCR experiments did not produce DNA fragments of the predicted size and sequence (data not shown), and we explain a few incongruent RT-PCR sequences that we obtained by a low level of template-switching activity of RT. We have also been unsuccessful in folding the flanking regions into a bona fide intron RNA structures (either by inference with intron models that are part of the MFannot annotation or manually).

How Many True Introns in *rns* and *rnl*?

Modeling of rRNA structures and prediction of introns are problematic, without confirmation at the RNA sequence level. It is well known that mitochondrial rRNAs may carry inserts in variable regions of the structure that are not excised at the RNA level (e.g., Lang et al. 1987). In *G. rosea rns* and *rnl*, we find numerous inserts of this kind (figs. 2 and 4) that do not carry evidence for sound group I or group II intron structures. RT-PCR

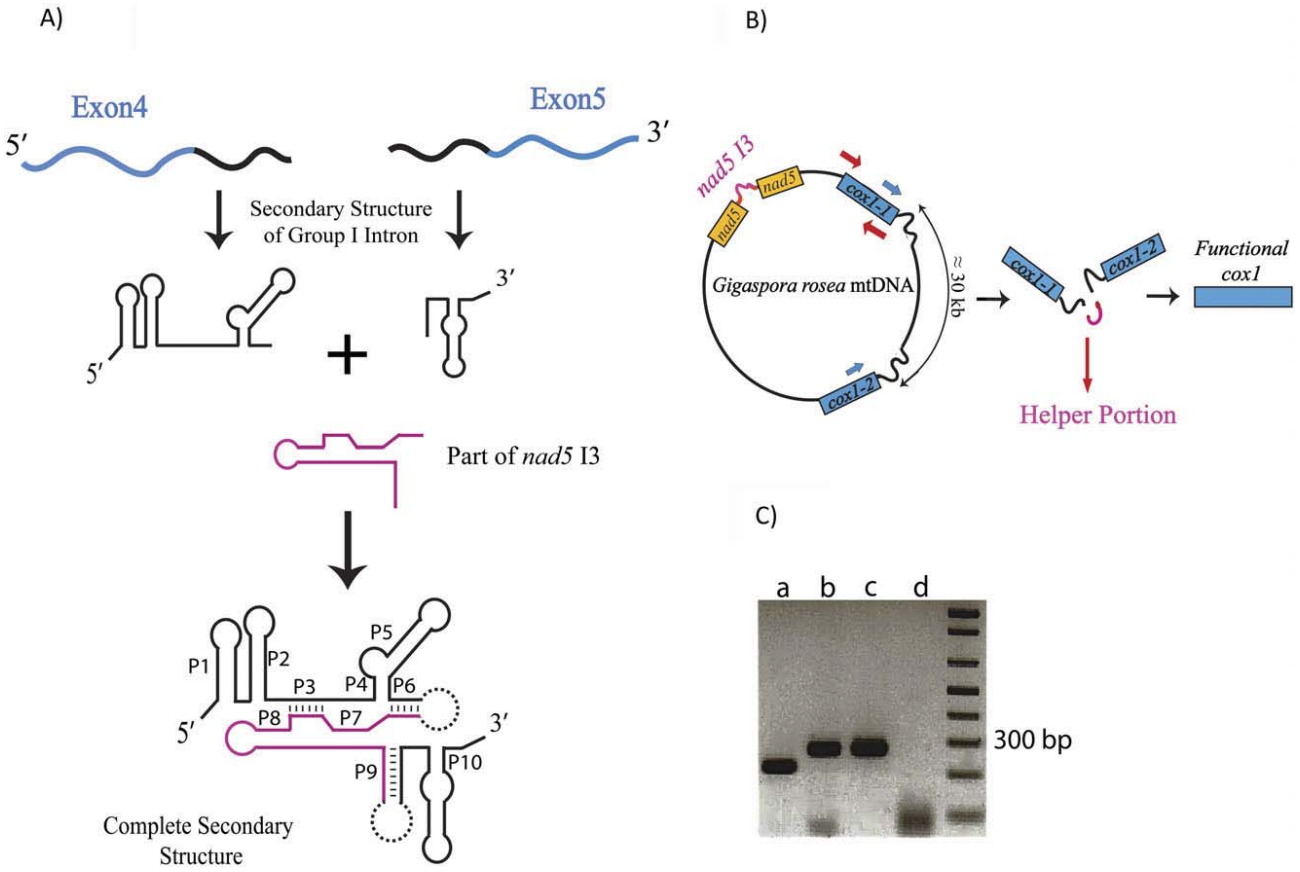


Fig. 3. Model of group I intron–mediated *trans*-splicing in *G. rosea* and demonstration of mRNA *trans*-splicing by RT-PCR. (*A*) Schematic view of the *trans*-spliced *cox1* group IB intron RNA secondary structure. The pink-colored RNA corresponds to the predicted *nad5* intron 3 helper sequence, and dashed loops indicate its interaction with the two intron fragments flanking *cox1* exons 4 and 5 (shown in blue). P1–P10 illustrate the conserved stems of the intron secondary structure according to Burke et al. (1987). Canonical Watson–Crick base pairing is shown by dashes. (*B*) Genomic view of RNAs involved in *cox*1 *trans*-splicing. Blue and red arrows indicate PCR and RT-PCR primers (see Materials and Methods). (*C*) Amplification products obtained with *G. rosea* cDNA and genomic DNA. a, RT-PCR amplification in the presence of total RNA (demonstration of *trans*-splicing and that mitochondrial *cox1* is not a pseudogene); b, RT-PCR amplification of an internal section of exon 4 (positive control); c, PCR amplification of total genomic DNA, with the same primers as in "b" (positive control); and d, PCR amplification of total genomic DNA, with the primers used in "a" (negative control, no presence of a contiguous gene in the nuclear genome). Further negative controls (see Materials and Methods) not shown.

experiments confirm this view: the insert in *rns_2* is not excised at the RNA level nor are five intron-size inserts in *rnl*. According to our results, RNA splicing occurs only for the two *rnl* introns that are predicted by RNAweasel (Beck and Lang 2009; one group II followed by a group I intron; fig. 2). It is noteworthy that one of the *rnl* inserts occurs precisely at a position where *G. irregulare* carries a true intron, as verified in that case by cDNA sequencing (Lee and Young 2009). Evidently, the nature and size of gene and genome insertions in glomeromycotan mtDNAs undergo the most rapid change, similar to what we observe with mitochondrial plasmid insertions.

Conclusion

The *Gigaspora* mitochondrial genome has turned out to be most unusual, with fragmented genes, *trans*-splicing, a ribosomal RNA gene in pieces, and an inflated genome size. The presence of substantial gene order differences between *Glomus* and *Gigaspora* is another intriguing observation, suggesting the capacity for nonhomologous recombination repair. The repeated insertion of plastid-like sequences into the mtDNA is in support of this view, as is a high but variable number of (mostly group I) introns. Group I introns propagate by a homing mechanism that relies on homologous recombination (Perrin et al. 1993; Chevalier et al. 2003), that is, we hypothesize that the biochemical machineries for both homologous and nonhomologous recombination are present in glomeromycotan mitochondria. The effectiveness and mechanisms of recombination may be tested by combining, for instance, compatible *Glomus* species with differences in sequence or intron content via anastomosis and analysis for recombination products in the segregated offspring. Evidently, mtDNAs do segregate effectively in Glomeromycota, as any mtDNA so far sequenced has had a perfectly homogeneous sequence: in contrast to their nuclear genomes, which are multinucleate, most divergent in

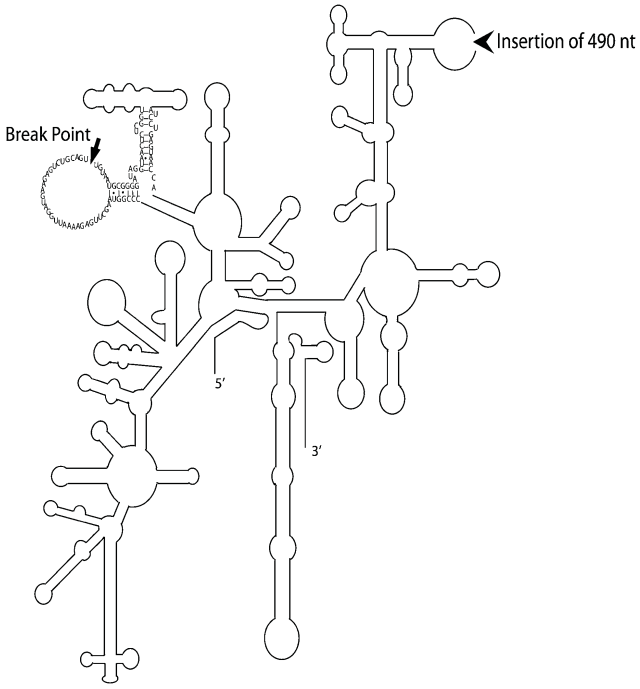


FIG. 4. Secondary structure model of the fragmented *G. rosea* small subunit rRNA. Schematic *rns* structure in which only sequences close to the breakpoint are indicated. The 5′ and 3′ termini at the breakpoint are precisely inferred based on (significant) sequence similarity with the *Glomus irregulare* counterpart. Dashes indicate canonical Watson–Crick base pairings, and dots denote guanine–uracil pairings. The arrowhead points to a 490-nt (nonintron) insertion in a variable loop.

sequence, and seemingly impossible to assemble (Martin et al. 2008). The large differences among glomeromycotan mitochondrial genomes should motivate the future inclusion of other distant lineages, such as *Acaulospora* and *Scutellospora*, for further comparisons. These new sequences will at the same time serve in confirming the phylogenetic positioning proposed here of Glomeromycota within Fungi.

Supplementary Material

Supplementary figures S1 and S2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Dr Michael Gray (Halifax, NS) for critical reading of the manuscript, Natacha Beck for continued development of automated organelle genome annotation software (MFannot), and Dr Yves Piché for kindly providing *G. rosea*. We thank NSERC-CRD and Genome Quebec/Genome Canada for providing financial support to B.F.L. and M.H.

References

Adams KL, Clements MJ, Vaughn JC. 1998. The *Peperomia* mitochondrial coxI group I intron: timing of horizontal transfer and subsequent evolution of the intron. *J Mol Evol.* 46:689–696.

Anziano PQ, Hanson DK, Mahler HR, Perlman PS. 1982. Functional domains in introns: trans-acting and cis-acting regions of intron 4 of the cob gene. *Cell* 30:925–932.

Baidyaroy D, Glynn JM, Bertrand H. 2000. Dynamics of asexual transmission of a mitochondrial plasmid in *Cryphonectria parasitica*. *Curr Genet.* 37:257–267.

Beck N, Lang BF. 2009. RNAweasel, a webserver for identification of mitochondrial, structured RNAs [Internet]. Montreal (Quebec): University of Montreal; 2006. Available from: http://megasun.bch.umontreal.ca/RNAweasel

Bendich AJ. 1996. Structural analysis of mitochondrial DNA molecules from fungi and plants using moving pictures and pulsed-field gel electrophoresis. *J Mol Biol.* 255:564–588.

Bertrand H, Griffiths AJ, Court DA, Cheng CK. 1986. An extrachromosomal plasmid is the etiological precursor of kalDNA insertion sequences in the mitochondrial chromosome of senescent *Neurospora*. *Cell* 47:829–837.

Boer PH, Gray MW. 1988. Scrambled ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. *Cell* 55: 399–411.

Bonen L. 2008. Cis- and trans-splicing of group II introns in plant mitochondria. *Mitochondrion* 8:26–34.

Boon E, Zimmerman E, Lang BF, Hijri M. 2010. Intra-isolate genome variation in arbuscular mycorrhizal fungi persists in the transcriptome. *J Evol Biol.* 23:1519–1527.

Bullerwell CE, Forget L, Lang BF. 2003. Evolution of monoblepharidalean fungi based on complete mitochondrial genome sequences. *Nucleic Acids Res.* 31:1614–1623.

Burger G, Yan Y, Javadi P, Lang BF. 2009. Group I-intron trans-splicing and mRNA editing in the mitochondria of placozoan animals. *Trends Genet.* 25:381–386.

Burke JM, Belfort M, Cech TR, Davies RW, Schweyen RJ, Shub DA, Szostak JW, Tabak HF. 1987. Structural conventions for group I introns. *Nucleic Acids Res.* 15:7217–7221.

Cech TR, Tanner NK, Tinoco I Jr, Weir BR, Zuker M, Perlman PS. 1983. Secondary structure of the *Tetrahymena* ribosomal RNA intervening sequence: structural homology with fungal mitochondrial intervening sequences. *Proc Natl Acad Sci U S A.* 80:3903–3907.

Chevalier B, Turmel M, Lemieux C, Monnat RJ Jr, Stoddard BL. 2003. Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-CreI and I-MsoI. *J Mol Biol.* 329:253–269.

Colleaux L, d'Auriol L, Betermier M, Cottarel G, Jacquier A, Galibert F, Dujon B. 1986. Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into E. coli as a specific double strand endonuclease. *Cell* 44:521–533.

Croll D, Giovannetti M, Koch AM, Sbrana C, Ehinger M, Lammers PJ, Sanders IR. 2009. Nonself vegetative fusion and genetic exchange in the arbuscular mycorrhizal fungus *Glomus intraradices*. *New Phytol.* 181:924–937.

Croll D, Sanders IR. 2009. Recombination in *Glomus intraradices*, a supposed ancient asexual arbuscular mycorrhizal fungus. *BMC Evol Biol.* 9:13.

Dombrovska O, Qiu YL. 2004. Distribution of introns in the mitochondrial gene nad1 in land plants: phylogenetic and molecular evolutionary implications. *Mol Phylogenet Evol.* 32:246–263.

Ebersberger I, de Matos Simoes R, Kupczok A, Gube M, Kothe E, Voigt K, von Haeseler A. A consistent phylogenetic backbone for the fungi. *Mol Biol Evol.* Advance Access published March 12, 2012, doi:10.1093/molbev/msr285

Eddy S. 2008. HMMER website [Internet]. Ashburn (VA): HHMI Janelia Farm Research Campus. Available from: http://hmmer.janelia.org

Edgar RC. 2004. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
Esch H, Hundeshagen B, Schneider-Poetsch H, Bothe H. 1994. Demonstration of abscisic acid in spores and hyphae of the arbuscular-mycorrhizal fungus *Glomus* and in the N2-fixing cyanobacterium *Anabaena variabilis*. *Plant Sci.* 99:9–16.
Felsenstein J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. *Syst Zool.* 27:27–33.
Foury F, Roganti T, Lecrenier N, Purnelle B. 1998. The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Lett.* 440:325–331.
Gautheret D, Lambert A. 2001. Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *J Mol Biol.* 313:1003–1011.
Gobert A, Gutmann B, Taschner A, Gossringer M, Holzmann J, Hartmann RK, Rossmanith W, Giegé P. 2010. A single *Arabidopsis* organellar protein has RNase P activity. *Nat Struct Mol Biol.* 17:740–744.
Goldschmidt-Clermont M, Choquet Y, Girard-Bascou J, Michel F, Schirmer-Rahire M, Rochaix JD. 1991. A small chloroplast RNA may be required for trans-splicing in *Chlamydomonas reinhardtii*. *Cell* 65:135–143.
Grewe F, Viehoever P, Weisshaar B, Knoop V. 2009. A trans-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Res.* 37:5093–5104.
Hausner G. 2011. Introns, mobile elements, and plasmids. In: Bullerwell CE, editor. Organelle genetics. Berlin (Germany): Springer.
Hecht J, Grewe F, Knoop V. 2011. Extreme RNA editing in coding islands and abundant microsatellites in repeat sequences of *Selaginella moellendorffii* mitochondria: the root of frequent plant mtDNA recombination in early tracheophytes. *Genome Biol Evol.* 3:344–358.
Hibbett DS, Binder M, Bischoff JF, et al. 2007. A higher-level phylogenetic classification of the fungi. *Mycol Res.* 111:509–547.
Hijri M, Sanders IR. 2005. Low gene copy number shows that arbuscular mycorrhizal fungi inherit genetically different nuclei. *Nature* 433:160–163.
Holzmann J, Frank P, Loffler E, Bennett KL, Gerner C, Rossmanith W. 2008. RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell* 135:462–474.
Jacquier A, Rosbash M. 1986. Efficient trans-splicing of a yeast mitochondrial RNA group II intron implicates a strong 5′ exon-intron interaction. *Science* 234:1099–1104.
James TY, Kauff F, Schoch CL, et al. 2006. Reconstructing the early evolution of fungi using a six-gene phylogeny. *Nature* 443:818–822.
Knoop V, Altwasser M, Brennicke A. 1997. A tripartite group II intron in mitochondria of an angiosperm plant. *Mol Gen Genet.* 255:269–276.
Lambowitz AM, Zimmerly S. 2004. Mobile group II introns. *Annu Rev Genet.* 38:1–35.
Lang BF. 1984. The mitochondrial genome of the fission yeast *Schizosaccharomyces pombe*: highly homologous introns are inserted at the same position of the otherwise less conserved *cox1* genes in *Schizosaccharomyces pombe* and *Aspergillus nidulans*. *EMBO J.* 3:2129–2136.
Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Gray MW. 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* 387:493–497.
Lang BF, Cedergren R, Gray MW. 1987. The mitochondrial genome of the fission yeast, *Schizosaccharomyces pombe*. Sequence of the large-subunit ribosomal RNA gene, comparison of potential secondary structure in fungal mitochondrial large-subunit rRNAs and evolutionary considerations. *Eur J Biochem.* 169:527–537.
Lang BF, Hijri M. 2009. The complete *Glomus intraradices* mitochondrial genome sequence—a milestone in mycorrhizal research. *New Phytol.* 183:3–6.
Lang BF, Laforest MJ, Burger G. 2007. Mitochondrial introns: a critical view. *Trends Genet.* 23:119–125.
Lang BF, Lavrov D, Beck N, Steinberg V. 2011. Mitochondrial tRNA structure, identity and evolution of the genetic code. In: Bullerwell CE, editor. Organelle genetics. Berlin (Germany): Springer.
Lartillot N, Brinkmann H, Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol.* 7(Suppl 1):S4.
Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol.* 21:1095–1109.
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 Phytol.* 183:200–211.
Liu Y, Leigh JW, Brinkmann H, Cushion MT, Rodriguez-Ezpeleta N, Philippe H, Lang BF. 2009. Phylogenomic analyses support the monophyly of Taphrinomycotina, including *Schizosaccharomyces* fission yeasts. *Mol Biol Evol.* 26:27–34.
Liu Y, Steenkamp ET, Brinkmann H, Forget L, Philippe H, Lang BF. 2009. Phylogenomic analyses predict sistergroup relationship of nucleariids and fungi and paraphyly of zygomycetes with significant support. *BMC Evol Biol.* 9:272.
Marleau J, Dalpe Y, St-Arnaud M, Hijri M. 2011. Spore development and nuclear inheritance in arbuscular mycorrhizal fungi. *BMC Evol Biol.* 11:51.
Martin F, Gianinazzi-Pearson V, Hijri M, Lammers P, Requena N, Sanders IR, Shachar-Hill Y, Shapiro H, Tuskan GA, Young JP. 2008. The long hard road to a completed *Glomus intraradices* genome. *New Phytol.* 180:747–750.
Michel F, Jacquier A, Dujon B. 1982. Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie* 64:867–881.
Muramatsu T, Yokoyama S, Horie N, Matsuda A, Ueda T, Yamaizumi Z, Kuchino Y, Nishimura S, Miyazawa T. 1988. A novel lysine-substituted nucleoside in the first position of the anticodon of minor isoleucine tRNA from Escherichia coli. *J Biol Chem.* 263:9261–9267.
Parniske M. 2008. Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nat Rev Microbiol.* 6:763–775.
Perrin A, Buckle M, Dujon B. 1993. Asymmetrical recognition and activity of the I-Scei endonuclease on its site and on intron-exon junctions. *EMBO J.* 12:2939–2947.
Pombert JF, Keeling PJ. 2010. The mitochondrial genome of the entomoparasitic green alga *Helicosporidium*. *PLoS One* 5:e8954.
Schmelzer C, Schweyen RJ. 1986. Self-splicing of group II introns in vitro: mapping of the branch point and mutational inhibition of lariat formation. *Cell* 46:557–565.
Schnare MN, Gray MW. 1990. Sixteen discrete RNA components in the cytoplasmic ribosome of *Euglena gracilis*. *J Mol Biol.* 215:73–83.
Schwarzott D, Walker C, Schüssler A. 2001. *Glomus*, the largest genus of the arbuscular mycorrhizal fungi (glomales), is nonmonophyletic. *Mol Phylogenet Evol.* 21:190–197.
Seif E, Cadieux A, Lang BF. 2006. Hybrid *E. coli*—mitochondrial ribonuclease P RNAs are catalytically active. *RNA* 12:1661–1670.
Seif E, Leigh J, Liu Y, Roewer I, Forget L, Lang BF. 2005. Comparative mitochondrial genomics in zygomycetes: bacteria-like RNase P RNAs, mobile elements and a close source of the group I intron invasion in angiosperms. *Nucleic Acids Res.* 33:734–744.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst Biol.* 51:492–508.

Slater GS, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:31.

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

Strack D, Fester T, Hause B, Schliemann W, Walter MH. 2003. Arbuscular mycorrhiza: biological, chemical, and molecular aspects. *J Chem Ecol.* 29:1955–1979.

Van der Veen R, Arnberg AC, van der Horst G, Bonen L, Tabak HF, Grivell LA. 1986. Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. *Cell* 44:225–234.

Vaughn JC, Mason MT, Sper-Whitis GL, Kuhlman P, Palmer JD. 1995. Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric *cox1* gene of *Peperomia*. *J Mol Evol.* 41:563–572.

Vierula PJ, Bertrand H. 1992. A deletion derivative of the kalilo senescence plasmid forms hairpin and duplex DNA structures in the mitochondria of *Neurospora*. *Mol Gen Genet.* 234:361–368.

Waring RB, Davies RW, Scazzocchio C, Brown TA. 1982. Internal structure of a mitochondrial intron of *Aspergillus nidulans*. *Proc Natl Acad Sci U S A.* 79:6332–6336.

Weber F, Dietrich A, Weil JH, Marechal-Drouard L. 1990. A potato mitochondrial isoleucine tRNA is coded for by a mitochondrial gene possessing a methionine anticodon. *Nucleic Acids Res.* 18:5027–5030.

7 - Article 2 - de la Providencia *et al.*, 2012

Detection of a transient mitochondrial DNA heteroplasmy in the progeny of crossed genetically divergent isolates of arbuscular mycorrhizal fungi

Ivan Enrique de la Providencia, Maryam Nadimi, Denis Beaudet, Gabriela Rodriguez Morales and Mohamed Hijri

Institut de Recherche en Biologie Végétale, Département de Sciences Biologiques, Université de Montréal, 4101 Rue Sherbrooke Est, Montréal, Québec, H1X 2B2, Canada

Authors for correspondence:
Ivan Enrique de la Providencia
[redacted]

Mohamed Hijri
[redacted]

Received: *21 March 2013*
Accepted: *14 May 2013*

New Phytologist (2013) **200:** 211–221
doi: 10.1111/nph.12372

Key words: anastomosis, arbuscular mycorrhizal fungi (AMF), heteroplasmy, mitochondrial genome, nonself fusion, specific molecular marker.

Summary

- Nonself fusion and nuclear genetic exchange have been documented in arbuscular mycorrhizal fungi (AMF), particularly in *Rhizophagus irregularis*. However, mitochondrial transmission accompanying nonself fusion of genetically divergent isolates remains unknown.
- Here, we tested the hypothesis that mitochondrial DNA (mtDNA) heteroplasmy occurs in the progeny of spores, obtained by crossing genetically divergent mtDNAs in *R. irregularis* isolates.
- Three isolates of geographically distant locations were used to investigate nonself fusions and mtDNA transmission to the progeny. We sequenced two additional mtDNAs of two *R. irregularis* isolates and developed isolate-specific size-variable markers in intergenic regions of these isolates and those of DAOM-197198. We achieved three crossing combinations in pre-symbiotic and symbiotic phases. Progeny spores per crossing combination were genotyped using isolate-specific markers.
- We found evidence that nonself recognition occurs between isolates originating from different continents both in pre-symbiotic and symbiotic phases. Genotyping patterns of individual spores from the progeny clearly showed the presence of markers of the two parental mtDNA haplotypes. Our results demonstrate that mtDNA heteroplasmy occurs in the progeny of the crossed isolates. However, this heteroplasmy appears to be a transient stage because all the live progeny spores that were able to germinate showed only one mtDNA haplotype.

Introduction

The Glomeromycota phylum encompasses an ancient group of obligate symbiotic fungi colonizing *c.* 80% of plants to form arbuscular mycorrhizae (Smith & Read, 2008). Long considered asexual organisms, this paradigm has recently been challenged by the discovery of intact meiotic machinery genes that are putatively functional (Halary *et al.*, 2011), although apparent reproductive structures have yet to be observed. Arbuscular mycorrhizal fungi (AMF) reproduce clonally through large and multinucleated spores, and due to their coenocytic nature (i.e. living hyphae lacking septa and thereby nuclei, mitochondria and other organelles sharing a common cytoplasm), spores and mycelium can host communities of hundreds to thousands of nuclei showing substantial genetic variability (Kuhn *et al.*, 2001; Hijri & Sanders, 2005; Jany & Pawlowska, 2010; Marleau *et al.*, 2011), which persist at the transcriptional level (Boon *et al.*, 2010).

Several studies have shown the impact of AMF on plant growth and health (Jeffries *et al.*, 2003; van der Heijden *et al.*, 2008; Smith & Read, 2008; Ismail & Hijri, 2012), soil quality and structure (Rillig & Mummey, 2006), and ecosystem functioning and biodiversity (van der Heijden *et al.*, 2006). When associated with plants, they often extend beyond roots into the surrounding environment to form an intermingled coenocytic and multinucleated extraradical mycelium network (Leake *et al.*, 2004) that facilitates up to 90% of phosphorus (P) plant-uptake and also enhances nitrogen (N)-acquisition under natural conditions (reviewed in Bolduc & Hijri (2011)). AMF structures can also link plants from the same or different species and/or families via the formation of hyphal fusions (i.e. anastomosis) (Giovannetti *et al.*, 2004), creating an extensive common mycorrhizal network. As previously demonstrated by Giovannetti *et al.* (1999), anastomosis also plays a key role in genetic exchange, either within the same isolate or between closely related isolates that are genetically divergent (Croll *et al.*, 2009). This mechanism is likely one of the most important factors contributing to the maintenance of genomic composition among members of the Glomeromycota (Bever & Wang, 2005; Corradi *et al.*, 2007; Croll *et al.*, 2009; Angelard & Sanders, 2011; Colard *et al.*, 2011). However, such huge genetic diversity within AMF populations (Corradi *et al.*, 2007; Boon *et al.*, 2010) or even in successive monosporal subcultures (Ehinger *et al.*, 2012) makes it difficult to use only nuclear markers to address questions related to AMF community structure, diversity and function. In contrast to nuclear genes, mitochondrial genomes appear to be homogeneous within AMF isolates.

This was clearly demonstrated by sequencing the first complete mitochondrial genome (Lee & Young, 2009) of the *Rhizophagus irregularis* isolate 494 (Błaszk., Wubet, Renker & Buscot) C. Walker & A. Schussler – syn. *Glomus irregulare* Błaszk., Wubet, Renker & Buscot (Blaszkowski *et al.*, 2008; Stockinger *et al.*, 2009; Schussler & Walker, 2010), followed by many other genomes such as *Gigaspora margarita, Gi. rosea*, four isolates of *R. irregularis* and a closely related *Rhizophagus* species (Formey *et al.*, 2012; Nadimi *et al.*, 2012; Pelin *et al.*, 2012; Beaudet *et al.*, 2013). These genomes offer exceptional opportunities for designing specific markers to identify AMF strains either in natural populations or in mesocosm/microscosm based-experiments (Lang & Hijri, 2009; Formey *et al.*, 2012; Beaudet *et al.*, 2013).

It was previously assumed that, based on incompatibility and nonself recognition mechanisms, anastomosis cannot occur between genetically distinct isolates (Giovannetti *et al.*, 2003). However, many studies have challenged this point of view and have provided convincing contrary evidence (Croll *et al.*, 2009; Angelard *et al.*, 2010; Angelard & Sanders, 2011). These studies used crossing experiments and specific genetic markers to show that nuclear material from both parents is passed on to their progeny, thus demonstrating genetic exchange and segregation between genetically divergent isolates of the AMF *R. irregularis*. In addition, a recent study has clearly demonstrated that genetically different isolates of *Rhizophagus clarus* (synonym, *Glomus clarum*) originating from the same habitat and sampled close to each other underwent anatomosis (Purin & Morton, 2012). This once again raises the questions of whether mitochondria are also exchanged via anastomosis between isolates of *R. irregularis* of different geographical origin, that is, heteroplasmy (i.e. the mixture of genetically different mtDNAs in a common cytoplasm) and whether this state is transient or is maintained into the progeny. Heteroplasmy is not rare in the fungal kingdom (Lesemann *et al.*, 2006) and has been observed in several eukaryotes (Barr *et al.*, 2005; White *et al.*, 2008). As suggested by researchers working on natural populations of the plant pathogen *Armillaria* sp. (Smith *et al.*, 1990), possessing multiple types of mtDNA might confer an advantage under changing environmental conditions. This hypothesis has already been demonstrated for the model fungus *Saccharomyces cerevisiae* (Taylor *et al.*, 2002; Hnatova *et al.*, 2003) and the plant pathogen *Podosphaera leucotricha* (Lesemann *et al.*, 2006).

Heteroplasmy has never been studied in AMF, presumably due to the lack of reliable and appropriate mitochondrial markers (Lang & Hijri, 2009; Lee & Young, 2009; Formey *et al.*, 2012). Consequently, we employed three genetically well-characterized *R. irregularis* isolates (DAOM-197198, DAOM-234328 and DAOM-240415) to address this question, which is an important one for advancing our knowledge of genetic exchange between AMF. These isolates were selected because they originate from distant geographical locations. In addition, *de novo* genome sequencing, assembly and annotation were performed for two other *R. irregularis* isolates, DAOM-234328 (Finland) and DAOM-240415 (Manitoba, Canada), and their results were compared with published mtDNAs (Lee & Young, 2009; Formey *et al.*, 2012; Nadimi *et al.*, 2012). Subsequently, specific size-variable markers were designed for each isolate to detect mtDNA mixture after anastomosis between the strains in microcosm experiments.

Materials and Methods

Growth conditions and maintenance of fungal cultures and roots

Monoxenically produced spores of *Rhizophagus irregularis*; DAOM-197198 (Pont-Rouge, Quebec, Canada) DAOM-234328 (Finland) and DAOM-240415 (Dufrost, Manitoba, Canada) were provided by the DAOM collection (Ottawa, Ontario, Canada). These three isolates were selected because they have different geographical origins and their mitochondrial genomes have been fully sequenced. Spores were subcultured in association with Ri T-DNA transformed chicory (*Cichorium intybus*) roots on a modified minimal (MM) medium (Bécard & Fortin, 1988) solidified with 0.4% (w/v) gellan gum (Sigma). Plates were incubated in the dark in an inverted position at 25°C. Several thousand spores and extraradical mycelia were obtained in a period of 12 wk. Ri T-DNA transformed chicory roots were routinely propagated by placing actively growing root apexes on MM medium and subsequent incubation at 25°C in the dark.

DNA extraction and sequencing

For DNA extraction, the extracted fungal spores and mycelia from the MM medium were dissolved in buffer citrate (pH 6.0) (Doner & Bécard, 1991) and washed with sterile water. Extracted fungal material from the MM medium was observed under a binocular microscope in order to detect and remove any root contaminants. DNA extraction was carried out using the DNeasy Plant Mini Kit (Qiagen), following the manufacturer's instructions. Whole genome shotgun sequencing was performed at McGill University's Genome Quebec Innovation Centre using a Roche GS-FLX-Titanium sequencing platform (454 Life Sciences, Branford, CT, USA).

de novo assembly and sequence analysis

The total DNA of *R. irregularis* isolates DAOM-240415 and DAOM-234328 was sequenced and the resulting reads (257 880 and 238 826, respectively) were assembled using Newbler (Genome Quebec Innovation Centre, McGill University, Montreal, PQ, Canada). Gene annotation was performed with MFannot (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl), followed by manual inspection and notation of missing gene features. More information on MFannot is described in Nadimi *et al.* (2012). The annotated mitochondrial genomes of *R. irregularis* isolates DAOM-240415 and DAOM-234328 were deposited in GenBank under the respective accession numbers JX993113 and JX993114.

Molecular marker development

In order to develop isolate size-specific molecular markers for the mtDNA of three different *R. irregularis* isolates, the variable intergenic regions of their complete sequenced mitochondrial genomes were used for multiple alignments. Consequently, we developed specific primers for each isolate, which resulted in three pairs of primers. In addition, these primers produced PCR fragments of different lengths (Table 1). All primers were tested for hairpin and dimer formation using AmplifX software and a SIGMA DNA calculator (http://www.sigma-genosys.com/calc/DNACalc.asp). The specificity of these markers was achieved using PCR on amplified DNA of each isolate (Fig. 1). These markers made it possible to trace parental mtDNA haplotypes in crossing experiments.

Experimental set up

We set up two experiments in pre-symbiotic and symbiotic phases, respectively, to study the occurrence of anastomosis and subsequent exchange of mitochondria among the three isolates in crossing experiments. We also performed a single symbiotic experiment to evaluate the inheritance pattern of mitochondria from monosporal culture lines obtained from spores harvested from the interaction zone. Hyphal interactions were classified into four categories as described in Croll *et al.* (2009): (1) perfect fusion meant that hyphae entered into contact, fused and protoplasmic flow was established following anastomosis. No sign of incompatibility reaction was observed. (2) Pre-fusion incompatibility was characterized by profuse mycelium ramification at the elongation zone of the hypha that could be followed by septa formation at the hyphal apex, thus preventing contact between the approaching hyphae. (3) No interaction means that two hyphae (regardless of the type of interaction) did not show signs of rejection (i.e. wall thickening at the apex region, septa formation or change of growth direction) but intermingled (i.e. hyphae from different isolates or the same isolate grew in different directions and crossed, but did not touch each other) or overlapped (i.e. hyphae from different isolates or the same isolate entered into contact and continued to grow over each other). (4) Post-fusion incompatibility means that following homing, contact and fusion, the protoplasm of one hypha withdrew and a septum was formed between the two fused hyphae. Percentage of anastomosis for the categories perfect fusion and nonself fusion incompatibility, was calculated by dividing the number of fusions by the total number of observed hyphal contacts (Purin & Morton, 2012).

Pre-symbiotic experiment Spores of the three isolates were extracted from the gel by solubilisation in buffer citrate (pH 6) (Doner & Bécard, 1991). A mono-compartment plate (35 × 10 mm) filled with 5 ml of MM medium was divided into four equal compartments. Before inoculation a colour-coded circle was marked at the bottom of the plate on each

Table 1 Isolate-specific primers used to discriminate the three *Rhizophagus irregularis* isolates

Primers	Codes	Primer sequences (5′- 3′)	Size (bp)	Position of primers in mtDNA		
				DAOM-197198	DAOM-240415	DAOM-234328
197198F	A	AGCAAATCTAAGTTCCTCAGAG	120	69523–69544		
197198R		TCCTGCCCCAGATACTCCAG		69624–69643		
240415F	B	AGAAGAGTTAATAAGTTCAACTGT	147		68561–68584	
240415R		AGACTAATAATACGAGATGCAG			68687–68708	
234328F	C	AAGAGGGGTGCATTCCGAAG	209			67641–67660
234328R		AGCTACGCTTTGCCTTAGCA				67831–67850

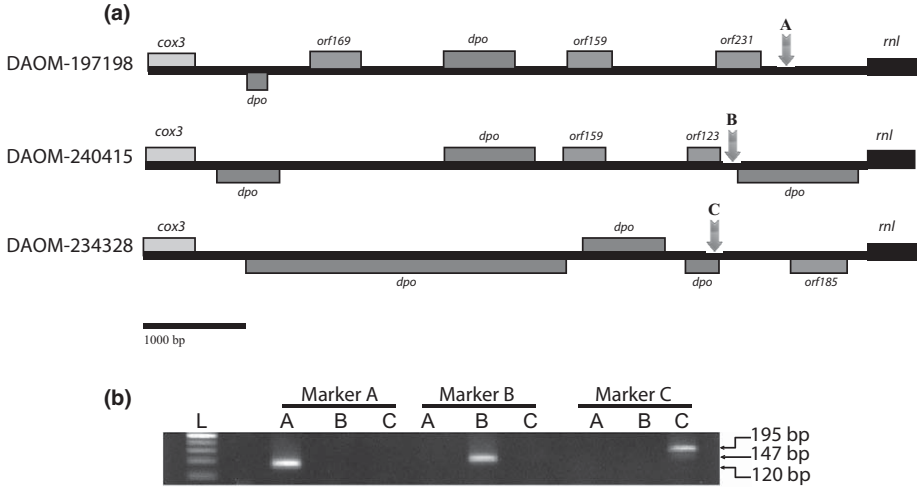


Fig. 1 Isolate-specific mtDNA markers. (a) Comparative view of the *cox3-rnl* intergenic regions of the three *Rhizophagus irregularis* isolates where primers were designed (arrows). Upper boxes (open reading frames, ORF) show direct orientation while bottom boxes are opposite orientation. (b) Gel electrophoresis showing the specificity of length-specific molecular markers developed for the *R. irregularis* isolates DAOM-197198 (A), DAOM-240415 (B) and DAOM-234328 (C). L, ladder (molecular marker).

compartment and used to identify clusters belonging to the same or different isolates. On each compartment, ten clusters (a cluster was defined in this study as a group of spores bearing the same mycelium (Fig. 2)) of up to five spores from each isolate were gently placed over the medium in the following combinations: (DAOM-197198/DAOM-234328; DAOM-197198/DAOM-240415; DAOM-234328/DAOM-240415). Clusters belonging to different isolates were separated by a maximum distance of 2 mm. Plates were then sealed and incubated at 25°C in the dark until germination. Following cluster germination, the occurrence of hyphal fusion was checked weekly for a 45-d period, and results were grouped into four categories (see above, 'Experimental set up' section) (Croll *et al.*, 2009). Each experimental unit (i.e. one replicate) consisted of one plate with four repetitions of the same pairing (combination) and ten replicates were performed for each combination. Ten replicates constituted from pairings of the same isolate were used as controls.

Symbiotic experiment An autoclaved (121°C for 15 min) slide (25 × 75 × 1 mm) was placed in the middle of an empty monocompartment Petri dish (100 × 15 mm). Thirty millilitres of the MM medium was then poured in to fill the plate and cover the slide. Following jellification, only the medium covering the slide was gently removed in such a way that *c.* 1 mm of each of the slide edges was still embedded in the medium. A flat surface where the development of symbiotic mycelia could easily be traced, adjoined by two identical distal and proximal compartments, was created (Fig. 3). Subsequently, 2-cm long Ri T-DNA transformed chicory roots were placed in both compartments and 100 spores of each isolate taken up from monoxenic cultures by solubilisation were inoculated in one of the two compartments using the same combinations described above. Plates were sealed and incubated upside down at 25°C in the dark. Plates were checked weekly and chicory roots growing towards the slide were cut and gently removed using scalpel and forceps. Because mycelia growing on both edges of the slide began to spread across it, plates were checked weekly over 10 wk for hyphal interactions, and results were grouped into four categories according to Croll *et al.* (2009). Each experimental unit (i.e. one replicate) consisted of one plate inoculated with two isolates in association with a Ri

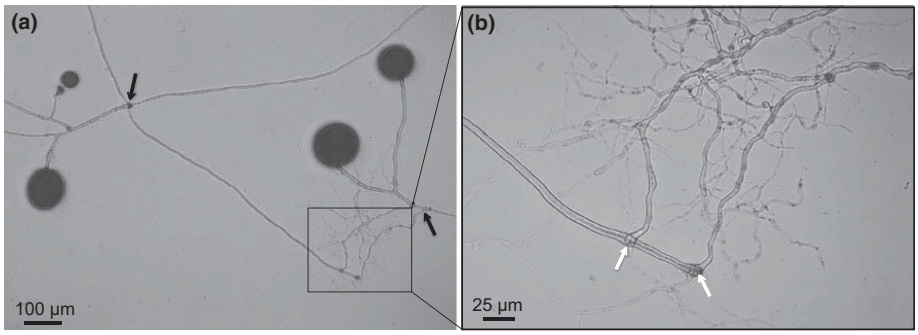


Fig. 2 Self-fusion between spore clusters belonging to *Rhizophagus irregularis* DAOM-4240415. (a) Emergence of several germ tubes through the cut extremities (black arrow) and anastomosis formation. (b) Inset showing details of the anastomosis (white arrows); note the profuse growth of thin hyphae at the interaction zone. Active bidirectional protoplasmic flow, is shown in Video S1.

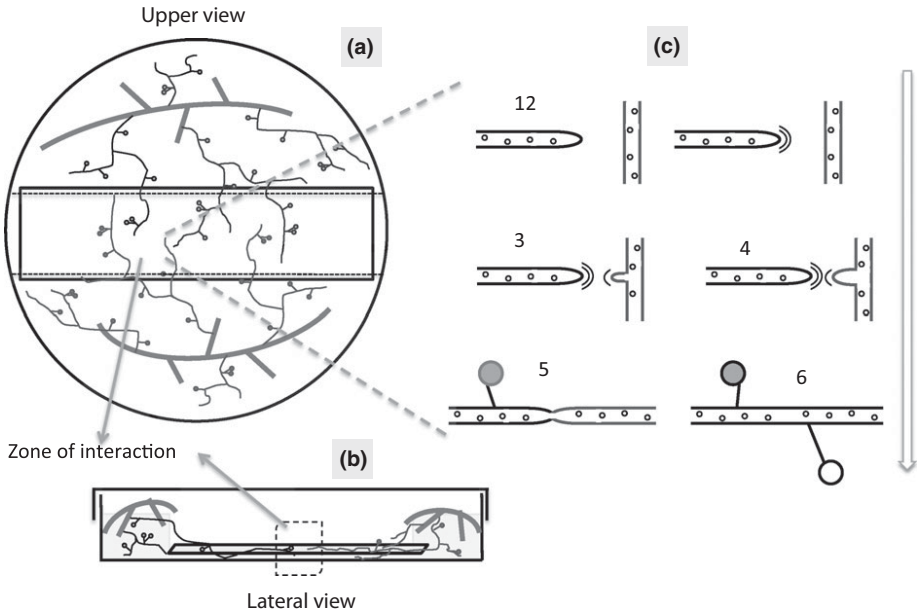


Fig. 3 Diagram presentation of the experimental set up for the study of anastomosis between geographically distinct isolates of *Rhizophagus irregularis* in the symbiotic mycelium. (a, b) Different views of the experimental set up showing mycelium (black and gray) from both sides growing towards the slide (i.e. zone of interaction) and (c) interacting with each other.

T-DNA transformed chicory root. Fourteen replicates were performed for each combination, and five replicates for each isolate inoculated in both compartments with the same fungal isolate were used as control.

Monosporal cultures lines with progeny from crossed experiments In order to test the hypothesis of whether heteroplasmy is a transient state or is maintained into the progeny, twenty full lipid spores (i.e. progenies) harvested from the interaction zone for each combination were randomly chosen and individually cut out from the mycelium and extracted in 2 µl sterile ultra-pure water (PureLab, Elga, Canada). Spores were placed in a new mono-compartment Petri dish (90 mm) containing M medium in the close vicinity of a Ri-T transformed chicory root. Sixty replicates, each consisting of one single spore associated with a chicory root were maintained as above. Each plate was checked weekly for germination, root colonization and colony development over 11 wk.

Microscopy, data collection, harvest and statistical analysis For all experiments, the occurrence of anastomosis and/or spore germination was observed under a Discovery V12 stereomicroscope (Carl Zeiss, Toronto, ON, Canada) at magnifications of ×6.7–×40. Bright-field microscopy (Axio Imager M1; Carl Zeiss) was also used to observe details of hyphal interactions at higher magnifications. Images and movies of the hyphal interactions, as well as protoplasmic streaming through hyphal fusions, were captured with an AxioCam HRC using Axiovision software v4.8.1.0 (Carl Zeiss).

For the second experiment, only the cultures showing synchronised growth of crossed isolates from both sides towards the slide were considered in order to trace the origin of hyphae and to avoid self-fusions. This restriction reduced the experimental replicates to three for some crossing combinations. The weekly growth of the mycelia over the slide was marked with coloured dots according to the isolate, making it easy to trace development as well as interactions between both mycelia. Hyphal interactions, number and phenotype of spores (e.g. shape, lipid droplet content, aborted-like structures) formed at the zone of interaction (slide) at both sides of the fungal cultures were evaluated in addition to measuring total hyphal length and hyphal density. Thirty full lipid spores (20 spores out of 30 were used to initiate monosporal cultures lines; see experiment 3) formed at the zone of interaction for each combination were randomly chosen, individually cut out from the mycelium and extracted in 2 µl sterile ultra-pure water (PureLab). Whole genome amplification was performed for all spores using the GenomiPhi V2 DNA Amplification Kit (GE Healthcare, Mississauga, ON, Canada) following the manufacturer's instructions.

Monosporal culture lines initiated from progenies did not produce any further visible extraradical structures, even though germination and root contact of several spores was observed. Under these circumstances, all germinated and nongerminated spores were harvested and their genome amplified following the same protocol described. The root segment in contact with the germinative mycelium was removed and stained following the protocol described in (Phillips & Hayman, 1970) to observe hyphal attachment and also assess the development of intraradical structures.

Data analysis was performed using STATISTICA software. Data were subjected to an ANOVA. The Tukey honestly significant difference test was used to identify significant differences ($P \geq 0.05$).

PCR genotyping and sequencing of progeny spores As already described, isolate-specific markers were designed for each of the three *R. irregularis* isolates, DAOM-240415, DAOM-234328 and DAOM-197198. Primer sequences are summarized in Table 1. DNA from individual spore progeny for the second and third experiments was subjected to PCR using the isolate-specific primers corresponding to their parents. PCR conditions were as follows: PCR reactions in a 50-µl volume contained 0.2 mM dNTPs (Fermentas, Canada), 0.5 µM of each primer, 0.5 U of *Taq* DNA polymerase (BioBasic, Montreal, Quebec, Canada) and 2 µl of the whole genome amplification product of each spore. PCRs were run on Mastercycler pro S (Eppendorf, Canada). PCR cycling conditions included an initial denaturation at 95°C for 2 min followed by 30 cycles at 95°C for 30 s; 55°C or 60°C for 30 s and 72°C for 30 s. An elongation at 72°C for 10 min was performed at the end of PCR. Hybridization temperature was 55°C for DAOM-240415 and 60°C for DAOM-97198 and DAOM-234328. PCR products were revealed using GelRed dye (Invitrogen) and images were recorded by GelDoc system (BioRad, Canada). PCR was performed according to the crossing combinations summarized in Table 1. All the amplification products obtained from progeny were sent for sequencing to McGill University and Genome Quebec Innovation Center facility (Montreal, Canada). Nucleotide Blast searches were performed using the NCBI website to check the presence of the target sequence.

Results

Mitochondrial genome comparison and marker development

The mtDNAs of the two newly sequenced *R. irregularis* isolates (DAOM-240415 and DAOM-234328) are almost identical, both to each other and to those retrieved from GenBank, *R. irregularis* isolates: 494 (Lee & Young, 2009), DAOM-197198 (Nadimi *et al.*, 2012), MUCL-46239, MUCL-46240 and MUCL-46241 (Formey *et al.*, 2012). *Rhizophagus irregularis* isolates have a mitochondrial genome ranging in size from 68 995 bp to 74 799 bp, except for the isolate MUCL-43204, which is outside of this range (87 754 bp). This variation in size is caused by a divergence in the presence/absence and size of mobile elements such as plasmid related DNA polymerase genes (*dpo*), open reading frames encoding homing endonuclease genes in defined intergenic regions, as previously described in Formey *et al.* (2012).

In order to develop the mtDNA isolate-specific marker for DAOM-197198, DAOM-240415 and DAOM-234328, multiple alignments of their complete mtDNAs and those of five

other *R. irregularis* isolates were performed. We focused on the *cox3-rnl* intergene, which is a large region comprising substantial sequence divergence, as a target to design size-variable isolate-specific markers. The specific markers that were developed were used to distinguish between the three *R. irregularis* isolates selected in our crossing experiments, although the presence of a large number of *dpo* elements of different lengths and orientations in the *cox3-rnl* intergenic region made it challenging to perform the multiple sequence alignments. Each specific marker amplified a single band of an expected size for the corresponding isolate (Fig. 1). In addition to the specificity, the three markers were designed to produce different PCR product sizes, in order to facilitate tracing each marker in spore progeny. For example, marker A, specific for isolate DAOM-197198, gave a PCR band of 120 bp only for DNA of its corresponding isolate and no band was observed for isolates DAOM-240415 and DAOM-234328. Markers B and C, specific for isolates DAOM-240415 and DAOM-234328, gave PCR bands of 147 bp and 209 bp, respectively (Fig. 1, Table 1). Length polymorphism can also be used to trace each haplotype in multiplex PCR reactions. However, primers of isolate DAOM-234328 can also amplify mtDNA of the isolate 494, producing a 209-bp PCR product. Because isolate 494 was not included in our crossing experiment, the primer pair developed for DAOM-234328 was valid.

Pre-symbiotic interactions

In this experiment, we recorded both self and nonself fusion between all the isolates tested and characterized them by total fusion of hyphal membranes and cell wall followed by bidirectional protoplasm streaming between the connected hyphae (Video S1). Cluster germination for all isolates was observed during the first 24 h. They germinated as a syncytium, which means that the germ tube never emerged through the subtending hypha when the spores were sharing a common mycelium, but one or several germ tubes re-grew from the cut extremities (Fig. 2). After 1 wk, several spores within the cluster lost their cytoplasmic content and became translucent, while others within the same cluster seemed to be dormant, as no change in color and lipid content was observed.

Self fusion was a common event within the same isolate (Fig. 2), and ranged from 60% to 38% in DAOM-197198 and DAOM-2404515, respectively. Pre-fusion as well as post-fusion incompatibility was never recorded in pairings within the same isolate. On the contrary, hyphal interaction between pairings of different isolates (i.e. nonself interaction) showed heterogeneous responses. Nonself hyphal fusion was rare between different isolates, and was recorded only two times in the combination DAOM-234328/DAOM-240415 out of 157 contacts observed (Supporting Information Table S1). Nonself post-fusion incompatibility was recorded for all combinations and the highest value was 3 (2%) out of 148 contacts observed in DAOM-197198/DAOM-240415. Microscopic observations of these hyphal fusions revealed slow protoplasm movement alongside the connected hyphae followed by protoplasm withdrawal and septa formation (Video S2).

Symbiotic interactions

Growth of the mycelium during the symbiotic phase showed the same developmental patterns for *Rhizophagus* species (formerly *Glomus* genus) (Bago *et al.*, 1998). After germination and root colonization, runner hyphae (RH) rapidly colonized the medium, and extended towards the interaction zone. They divided dichotomously into thinner hyphae of increasing branch order. Hyphae subsequently branched profusely to form the so-called branching absorbing structures (BAS). Analysis of the hyphal interactions between isolates also showed a heterogeneous response, as in the first experiment. No differences in hyphal density, number of contacts or anastomosis formation between the treatments were detected ($P \geq 0.05$) (Table 2). Perfect hyphal fusions between colonies of the same isolate ranged from 19 ± 8.2 in DAOM-240415 to 49.5 ± 17.5 in DAOM-234328 (mean $\pm$ SE) Pre-fusion as well as post-fusion incompatibility was never recorded, as observed in the pre-symbiotic interactions.

Nonself fusions were only observed at the zone of interaction for the combination DAOM-240415/DAOM-234328. Because anastomosis traceability was solely confined to the zone of the interaction, the occurrence of nonself fusions far beyond the interaction zone might not be ruled out. Pre-fusion incompatibility was observed for all combinations characterized by hyphal burst at the hyphal apex or septa formation and cytoplasm retraction before contact. No signs of wall thickening at the hyphal-apex region were observed when hyphae from different isolates came into contact, but overlapped mycelium was the rule rather than the exception for all combinations.

Sporogenesis was observed during the first week after root colonization on the RH. Most spores were formed into lower order hyphae after extensive growth of the fungal colony. Full lipids and empty spores as well as aborted-like structures were observed and recorded in all the isolates tested, regardless of the combination. The percentage of aborted-like structures in the interaction zone ranged from absence to 2 ± 1 (mean $\pm$ SE) among crossed experiments and from 2 ± 2 to 5 ± 2 in controls. The same pattern was observed at each side, when the plate was inoculated with both isolates in crossing (i.e. nonself interaction), or within controls (i.e. self interactions). Statistical analysis did not show any significant difference in the number and percentage of spores produced for any of the categories ($P \geq 0.05$) mentioned above (Table S2). Although in the combination DAOM-197198/DAOM-240415 (Fig. 4, Video S3), several aborted-like structures (data not shown) often arrested their development and burst, after which several hyphae grew out of the spore lumen (Fig. 5) Observations of these hyphae showed that cytoplasm retracted and several septa were then formed.

mtDNA genotyping and sequencing in spore progeny

For each crossing combination in a symbiotic interaction, 10 full-lipid spores of the progeny were individually genotyped using isolate-specific markers corresponding to their parents, as well as the progeny of the monosporal cultures lines experiment. The

Table 2 Anastomosis[1,2] frequency from the interaction of *Rhizophagus irregularis* isolates in the symbiotic experiment

Type of interaction	Hyphal density at the interaction zone[3] (cm cm^{-2})	Number of contacts	Perfect fusion (self fusion)[4]	No interaction	Pre-fusion incompatibility or hyphal avoidance	Post-fusion incompatibility
Within the same isolate						
A–A	10.8 ± 1	42 ± 25	26 ± 23	16 ± 2	0	0
B–B	13.4 ± 3.1	55 ± 14	19 ± 8.2	36.3 ± 6.6	0	0
C–C	14.6 ± 1.2	77 ± 18	49.5 ± 17.5	30 ± 3	0	0

Between isolates	Hyphal density at the interaction zone (cm cm^{-2})	Number of contacts	Nonself fusion[4]	No interaction	Nonself pre-fusion incompatibility	Nonself post-fusion incompatibility
A–C	9.4 ± 1.4	25 ± 7.1	0	23.7 ± 7.8	1.2 ± 0.8	0
A–B	12 ± 2	25 ± 7.1	0	23.8 ± 7.7	1.25 ± 0.7	0
B–C	11.2 ± 1.6	26.7 ± 4.3	1.3 ± 0.9	25.3 ± 3.9	0.33 ± 0.33	0

A, DAOM 197198; B, DAOM 240415; C, DAOM 234328.
[1]Anastomoses within the same mycelium in crossing treatment were not recorded.
[2]No significance differences were found within the treatments and controls ($P \geq 0.05$).
[3]Mean ± SE.
[4]Fusions were followed for 3 wk and no septa were formed between the connected hyphae.

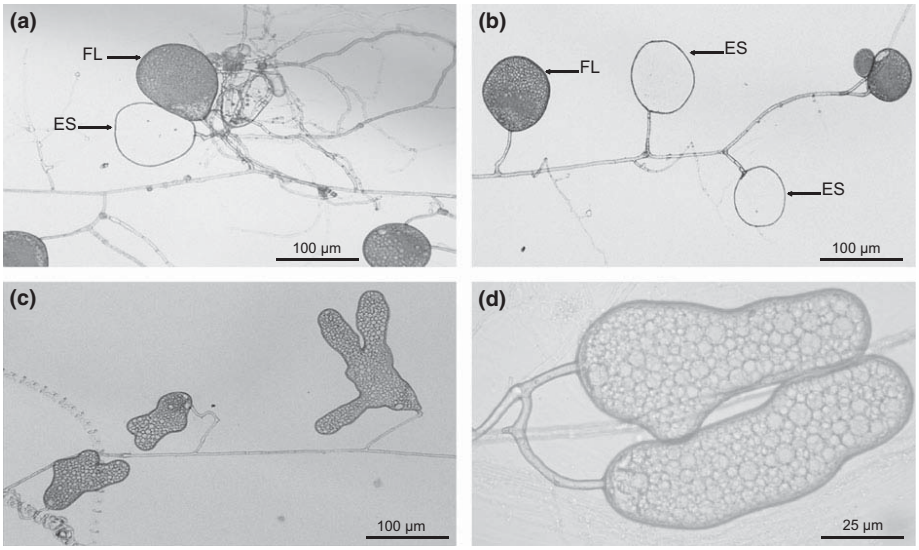


Fig. 4 Variability in spore morphology of *Rhizophagus irregularis* at the interaction zone in the combination DAOM197198/DAOM240415. (a) Spore clusters showing empty, full lipid droplets and crushed spores. (b) Extraradical mycelium bearing both full lipid droplets (FL) and emptied spores (ES). (c, d), Irregular shaped spores within the same mycelium. For a detailed view of the spore variability see Video S3.

progeny of the three crossing combinations showed different PCR amplification patterns (Fig. 6). For example, the crossing of isolates DAOM-197198 and DAOM-240415 showed that nine progeny spores (S) out of ten contained both parental markers and one spore (S6) did not produce a PCR amplification product using mtDNA markers, while *18S* rDNA primers AML1 and AML2 showed a PCR band of an expected size (790 bp; data not shown). Five spores of the crossing combination between DAOM-197198 and DAOM-234328 showed both mtDNA haplotypes, while five contained only 197198-mtDNA haplotype. Three spores (S3, S6 and S9) of the crossing combination between DAOM-240415 and DAOM-234328 showed both mtDNA haplotypes, six spores showed one mtDNA haplotype whereas DAOM-240415 mtDNA haplotype was predominant (S1, S2, S5, S7 and S10) and spore (S8) did not produce a PCR amplification product using mtDNA markers, while *18S* rDNA primers did produce a PCR band (data not shown). Interestingly, the PCR bands varied greatly in intensity among spores and independently of the length of the marker, probably due to the frequency of each mtDNA haplotype.

Monosporal cultures with progeny from crossed experiments

Even though attempts to obtain monosporal culture lines failed, spore germination was observed for progenies from all crossed experiments and ranged from 40% in DAOM-197198/DAOM-234328 to 60% in DAOM-240415/DAOM-234328 (Table S3). The germination process was characterized by the emission of one to several germinative hyphae through the subtending

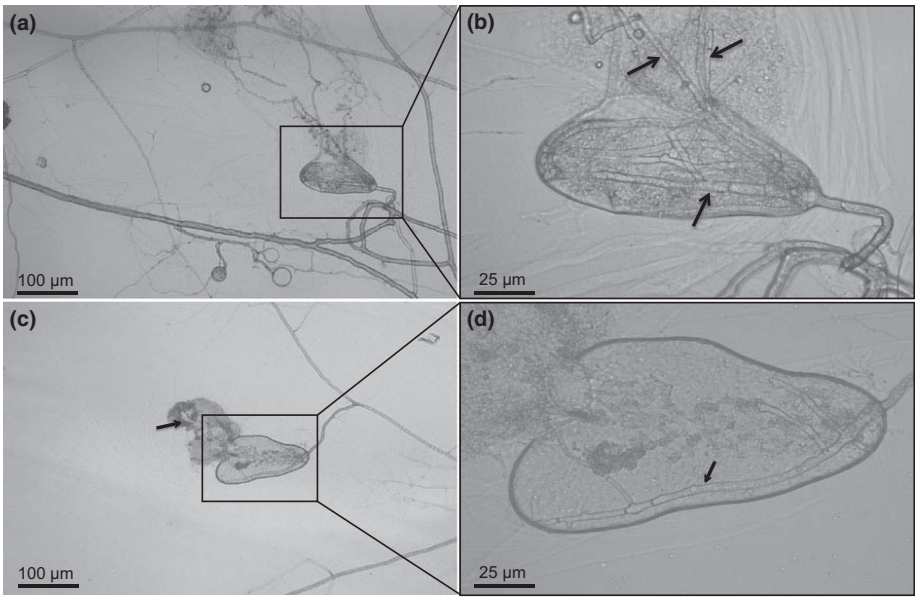


Fig. 5 Aborted-like structures produced by the *Rhizophagus irregularis* isolates at the interaction zone, (a, b) DAOM-197198/DAOM-240415; (c, d) DAOM-234328/DAOM-240415. (a) Several hyphae exit from the aborted-like structures lumen and continue growing following cytoplasm retraction and multiple septa formation. (b) Inset showing details of the lumen and septated hyphae (arrows). (c) Burst and cytoplasm exit (black arrow). (d) Inset showing details of the septated hypha inside the aborted-like structures.

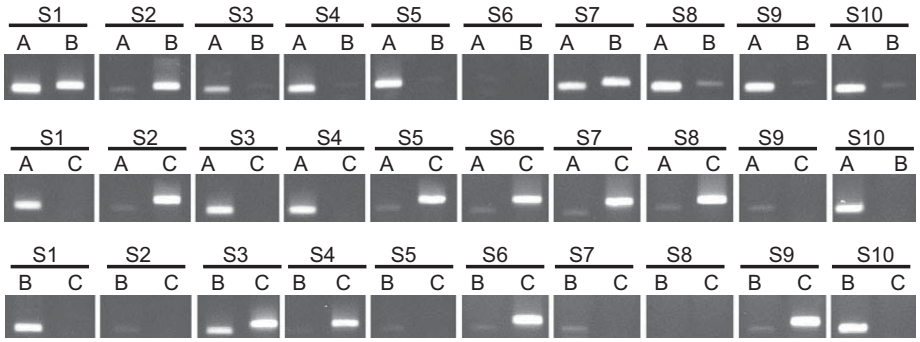


Fig. 6 Gel electrophoresis showing patterns of mtDNA genotyping of ten progeny spores (S1–S10) of three combinations using three *Rhizophagus irregularis* isolates DAOM-197191 (A) DAOM-240415 (B) and DAOM-234328 (C). mtDNA marker sizes are 120, 147 and 209 bp corresponding to DAOM-197198, DAOM-240415 and DAOM-234328, respectively.

hyphae (Fig. S1). Some germinative hyphae showed a marked rhizo-tropism (Fig. S1); however, a nonorientated growth was also observed. Germinative hyphae also underwent contact with roots but after several weeks no progress of the fungal growth was observed (i.e. runner hyphae bearing spores and branching absorbing structures).

Genotyping of the progeny showed that 14 spores gave positive PCR amplifications while the rest of the progeny did not produce any amplification product. All PCR products were sequenced and matched to the isolate-specific parental mtDNA markers. However, genotyping results showed that all the progeny amplified only one parental mtDNA haplotype (Fig. S2). For example, in the combination DAOM-197198/DAOM-234328, three spores were genotyped of which two contained the DAOM-197198-mtDNA haplotype and the other showed the DAOM-234328 mtDNA haplotype. Five spores in the combination DAOM-240415/DAOM-234328 were genotyped where two spores contained the DAOM-234328 mtDNA and three presented the DAOM-240415 mtDNA haplotype. Finally, six spores were genotyped in combination DAOM-197198/DAOM-240415, where only one showed the DAOM-197198 genotype while the others contained the DAOM-240415 mtDNA haplotype.

Colonization detection

Microscopic observation of stained root fragments, which were suspected for being colonized by progenies, revealed the absence of hyphopodia formation. Consequently, no intraradical structure was observed.

Discussion

Nonself recognition between members of the *Rhizophagus* genus has been documented (Croll *et al.*, 2009; Angelard *et al.*, 2010; Colard *et al.*, 2011; Purin & Morton, 2012) and has resulted in nuclear segregation to new individuals affecting plant fitness (Angelard *et al.*, 2010; Colard *et al.*, 2011). Interestingly, combination B-C (isolates DAOM-240415 originating from Manitoba, Canada, and DAOM-234328 originating from Finland) showed perfect nonself fusion and the occurrence of heteroplasmy. Moreover, in the symbiotic crossing experiment of combinations A-B and A-C, we did not observe perfect nonself fusion, but a few proportions indicative of pre-fusion incompatibility; genotyping of mtDNA clearly showed the presence of both parental haplotypes in the spore progeny. These results can likely be explained on the basis of the coenocytic nature of the AMF

mycelium allowing protoplasm mixture before hyphal septation, suggesting the existence, as in higher fungi (Fu *et al.*, 2011), of a putative molecular mechanism regulating nonself recognition that has yet to be demonstrated.

The observed mtDNA heteroplasmy in the progeny following nonself recognition (i.e. unrelated to apparent sexual reproduction) between genetically divergent isolates of the same species originating from different continents is somewhat surprising. This assumption is made on the basis that all mtDNA sequenced so far in AMF (Borstler *et al.*, 2008; Lang & Hijri, 2009; Lee & Young, 2009; Formey *et al.*, 2012; Nadimi *et al.*, 2012; Pelin *et al.*, 2012; Beaudet *et al.*, 2013) and those presented in this study, have been shown to be homogeneous. This apparent contradiction is evidence of the putative transient heteroplasmic state in AMF, and is consistent with our results as both germinated progeny and spores that do not break their dormancy after long periods of sub-cultivation showed the presence of one parental mtDNA marker. This evidence suggests the presence of an effective segregation mechanism, which might regulate mtDNA inheritance in these organisms.

Inheritance of mitochondria in fungi during the mating process is mostly uniparental, with only one parental haplotype of the mitochondrial genome inherited by the progeny (reviewed in Ni *et al.*, 2011). However, biparental inheritance of mtDNA has also been shown in some fungal species, such as *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, because the two mating cells contribute equally through their cytoplasm organelles to form a mixed cytoplasm in the zygote. Interestingly, the coexistence of different mitochondrial genome haplotypes has been shown to be transient in these organisms (Birky, 2001; Barr *et al.*, 2005; Ni *et al.*, 2011). Mechanisms contributing to homoplasmy are often related to the recombination of both parental mtDNAs or by the selection of one of the parental haplotypes due to the presence of segregation mechanisms, such as genetic bottleneck followed by a stochastic drift (Birky, 2001). As proposed by White *et al.* (2008), heteroplasmy may either persist, become fixed, or be lost due to either selection forces or drift, or both. The extent to which such mechanisms may occur in AMF and the resulting haplotype to be fixed in the population will need to be assessed.

In arbuscular mycorrhizal fungi, no studies have been addressed to investigate patterns of mtDNA inheritance and segregation to the progeny, however, the occurrence of distinct mtDNA haplotypes within an AMF individual, with divergence in mobile elements and *dpo* insertions, points to the possible occurrence of interhaplotype recombination and mtDNA segregation mechanisms in these fungi (D. Beaudet *et al.*, unpublished data). In budding yeasts mtDNA segregation occurs via nucleoids (i.e. packages of identical mtDNAs maintained together by Holliday junctions and proteins; Lockshon *et al.*, 1995; White & Lilley, 1997; MacAlpine *et al.*, 1998), creating a genetic bottleneck responsible for further rapid segregation. Due to this mechanism, heteroplasmy can last over 20 generations in these organisms. It would be interesting to investigate whether the formation of such mitochondrial nucleoids occurs in AMF and, if so, how long this heteroplasmic state is maintained. The efficiency of the mitochondrial segregation process is directly correlated to the initial population size and the relative proportion of one given haplotype in a population (White *et al.*, 2008). In this study, only two different mtDNA haplotypes were mixed *in vitro*, giving rise to limited initial genetic variants. Also, we only have evidence of the mtDNA haplotypes frequency from the intensity of the PCR bands we observed, which make it difficult to predict the effectiveness of the segregation process.

Other studies have shown that coexistence of two mtDNAs within a common cytoplasm may be disadvantageous, likely due to genetic conflicts or competition between divergent mtDNA haplotypes (reviewed in Ni *et al.* (2011), references therein). It has been proposed that uniparental inheritance of mtDNA in fungi as well as other eukaryotes might be maintained by selection forces to avoid such conflicts (White *et al.*, 2008). Some authors have hypothesized that uniparental inheritance of mtDNA and chloroplastic DNA may avoid dispersion of selfish and deleterious mobile elements (Hoekstra, 2000; Mamirova *et al.*, 2007). Even if the divergence between the mtDNA haplotypes used in this study appeared to be neutral, because they are located in intergenic regions, it is difficult to predict with confidence the implied function in the genome and its role in the segregation and inheritance process. In any case, selection of one mtDNA haplotype due to potential mtDNA conflicts or stochastic segregation mechanisms might explain the transient character of heteroplasmy in our experiment and also provides a rational explanation for why homoplasmy is more often observed in spores from field (Borstler *et al.*, 2008) and/or *in vitro* experiments than heteroplasmy (Lee & Young, 2009; Formey *et al.*, 2012; Nadimi *et al.*, 2012; Pelin *et al.*, 2012; Beaudet *et al.*, 2013).

Previous studies on AMF mtDNA have already demonstrated the great potential for developing isolate-specific molecular markers (Borstler *et al.*, 2008; Lang & Hijri, 2009; Formey *et al.*, 2012; Beaudet *et al.*, 2013). For example, Borstler *et al.* (2008) used the *rnl* gene encoding the mitochondrial ribosomal large subunit (mtLSU) to discriminate different haplotypes among cultivated isolates of *R. irregularis* (formerly *G. intraradices*) in soil and within root samples from the field. These authors found 12 *rnl* haplotypes among 16 isolates of *R. irregularis* originating from five continents. Formey *et al.* (2012) recently compared the complete mtDNA of six *R. irregularis* isolates and were able to develop mtDNA markers using the *cox3-rnl* intergene as well as other regions to discriminate isolates using length polymorphism markers. These mtDNA markers can be useful for other studies, both in population genetics as well as in tracing isolates in agriculture soil and plant roots to test co-inoculation of different *R. irregularis* isolates. However, it will be important to have a better understanding of the mitochondrial inheritance process and the extent of the mtDNA variability present in a natural population in order to benefit from the full potential of such markers.

In this particular study, we believe it was crucial to use size-variable markers combined with a whole genome amplification step, to easily test the occurrence of heteroplasmy between the crossed parental cultures. Some studies have suggested that during their formation, spores can randomly inherit different nuclear haplotypes following hyphal fusion (Marleau *et al.*, 2011). The same

pattern could be expected for mtDNA, however they could be critically underrepresented in daughter spores, given the competition with the endogenous haplotypes and the putative process of mtDNA segregation. If the isolate-specific markers were variable in single nucleotide polymorphisms (SNPs), it would have required a large number of clones for each tested spore to find even one different locus. PCR and cloning-based studies with such markers could fail to identify occurrences of heteroplasmy. Despite its potential as a DNA barcode, mtDNA has never been considered as a unique and reliable highly-sensitive target to address studies on fungal individualism in AMF (Rayner, 1991).

Conclusions

We have demonstrated a potential use of mtDNA markers to address basic questions of evolutionary biology in AMF using *R. irregularis* isolates as a model. Our study provided additional information in the AMF mitochondrial genome collection with the complete mtDNA of two *R. irregularis* isolates. Currently, there are eight *R. irregularis* isolates with annotated mitochondrial genomes that differ substantially in some intergenic regions, making them useful sites for developing isolate-specific markers. We also provided additional markers that were specific to some isolates, and demonstrated their potential use in tracing mtDNA in a crossing experiment.

We also showed that isolates of *R. irregularis* from different geographical locations, even from different continents, were able to exchange their genetic material regarding mtDNA, resulting in heteroplasmy in their spore progeny. The extent to which such a state may occur in natural communities and the mechanisms responsible for the fixation of a homogeneous haplotype in an AMF individual need to be assessed. Further studies on vegetative compatibility and incompatibility as well as putative sex machinery in AMF, are needed to advance our understanding of the evolution of AMF and to better understand their population dynamics.

Acknowledgements

This study was conducted as part of a research project organized and coordinated by Premier Tech. The authors are grateful for financial supports from NSERC Cooperative Research and Development (grant number: RDCPJ 395241-09), NSERC Discovery (grant number: 328098-2012), Premier Tech and CRIBIQ. I.D.P. was supported by Genome Canada and Genome Quebec funds. We would also like to thank Biopterre *centre du développement des bioproduits* and CRBM for assistance and Nathalie De Jaeger, Terrence Bell, Karen Fisher Favret and three anonymous referees for their useful reading and comments on the manuscript.

References

Angelard C, Colard A, Niculita-Hirzel H, Croll D, Sanders IR. 2010. Segregation in a mycorrhizal fungus alters rice growth and symbiosis-specific gene transcription. *Current Biology* **20**: 1216–1221.

Angelard C, Sanders IR. 2011. Effect of segregation and genetic exchange on arbuscular mycorrhizal fungi in colonization of roots. *New Phytologist* **189**: 652–657.

Bago B, Azcón-Aguilar C, Piché Y. 1998. Architecture and developmental dynamics of the external mycelium of the arbuscular mycorrhizal fungus *Glomus intraradices* grown under monoxenic conditions. *Mycologia* **90**: 52–62.

Barr CM, Neiman M, Taylor DR. 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytologist* **168**: 39–50.

Beaudet D, Nadimi M, Iffis B, Hijri M. 2013. Rapid mitochondrial genome evolution through invasion of mobile elements in two closely related species of arbuscular mycorrhizal fungi. *PLoS ONE* **8**: e60768.

Bécard G, Fortin AJ. 1988. Early events of vesicular-arbuscular mycorrhiza formation on Ri T-DNA transformed roots. *New Phytologist* **108**: 211–218.

Bever JD, Wang M. 2005. Arbuscular mycorrhizal fungi: hyphal fusion and multigenomic structure. *Nature* **433**: E3–E4; discussion E4.

Birky WC. 2001. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annual Review of Genetics* **35**: 125–148.

Blaszkowski J, Czerniawska B, Wubet T, Schuessler T, Buscot F, Renker C. 2008. *Glomus irregulare*, a new arbuscular mycorrhizal fungus in the Glomeromycota. *Mycotaxon* **106**: 247–267.

Bolduc AR, Hijri M. 2011. The use of mycorrhizae to enhance phosphorus uptake: a way out the phosphorus crisis. *Journal of Biofertilizers & Biopesticides* **2**: 104 doi: 10.4172/2155-6202.1000104.

Boon E, Zimmerman E, Lang BF, Hijri M. 2010. Intra-isolate genome variation in arbuscular mycorrhizal fungi persists in the transcriptome. *Journal of Evolutionary Biology* **23**: 1519–1527.

Borstler B, Raab PA, Thiery 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**: 452–465.

Colard A, Angelard C, Sanders IR. 2011. Genetic exchange in an arbuscular mycorrhizal fungus results in increased rice growth and altered mycorrhiza-specific gene transcription. *Applied and Environmental Microbiology* **77**: 6510–6515.

Corradi N, Croll D, Colard A, Kuhn G, Ehinger M, Sanders IR. 2007. Gene copy number polymorphisms in an arbuscular mycorrhizal fungal population. *Applied and Environmental Microbiology* **73**: 366–369.

Croll D, Giovannetti M, Koch AM, Sbrana C, Ehinger M, Lammers PJ, Sanders IR. 2009. Nonself vegetative fusion and genetic exchange in the arbuscular mycorrhizal fungus *Glomus intraradices*. *New Phytologist* **181**: 924–937.

Doner LW, Bécard G. 1991. Solubilization of gellan gels by chelation of cations. *Biotechnology Techniques* **5**: 25–28.

Ehinger MO, Croll D, Koch AM, Sanders IR. 2012. Significant genetic and phenotypic changes arising from clonal growth of a single spore of an arbuscular mycorrhizal fungus over multiple generations. *New Phytologist* **196**: 853–861.

Formey D, Moles M, Haouy A, Savelli B, Bouchez O, Becard G, Roux C. 2012. Comparative analysis of mitochondrial genomes of *Rhizophagus irregularis* – syn. *Glomus irregulare* – reveals a polymorphism induced by variability generating elements. *New Phytologist* **196**: 1217–1227.

Fu C, Iyer P, Herkal A, Abdullah J, Stout A, Free SJ. 2011. Identification and characterization of genes required for cell-to-cell fusion in *Neurospora crassa*. *Eukaryotic Cell* **10**: 1100–1109.

Giovannetti M, Azzolini D, Citernesi AS. 1999. Anastomosis formation and nuclear and protoplasmic exchange in arbuscular mycorrhizal fungi. *Applied and Environmental Microbiology* **65**: 5571–5575.

Giovannetti M, Sbrana C, Avio L, Strani P. 2004. Patterns of below-ground plant interconnections established by means of arbuscular mycorrhizal networks. *New Phytologist* **164**: 175–181.

Giovannetti M, Sbrana C, Strani P, Agnolucci M, Rinaudo V, Avio L. 2003. Genetic diversity of isolates of *Glomus mosseae* from different geographic areas detected by vegetative compatibility testing and biochemical and molecular analysis. *Applied and Environmental Microbiology* **69**: 616–624.

Halary S, Malik S-B, Lildhar L, Slamovits CH, Hijri M, Corradi N. 2011. Conserved meiotic machinery in *Glomus* spp., a putatively ancient asexual fungal lineage. *Genome Biology and Evolution* **3**: 950–958.

van der Heijden MGA, Bardgett RD, van Straalen NM. 2008. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology Letters* **11**: 296–310.

van der Heijden MG, Streitwolf-Engel R, Riedl R, Siegrist S, Neudecker A, Ineichen K, Boller T, Wiemken A, Sanders IR. 2006. The mycorrhizal contribution to plant productivity, plant nutrition and soil structure in experimental grassland. *New Phytologist* **172**: 739–752.

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

Hnatova M, Gbelska Y, Obernauerova M, Subikova V, Subik J. 2003. Cross-resistance to strobilurin fungicides in mitochondrial and nuclear mutants of *Saccharomyces cerevisiae*. *Folia Microbiologica* **48**: 496–500.

Hoekstra R. 2000. Evolutionary origin and consequences of uniparental mitochondrial inheritance. *Human Reproduction* **15**: 102–111.

Ismail Y, Hijri M. 2012. Arbuscular mycorrhisation with *Glomus irregulare* induces expression of potato PR homologues genes in response to infection by *Fusarium sambucinum*. *Functional Plant Biology* **39**: 236–245.

Jany JL, Pawlowska TE. 2010. Multinucleate spores contribute to evolutionary longevity of asexual glomeromycota. *American Naturalist* **175**: 424–435.

Jeffries P, Gianinazzi S, Perotto S, Turnau K, Barea J-M. 2003. The contribution of arbuscular mycorrhizal fungi in sustainable maintenance of plant health and soil fertility. *Biology and Fertility of Soils* **37**: 1–16.

Kuhn G, Hijri M, Sanders IR. 2001. Evidence for the evolution of multiple genomes in arbuscular mycorrhizal fungi. *Nature* **414**: 745–748.

Lang BF, Hijri M. 2009. The complete *Glomus intraradices* mitochondrial genome sequence – a milestone in mycorrhizal research. *New Phytologist* **183**: 3–6.

Leake J, Johnson D, Donnelly D, Muckle G, Boddy L, Read DJ. 2004. Networks of power and influence: the role of mycorrhizal mycelium in controlling plant communities and agroecosystem functioning. *Canadian Journal of Botany* **82**: 1016–1045.

Lee J, Young JPW. 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**: 200–211.

Lesemann SS, Schimpke S, Dunemann F, Deising HB. 2006. Mitochondrial heteroplasmy for the cytochrome b gene controls the level of strobilurin resistance in the apple powdery mildew fungus *Podosphaera leucotricha* (Ell. & Ev.) E.S. Salmon. *Journal of Plant Diseases and Protection* **113**: 259–266.

Lockshon D, Zweifel SG, Freeman-Cook LL, Lorimer HE, Brewer BJ, Fangman WL. 1995. A role for recombination junctions in the segregation of mitochondrial DNA in yeast. *Cell* **81**: 947–955.

MacAlpine DM, Perlman PS, Butow RA. 1998. The high mobility group protein Abf2p influences the level of yeast mitochondrial DNA recombination intermediates *in vivo*. *Proceedings of the National Academy of Sciences, USA* **95**: 6739–6743.

Mamirova L, Popadin K, Gelfand MS. 2007. Purifying selection in mitochondria, free-living and obligate intracellular proteobacteria. *BMC Evolutionary Biology* **7**: 17.

Marleau J, Dalpe Y, St-Arnaud M, Hijri M. 2011. Spore development and nuclear inheritance in arbuscular mycorrhizal fungi. *BMC Evolutionary Biology* **11**: 51.

Nadimi M, Beaudet D, Forget L, Hijri M, Lang BF. 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**: 2199–2210.

Ni M, Feretzaki M, Sun S, Wang X, Heitman J. 2011. Sex in fungi. *Annual Review of Genetics* **45**: 405–430.

Pelin A, Pombert J-F, Salvioli A, Bonen L, Bonfante P, Corradi N. 2012. The mitochondrial genome of the arbuscular mycorrhizal fungus *Gigaspora margarita* reveals two unsuspected trans-splicing events of group I introns. *New Phytologist* **194**: 836–845.

Phillips JM, Hayman DS. 1970. Improved procedures for clearing roots and staining parasitic and vesicular-arbuscular mycorrhizal fungi for rapid assessment of infection. *Transactions of The British Mycological Society* **55**: 157–160.

Purin S, Morton JB. 2012. Anastomosis behaviour differs between asymbiotic and symbiotic hyphae of *Rhizophagus clarus*. *Mycologia* **12**: 589–602.

Rayner ADM. 1991. The challenge of the individualistic mycelium. *Mycologia* **83**: 48–71.

Rillig MC, Mummey DL. 2006. Mycorrhizas and soil structure. *New Phytologist* **171**: 41–53.

Schussler A, Walker C. 2010. *The Glomeromycota: a species list with new families and genera*. Edinburgh & Kew, UK: The Royal Botanic Garden; Munich, Germany: Botanische Staatssammlung Munich; OR, USA: Oregon State University (http://www.amf-phylogeny.com).

Smith SE, Read DJ. 2008. *Mycorrhizal symbiosis*. London, UK: Academic Press.

Smith ML, Duchesne LC, Bruhn JN, Anderson JB. 1990. Mitochondrial genetics in a natural population of the plant pathogen armillaria. *Genetics* **126**: 575–582.

Stockinger H, Walker C, Schussler A. 2009. '*Glomus intraradices* DAOM197198', a model fungus in arbuscular mycorrhiza research, is not *Glomus intraradices*. *New Phytologist* **183**: 1176–1187.

Taylor DR, Zeyl C, Cooke E. 2002. Conflicting levels of selection in the accumulation of mitochondrial defects in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences, USA* **99**: 3690–3694.

White DJ, Wolff JN, Pierson M, Gemmell NJ. 2008. Revealing the hidden complexities of mtDNA inheritance. *Molecular Ecology* **17**: 4925–4942.

White MF, Lilley DM. 1997. Characterization of a Holliday junction-resolving enzyme from *Schizosaccharomyces pombe*. *Molecular and Cellular Biology* **17**: 6465–6471.

Supporting Information

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

Fig. S1 Spore germination features.

Fig. S2 Genotyping germinated and nongerminated.

Table S1 Number and percentage of anastomosis in pre-symbiotic mycelium

Table S2 Total number of spores by category

Table S3 Monosporal culture lines

Video S1 Protoplasmic flow through anastomosis.

Video S2 Slow protoplasm flow after fusion.

Video S3 Variability in spore morphology.

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