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

¹Institut de Recherche en Biologie Végétale, Département de Sciences Biologiques, Université de Montréal, Montréal, Québec, Canada
²Département de Biochimie, Centre Robert-Cedergren, Université de Montréal, Montréal, Québec, Canada

†These authors contributed equally to this work.

*Corresponding author: Email: franz.lang@umontreal.ca.

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-encoding 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...
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 quillwort *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 Glomeromyctota 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'-CTGTATTGTCATGCCTGT-3') and exon 5 (5'-AAAGGCTGAATAACATGGCT-3'). For rns, the two respective primers were 5'-ACCGAAATCAGGA-3' and 5'-CACACTATCATGTTACG-3'. To test for functionality of the RT-PCR assay (positive control), an internal region of cox1 exon 4 was amplified (primers 5'-CTCTAGACGGGACCAGTC-3' and 5'-CCCCGATCTAGAAAGCT-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'-GGCGATGCCTGAGAAACATGGCT-3', 5'-CGAGTATGACAGCACTGTC-3', and 5'-GTCTGTCATAACAGCCTTCCTA-3'. To test for functionality of the RT-PCR assay (positive control), an internal region of cox1 exon 4 was amplified (primers 5'-CTCTAGACGGGACCAGTC-3' and 5'-CCCCGATCTAGAAAGCT-3').

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, 1/4 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). Minixons 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,
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 minintrons, 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 $-d$C 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 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

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

<table>
<thead>
<tr>
<th>Species</th>
<th>Genes and Introns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rnl, ms</td>
</tr>
<tr>
<td>Gigaspora rosea</td>
<td>2</td>
</tr>
<tr>
<td>Glomus irregulare 494</td>
<td>2</td>
</tr>
<tr>
<td>G. irregulare 197198</td>
<td>2</td>
</tr>
<tr>
<td>Smitium culcetae</td>
<td>2</td>
</tr>
<tr>
<td>Mortierella verticillata</td>
<td>2</td>
</tr>
<tr>
<td>Rhizopus oryzae</td>
<td>2</td>
</tr>
<tr>
<td>Allozymes macrogrunyus</td>
<td>2</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae</td>
<td>2</td>
</tr>
</tbody>
</table>

*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).
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. irregularare 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 Gigaspora and Glomus, 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.

**Fig. 1.** Comparison of Gigaspora rosea and Glomus irregularare mitochondrial genomes. The circular-mapping genomes of G. rosea and G. irregularare 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.
Evolution of the Mitochondrial Genetic Code in Glomeromyctota

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 Glomeromyctota, deviations from the standard genetic code also vary. A few clear-cut instances of UGA(Trp) codons are found in the two published *G. irregularare* 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 Glomeromyctota with Mortierellales

The Glomeromyctota are part of zygomycetes, a phylogenetically heterogeneous (paraphyletic) taxon (Seif et al. 2005; James et al. 2006; Liu, Steenkamp, et al. 2009). According to a current proposal, Glomeromyctota 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. irregularare* 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 Glomeromyctota merit the status of an independent phylum.

When *G. rosea* data are included in a data set containing all standard mtDNA-encoded proteins, Glomeromyctota 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 Glomeromyctota 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 Glomeromyctota 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 Glomeromyctota increases substantially (PP 1.0; 93% JV).

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

<table>
<thead>
<tr>
<th>Species</th>
<th>Accession</th>
<th>Introns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Glomus irregulariae</em></td>
<td>FJ648425</td>
<td>adj1398, aden515</td>
</tr>
<tr>
<td><em>Gigaspora rosea</em></td>
<td></td>
<td>adj1398, aden515,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aden515</td>
</tr>
<tr>
<td>Basidiomycota</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Cryptococcus neoformans</em></td>
<td>NC004336</td>
<td>adj1398,</td>
</tr>
<tr>
<td><em>Aspergillus niger</em></td>
<td>EF204913</td>
<td>adj1398,</td>
</tr>
<tr>
<td><em>Pleurotus ostreatus</em></td>
<td>EF204913</td>
<td>adj1398,</td>
</tr>
<tr>
<td><em>Ascomycota</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Penicillium chrysogenum</em></td>
<td>GU133622</td>
<td>adj1398,</td>
</tr>
<tr>
<td><em>Aspergillus niger</em></td>
<td>NC001329</td>
<td>adj1398,</td>
</tr>
<tr>
<td><em>Zygomycota</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Rhizopus oryzae</em></td>
<td>CY663211</td>
<td>adj1398,</td>
</tr>
<tr>
<td><em>Chytridiomycota</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Monoblepharella</em></td>
<td>CY663212</td>
<td>adj1398,</td>
</tr>
<tr>
<td><em>A. macrognosus</em></td>
<td>NC001715</td>
<td>adj1398,</td>
</tr>
</tbody>
</table>

**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 Dombrowska 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. irregulariae* 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. irregulariae* 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* group matches perfectly the two partial discontinuous intron sequences, completing a comprehensive intron RNA secondary structure (fig. 3A 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. 3C). Sequencing of the PCR product confirmed the 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.
experiments confirm this view: the insert in \textit{rns} \textunderscore 2 is not excised at the RNA level nor are five intron-size inserts in \textit{rnl}. According to our results, RNA splicing occurs only for the two \textit{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 \textit{rnl} inserts occurs precisely at a position where \textit{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 \textit{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 \textit{Glomus} and \textit{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 \textit{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

\textbf{Fig. 3.} Model of group I intron-mediated trans-splicing in \textit{G. rosea} and demonstration of mRNA trans-splicing by RT-PCR. (A) Schematic view of the trans-spliced \textit{cox1} group IB intron RNA secondary structure. The pink-colored RNA corresponds to the predicted \textit{nad5} intron 3 helper sequence, and dashed loops indicate its interaction with the two intron fragments flanking \textit{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 \textit{cox1} trans-splicing. Blue and red arrows indicate PCR and RT-PCR primers (see Materials and Methods). (C) Amplification products obtained with \textit{G. rosea} cDNA and genomic DNA. a, RT-PCR amplification in the presence of total RNA (demonstration of trans-splicing and that mitochondrial \textit{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.
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


Mitochondrial Evolution in Gigaspora

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.


