J Mol Evol (2002) 55:74–84
DOI: 10.1007/s00239-001-2305-x

JOURNAL OF MOLECULAR EVOLUTION

Vertical Evolution and Intragenic Spread of Lichen-Fungal Group I Introns

Debashish Bhattacharya,[1] Thomas Friedl,[2] Gert Helms[2]

[1] Department of Biological Sciences and Center for Comparative Genomics, University of Iowa, 239 Biology Building, Iowa City, IA 52242-1324, USA
[2] Albrecht-von-Haller-Institut für Pflanzenwissenschaften, Abteilung Experimentelle Phykologie und Sammlung von Algenkulturen, Universität Göttingen, 37073 Göttingen, Germany

Received: 19 September 2001 / Accepted: 17 January 2002

Abstract. One family within the Euascomycetes (Ascomycota), the lichen-forming Physciaceae, is particularly rich in nuclear ribosomal [r]DNA group I introns. We used phylogenetic analyses of group I introns and lichen-fungal host cells to address four questions about group I intron evolution in lichens, and generally in all eukaryotes: 1) Is intron spread in the lichens associated with the intimate association of the fungal and photosynthetic cells that make up the lichen thallus? 2) Are the multiple group I introns in the lichen-fungi of independent origins, or have existing introns spread into novel sites in the rDNA? 3) If introns have moved to novel sites, then does the exon context of these sites provide insights into the mechanism of intron spread? and 4) What is the pattern of intron loss in the small subunit rDNA gene of lichen-fungi? Our analyses show that group I introns in the lichen-fungi and in the lichen-algae (and lichenized cyanobacteria) do not share a close evolutionary relationship, suggesting that these introns do not move between the symbionts. Many group I introns appear to have originated in the common ancestor of the Lecanorales, whereas others have spread within this lineage (particularly in the Physciaceae) putatively through reverse-splicing into novel rRNA sites. We suggest that the evolutionary history of most lichen-fungal group I introns is characterized by rare gains followed by extensive losses in descendants, resulting in a sporadic intron distribution. Detailed phylogenetic analyses of the introns and host cells are required, therefore, to distinguish this scenario from the alternative hypothesis of widespread and independent intron gains in the different lichen-fungal lineages.

Key words: Euascomycetes — Group I intron — Intron movement — Lecanorales — Lichens — Physciaceae — Reverse-splicing — Ribosomal RNA

Correspondence to: D. Bhattacharya; *email*: dbhattac@blue.weeg.uiowa.edu

Introduction

Group I introns are a distinct class of RNA-enzymes (ribozymes) that are characterized by a conserved RNA primary and secondary structure (see R. Gutell's Comparative RNA Web Site, http://www.rna.icmb.utexas.edu/) essential for splicing, and are often capable of "self-splicing" (Kruger et al. 1982; Cech 1985). The sporadic and wide distribution of group I introns in the nuclear-encoded ribosomal (r-)DNA and organellar genes of green algae, red algae, fungi, ciliates, and different amoebae suggests that these sequences are highly successful at invading and maintaining themselves in eukaryotic genomes (Lambowitz and Belfort 1993; Turmel et al. 1993; Bhattacharya et al. 1994; Gargas et al. 1995; Bhattacharya 1998; Nishida et al. 1998; Schroeder-Diedrich et al. 1998; Watanabe et al. 1998; Perotto et al. 2000). The mechanisms by which group I introns are able to "move" within and between genes and between na-

tural populations and species is, therefore, one of the most challenging problems in their evolution (see Bhattacharya et al. 2001).

To address intron movement, we focus on an assemblage of organisms that provides the most compelling example that is known of nuclear group I intron spread in eukaryotes, the lichen fungi within the family Physiaceae (Euascomycetes, Bhattacharya et al. 2000). Some of these taxa (e.g., *Physconia*) contain up to eight introns in the SSU rDNA gene (see below). Lichens are symbioses of photosynthetic cells and fungi. Lichen symbioses originate through the long-term association of previously free-living fungi and cyanobacteria or coccoid green algae and are widespread in the Euascomycetes (Gargas et al. 1995; Lutzoni et al. 2001). Though in close proximity, the fungal and photosynthetic cells normally maintain tissue independence. The lichen symbiosis has been characterized both as a controlled parasitism because the fungus often penetrates the cell wall (not the cytoplasm) of the lichen alga with structures called haustoria to facilitate carbohydrate transfer (Honegger 1991), to a partnership that is beneficial to both symbionts. The fungal thallus, for example, affords the lichen alga protection from desiccation and high light levels, allowing the photosynthetic cells to colonize habitats that lie outside their normal range of tolerance.

Here, we take advantage of the primary and secondary structure conservation of group I introns to align intron sequences from different insertion sites in Physiaceae and other lichen-fungal rDNA genes for phylogenetic analyses (e.g., De Jonckheere and Brown 1994; Gielly and Taberlet 1994; Hibbett 1996; Bhattacharya et al. 1996; Bhattacharya 1998; Besendahl and Bhattacharya 1999; Besendahl et al. 2000). Our analyses elucidate the evolutionary history of the different intron lineages in the Physciaceae and specifically address the following questions regarding group I intron evolution: 1) Is intron spread in the lichens associated with the intimate association of the fungal and photosynthetic cells that make up the lichen thallus; i.e., can introns move from one symbiont to the other? 2) Are the multiple group I introns in the lichen-fungi of independent origins, or have existing introns spread into novel sites in the rDNA? 3) If introns have moved to novel sites, then does the exon context of these sites provide insights into the mechanism of intron spread? 4) What is the pattern of intron loss in the SSU rDNA of lichen-fungi?

Materials and Methods

DNA Extraction and PCR Sequencing

DNA was isolated from members of the ascomycetes family Physciaceae (Table 1) as previously described (Helms et al. 2001). Mycobiont-specific PCR primers (for SSU rDNA: MY60F, for ITS rDNA: MY1200F) were combined with primers MY1200R/MY1300R (G. Helms unpublished data) and ITS4/LR3 (White et al. 1990) to amplify SSU and/or ITS rDNA sequences. PCR conditions were as in Helms et al. (2001). PCR products were cleaned with an ethanol precipitation and sequenced directly. Alternatively, PCR products were cloned using the pGEM-TVector-System I (Promega) and competent DH 5α cells of *Escherichia coli*. The inserts were then amplified from transformant colonies using M13 primers and sequenced directly. A set of nested sequencing primers was used for the SSU and primers LR1850 (Friedl 1996) and ITS4 (White et al. 1990) for the ITS rDNAs. The sequencing primers were IRD-labeled and used for cycle-sequencing with the Thermo Sequenase Sequencing Kit with 7-deaza-dGTP (Amersham). Reactions were run on a L4200 automated sequencer (LI-COR).

Construction of Group I Intron and "Host" Cell Phylogenies

For the group I intron data, a set of 90 group I and IE intron sequences from nuclear (16S and 16S-like) SSU rDNA in lichen-fungi were manually aligned using primary and secondary structure similarity to juxtapose homologous regions (e.g., Michel and

Table 1. Taxa of the Physciaceae for which SSU rDNAs, group I introns, and ITS rDNA regions were determined in this study

Taxon	Source of rDNA	ITS GenBank No.	SSU GenBank No.
*Anaptychia runcinata**	Triebel and Rambold 6162 (M)[a]	AJ421249	AJ421692
Buellia elegans	Trinkaus 439 (GZU)	—	AJ421680
*Buellia georgei**	Trinkaus 356a (GZU)	AJ421416	AJ421681
Buellia zoharyi	Trinkaus 450 (GZU)	—	AJ421682
Dimelaena oreina	Mayrhofer 13.737 (GZU)	—	AJ421683
*Diploicia canescens**	Triebel and Rambold 6188 (M)	AJ421992	AJ421684
*Heterodermia boryi**	Spain, Canary Islands	AJ421419	AJ421685
*Physcia adscendens**	Germany, Müllheim	AJ421414	AJ421686
*Physcia aipolia**	Germany, Pullach	AJ421420	AJ421687
*Physcia stellaris**	Germany, Munich	AJ421421	AJ421688
*Physconia perisidiosa**	Germany, Pullach	AJ421422	AJ421689
*Rinodina cacuminum**	Mayrhofer 13.706 (GZU)	AJ421693	AJ421690
*Rinodinella controversa**	Mayrhofer and Ertl 13.747 (GZU)	AJ421423	AJ421691

M and GZU indicate samples for which herbaria vouchers have been deposited in the herbaria GZU (Universität Graz, Austria) and M (Botanische Staatssammlung München, Germany), respectively. Other lichen material used in this study has not been deposited in a public herbarium.

*Taxa that contained SSU rDNA introns.

Westhof 1990; Bhattacharya et al. 1994; Damberger and Gutell 1994; Bhattacharya et al. 2001). An alignment with a total length of 143 characters was submitted to a pairwise distance analysis using the Jukes-Cantor model (Jukes and Cantor 1969) with equal divergence rates across sites and the transition/transversion ratio = 2. Missing data and gaps were excluded from each pairwise comparison. This distance matrix was used as input for neighbor-joining tree building. One-thousand bootstrap replicates (Felsenstein 1985) were analyzed with the distance method. Hierarchical likelihood ratio tests were done to identify the best-fit model for the intron data set (MODELTEST v3.06; Posada and Crandall 1998). This analysis showed the general time reversible model (Rodriguez et al. 1990) with estimations of nucleotide frequencies (A = 0.192, C = 0.297, G = 0.293, T = 0.219), the shape parameter of the gamma distribution (α = 1.370) to accommodate rate variations across sites, and the proportion of invariant sites (0.024) that are unable to accept substitutions as fitting best the intron data. The GTR + G + I model was used to calculate bootstrap values (1000 replications) for monophyletic groups identified by the Jukes-Cantor distance tree. In this way, we had estimates of support for different groups in the intron tree with evolutionary models that were relatively simple (Jukes-Cantor) or parameter-rich (GTR + G + 1). The IE introns were used to root the intron tree.

For the SSU rDNA sequence data, an initial alignment as constructed by Lumbsch et al. (2001) was obtained from TreeBase (http://herbaria.harvard.edu/treebase/) to which were added additional available and newly determined fungal SSU rDNA sequences (Tables 1 and 2). The alignment was then manually refined using BioEdit 5.0.6 (Hall 1999). The final alignment contained 47 fungal SSU rDNA sequences and (after exclusion of regions that could not be aligned unambiguously) was 1285 nt in length, 389 nt of which were variable with 222 parsimony informative sites. *Saccharomyces cerevisiae* and *Candida albicans* (Saccharomycetales) were used as outgroup taxa. To further examine the phylogenetic relationships of group I intron-containing lichenized species of the ascomycetes family Physciaceae (Lecanorales and Tehler 1996), a second data set of ITS rDNA sequences was constructed that contained the same Physciaceae taxa as in the SSU rDNA data set and additional taxa from this group. Newly determined (Table 1) and available (Table 2) ITS rDNA sequences for the Physciaceae were aligned using CLUSTAL W (Thompson et al. 1994) and then manually refined using BioEdit. The ITS data set contained 29 taxa and (after exclusion of regions of ambiguous alignment) 271 sequence positions with 183/129 variable/parsimony- informative sites. No outgroup taxa were defined for the ITS data set.

Both SSU and ITS rDNA data sets were subjected to maximum-parsimony, distance, and maximum likelihood analyses. Modeltest was used to identify the model of DNA substitution that best fits the SSU and ITS rDNA data. For the SSU rDNA data set, these analyses showed the time reversible model with equal base frequencies (TrNef + I + G, Tamura and Nei 1993) with the shape parameter of the gamma distribution (α = 0.5892) to accommodate rate variations across sites, and the proportion of invariant sites (0.4775) that are unable to accept substitutions as fitting best the data. For the ITS data set also the time reversible model (TrN + G, Tamura and Nei 1993) with estimations of nucleotide frequencies (A = 0.236, C = 0.297, G = 0.232, T = 0.235), the shape parameter of the gamma distribution (α = 0.7873), and the proportion of invariant sites (0.024) was selected. For maximum parsimony analyses of both data sets, the sites were weighted (rescaled consistency index [RC] over an interval of 1–1000) and then used as input for bootstrap analyses (1000/2000 replications for the SSU/ITS rDNA data sets). Starting trees were built stepwise with 10 random additions of taxa, using the tree bisection-reconnection branch-swapping algorithm to find the best tree. Best scoring trees were held at each step. Gaps were treated as missing data. Distance analyses were done using the parameter-rich models (TrNef + I + G/TrN + G) and the LogDet (LD) transformation (Lockhart et al. 1994) with proportion of invariant sites as estimated from Modeltest runs. To build the distance trees, the neighbor-joining (Saitou and Nei 1987) and the minimum evolution (Rzhetsky and Nei 1992) methods were used and bootstrap analyses (1000/2000 replications for the SSU/ITS rDNA data sets) were done with these models. Maximum likelihood (ML) trees were constructed using the same models (TrNef + I + G/TrN + G). Initial ML trees were used as input for subsequent ML runs to search for trees with better-Ln likelihoods. No bootstrap resampling was done on ML trees. All phylogenetic analyses were done with PAUP* V4.0b8 (Swofford 2001).

Results and Discussion

Phylogeny of Lichen-Fungal Group I Introns and Host Cells

Two different distance models were used to infer a neighbor-joining tree of group I introns in the SSU rDNA of lichen-fungi. The tree, inferred using the Jukes-Cantor correction, is shown in Fig. 1. The second model (GTR + G + I) resulted in a nearly identical phylogeny. The bootstrap analyses using these methods showed significant support for a number of nodes in the tree. The I and IE introns form two independent lineages, suggesting a long evolutionary separation of these sequences (Suh et al. 1999; Bhattacharya et al. 2001) and group I introns at nearly all rDNA insertion sites form monophyletic groups with moderate to strong bootstrap support [e.g., sites 516, 1199, 1210, 1389, (the intron numbering reflects their homologous position in the *E. coli* gene)]. As an example of this result, the distribution of the eight different group I introns in the SSU rDNA of *Physconia perisidiosa* is shown with filled circles in Fig. 1. These introns all cluster with other sequences at their respective insertion sites suggesting that each intron has been vertically inherited for some time within the different lichen fungi. The eight introns are not, therefore, the result of intron invasions exclusively in *P. perisidiosa* rDNA.

Our phylogenetic analyses show, however, that there are a number of introns at heterologous sites that form monophyletic groups with strong bootstrap support (i.e., 114 + 303 and 1046 + 1052 introns, Fig 1). The 1506 + 1516 and 287 + 1199 introns are also closely related although not with strong bootstrap support. Note that the 1506 intron in the lichen-fungi has an independent evolutionary origin from the more common 1506 intron found in non-lichenized fungi (e.g., *Cladosporium caryigenum, Fusicladium effusum*) and in many other algae and protists (Bhattacharya 1998; Bhattacharya et al. 2001). The 1506 position may, therefore, be a favored site for intron insertion in SSU rRNA. Similarly, the 1046 and 1052 positions are also favored for intron insertion because the lichen-fungal introns at these sites are evolutionaily distinct from the 1046 and 1052 green algal group I introns (see Bhattacharya 1998, and the position in

Table 2. Reference taxa used for the construction of SSU rDNA and ITS rDNA host trees

Classification	Taxon	SSU GenBank No.	ITS GenBank No.
Agyriales	*Placopsis gelida*	AF119502	
Eurotiales	*Eremascus albus*	M83258	
	Eurotium rubrum	U00970	
Lecanorales			
Bacidiaceae	*Bacidia rosella*	AF091585	
	Cliostomum griffithii	AF091590	
	Lecania cyrtella	AF091589	
	Squamarina lentigera	AF088250	
	Toninia sedifolia	AF091591	
Cladoniaceae	*Cladia retipora*	AF184751	
	Cladonia rangiferina[*]	AF184753	
	Cladonia subcervicornis[*]	AF08547	
	Cladonia sulphurina[*]	AF241544	
	Stereocaulon paschale	AF140236	
Lecanoraceae	*Gymnoderma lineare*	AF085470	
	Lecanora intumescens[*]	AF091586	
	Lecanora dispersa	L37734	
	Lecidella meiococca	AF091583	
	Pyrrhospora quernea	AF091584	
Parmeliaceae	*Alectoria ochroleuca*	AF140234	—
	Cetraria islandica	AF117986	—
	Cornicularia normoerica	AF140235	—
	Hypogymnia physodes	AF117984	—
	Usnea florida	AF117988	—
	Vulpicina pinastri	AF117989	—
Physciaceae	*Buellia papillata*	—	AF250790
	Buellia schaereri	—	AF250791
	Heterodermia speciosa	—	AF224360
	Phaeophyscia ciliata	—	AF224457
	Phaeophyscia constipata	—	AF224374
	Phaeophyscia endococcinea	—	AF224444
	Phaeophyscia endophoenicea	—	AF224445
	Phaeophyscia nigricans	—	AF224375
	Phaeophyscia orbicularis	—	AF224452
	Phaeophyscia sciastra	—	AF224357
	Phaeophyscia sp. PSD1	—	AF278752
	Physcia caesia	—	AF224435
	Physcia dubia	—	AF224403
	Physcia tenella	—	AF224425
	Physconia enteroxantha	—	AF224370
	Physconia grisea	—	AF224368
	Rinodina plana	—	AF250812
	Rinodina tunicata	—	AF250816
Rhizocarpaceae	*Rhizocarpon geographicum*	AF088246	—
Sphaerophoraceae	*Austropeltum glareosum*	AF117982	—
	Sphaerophorus stereocauloides	AF184760	—
Teloschistaceae	*Caloplaca flavorubescens*	AF241540	—
	Xanthoria elegans	AF088254	—
	Xanthoria parietina	AF241541	—
Pertusariales	*Pertusaria trachythallina*	AF088242	—
Pleosporales	*Pleospora herbarum*	U43448	—
Saccharomycetales	*Candida albicans*	X53497	—
	Saccharomyces cerevisiae	M27606	—

[*]Taxa that contain introns.

Fig. 1 of the 1052 intron from the non-lichenized green alga, *Chlorella luteoviridis*). These results suggest that evolutionarily distinct introns have independently inserted into the 1506, 1046, and 1052 SSU rDNA sites in lichens and other protists.

One explanation for the origin of the "mixed" intron lineages identified in Fig. 1 (i.e., 114 + 303, 287 + 1199, 1046 + 1052, 1506 + 1516) is the intragenic movement of an intron to a heterologous rDNA site in one organism. A prediction of this hypothesis is that, given the establishment of intragenic intron movement in a common ancestor, some descendants may maintain both introns, whereas others may contain one or both sequences, based on the

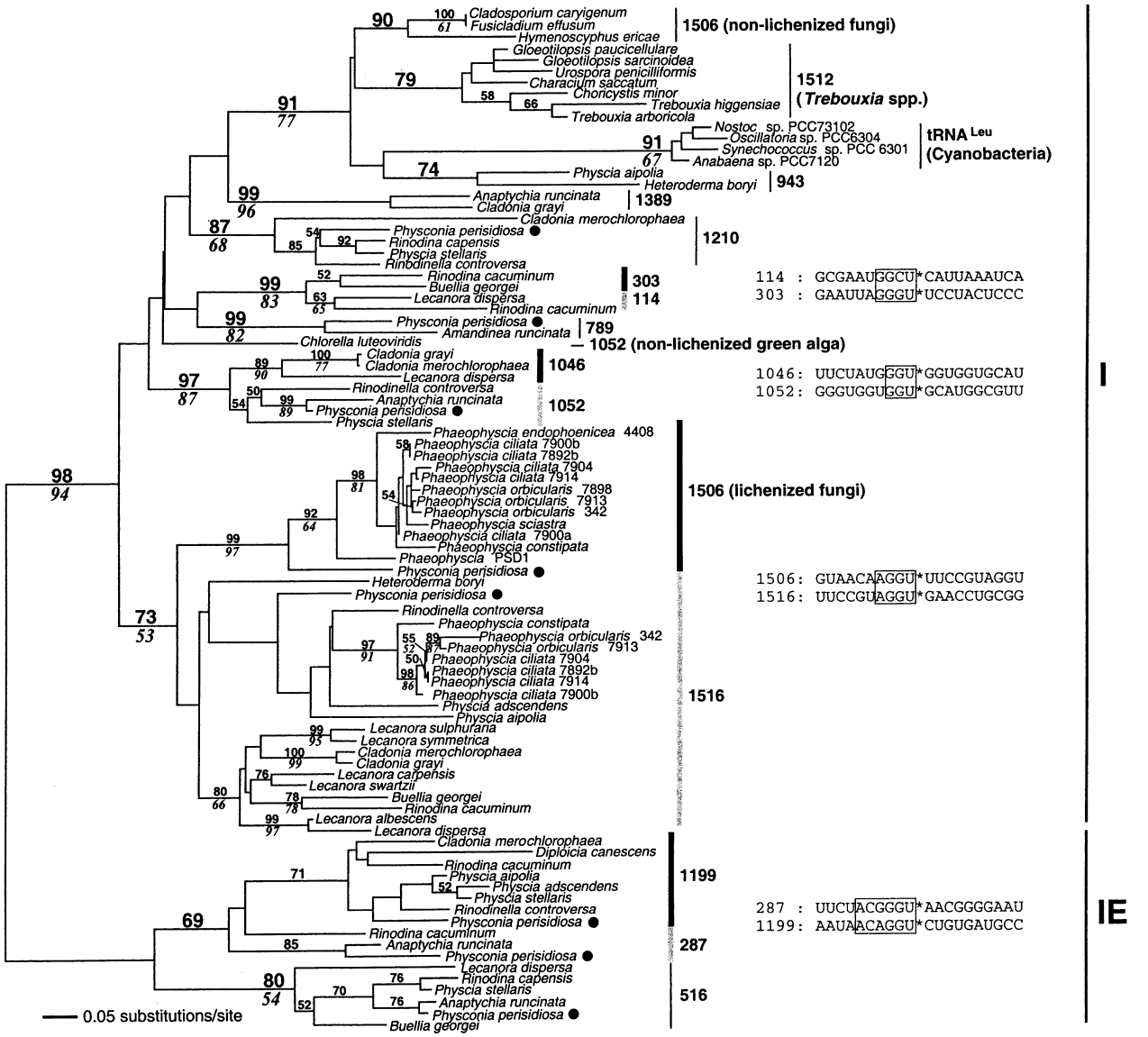

Fig. 1. Neighbor-joining tree of rDNA group I introns from lichen-fungi including introns from non-lichenized fungi, green algae (e.g., *Trebouxia* spp., which are lichen algal-symbionts), and cyanobacteria. This tree has been rooted with the group IE introns. The values above the branches show the results of a distance bootstrap analysis (2000 replications) using the Jukes-Cantor correction, whereas the values shown below the branches are from a distance bootstrap analysis using the GTR + G + I model. Only bootstrap values ≥50% are shown. The coding region positions are shown for each intron lineage. The intron lineages marked with the thicker solid/dotted vertical lines indicate four putative cases of intragenic intron spread. The proximal flanking exon sequences are shown for these mixed intron lineages. The eight introns in the *Physconia perisidiosa* SSU rDNA are marked with the filled circles. The accession numbers for the lichen-fungal introns are found in Tables 1 and 2, whereas the numbers for introns in non-lichenized taxa are available from GenBank.

number of intron losses that have occurred during their evolution. Inspection of the SSU rDNA tree in Fig. 2 shows that there exist taxa that contain both 114 + 303 introns (*Rinodina cacuminum*), both 287 + 1199 introns (*Physconia perisidiosa, R. cacuminum*), and both 1506 + 1516 introns (*P. perisidiosa, Phaeophyscia* spp.). Both 1046 + 1052 introns are not, however, found in any of the taxa in our study. Given the possibility of intragenic intron movement, we then analyzed the exon sequences flanking the mixed intron sites to look for clues to their mechanism of spread. This analysis showed little conservation of exon sequence apart from a short 3–6 nt stretch in the 5′-proximal region (Fig. 1 and see below).

Intron Distribution on Host Trees

The mapping of the presence/absence of group I introns on the SSU rDNA phylogeny of Physciaceae and other fungi shows that group I introns are concentrated in the Physciaceae although *Cladonia merochlorophaea*, *Lecanora dispersa*, and *Cordyceps*

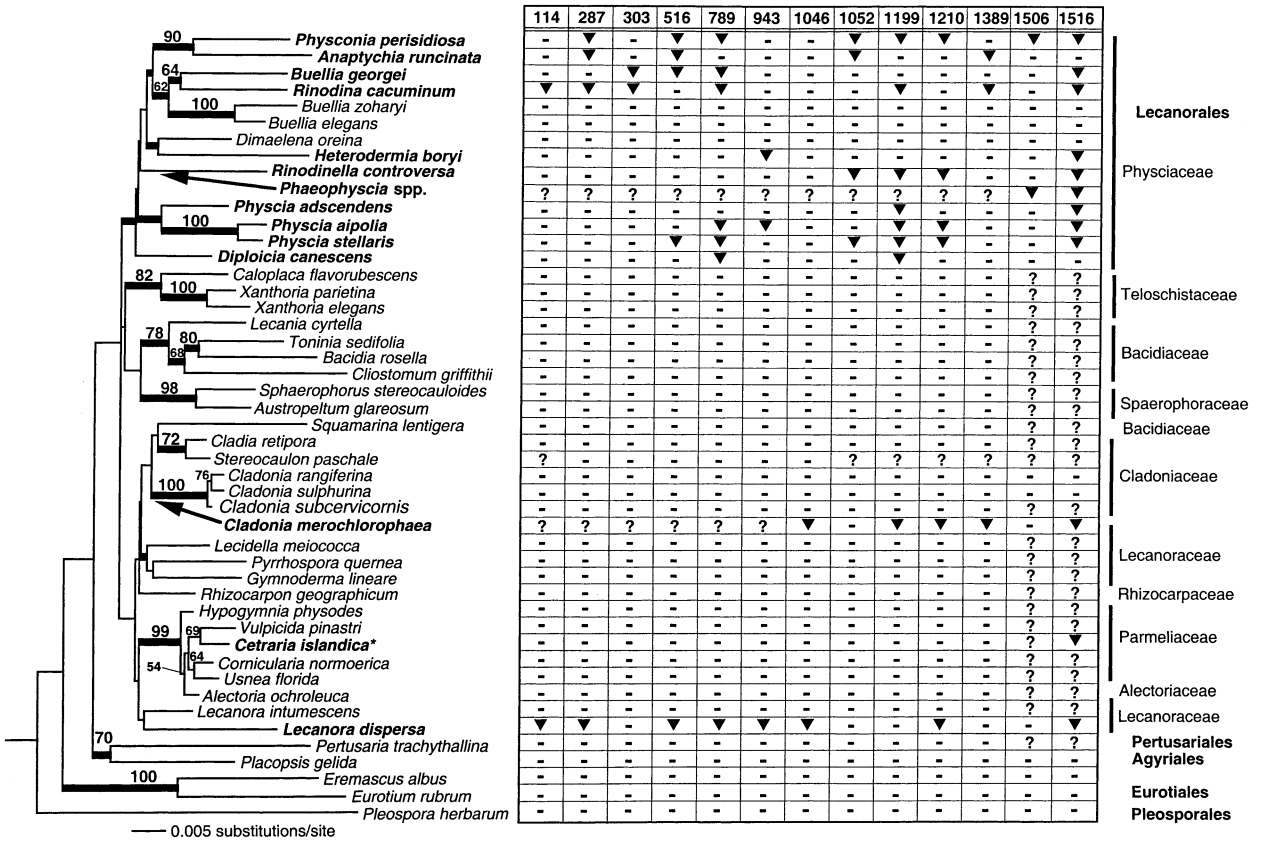

Fig. 2. Small-subunit rDNA tree of lichen-fungi families of the Lecanorales and other ascomycetes orders. A LogDet neighbor-joining tree is shown with bootstrap values (above nodes) from 1000 replicates using neighbor-joining in connection with the TrNEF+I+G model (see text). The thick lines indicate branches that were resolved in distance, maximum parsimony, and maximum likelihood analyses. The presence/absence of 13 different group I introns is shown; intron presence is denoted with a filled triangle and intron absence with a dash. Missing data for the SSU rDNA is shown with a question mark. Taxa containing introns are in bold face. The *Cetraria islandica* strain used in this analysis [marked with an asterisk (GenBank AF117986)] has not been sequenced in the region of the 1506–1516 introns, however, other subspecies, which have only been partially sequenced at the 3′ terminus of the SSU rDNA [e.g., *C. islandica* subsp. *islandica* (GenBank AF228297)], contain the 1516 intron. The approximate positions of taxa that are known from ITS sequence comparisons [*Phaeophyscia* spp. (see Fig. 3), *Cladonia merochlorophaea* (Piercey-Nomore and DePriest 2001)] for which a complete SSU rDNA sequence does not exist, are shown with the arrows.

spp. are also intron-rich (DePriest and Been 1992; Gargas et al. 1995; Nikoh and Fukatsu 2001). Some of the introns (e.g., at positions 114, 287, 516, 789, 943, 1199, 1389) appear to have been present (at least) in the common ancestor of *Cladonia*/*Lecanora* and the Physciaceae (Fig. 2). Under a hypothesis of rare intron gains, intron loss appears to be much more common with multiple losses occurring in lineages such as in the Alectoriaceae, Bacidiaceae, Sphaerophoraceae, Rhizocarpaceae, and the Teloschistaceae. Although, the missing data for the extreme 3′ terminus of the SSU rDNA in many of these latter taxa makes this a preliminary result, the majority of the coding regions are intron-free.

To understand better intron loss, we limited our host cell phylogenetic analysis to ITS regions in the Physiaceae (Fig. 3). Most nodes in this tree are well-resolved and provide more concrete insights into intron evolution. The region interrupted by the 1506 and 1516 SSU rDNA introns have been sequenced in these taxa. To identify cases of intron loss, we assumed that the 1516 intron was present in the common ancestor of the Physciaceae (i.e., homologs are present in *L. dispersa* or *C. merochlorophea*) and then calculated the number of putative losses on the tree shown in Fig. 3. We also counted the number of independent losses of the 1506 intron that was present at least in the common ancestor of the taxa that diverge after *Diploicia canescens* (see Fig. 3). The parsimony principle was used so that we chose the loss pattern that minimized the total number of events. This analysis resulted in the following numbers of intron presence/loss: 1506 (7/4) and 1516 (15/11). In summary, of the 29 taxa included in Fig. 3, only 18 contained one or both of these introns. Overall, our data suggest that loss may be a pervasive phenomenon in group I intron evolution. There are two important caveats of this analysis. The first is that we have presumed the tree shown in Fig. 3

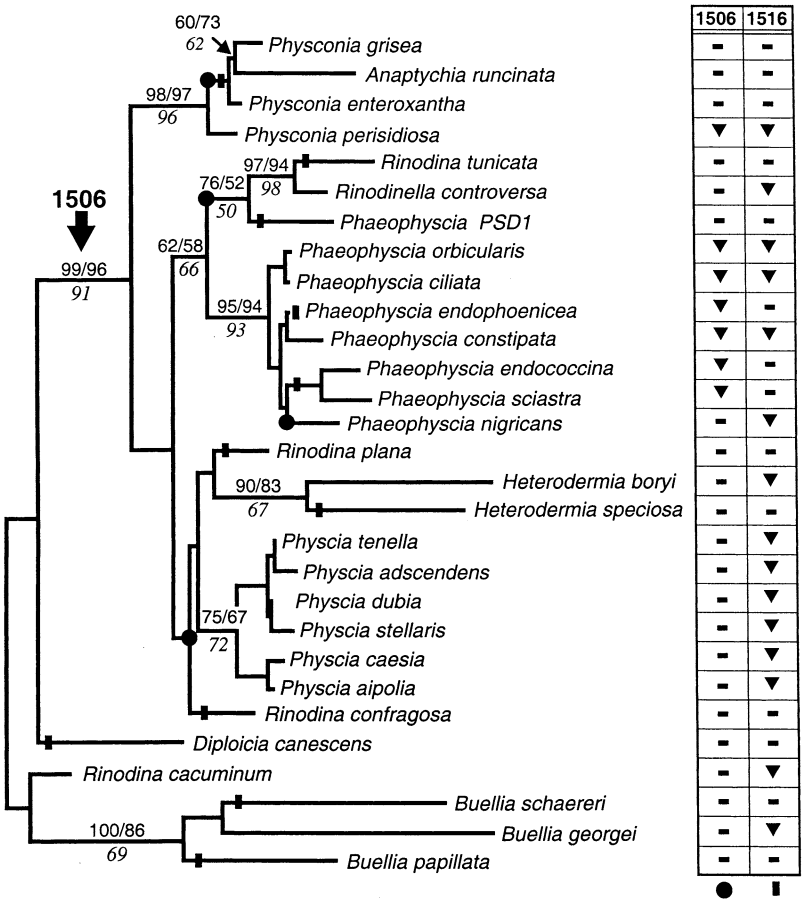

Fig. 3. An ITS rDNA tree of the Physciaceae; one of 10 equally parsimonious trees is presented. Bootstrap values (from 2000 replications) are from weighted parsimony (above lines, left), minimum evolution using the TrN + G model (above lines, right), and LogDet (below lines) analyses. Bootstrap values below 50% are not shown. The distribution of the 1506 and 1516 group I introns is shown. Intron presence is denoted with a filled triangle and intron absence with a dash. The patterns of independent intron loss are marked in the tree with filled circles for the 1506 intron [four losses (after the divergence of *Diploicia canescens*)] and filled rectangles for the 1516 intron (11 losses).

to be the "true" tree, although some nodes do not have significant bootstrap support. It is, therefore, entirely possible that the numbers of losses may change with the use of a more complete and resolved phylogeny of the Physiaceae. The second caveat is that we are using a minimum estimate of intron presence because we have sampled single individuals of each species, and there may be other individuals in the same or other populations which contain the intron. Although a daunting task with lichens, the gain/loss of group I introns would best be studied in a population context with multiple individuals of each species being sampled from a large number of different populations. These data would allow us to estimate frequencies of intron presence in the different populations or to establish complete loss from the species.

Do Group I Introns Move Between Lichen-Fungal and Lichen-Algal Symbionts?

Given the success of lichens (over 17,000 species are estimated to exist), it is surprising that relatively little is known about the genetic basis for this symbiosis or whether genetic material can be, or has been, transferred between the symbionts (Nash 1996). Recent data show that lichen symbioses are very dynamic with algal symbionts frequently switching their lichen-fungal partners (Friedl et al. 2000; Piercey-Nomore and DePriest 2001). This switching may promote group I intron movement between different *Trebouxia* strains/species (Friedl et al. 2000) and could potentially facilitate intron lateral transfer between lichen-fungi and between lichen-algae.

The results of our phylogenetic analysis, suggest, however, that in spite of their close physical proximity, group I introns are not transferred between lichen-fungi and lichen-algae (or cyanobacteria). The $tRNA^{Leu}$ intron in cyanobacteria (e.g., *Nostoc* spp. are often lichenized [Paulsrud and Lindblad 1998; Besendahl et al. 2000]) and the 1512 intron in the SSU rDNA of lichen-algae in the genus *Trebouxia* (Bhattacharya et al. 1996; Friedl et al. 2000) are distantly related to the fungal group I introns. Within the 1512 lineage, the *Trebouxia* introns are most closely related to the homologous intron in non-lichenized green algae (e.g., *Urospora penicilliformis*, *Protoderma sarcinoidea*). We have previously speculated that the transfer of lichen group I introns may be facilitated by viruses (e.g., as between *Chlorella* viruses, Nishida et al. 1998) which may infect different strains of either lichen-fungi or lichen-algae but not both phylogenetically distantly related symbiont partners (Bhattacharya et al. 1996). This

prediction would explain why, in spite of intron-richness in both lichen partners, the introns in these taxa are only distantly related to each other (Fig. 1). There is, however, not yet any direct proof for virus-mediated group I intron transfer in lichens.

How Do Group I Introns Spread in Lichen-Fungal rDNA?

Two general models exist for group I intron spread. In the first, mobility at the DNA-level is facilitated by an endonuclease (ENase) that recognizes a long sequence (14–40 nt) within a homologous intronless copy of the coding region (Lambowitz and Belfort 1993). Initiation of a double-strand break in the DNA of the intronless allele is followed by a unidirectional conversion event that results in intron insertion (Belfort and Perlman 1995; Johansen et al. 1997). The intron-containing allele is then protected from the endonuclease because the intron interrupts the ENase recognition sequence (Edgell et al. 2000). This process, called intron "homing", is highly efficient at introducing introns into the alleles of a gene but is not expected to lead to the frequent, lateral transfer of group I introns into heterologous sites (intron transposition) because of the high ENase sequence-specificity (Woodson and Cech 1989; but see Bryk et al. 1993).

The role of homing in the spread of group I introns in the nucleus, particularly in lichen-fungi, is unclear because of the paucity of ENase coding regions (Bhattacharya et al. 1996). None of the introns shown in Fig. 1, for example, encodes an ENase coding region (either functional or inactivated). If ENases (encoded within the intron sequences) did, however, mediate the lateral transfer of lichen-fungal nuclear rDNA introns, then one would have to postulate widespread and perfect deletion of these coding regions after the introns had attained their present distribution. In contrast, when ENases have rarely been found in nuclear group I introns, they did not have extensive deletions but rather frame-shift mutations or short truncations that resulted in their inactivation (e.g., Haugen et al. 1999; Johansen and Haugen 1999).

Group I introns may also spread at the RNA-level through a reversal of the splicing process. In "reverse-splicing," a free intron inserts into a homologous or a heterologous RNA by recognizing a short (4–6 nt) sequence at the 5′ splice site, the internal guide sequence (IGS, Woodson and Cech 1989; Cech et al. 1994; Thompson and Herrin 1994; Roman and Woodson 1995, 1998). Upon incorporation into the nuclear genome, the spread of intron-containing rDNA coding regions could result from unequal crossing over, or gene conversion, phenomena that are known to lead to the stochastic or directional fixation of mutations in multiple-copy gene families (Dover 1994). Loss of group I introns is also thought to occur at the RNA-level through reverse-transcription of an RNA after intron excision, and general recombination with the intron-containing coding region in the genome. These scenarios of intron gain and loss are, therefore, both dependent on reverse transcription followed by recombination. The extra step (i.e., reverse-splicing) that is required for intron gain suggests these events should be rarer than losses. Another possibility for intron gain or loss is through homologous exon recombination between intron-plus and intron-minus forms of the genes in a genetic cross (see Burke 1988). Given, however, the absence of ENase open reading frames in the introns and the lack of extensive flanking exon sequence identity at the different sites of lichen-fungal group I intron insertion, the reverse-splicing model is presently the best working model for the spread of these sequences. This hypothesis is also supported by the limited 3–6 nt conservation within the proximal 3′-exon sequence among mixed intron lineage pairs (see Fig. 1), a pattern of sequence conservation that is predicted for reverse-splicing-mediated intron movement (Woodson and Cech 1989; Roman and Woodson 1998).

To assess the likelihood of reverse-splicing spread of the lichen-fungal mixed intron lineages, we built putative P1 and P10 helices for the 114 + 303, 1046 + 1052, and 1506 + 1516 introns using a comparative approach to determine whether they could form stable pairings at both the native and heterologous sites. The conserved exon-intron interaction (e.g., G • U bulged base pair) at the 5′ splice site was used to guide pairing of this proximal region of the helix, whereas the folding of the distal region was determined using mfold V3.1 (http://bioinfo.math.rpi.edu/-mfold/rna/form1.cgi, Zuker et al. 1999). The ability to build the P1 and P10 domains with rRNA exon flanking sequences is required for both group I inton forward-and reverse-splicing (Cech 1988; Woodson and Cech 1989; Cech et al. 1994). This analysis showed that, consistent with reverse-splicing spread, the *R. cacuminum* and *P. perisidiosa* introns can build stable pairings at the native and heterologous SSU rRNA sites (Fig. 4). Splicing analyses of these introns at the native and heterologous sites is required to test this result.

Summary

In summary, our data show that many group I introns (e.g., 114, 287, lichen fungal 1046, 1506) have been gained in the common ancestor of the lichen-fungi (*Physciaceae* + *Cladonia*/*Lecanora*), that lichen-fungal and lichen-algal introns are not specifically related, that there may have been intragenic movement of some introns from one rDNA site to another,

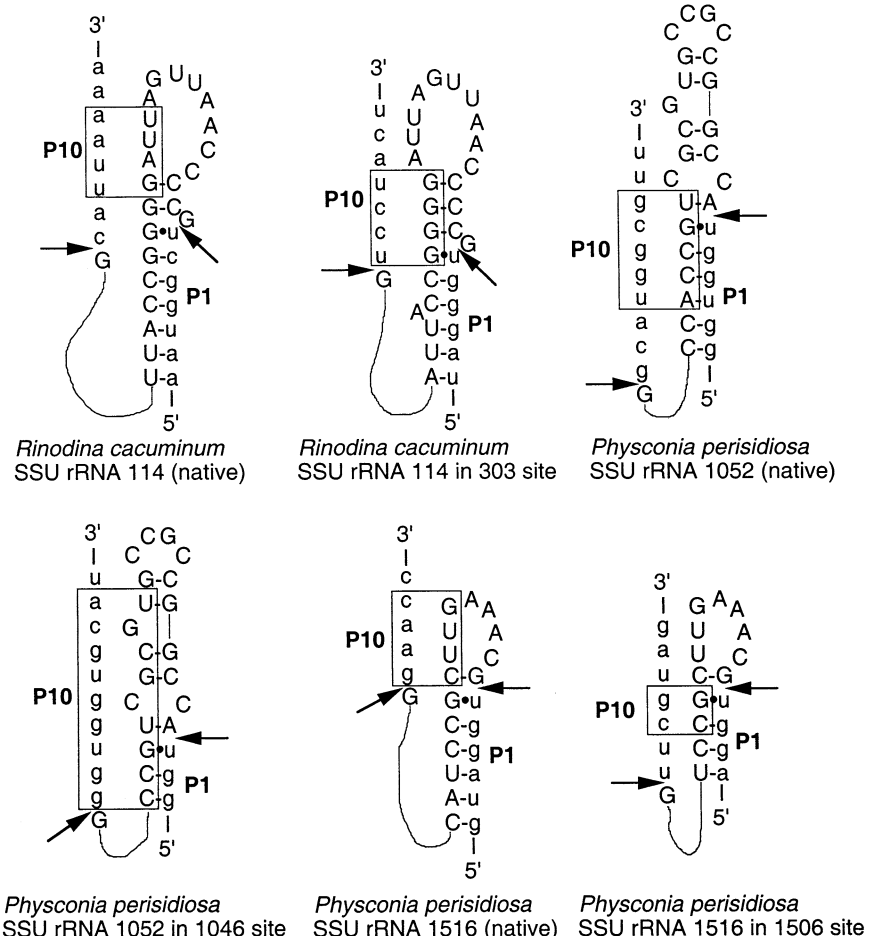

Fig. 4. Secondary structure models of helices P1 and P10 in lichen-fungal SSU rDNA group I introns. The P1 and P10 (boxed) helices of the *Rinodina cacuminum* 114 intron in its native site and in the heterologous 303 site, the *Physconia perisidiosa* 1052 intron in its native and in the heterologous 1046 site, and the *P. perisidiosa* intron in its native 1516 and in the heterologous 1506 site are shown. The arrows indicate the 5′ and 3′ splice sites in these introns, with exon sequences shown in lower case and intron sequences in upper case. The intron sequences between P1 and P10 are represented with a solid line.

and that a small proportion of lichen-fungal group I introns have been maintained vertically, whereas most have been lost during evolution. The hypothesis of rare intron gains and widespread intron loss is supported by the extra step (reverse-splicing) that is required for intron gain and the intron phylogeny that shows strong support for the monophyletic origin of the different intron lineages. Consistent with the host tree that shows the Cladoniaceae and the Lecanoraceae to only be distantly related to the Physciaceae, the subtrees of the different intron insertion sites show that the *Cladonia merochlorophaea*/*Lecanora dispersa* introns form a sister group to the different Physiaceae introns [e.g., at sites 516, 1210, 1516- (Fig. 1)]. This pattern is consistent with the presence of the different introns in the common ancestor of the Lecanorales (see Fig. 2), and presently argues against widespread, independent intron gains at the homologous rDNA sites in the Physciaceae. This hypothesis needs to tested with a detailed study of the congruence of intron and host trees for the different insertion sites (e.g., Bhattacharya et al. 1994; Friedl et al. 2000). Our results also underline the importance of extensive taxon sampling in understanding whether the sporadic intron distribution within a lineage is the result of multiple independent gains or the widespread independent losses of more ancient, vertically inherited sequences. Consistent with our findings, a recent phylogenetic analysis of multiple group I introns in the endoparasitic fungus *Claviceps* spp. provides strong evidence for few intron gains followed by many intron losses in derived taxa (see Nikoh and Futaksu 2001).

Acknowledgments. D.B. acknowledges financial support from the Carver Foundation, the Iowa Biosciences Initiative, and the National Science Foundation (MCB 01-10252). This research was also supported in part by a grant from the Deutsche Forschungsgemeinschaft to T.F. (Fr 905/7-1,2). We thank two anonymous reviewers for their helpful comments.

References

Belfort M, Perlman PS (1995) Mechanisms of intron mobility. J Biol Chem 270:30237–30240

Besendahl A, Bhattacharya D (1999) Evolutionary analyses of small subunit rDNA coding regions and the 1506 group I introns of the Zygnematales (Charophyceae, Streptophyta). J Phycol 35:560–569

Besendahl A, Qiu Y-L, Lee J, Palmer JD, Bhattacharya D (2000) The cyanobacterial origin and vertical transmission of the plastid $tRNA^{Leu}$ group I intron. Curr Genet 37:12–23

Bhattacharya D (1998) The origin and evolution of protist group I introns. Protist 149:113–122

Bhattacharya D, Surek B, Ruesing M, Melkonian M (1994) Group I introns are inherited through common ancestry in the nuclear-encoded rRNA of Zygnematales (Charophyceae). Proc Natl Acad Sci USA 91:9916–9920

Bhattacharya D, Friedl T, Damberger S (1996) Nuclear-encoded rDNA group I introns: Origin and evolutionary relationships of insertion site lineages in the green algae/charophytes. Mol Biol Evol 13:978–989

Bhattacharya D, Lutzoni F, Reeb V, Simon D, Fernandez F (2000) Widespread occurrence of spliceosomal introns in the rDNA genes of ascomycetes. Mol Biol Evol 17:1971–1984

Bhattacharya D, Cannone JJ, Gutell RR (2001) Group I intron lateral transfer between red and brown algal ribosomal RNA. Curr Genet 40:82–90

Bryk M, Quirk SM, Mueller JE, Loizos N, Lawrence C, Belfort M (1993) The *td* intron endonuclease I-*Tev*I makes extensive sequence-tolerant contacts across the minor groove of its DNA target. EMBO J 12:2141–2149

Burke JM (1988) Molecular genetics of group I introns: RNA structures and protein factors required for splicing—a review. Gene 73:273–294

Cech TR (1985) Self-splicing RNA: implications for evolution. Int Rev Cytol 93:3–22

Cech TR (1988) Conserved sequences and structures of group I introns: building an active site for RNA catalysis—a review. Gene 73:259–271

Cech TR, Damberger SH, Gutell RR (1994) Representation of the secondary and tertiary structure of group I introns. Nature Struc Biol 1:273–280

Damberger SH, Gutell RR (1994) A comparative database of group I intron structures. Nucleic Acids Res 22:3508–3510

DePriest PT, Been MD (1992) Numerous group I introns with variable distributions in the ribosomal DNA of a lichen fungus. J Mol Biol 228:315–321

De Jonckheere JF, Brown S (1994) Loss of the ORF in the SSU rDNA group I intron of one *Naegleria* lineage. Nucleic Acids Res 22:3925–3927

Dover G (1994) Concerted evolution, molecular drive and natural selection. Curr Biol 4:1165–1166

Edgell DR, Belfort M, Shub DA (2000) Barriers to intron promiscuity in bacteria. J Bacteriol 182:5281–5289

Felsenstein J (1985) Confidence intervals on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Friedl T (1996) Evolution of the polyphyletic genus *Pleurastrum* (Chlorophyta): inferences from nuclear-encoded ribosomal DNA sequences and motile ultrastructure. Phycologia 35:456–469

Friedl T, Besendahl A, Pfeiffer P, Bhattacharya D (2000) The distribution of group I introns in lichen algae suggests that lichenization facilitates intron lateral transfer. Mol Phylogenet Evol 14:342–352

Gargas A, DePriest PT, Taylor JW (1995) Positions of multiple insertions in SSU rDNA of lichen-forming fungi. Mol Biol Evol 12:208–218

Gielly L, Taberlet P (1994) The use of chloroplast DNA to resolve plant phylogenies: non-coding versus *rbcL* sequences. Mol Biol Evol 11:769–777

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Haugen P, Huss VAR, Nielsen H, Johansen S (1999) Complex group-I introns in nuclear SSU rDNA of red and green algae: evidence of homing endonuclease pseudogenes in the Bangiophyceae. Curr Genet 36:345–353

Helms G, Friedl T, Rambold G, Mayhofer H (2001) Identification of photobionts from the lichen family Physciaceae using algal-specific ITS rDNA sequencing. Lichenologist 33:73–86

Hibbett DS (1996) Phylogenetic evidence for horizontal transmission of group I introns in the nuclear ribosomal DNA of mushroom-forming fungi. Mol Biol Evol 13:903–917

Honegger R (1991) Functional aspects of the lichen symbiosis. Ann Rev Plant Physiol Plant Mol Biol 42:553–578

Johansen S, Haugen P (1999) A complex group I intron in *Nectria galligena* rDNA. Microbiology 145:516–517

Johansen S, Elde M, Vader A, Haugen P, Haugli K, Haugli F (1997) In vivo mobility of a group I twintron in nuclear ribosomal DNA of the myxomycete *Didymium iridis*. Mol Microbiol 24:737–745

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, p 21–132

Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR (1982) Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. Cell 31:147–l57

Lambowitz AM, Belfort M (1993) Introns as mobile genetic elements. Annu Rev Biochem 62:587–622

Lockhart PJ, Steel MA, Hendy MD, Penny D (1994) Recovering evolutionary trees under a more realistic model of sequence evolution. Mol Biol Evol 11:605–612

Lumbsch HT, Schmitt I, Doering H, Wedin M (2001) Molecular systematics supports the recognition of an additional order of Ascomycota: the Agyriales. Mycol Res 105:16–23

Lutzoni F, Pagel M, Reeb V (2001) Major fungal lineages are derived from lichen symbiotic ancestors. Nature 411:937–940

Michel F, Westhof E (1990) Modelling the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J Mol Biol 216:585–610

Nash T (1996) Introduction. In: Nash T (ed) Lichen biology. Cambridge University Press, Cambridge, p 1–7

Nikoh N, Fukatsu T (2001) Evolutionary dynamics of multiple group I introns in nuclear ribosomal RNA genes of endoparasitic fungi of the genus *Cordyceps*. Mol Biol Evol 18:1631–1642

Nishida K, Suzuki S, Kimura Y, Nomura N, Fujie M, Yamada T (1998) Group I introns found in *Chlorella* viruses: biological implications. Virology 242:319–326

Paulsrud P, Lindblad P (1998) Sequence variation of the tRNA (Leu) intron as a marker for genetic diversity and specificity of symbiotic cyanobacteria in some lichens. Appl Environ Microbiol 64:310–315

Perotto S, Nepote-Fus P, Saletta L, Bandi C, Young JP (2000) A diverse population of introns in the nuclear ribosomal genes of ericoid mycorrhizal fungi includes elements with sequence similarity to endonuclease-coding genes. Mol Biol Evol 17:44–59

Piercey-Normore MD, DePriest PT (2001) Algal switching among lichen symbioses. Am J Bot 88:1490–1498

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818

Rodriguez F, Oliver JL, Marin A, Medina JR (1990) The general stochastic model of nucleotide substitution. J Theor Biol 142: 485–501

Roman J, Woodson SA (1995) Reverse splicing of the *Tetrahymena* IVS: evidence for multiple reaction sites in the 23S rRNA. RNA 1:478–490

Roman J, Woodson SA (1998) Integration of the *Tetrahymena* group I intron into bacterial rRNA by reverse splicing in vivo. Proc Natl Acad Sci USA 95:2134–2139

Rzhetsky A, Nei M (1992) Statistical properties of the ordinary least-squares, generalized least-squares, and minimum-evolution methods of phylogenetic inference. J Mol Evol 35:367–375

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

Schroeder-Diedrich JM, Fuerst P, Byers TJ (1998) Group-I introns with unusual sequences occur at three sites in nuclear 18S rRNA genes of *Acanthamaeba lenticulata*. Curr Genet 34:71–78

Suh S-O, Jones KG, Blackwell M (1999) A group I intron in the nuclear small subunit rRNA gene of *Cryptendoxyla hypophloia,* an ascomycetous fungus: evidence for a new major class of group I introns. J Mol Evol 48:493–500

Swofford DL (2001) PAUP* V4.0b6. Phylogenetic analysis using parsimony (* and other methods), version 4.0. Sinauer Associates, Sunderland, MA

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526

Tehler A (1996) Systematics, phylogeny and classification. In: Nash T (ed) Lichen biology. Cambridge University Press, Cambridge, pp 217–239

Thompson AJ, Herrin DL (1994) A chloroplast group I intron undergoes the first step of reverse splicing into host cytoplasmic 5.8 S rRNA: implications for intron-mediated RNA recombination, intron transposition and 5.8 S rRNA structure. J Mol Biol 236:455–468

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Tunnel M, Gutell RR, Mercier JP, Otis C, Lemieux C (1993) Analysis of the chloroplast large subunit ribosomal RNA gene from 17 *Chlamydomonas* taxa. Three internal transcribed spacers and 12 group I intron insertion sites. J Mol Biol 232:446–467

Watanabe KI, Ehara M, Inagaki Y, Ohama T (1998) Distinctive origins of group I introns in the *COXI* genes of three green algae. Gene 213:1–7

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA gene for phylogenetics. In: Innes MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR protocols: a guide to methods and applications. Academic Press, San Diego, pp 315–322

Woodson SA, Cech TR (1989) Reverse self-splicing of the *Tetrahymena* group I intron: implication for the directionality of splicing and for intron transposition. Cell 57:335–345

Zuker M, Mathews DH, Turner DH (1999) Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In: Barciszewski J, Clark BFC (eds) RNA biochemistry and biotechnology. NATO ASI Series, Kluwer Academic Publishers, Dordrecht, pp 11–43