Molecular Phylogenetics and Evolution
Vol. 13, No. 3, December, pp. 483–492, 1999
Article ID mpev.1999.0680, available online at http://www.idealibrary.com on IDEAL®

Use of *atp*6 in Fungal Phylogenetics: An Example from the Boletales

Annette M. Kretzer[1] and Thomas D. Bruns

Department of Plant and Microbial Biology, University of California at Berkeley, 111 Koshland Hall, Berkeley, California 94720-3102

Received December 24, 1998; revised April 28, 1999

Complete nucleotide sequences have been determined for *atp*6 from *Suillus luteus* and *cox*3 from *Suillus sinuspaulianus* (Boletales, Hymenomycetes, Basidiomycota), which code for ATPase subunit 6 and cytochrome oxidase subunit 3, respectively. These sequences were used to design PCR primers for the amplification of partial *atp*6 and *cox*3 sequences from other members of the Boletales and outgroup taxa. In *atp*6 and *cox*3 from *Russula rosacea*, one of the outgroup taxa, we observed a number of in-frame TGA_{trp} codons, which imply a *Neurospora crassa*-type mitochondrial code in *R. rosacea* and possibly in basidiomycetes in general. Interestingly, however, most basidiomycetes other than *R. rosacea* appear to strongly prefer the TGG_{trp} codon, which is unusual, given the strong A + T bias in fungal mitochondrial genomes. Pairwise comparisons were performed between *atp*6 sequences from increasingly divergent fungal lineages, and results show that all three codon positions become saturated in substitutions after an estimated divergence time of approx 300 Ma. This means that *atp*6 is likely to provide phylogenetic resolution within fungal classes but not at higher taxonomic levels. Also, because of the strong A + T bias in fungal mitochondrial genomes, A/T transversions were found to be more common than any other type of substitution, resulting in transversions being about two to three times more common in most pairwise sequence comparisons. Finally, *atp*6 sequences were used to infer phylogenetic relationships between 27 taxa from the Boletales and 4 outgroup taxa. Analyses were performed (i) on nucleotide sequence data using parsimony (successive approximation) as well as maximum likelihood methods and (ii) on deduced amino acid sequences using distance methods based on empirical substitution probabilities. Results from the various analyses are largely concordant with each other as well as with prior analyses of partial mitochondrial large-subunit rDNA (mtLSU rDNA). Analysis of the combined *atp*6 and mtLSU rDNA sequences results in increased bootstrap support for several key branches. Relationships that have been resolved for the first time in the current analysis are discussed.

INTRODUCTION

The vast majority of molecular systematic studies in fungi have utilized ribosomal RNA genes or the spacers between them (e.g., Hibbett *et al.*, 1997; Begerow *et al.*, 1997; Binder *et al.*, 1997; Lutzoni, 1997; Ahren *et al.*, 1998; Vogler and Bruns, 1998; and many others). This choice has been based in part on the availability of universal primers and the ease with which these multicopy regions can be amplified, even from samples with limited and/or degraded DNA, such as herbarium specimens. The fact that the ribosomal repeats undergo concerted evolution has also increased their attractiveness, as it generally avoids the problem of paralogous comparisons (but see O'Donnell and Cigelnik, 1997). Unfortunately, the rDNA regions have not been able to answer all the phylogenetic questions to which they have been applied. One example of their limitation is within the mushroom-forming basidiomycetes (Hymenomycetes). Several studies using nuclear rDNA sequences have shown that the basal branches that connect distantly related genera or family-sized clades are not well resolved (Hibbett *et al.*, 1997; Moncalvo *et al.*, 1997). The mitochondrial rRNA genes, though useful because of their independence from the nuclear rDNA locus, pose greater alignment and saturation problems because of their faster rate of evolution, which reduces their utility at higher taxonomic levels (Bruns and Szaro, 1992; Hibbett and Donoghue, 1995; Bruns *et al.*, 1998).

Protein coding sequences from organellar genomes share the advantage of high copy number and concerted evolution. Additionally, protein coding sequences are constrained by a reading frame, which facilitates alignments by reducing the frequency of length mutations and provides three natural rate classes of sites (i.e., codon positions) that may span different levels of analysis. Animal and plant studies have often used

[1] To whom correspondence should be addressed at current address: Department of Botany and Plant Pathology, Oregon State University, 2082 Cordley Hall, Corvallis, OR 97331-2902. Fax: (541) 737-3573. E-mail: kretzera@bcc.orst.edu.

1055-7903/99 $30.00

organellar protein coding sequences for these reasons (e.g., Chase *et al.*, 1993; Kellog and Juliano, 1997; Zardoya and Meyer, 1996; Russo *et al.*, 1996). In contrast, mitochondrial protein genes have rarely been used in fungal studies (for exceptions see Paquin *et al.*, 1995a,b; Wetterer *et al.*, 1998; Wang *et al.*, 1998). This is due partly to the lack of universal primers and partly to the widespread presence of large introns in most of the likely target genes. In nuclear genes, introns have proven highly useful for phylogenetic analysis of related taxa (O'Donnell and Cigelnik, 1997; Schardl *et al.*, 1997). But fungal mitochondrial introns, which are exclusively group I and group II introns, vary in their presence within particular genes, and, more importantly, group I introns appear to be horizontally transmitted (Paquin *et al.*, 1997; Nishida and Sugiyama, 1995; Vaughan *et al.*, 1995; Hibbett, 1996).

Despite the potential advantages described above, phylogenetic analysis of protein coding sequences encounters several challenges. (i) Because protein coding sequences align nicely even between distant taxa, saturation may be more cryptic than in rDNA sequences, which quickly become unalignable in the most variable regions. (ii) Rate differences between codon positions are too well documented to be ignored but exactly how to weight positions under the parsimony criterion is not straight forward. (iii) Fungal mitochondrial genes confront us with the additional challenge of being very (A + T)-rich; this bias in base frequencies likely effects the frequencies and costs of various types of substitutions. (iv) Finally, codon bias can also change the way in which protein sequences evolve.

In the current study, we have examined the use of two mitochondrial protein genes for phylogenetic analyses, *atp*6 and *cox*3, which code for ATPase subunit 6 and cytochrome oxidase subunit 3, respectively. We chose *atp*6 and *cox*3 because we did not expect large introns to be common in those genes. This expectation was based on known *atp*6 and *cox*3 sequences, which were primarily from ascomycetes, and on an early mitochondrial mapping study within the Boletales (Bruns *et al.*, 1988; Bruns and Palmer, 1989). We also used *atp*6 to examine phylogenetic relationships in the Boletales. The Boletales is a large and diverse order of mushroom-forming basidiomycetes for which we have some preliminary phylogenetic estimates based mostly on partial mitochondrial large-subunit rDNA (mtLSU rDNA); we also have a long-standing interest in the group (Bruns *et al.*, 1989; Bruns and Szaro, 1992; Baura *et al.*, 1992; Kretzer *et al.*, 1996; Kretzer and Bruns, 1997). As is true for many fungal orders, such as the Agaricales and Russulales, the key basal branches in the Boletales, unfortunately, remain unresolved (Bruns *et al.*, 1998).

Our main objectives were (i) to design PCR primers for convenient amplification and sequencing of the *atp*6 and *cox*3 genes, (ii) to explore sequence divergence in both genes at different taxonomic levels, (iii) to document character and character state evolution as relevant for phylogenetic analyses, and finally (iv) to use *atp*6 and/or *cox*3 to reconstruct phylogenetic relationships in the Boletales (Hymenomycetes, Basidiomycota).

MATERIALS AND METHODS

Clones. Cloning of a 5.5-kb *Hin*dIII fragment from *Suillus luteus* strain TDB-571 and a *Cla*I/*Bam*HI fragment from *Suillus sinuspaulianus* strain DAOM-66995 coding for ATPase subunit 6 and cytochrome oxidase subunit 3, respectively, has been described in a previous paper (Bruns and Palmer, 1989). From the *atp*6 clone, an approx 2.4-kb *Eco*RI fragment was subcloned into pUC18/DH5α using an internal *Eco*RI site. Sequencing of the insert revealed that both the original *Hin*dIII clone as well as the *Eco*RI subclone contained an *atp*6 gene that was truncated at the carboxy terminus. The missing sequence was recovered using inverse PCR (Ochman *et al.*, 1988; Triglia *et al.*, 1988): mtDNA of *Suillus luteus* TDB-571 prepared as described before (Bruns and Palmer, 1989) was digested with *Bfa*I, which was known to cut 191 bp upstream of the truncation. After extraction with phenol/chloroform and diethyl ether, the restriction fragments were religated with themselves. The self-ligated *Bfa*I fragments were subsequently used as templates in a PCR together with the following primers: 5′-GGTACTAAAGTTACAATAAATATA-3′ and 5′-TTCTTAGCTATAACTGGATTAG-3′. PCR conditions were 35 cycles of 35 s at 94°C, 55 s at 50°C, and 45 s at 72°C, the latter increasing by 4 s on every cycle. A specific, approx 400-bp-large fragment was obtained and reamplified for sequencing.

Extraction of genomic DNA. DNA was extracted from voucher collections by a CTAB extraction and glassmilk purification method. Three hundred microliters of CTAB buffer (100 mM Tris, pH 8–9; 1.4 M NaCl; 20 mM EDTA; 2% CTAB) were added to 10–50 mg of dried material. After softening the fungal tissue by freezing and thawing three times, it was crushed with a micropestle and incubated for 30–60 min at 65°C. After the incubation, the cloudy suspension was extracted with an equal volume of chloroform and the clear supernatant was transferred to a new tube. DNA was recovered from the clear buffer phase by binding to glassmilk according to the instructions of the BIO 101 Geneclean II kit.

Polymerase chain reaction. PCRs contained 10 mM Tris/HCl, pH 8.3; 50 mM KCl; 2.5 mM $MgCl_2$; 0.1 mg/ml gelatin; 200 μM each of the four deoxyribonucleotide triphosphates; 0.5 μM each of two different primers; 25 U/ml *Taq* polymerase; and empirical amounts of DNA. Unless otherwise specified, reaction conditions were as follows: 5 cycles of 35 s at 94°C, 55 s at 37°C, slow ramp up to 72°C, and 1 min at 72°C, followed by 30 cycles of 35 s at 94°C, 55 s at 45°C, and 1 min at 72°C, the latter

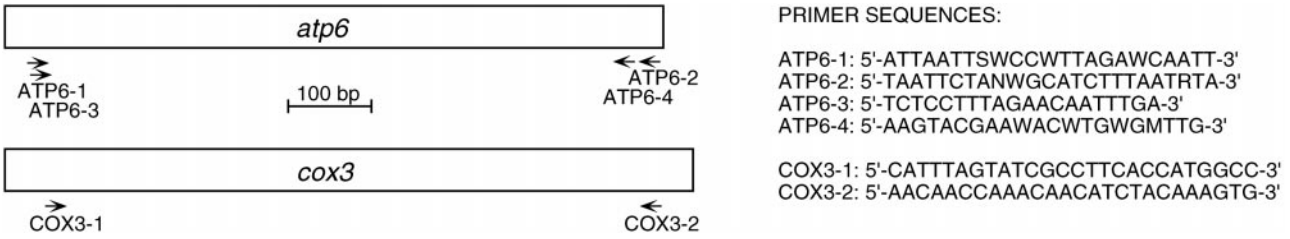

FIG. 1. Sequences and approximate annealing sites of the primers used in this study.

increasing by 4 s on every cycle. Successful amplification was verified by electrophoresis on an agarose gel stained with ethidium bromide. If initial amplifications resulted in weak bands, they were cut out from the gel, melted in sterile water, and reamplified under the same conditions.

Sequencing. Plasmid DNA was prepared for sequencing using a modified alkaline-lysis and PEG precipitation procedure as described in the manual for the PRISM Ready Reaction DyeDeoxy Terminator Cycle Sequencing Kit from Applied Biosystems Inc. PCR products were purified before sequencing by one of

TABLE 1

Partial *atp*6, *cox*3, and mt LSU rDNA Sequences Determined in This Study

Species	Collection no.	Sequence accession no.		
		*atp*6[a]	*cox*3	Partial mt LSU rDNA[b]
Ingroup taxa				
Boletinellus merulioides	HDT-50394	AF114454	—	—
Boletus edulis	TDB-1002	AF002141	AF002153	—
Boletus pallidus	TDB-1231	AF002142	AF002154	—
Brauniellula albidipes	HDT-52855	AF114463	—	—
Chalciporus piperatoides	TDB-1235	AF114445	—	—
Chamonixia caespitosa	F-945	AF114444	—	—
Chamonixia caespitosa	TJB-6640	—	—	AF114464
Coniophora arida	FP-104367-SP	AF114452	—	—
Gyrodon lividus	HDT-46035	AF114455	—	AF114466
Gyroporus cyanescens	TDB-194	AF114459	—	—
Hygrophoropsis aurantiaca	DED-5560	AF114451	—	—
Phylloporus rhodoxanthus	TDB-540	AF114443	—	—
Paxillus filamentosus	UME-25834	AF114446	—	AF114465
Paxillus involutus	HDT-53590	AF114447	—	—
Paxillus atrotomentosus	HDT-50423	AF114448	—	—
Phaeogyroporus portentosus	HDT-42534	AF114453	—	—
Pisolithus arhizus	TDB-2197	AF114456	—	—
Rhizopogon olivaceotinctus	S-4432	AF114462	—	—
Scleroderma laeve	TDB-1050	AF114457	—	AF114467
Scleroderma hypogaeum	TDB-1638	AF114458	—	AF114468
Serpula himantoides	FP-94342-R	AF114450	—	—
Suillus granulatus	TDB-878	AF002137	AF002149	—
Suillus lakei	T-20854B	AF114460	—	AF114469
Suillus ochraceoroseus	SAR-84-137	—	AF002150	—
Suillus variegatus	HB-325	AF002140	AF002152	—
Tapinella panuoides	HDT-54533	AF114449	—	—
Truncocolumella citrina	TDB-2001	AF114461	—	—
Xerocomus chrysenteron	TDB-1870	AF002143	AF002155	—
Outgroup taxa				
Gomphus floccosus	TDB-1310	AF002146	AF002158	—
Russula rosacea	TDB-2074	AF002148	AF002160	—
Sarcodon imbricatum	TDB-1774	AF002147	AF002159	—

[a] All partial *atp*6 sequences were determined using the primer combination ATP6-1/2, except for *atp*6 from *Coniophora arida*, which was amplified and sequenced using ATP6-3/4.

[b] Partial mtLSU rDNA sequences were determined using the primers ML5 and 6 as described in Bruns *et al.* (1998).

several methods. (i) PCR products were diluted several times with H_2O and reconcentrated using centrifugal filters (Millipore Ultrafree-MC filters). (ii) PCR products were run on an agarose gel (1% normal agarose and 2% NuSieve agarose from FMC BioProducts); bands were cut out and subsequently cleaned with the BIO 101 Geneclean II kit according to the kit's instructions; this method proved to be especially useful to remove nonspecific amplification by-products, which were occasionally produced by the *cox*3 primers. (iii) PCR products were cleaned using the QIAquick PCR Purification Kit (QIAGEN); this method gave the highest yields.

Nucleotide sequences were determined by the cyclic reaction termination method using fluorescence-labeled dideoxyribonucleotide triphosphates. The sequence reaction and the processing of the reaction products for electrophoresis were performed following the instructions for the sequencing kit (ABI PRISM Dye Terminator Cycle Sequencing Core Kit, Perkin–Elmer Corp.). Electrophoresis and data collection were done on an ABI Model 377 DNA Sequencer (Perkin–Elmer Corp.). DNA Sequencing Analysis (version 2.12) and Sequence Navigator (version 1.0.1) were used for processing the raw data.

Pairwise sequence comparisons. Pairwise sequence comparisons were performed using the "show distance matrix" command of the computer program PAUP (version 3.1.1). "Mean distances" as calculated under that command are essentially numbers of changes per sites (Swofford, 1993). These were calculated individually for 1st, 2nd, and 3rd codon positions by respectively excluding the other codon positions. Sequence differences due to individual types of substitutions were calculated in the same way, except that various symmetric stepmatrices were applied that defined a cost of 1 to the character change(s) under investigation and a cost of 0 to all others.

Sequence differences were plotted over the estimated divergence time as follows. The average "mean distance" was calculated from all pairwise comparisons (i) between members of the genus *Suillus,* (ii) between members of the suilloid and boletoid radiations in the Boletales (for definition see Bruns *et al.*, 1998 and Fig. 5A), (iii) between yeasts and filamentous ascomycetes, (iv) between ascomycetes and basidiomycetes, and (v) between the two chytridiomycete taxa and all other fungi. The average "mean distances" were then plotted over estimated divergence times of (i) 2 Ma, (ii) 100 Ma, (iii) 310 Ma, (iv) 395 Ma, and (v) 560 Ma, respectively, in accordance with molecular clock estimates based on nuclear small subunit rDNA (Berbee and Taylor, 1993; Bruns *et al.*, 1998). We are aware of a number of problems associated with those molecular clock estimates as well as the biased selection of taxa representing certain radiation events in this study. But our conclusions rely only on the assumed hierarchy of events and the relative evolutionary rates and not on the accuracy of the absolute time values.

TABLE 2

Number of In-Frame TGA versus TGG Codons Found in *atp*6 and *cox*3 from Fungi

	No. of TGG codons found	No. of TGA codons found	Preferred trp-codon
Basidiomycota			
Suillus sp.[a]	7	0	TGG
Russula rosacea[b]	0	4	TGA
Gomphus floccosus[b]	4	0	TGG
Sarcodon imbricatum[b]	4	0	TGG
Schizophyllum commune	7	0	TGG
Ascomycota			
Neurospora crassa	0	9	TGA
Podospora anserina	1	9	TGA
Aspergillus nidulans	0	9	TGA
Saccharomyces cerevisiae	0	14	TGA
Yarrowia lipolytica	0	13	TGA
Candida parapsilosis	0	11	TGA
Schizosaccharomyces pombe	9	0	TGG
Chytridiomycota			
Allomyces macrogynus	13	0	TGG

[a] The values for *Suillus* sp. are based on the combined sequences of the *atp*6 gene from *S. luteus* and the *cox*3 gene from *S. sinuspaulianus.*

[b] The values are based on partial *atp*6 and *cox*3 sequences as determined in this study.

Phylogenetic analyses. Parsimony analyses as well as maximum likelihood analyses of nucleotide sequence data were performed in PAUP*d64. The partition homogeneity test was based on 100 random partitions and used equal weights for all characters and character changes. Most parsimonious trees were re-

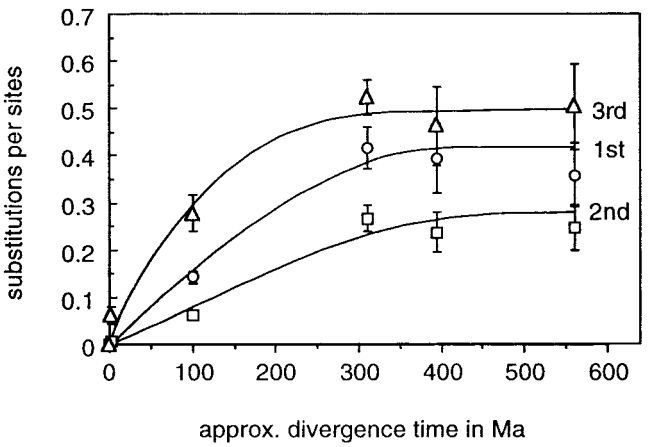

FIG. 2. Number of substitutions at 1st, 2nd, and 3rd codon positions per number of sites was calculated from pairwise comparisons of *atp*6 sequences and plotted over estimated divergence time as described under Materials and Methods. Error bars represent standard deviations.

trieved using 10 heuristic searches with random sequence addition. Successive approximation was performed by reweighting characters according to their mean consistency index (Farris, 1969; Carpenter, 1988). Bootstrap values are based on 500 replicates using five random sequence additions each and equal weights for all characters and character changes. For maximum likelihood analyses of nucleotide sequence data, a six-step substitution model was implemented with the substitution rate matrix estimated via maximum likelihood. Base frequencies were empirically determined from the dataset to be A = 0.306, C = 0.130, G = 0.114, and T = 0.449. Among-site rate variation was accounted for by a discreet approximation to the gamma distribution using four rate classes and a shape parameter estimated by maximum likelihood. The proportion of sites assumed to be invariable was zero. The same model was implemented in the Kishino–Hasegawa test.

Amino acid sequences were deduced from nucleotide sequences assuming a *Neurospora crassa* type mitochondrial code (Fox, 1987). Neighbor joining and bootstrap analyses of the deduced amino acid sequences were performed in PHYLIP 3.573 using distances based on the Dayhoff *et al.* (1978) substitution probability matrix. Bootstrap values are based on 500 replicates.

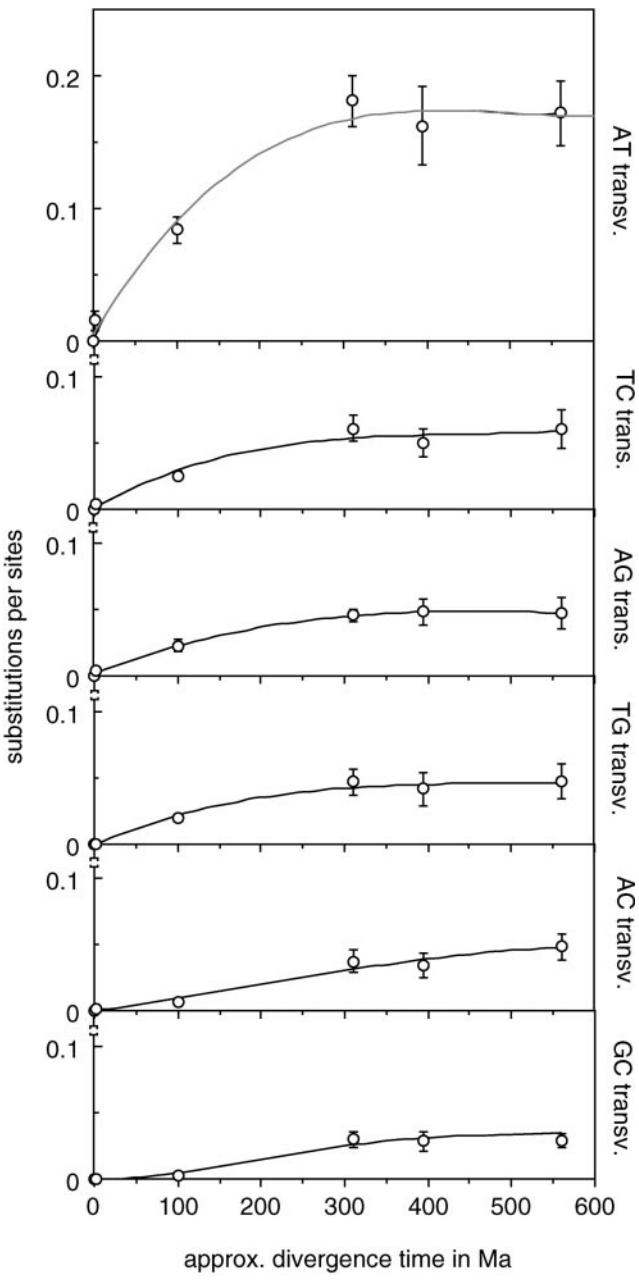

FIG. 3. Number of different substitutions per number of sites was calculated from pairwise comparisons of *atp*6 sequences and plotted over estimated divergence time as described under Materials and Methods. Error bars represent standard deviations.

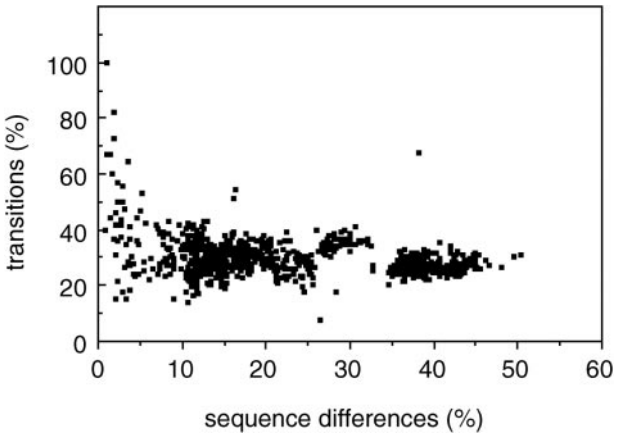

FIG. 4. Number of transitions per total number of substitutions was derived through pairwise comparison of *atp*6 sequences and plotted over sequence differences expressed as percentages.

RESULTS

Design of PCR primers. When the study was started, very few *atp*6 and *cox*3 sequences were known from basidiomycetes and none from the Boletales. We therefore sequenced both the *atp*6 gene from *Suillus luteus* TDB-571 and the *cox*3 gene from *Suillus sinuspaulianus* DAOM-66995 that had been cloned in a previous study (Bruns and Palmer, 1989). The complete, two-directional sequences have been deposited at GenBank under the Accession Nos. AF002135 and AF002136, respectively. They were subsequently aligned with homologous sequences from higher fungi in order to identify conserved areas as potential target sites for PCR primers. Available homologous sequences were *atp*6 from *Agaricus bitorquis* (GenBank Accession No. U60235), *Neurospora crassa* (GenBank K02655), *Aspergillus nidulans* (GenBank K01799 + X04161), *Podospora anserina* (GenBank X55026), *Saccharomyces cerevisiae* (GenBank V00683), *Candida parapsilosis* (GenBank X55653), *Yarrowia lipolytica* (GenBank

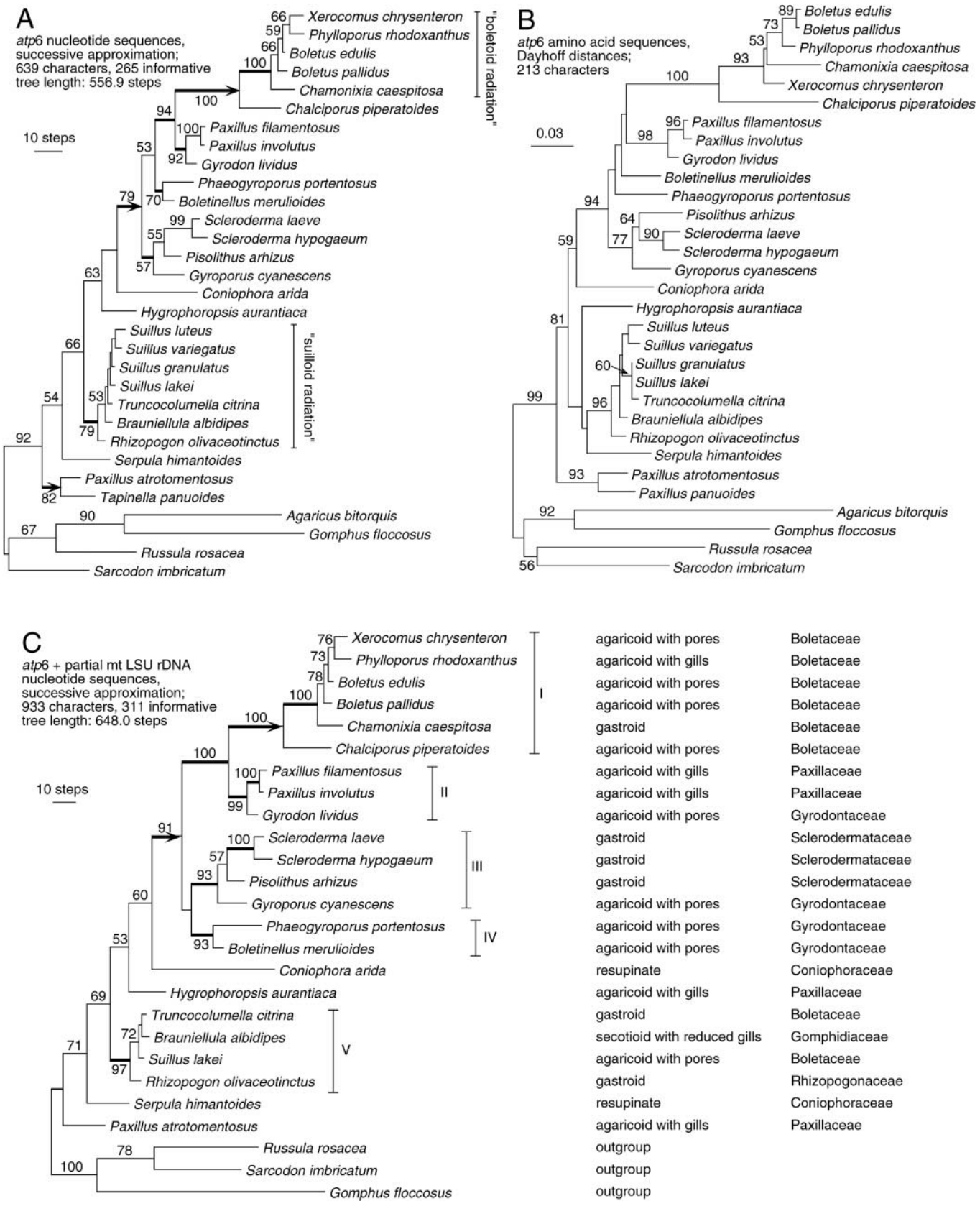

FIG. 5. (A) Most parsimonious tree derived from *atp*6 nucleotide sequences after one round of successive approximation. Numbers indicate support for individual branches from bootstrap analysis under conditions of equal weights. Bold branches received moderate to strong support in previous analyses (Bruns *et al.*, 1998); arrows indicate branches that received moderate to strong support for the first time in this analysis.

L15359), *Schizosaccharomyces pombe* (GenBank X54421), *Allomyces arbusculus* (GenBank U02038), and *Allomyces macrogynus* (GenBank U02039); and *cox*3 from *Schizophyllum commune* (GenBank M36270), *Neurospora crassa* (GenBank J01430), *Aspergillus nidulans* (GenBank X06960), *Podospora anserina* (GenBank X55026), *Saccharomyces cerevisiae* (GenBank J01478), *Candida parapsilosis* (GenBank X75679), *Yarrowia lipolytica* (GenBank L15359), *Schizosaccharomyces pombe* (GenBank X16868), and *Allomyces macrogynus* (GenBank U41288).

Sequences and approximate annealing sites of the primers used in this study are shown in Fig. 1. Using different combinations of those primers, fragments of the *atp*6 and *cox*3 genes were amplified and sequenced from a number of boletoid taxa plus several outgroup taxa (see Table 1). Since amplifications of *atp*6 were more specific and more consistent across taxa, we did not continue to pursue development of *cox*3 as a target for phylogenetic analyses. In addition, *cox*3 sequences amplified from a number of *Suillus* species appeared to contain introns which we were also trying to avoid. We did, however, align the available *cox*3 sequences and examined rates of evolution (data not shown); all features turned out to be nearly identical to those presented below for *atp*6.

Mitochondrial code and codon usage. Between all *atp*6 and *cox*3 sequences determined in this study, a total of four in-frame TGA codons were found; all occurred in *Russula rosacea* and corresponded to conserved tryptophan positions in other fungi. All other basidiomycetes studied so far strongly prefer the TGG codon for tryptophan (see also Table 2). Other than the TGA codons found in *Russula rosacea*, no obvious indications were found for deviations from the "universal" genetic code.

Sequence alignment. For evolutionary and phylogenetic studies of *atp*6, two alignments were made, one including sequences from all fungal lineages and another including only basidiomycete sequences. According to the expectations, *atp*6 sequences were found to align reasonably well, given the broad range of taxa. Only one small region that was 9 bp long in most taxa but contained 18- to 27-bp-long insertions in *Boletus edulis*, *Boletus pallidus*, *Xerocomus chrysenteron*, *Phylloporus rhodoxanthus*, *Chamonixia caespitosa*, *Chalciporus piperatoides*, and *Gomphus floccosus* had to be excluded from the basidiomycete alignment due to ambiguities; 42 additional positions were excluded from the alignment of all fungal sequences. PAUP files are available at our website (http://plantbio.berkeley.edu/~bruns/) and have been deposited at TreeBASE.

Character evolution. When mean distances at 1st, 2nd, and 3rd codon positions were plotted against estimated divergence time, the relative rates of divergence were as expected, with substitutions at 3rd codon positions accumulating the fastest followed by 1st and 2nd codon positions (Fig. 2). Unexpectedly, the relative times for saturation were approximately the same for each codon position (roughly 300 Ma). This appears to be caused by differences in the number of positions within each class that are free to vary.

Since fungal mitochondrial genomes are known to have a strong A + T bias, the frequencies of different types of substitutions were examined in an analogous way. Figure 3 shows that A/T transversions are by far the most common type of substitution. The abundance of A/T transversions causes the dataset to exhibit an unusually low transition/transversion ratio, with transversions being about two to three times as abundant as transitions, even in comparisons between fairly closely related sequences (Fig. 4). The same pattern has been observed previously in the mitochondrial small subunit rDNA and has been attributed to base composition bias (Bruns and Szaro, 1992).

Phylogenetic analyses in the Boletales. The *atp*6 dataset consisted of 27 ingroup and 4 outgroup taxa. Since a most closely related outgroup has not been identified for the Boletales (Bruns *et al.*, 1998; Hibbett *et al.*, 1997), outgroup taxa were chosen from several different hymenomycete lineages. The rate heterogeneity documented above between nucleotide sites and between different types of substitution calls for a few comments on the method of nucleotide sequence analysis that was employed. Parsimony analysis deals with rate heterogeneity mostly through weights but appropriate a priori weights are often very difficult to rationalize. We therefore chose a posteriori weighting through successive approximation in addition to maximum likelihood analysis. For maximum likelihood analysis, a model was implemented that allowed for rate heterogeneity between sites (gamma distribution) as well as between six types of substitutions (for details see Materials and Methods). We finally also translated the nucleotide sequences into amino acid sequences; the latter were analyzed by distance methods based on

(B) Neighbor joining tree derived from deduced amino acid sequences using distances based on the Dayhoff *et al.* (1978) probability matrix. Numbers indicate support for individual branches from bootstrap analysis. (C) Most parsimonious tree derived from combined *atp*6 and partial mtLSU rDNA nucleotide sequences after one round of successive approximation. Numbers indicate support for individual branches from bootstrap analysis under conditions of equal weights. Roman numerals indicate five well-supported, monophyletic clades; two clades (I and V) contain gilled, pored, and gastroid taxa; one clade (II) contains gilled and pored taxa; and one clade (III) contains pored and gastroid taxa (for discussion see text).

empirical probabilities of amino acid substitution (Dayhoff *et al.*, 1978).

Parsimony analysis of the *atp*6 nucleotide sequence data under conditions of equal weights resulted in two most parsimonious trees, which differed only in the position of *Hygrophoropsis aurantiaca* and which converged on a single most parsimonious tree after one round of successive approximation (Fig. 5A). Maximum likelihood analysis of the nucleotide sequence data resulted in an extremely similar tree that was not significantly different from the parsimony tree(s) by the Kishino–Hasegawa test. In fact, differences were limited to branching orders within the "suilloid and boletoid radiations"; *atp*6 sequences are clearly too conserved to resolve relationships within these two radiations. Neighbor joining analysis of the deduced amino acid sequences under the Dayhoff *et al.* (1978) model of substitution resulted in a slightly different tree (Fig. 5B) but differences were again limited to branches that received weak bootstrap support (≤70%).

Phylogenetic analyses were also performed on a second dataset in which *atp*6 nucleotide sequences were combined with an approx 300-bp-long fragment of the mtLSU rDNA. Most of the mtLSU rDNA sequences were determined in a previous study (Bruns *et al.*, 1998) with a few exceptions given in Table 1. The null hypothesis of congruence between both datasets was not rejected by the partition homogeneity test ($P = 0.44$). The combined dataset consisted of 23 ingroup and 3 outgroup taxa. Parsimony analysis under conditions of equal weights again resulted in only two most parsimonious trees that differed in the grouping of *Truncocolumella citrina* with either *Brauniellula albidipes* or *Suillus lakei* (data not shown) and that converged on a single most parsimonious tree after one round of successive approximation (Fig. 5C). The only notable difference between the two nucleotide datasets is that *Phaeogyroporus* and *Boletinellus* group with *Gyroporus* and the Sclerodermataceae in the combined dataset but this grouping is not supported by either bootstrap analysis (Fig. 5C) or maximum likelihood analysis of the combined nucleotide sequence data. Other than the *Boletinellus*/*Phaeogyroporus* grouping, the maximum likelihood tree was identical to the parsimony tree shown in Fig. 5C and was not significantly different by the Kishino–Hasegawa test (data not shown).

DISCUSSION

Probably the main reason that mitochondrial protein coding sequences have not been used extensively in fungal phylogenetics is the lack of "universal" PCR primers. Primer design, however, is more difficult because, unlike rRNA genes, which have nucleotide regions that are almost universally conserved, protein genes are conserved only at the amino acid level. In addition, we found conserved areas in *atp*6 and *cox*3 typically limited to a few consecutive amino acids, which often did not provide a long enough nucleotide target sequence for primer annealing. Despite these problems, we were able to amplify specific *atp*6 sequences from most boletes that we tried plus several outgroup taxa using different combinations of the listed primers and/or low annealing temperatures (Fig. 1 and Table 1).

Since the dataset compiled in this study is the largest collection of protein coding sequences from basidiomycete mitochondria known to date, we have looked for obvious deviations from the "universal" genetic code. Deviations known to occur in mitochondria of other fungi include TGA coding for tryptophan rather than translation stop (most ascomycetes) and ATA and CTN coding for methionine and threonine, respectively, rather than isoleucine and leucine (some yeasts) (Fox, 1987). The only obvious deviation from the "universal" genetic code observed in this study was TGA coding for tryptophan in *atp*6 and *cox*3 of *Russula rosacea*. Interestingly, however, other than *Russula rosacea,* all basidiomycetes studied so far appear to strongly prefer the universal TGG_{trp} codon (Table 2) despite a high A + T content of their mitochondrial genomes that results in a strong bias for A or T in all other 3rd codon positions. This has led Phelps *et al.* (1988) to assume a "universal" genetic code in mitochondria of the basidiomycete *Schizophyllum commune;* Lang and co-workers, however, after sequencing the entire mitochondrial genome of *Schizophyllum commune,* found a few in-frame TGA codons in essential proteins (Paquin *et al.*, 1997).

Results from pairwise sequence comparisons showed that *atp*6 and *cox*3 (data for *cox*3 not shown) reach saturation in substitutions after an estimated divergence time of 300 Ma or, if this absolute time value is incorrect, at a timepoint corresponding to the split of the yeasts and filamentous ascomycetes. The time of saturation is similar for fast- and slow-evolving codon positions as well as for frequent and rare substitution types but the initial slopes vary (Figs. 2 and 3). These findings make *atp*6 and *cox*3 likely to be suitable for phylogenetic studies roughly within fungal classes but not on higher taxonomic levels. Our focus of interest is on relationships within the Boletales (Hymenomycetes), which are estimated to have diverged roughly 100 Ma ago (Bruns *et al.*, 1998). This timepoint appears to lie well within the nonsaturated area of the curves shown in Figs. 2 and 3.

Phylogenetic trees derived from *atp*6 nucleotide sequence data under different methods of analysis were extremely similar and not significantly different by the Kishino–Hasegawa test (see Results). The same test identified the neighbor joining tree derived from the deduced amino acid sequences as significantly worse but since the test used nucleotide sequence data, it is likely to be biased toward the trees derived from the

nucleotide dataset. Translation of nucleotide sequences into amino acid sequences constitutes a loss of information and therefore appears advisable only when synonymous nucleotide positions are likely to be saturated; the latter is unlikely to be the case in the present dataset (see Fig. 2). Conflicts between the trees derived from nucleotide and amino acid sequences, however, are also limited to branches that receive weak bootstrap support from either dataset (Figs. 5A and 5B).

Trees derived from *atp*6 sequences were quite similar to those derived previously from partial mtLSU rDNA (Bruns *et al.*, 1998). Combination of both datasets also resulted in extremely similar tree topologies but bootstrap support was generally increased in the combined analysis (Fig. 5C). Similarities between previous (Bruns *et al.*, 1998) and current (Fig. 5) analyses include recognition and strong to moderate support for the groups indicated in Figs. 5A and 5C and the finding that the Boletaceae, Paxillaceae, and Gyrodontaceae are polyphyletic. Two strongly supported, monophyletic groups contain gilled, tubed, and gastroid taxa (clades I and V in Fig. 5C), one contains gilled and tubed taxa (clade II in Fig. 5C), and one contains pored as well as gastroid taxa (clade III in Fig. 5C). Thus, the convergence in gross morphology of the basidiocarp, which has been noted previously (Bruns *et al.*, 1989; Bruns and Szaro, 1992; Hibbett, 1998; Bruns *et al.*, 1998), is further reinforced.

Several monophyletic groups are resolved and supported by the current analysis that were not resolved by the earlier study (Bruns *et al.*, 1998). These include the grouping (i) of *Chalciporus piperatoides* with the boletoid group (100% bootstrap and supported by an insertion), (ii) of *Paxillus atrotomentosus* and *Tapinella* (*Paxillus*) panuoides to the exclusion of *Serpula* and *Hygrophoropsis* (82%), and (iii) of all members of the Gyrodontaceae, the Sclerodermataceae, and *Paxillus sensu stricto* with the boletoid group (91% bootstrap support in the combined analysis). Also, previous molecular studies have not included *Gyrodon lividus* nor members of the genus *Scleroderma*. Molecular evidence presented in this paper strongly supports inclusion of the Sclerodermataceae in the Boletales, as indicated by secondary metabolites (Gill and Watling, 1986; Arnold *et al.*, 1996), and supports Smith and Thiers' (1971) view that *Gyrodon* and *Boletinellus* are not congeneric.

Overall, *atp*6 provides another useful mitochondrial locus for fungal phylogenetic studies that largely avoids the alignment problems common to ribosomal genes. With only two most parsimonious trees recovered from the *atp*6 dataset under conditions of equal weights, the achieved resolution is very reasonable. However, several basal branches in the Boletales, including the position of the root, continue to receive weak bootstrap support. The problem is likely to stem from (i) the lack of a closely related outgroup and (ii) the presence of fairly long branches leading to *Coniophora*, *Serpula*, and *Hygrophoropsis*, which are also the taxa whose positions are not well supported.

REFERENCES

Ahren, D., Ursing, B. M., and Tunlid, A. (1998). Phylogeny of nematode-trapping fungi based on 18S rDNA sequences. *FEMS Microbiol. Lett.* **158:** 179–184.

Arnold, N., Steglich, W., and Besl, H. (1996). Derivates of pulvinic acid in the genus *Scleroderma*. *Z. Mykol.* **62:** 69–73.

Baura, G., Szaro, T. M., and Bruns, T. D. (1992). *Gastrosuillus laricinus* is a recent derivative of *Suillus grevillei:* Molecular evidence. *Mycologia* **84:** 592–597.

Begerow, D., Bauer, R., and Oberwinkler, F. (1997). Phylogenetic studies on nuclear large subunit ribosomal DNA sequences of smut fungi and related taxa. *Can. J. Bot.* **75:** 2045–2056.

Berbee, M. L., and Taylor, J. W. (1993). Dating the evolutionary radiation of the true fungi. *Can. J. Bot.* **71:** 1114–1127.

Binder, M., Besl, H., and Bresinsky, A. (1997). Agaricales or Boletales? Molecular evidence towards the classification of some controversial taxa. *Z. Mykol.* **63:** 189–196.

Bruns, T. D., and Palmer, J. D. (1989). Evolution in mushroom mitochondrial DNA: *Suillus* and related genera. *J. Mol. Evol.* **28:** 349–362.

Bruns, T. D., and Szaro, T. M. (1992). Rate and mode differences between nuclear and mitochondrial small-subunit rRNA genes in mushrooms. *Mol. Biol. Evol.* **9:** 836–855.

Bruns, T. D., Palmer, J. D., Shumard, D. S., Grossman, L. I., and Hudspeth, M. E. S. (1988). Mitochondrial DNAs of *Suillus:* Three fold size change in molecules that share a common gene order. *Curr. Genet.* **13:** 49–65.

Bruns, T. D., Fogel, R., White, T. J., and Palmer, J. D. (1989). Accelerated evolution of a false-truffle from a mushroom ancestor. *Nature* **339:** 140–142.

Bruns, T. D., Szaro, T. M., Gardes, M., Cullings, K. W., Pan, J. J., Taylor, D. L., Horton, T. R., Kretzer, A., Garbelotto, M., and Li, Y. (1998). A sequence database for the identification of ectomycorrhizal basidiomycetes by phylogenetic analysis. *Mol. Ecol.* **7:** 257–272.

Carpenter, J. M. (1988). Choosing among multiple equally parsimonious cladograms. *Cladistics* **4:** 291–296.

Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, B. D., Duvall, M. R., Price, R. A., Hills, H. G., Qiu, Y.-L., and others. (1993). Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *Rbc*L. *Ann. Miss. Bot. Gard.* **80:** 528–580.

Dayhoff, M. O., Schwartz, R. M., and Orcutt, B. C. (1978). A model of evolutionary change in proteins. *In* "Atlas of Protein Sequence and Structure" (M. O. Dayhoff, Ed.), Vol. 5, Suppl. 3, pp. 345–352. National Biomedical Research Foundation, Washington, DC.

Fox, T. D. (1987). Natural variation in the genetic code. *Annu. Rev. Genet.* **21:** 67–91.

Farris, J. S. (1969). A successive approximation approach to character weighting. *Syst. Zool.* **18:** 374–385.

Gill, M., and Watling, R. (1986). The relationships of *Pisolithus* (Sclerodermataceae) to other fleshy fungi with particular reference to the occurrence and taxonomic significance of hydroxylated pulvinic acids. *Plant Syst. Evol.* **154:** 225–236.

Hibbett, D. S. (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.

Hibbett, D. S., and Donoghue, M. J. (1995). Progress toward a phylogenetic classification of the Polyporaceae through parsimony analysis of mitochondrial ribosomal DNA sequences. *Can. J. Bot.* **73:** S853–S861.

Hibbett, D. S., Pine, E. M., Langer, E., Langer, G., and Donoghue, M. J. (1997). Evolution of gilled mushrooms and puffballs inferred from ribosomal DNA sequences. *Proc. Natl. Acad. Sci. USA* **94:** 12002–12006.

Kellog, E. A., and Juliano, N. D. (1997). The structure and function of RuBisCO and their implications for systematic studies. *Am. J. Bot.* **84:** 413–428.

Kretzer, A., and Bruns, T. D. (1997). Molecular revisitation of the genus *Gastrosuillus*. *Mycologia* **89:** 586–589.

Kretzer, A., Li, Y., Szaro, T., and Bruns, T. D. (1996). Internal transcribed spacer sequences from 38 recognized species of *Suillus* sensu lato: Phylogenetic and taxonomic implications. *Mycologia* **88:** 776–785.

Lutzoni, F. M. (1997). Phylogeny of lichen- and non-lichen-forming omphalinoid mushrooms and the utility of testing for combinability among multiple data sets. *Syst. Biol.* **46:** 373–406.

Moncalvo, J. M., Lutzoni, F., Rehner, S., Johnson, J., and Vilgalys, R. (1997). Molecular phylogeny of the Agaricales based on 25S rDNA sequences. *Inoculum* **48:** 26.

Nishida, H., and Sugiyama, J. (1995). A common group I intron between a plant parasitic fungus and its host. *Mol. Biol. Evol.* **12:** 883–886.

Ochman, H., Gerber, A. S., and Hartl, D. L. (1988). Genetic application of an inverse polymerase chain reaction. *Genetics* **120:** 621–623.

O'Donnell, K., and Cigelnik, E. (1997). Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are non-orthologous. *Mol. Phylogenet. Evol.* **7:** 103–116.

Paquin, B., Forget, L., Roewer, I., and Lang, B. F. (1995a). Molecular phylogeny of *Allomyces macrogynus:* Congruency between nuclear ribosomal RNA- and mitochondrial protein-based trees. *J. Mol. Evol.* **41:** 657–665.

Paquin, B., Roewer, I., Wang, Z., and Lang, B. F. (1995b). A robust fungal phylogeny using the mitochondrially encoded NAD5 protein sequence. *Can. J. Bot.* **73:** S180–S185.

Paquin, B., Laforesst, M.-J., Forget, L., Roewer, I., Wang, Z., Longcore, J., and Lang, B. F. (1997). The fungal mitochondrial genome project: Evolution of fungal mitochondrial genomes and their gene expression. *Curr. Genet.* **31:** 380–395.

Phelps, L. G., Burke, J. M., Ullrich, R. C., and Novotny, C. P. (1988). Nucleotide base sequence of the mitochondrial COIII gene of *Schizophyllum commune*. *Curr. Genet.* **14:** 401–403.

Russo, C. A. M., Takezaki, N., and Nei, M. (1996). Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. *Mol. Biol. Evol.* **13:** 525–536.

Schardl, C. L., Leuchtmann, A., Chung, K.-R., Penny, D., and Siegel, M. R. (1997). Coevolution by common descent of fungal symbionts (*Epichloë* spp.) and grass hosts. *Mol. Biol. Evol.* **14:** 133–143.

Smith, A. H., and Thiers, H. D. (1971). "The Boletes of Michigan," Univ. of Michigan Press, Ann Arbor.

Swofford, D. L. (1993). PAUP: Phylogenetic analysis using parsimony, version 3.1. Computer program distributed by the Illinois Natural History Survey, Champaign, IL.

Triglia, T., Peterson, M. G., and Kemp, D. J. (1988). A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Res.* **16:** 8186.

Vaughan, J. C., Mason, M. T., Sper-Whitis, G. L., Kuhlman, P., and Palmer, J. D. (1995). Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric *Cox*I gene of *Peperomia*. *J. Mol. Evol.* **41:** 563–572.

Vogler, D. R., and Bruns, T. D. (1998). Phylogenetic relationships among the pine stem rust fungi (*Cronartium* and *Peridermium* spp.). *Mycologia* **90:** 244–257.

Wang, L., Yokoyama, K., Miyaji, M., and Nishimura, K. (1998). The identification and phylogenetic relationship of pathogenic species of *Aspergillus* based on the mitochondrial cytochrome *b* gene. *Med. Mycol.* **36:** 153–164.

Wetterer, J. K., Schultz, T. R., and Meier, R. (1998). Phylogeny of fungus-growing ants (Tribe Attini) based on mtDNA sequence and morphology. *Mol. Phylogenet. Evol.* **9:** 42–47.

Zardoya, R., and Meyer, A. (1996). Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Mol. Biol. Evol.* **13:** 933–942.