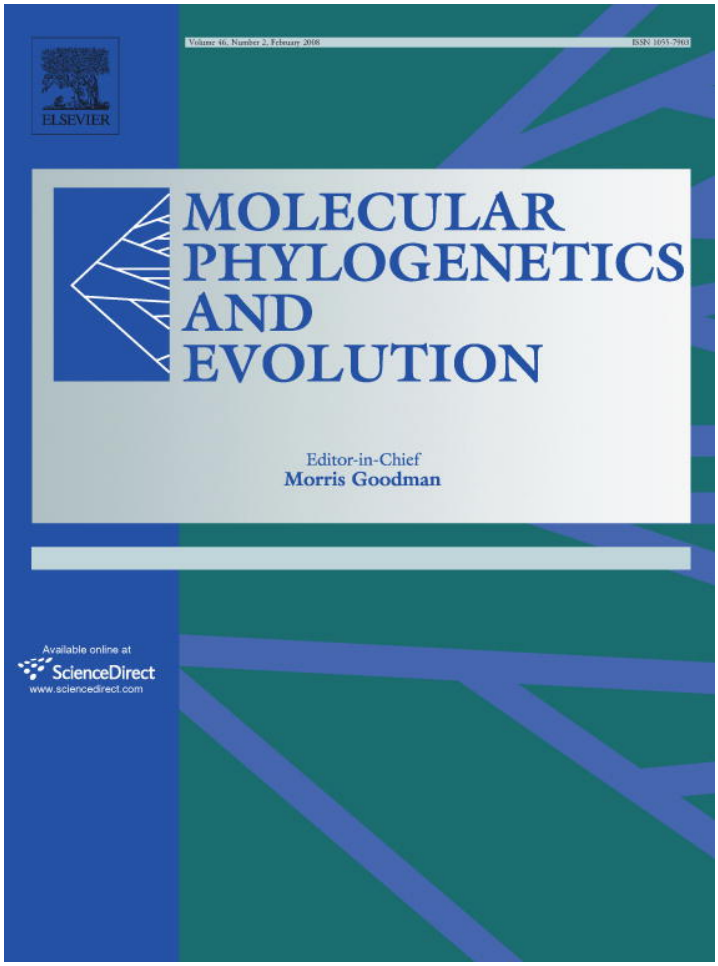


Available online at www.sciencedirect.com
ScienceDirect
Molecular Phylogenetics and Evolution 46 (2008) 635–644
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Widespread occurrence and phylogenetic placement of a soil clone group adds a prominent new branch to the fungal tree of life

T.M. Porter [a,*,1], C.W. Schadt [b,d,1], L. Rizvi [c], A.P. Martin [d], S.K. Schmidt [d], L. Scott-Denton [d], R. Vilgalys [e], J.M. Moncalvo [a,c]

[a] Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto, ON, Canada M5S 3B2
[b] Microbial Ecology and Physiology Group, Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6038, USA
[c] Department of Natural History, Royal Ontario Museum, 100 Queens Park, Toronto, ON, Canada M5S 2C6
[d] Department of Ecology and Evolutionary Biology, Campus Box 334, University of Colorado, Boulder, CO 80309, USA
[e] Department of Biology, Duke University, Durham, NC 27708, USA

Received 30 May 2007; revised 28 September 2007; accepted 6 October 2007
Available online 13 October 2007

Abstract

Fungi are one of the most diverse groups of Eukarya and play essential roles in terrestrial ecosystems as decomposers, pathogens and mutualists. This study unifies disparate reports of unclassified fungal sequences from soils of diverse origins and anchors many of them in a well-supported clade of the Ascomycota equivalent to a subphylum. We refer to this clade as Soil Clone Group I (SCGI). We expand the breadth of environments surveyed and develop a taxon-specific primer to amplify 2.4 kbp rDNA fragments directly from soil. Our results also expand the known range of this group from North America to Europe and Australia. The ancient origin of SCGI implies that it may represent an important transitional form among the basal Ascomycota groups. SCGI is unusual because it currently represents the only major fungal lineage known only from sequence data. This is an important contribution towards building a more complete fungal phylogeny and highlights the need for further work to determine the function and biology of SCGI taxa.

Keywords: Fungi; Ascomycota; Environmental sampling; Soil; Novel taxa; rDNA

1. Introduction

Diversity estimates suggest that less than 5% of the 1.5 million fungal species postulated to exist have been formally described (Hawksworth, 1991, 2001). Although higher-level novel lineages have been frequently detected in the *Bacteria* and *Archaea* using molecular methods (Borneman and Triplett, 1997; Pace, 1997; Handelsman, 2004; Hugenholtz and Pace, 1996; Woese et al., 1986; Dawson and Pace, 2002), this is a relatively rare occurrence in the fungi (Vandenkoornhuyse et al., 2002; Schadt et al., 2003) where species-level novelty is more commonly detected in systematic studies. This is due to the widespread application of culture-free methods to study natural microbial ecosystems by microbiologists.

It has been predicted that new fungal species or groups are more likely to be discovered from poorly studied habitats, particularly from the tropics (Hawksworth, 2001). It was thus surprising when unclassified groups of fungi were discovered from temperate soils, since mycologists have intensively studied them using direct fungal isolation of hyphae and indirect isolation of dormant propagules in culture for many years (Schadt et al., 2003; Malloch, 1981; Warcup, 1965). Environmental DNA sampling strategies have revolutionized our understanding of the diversity of *Bacteria* and *Archaea* in water, sludge, sediments and soil, but only occasionally have the diversity of microscopic eukaryotes been examined using culture and morphology

* Corresponding author.
E-mail address: terri.mclenon@utoronto.ca (T.M. Porter).
[1] These authors contributed equally to this work.

1055-7903/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2007.10.002

independent methods (Dawson and Pace, 2002; Venter et al., 2004; Luo et al., 2005; Moon-van der Staay et al., 2001; Lopez-Garcia et al., 2001). The development and refinement of molecular tools have also prompted a resurgence of fungal community studies (O'Brien et al., 2005; Horton and Bruns, 2001). Identification of environmental sequences is made possible by the expansion of multi-locus datasets and development of bioinformatic tools to facilitate high-throughput sequence analysis for phylogenetic identification (Pennisi, 2005; Bruns, 2006). Preliminary studies using this approach have revealed many "unclassified" sequences that cannot be directly compared with known fungi or each other, either because the regions under study are too small (0.3–0.8 kb) and divergent, or target non-overlapping rDNA regions such as SSU 18S rDNA, internal transcribed spacer (ITS) regions, or LSU 28S rDNA (Schadt et al., 2003; Vandenkoornhuyse et al., 2002; Jumpponen and Johnson, 2005).

This study is the first to attempt to reconcile the relationships among the many "unclassified" fungal sequences that have been recovered from independent studies. To address this issue, we chose to further characterize the LSU Soil Group I clade (Ascomycota, fungi) originally recorded from alpine tundra soil in Colorado, which was highly divergent from known fungal taxa (Schadt et al., 2003). To accomplish this we vastly broadened the diversity of soil environments surveyed to include four new locations in the Americas, and we developed a taxon-specific primer and a nested-PCR technique to generate a 2.4 kb rDNA fragment from Group I members that encompassed portions of the SSU, ITS, and LSU rDNA regions. From these long sequences we retrieved many similar fragments from GenBank and determined the phylogenetic placement of Group I in the fungal tree of life. This allows us to present a first assessment of the taxonomic diversity, evolutionary relationships, geographic distribution and ecology of these newly discovered fungi that we refer to as Soil Clone Group I (SCGI).

2. Materials and methods

2.1. Site characteristics and soil sampling

Upper soils (10–15 cm) were collected from four locations: *Treeline Forest*. This is a forested site dominated by *Picea engelmannii* and *Pseudotsuga menziesii* at 3325 m (10,900 ft) elevation at the forest-tundra ecotone in Colorado. This site is part of the Niwot Ridge Long Term Ecological Research Site (NWR-LTER) which is 35 km west of Boulder, Colorado. Soils originated from equal composites of samples taken randomly from within five, 5 m radius plots taken on three dates; January 25th, 2001 from under winter snow pack (SF102), during spring snowmelt on June 11th, 2002 (SF602) and in summer after the soils had dried out on July 10th (SF702) and were used for separate DNA extraction and rDNA clone library generation. *Montane Forest.* This is a lodgepole pine stand in an aggrading *Pinus contorta*, *Abies lasiocarpa*, *Picea engelmannii* forest at an elevation of 3050 m (10,000 ft) ~2 km east of the forest-tundra ecotone at the NWR-LTER. The site, soil characteristics and ongoing studies have been described (Monson et al., 2002; Scott-Denton et al., 2003). Samples from five control plots (TFC) and the five plots where the trees had been girdled (TFG) were taken randomly on July 31st, 2002, for DNA extraction and rDNA clone library generation. *Costa Rican Oxisol.* This site is located in tropical southwest Costa Rica at a 8°43′N, 83°37′W, ~5 km inland near the town of Drake (Agujitas). The characteristics of the site have been described (Townsend et al., 2002; Cleveland et al., 2002). Ten samples were taken August 31st, 2001, from a forested, P limited, highly-weathered oxisol (CROX) plot. *Temperate Coniferous Forest.* This is a mixed forest heavily dominated by Eastern Hemlock (*Tsuga canadensis*) in southern Ontario, Canada (44°37′N, 79°39′W). The study site is a 50 m × 100 m plot in the Koffler Scientific Reserve situated within the Oak Ridges Moraine. Soil was collected in July 2003 by taking ten soil cores across a 100 transect. For each soil core the B and C horizons were separated, corresponding soil horizons were pooled and mixed by sieving, removing rocks and large roots. In May 2004, a single soil core was sampled from within the plot and soil was separated into 2 cm deep portions. Separate DNA extractions and clone libraries were created for each pooled soil horizon sample and 2 cm deep portion.

2.2. DNA extraction, amplification, cloning and sequencing

DNA from the Colorado and Costa Rican soils was extracted and purified as described previously (Schadt et al., 2003). A portion of the LSU rDNA region was targeted for amplification using the primers ITS9 (Egger, 1995) and nLSU1221R (Schadt et al., 2003). PCR amplification reactions consisted of: 2.75 mM $MgCl_2$, 800 μM dNTPs, 25 μg BSA, 0.5 μM each primer, 1× Taq PCR buffer and 1.875 U Taq Polymerase (Promega, Maddison, WI, USA), and 45 ng template DNA. To avoid biases and artefacts in PCR amplification, eight replicate rxns (25 μL) of each sample were prepared and the total PCR cycle number kept low (Polz and Cavanaugh, 1998; Suzuki and Giovannoni, 1996; Qiu et al., 2001). Thermocycling used a Mastercycler Gradient (Eppendorf, Hamburg, Germany) with initial denaturation at 95 °C for 1 min; 28 cycles of denaturation at 94 °C for 45 s, annealing at 60 °C for 50 s, and extension at 72 °C for 1 min and 45 s (+1 s/cycle); and a final extension step of 72 °C for 5 min. Amplification products were combined and cloned using the TOPO-TA PCR 2.1 kit (Invitrogen, Carlsbad, CA, USA). At least 100 colonies were selected from each cloned sample. Plasmids were extracted using a standard miniprep modified for 96 well plates and inserts were amplified using vector primers T7 and M13R. PCR product was digested with 1 μL each of EXO1 and SAP (USB, Cleveland, OH, USA) to remove unincorporated primers and dNTPs.

2 μL of purified PCR product was used for sequencing reactions with 3.2 pmol of the primers LR0R (Rehner and Samuels, 1994) and TW13 (Taylor and Bruns, 1999) and 2 μL ABI BigDye Ready Reaction Mix (Foster City, CA, USA) in a total volume of 12 μL. Reactions were purified using DyeEx-96 kits (Qiagen, Valencia, CA, USA) and processed on an ABI3700 (Foster City, CA, USA) at the ISU DNA Sequencing Facility (Ames, IA, USA).

DNA for the Ontario samples were extracted using the UltraClean Soil DNA Isolation Kit (Mo Bio Laboratories, Inc., CA, USA). 0.5 g (wet weight) of soil was used per sample. Replicate extractions were performed with various portions of hemlock forest soil and used to generate a total of 12 clone libraries. The first 900 bp of the 5′-nLSU region was amplified using LR0R and LR5 (Vilgalys and Hester, 1990) primers. PCR amplification reactions consisted of: 5 mM dNTPs, 25 μg BSA, 0.4 μM each primer, 1× PCR buffer (Qiagen Inc., ON, Canada), 2.5 U Taq DNA polymerase and 0.1–10 ng template DNA. Thermalcycling used a PTC-100 (MJ Research, Waltham, MA, USA) or a GeneAmp 9700 (Applied Biosystems, CA, USA) with 25 cycles of denaturation at 95 °C for 45 s, annealing at 50 °C for 45 s, extension at 72 °C for 2 min and a final extension step of 72 °C for 5 min. Amplicons were purified using the Qiaquick PCR Purification Kit and cloned using Qiagen PCR Cloning Kit with Qiagen EZ Competent Cells (Qiagen Inc., ON, Canada). Colony PCR was used to amplify 50–100 colonies per plate by picking cells from a single colony using a pipette tip and adding this directly into the PCR cocktail above. The thermal cycler program used was as above, but with 36 cycles, with a 72 °C extension step for 1.5 min. BigDye Terminator v3.1 (Applied Biosystems, CA, USA) chemistry was used for cycle sequencing. Reactions were purified using gel filtration with 600 μL of Sephadex (G-50 fine) slurry in Centri-Sep columns (Princeton Separations, Inc., NJ, USA), dried in a heated vacuum centrifuge at 45–65 °C and run on an ABI Prism 3100 automated sequencer.

2.3. SCGI taxon-specific nested PCR and sequencing

Nested PCR was used to increase specificity, by reducing background amplification due to non-specific primer binding, with bulk DNA from soil. The first reaction used the primers NS1 (White et al., 1990) and LR5 to amplify a long stretch of rDNA spanning the SSU, ITS, and 5′-LSU regions. The PCR mixtures and thermal conditions used were as described above for Ontario soils except that the extension time was increased to 2.5 min. We designed a taxon-specific primer in the D2 divergent domain of the LSU rDNA region (4c26R = 5′-CAGCGTCCTAGGAAGAAC-3′). This primer was used with the SSU primer NS1 to specifically amplify ca. 2.4 kb rDNA. PCR products were purified, cloned and sequenced as described above for the Ontario soils using the following sequencing primers: NS1, SR1.5 and SR6 (James et al., 2000), ITS1 (White et al., 1990) and 4c26R.

2.4. Checking for chimeric sequences and pseudogenes

The presence of chimeric sequences in the data set was checked using two methods: (1) the Chimera Check program (Cole et al., 2003), after uploading additional sequences from our clone libraries; (2) we systematically performed BLAST searches in GenBank with the first-half, followed by the second-half of each sequence, and compared the two taxonomy reports to look for consistency. We were unable to detect the presence of any chimeric sequences in our final dataset using these methods. Also, SCGI sequences do not appear to be pseudogenes because rates of sequence evolution of the ITS1 and ITS2 regions were much faster than in the 5.8S gene, as expected (Bailey et al., 2003).

2.5. Phylogenetic analyses

Sequences were assembled using Sequencher 4.0 (Gene Codes, Ann Arbor MI), and alignments were manually optimized using Se-Al 2.0 (Rambaut, 1996). The placement of SCGI was determined by assembling a dataset of representative Ascomycota taxa from Lutzoni et al. (2004) with SSU and LSU data from our 2.4 kbp SCGI rDNA sequences. We conducted a Bayesian analysis using MrBayes 3.1.1 (Ronquist and Huelsenbeck, 2003) on a 28-processor Linux Beowulf cluster using a general time-reversible model of DNA substitution with the following settings: six classes of nucleotide substitutions, gamma rate amongst sites, four Monte-Carlo Markov chains run for five million generations starting from random trees, and sampling one tree every 100 generations. The first 1000 sampled trees were discarded (burn-in). The resulting 50% majority-rule tree was computed and visualized in PAUP* 4.0b10 (Swofford, 2002). Parsimony bootstrap values were calculated in PAUP* by running 1000 full heuristic bootstrap replicates with the following settings: 100 random addition sequence replicates with TBR branch swapping, and keeping one tree per replicate.

The structure within SCGI was determined by including the greatest number of SCGI taxa with overlapping sequences in a single analysis. We aligned 144 SCGI taxa with 7 reference taxa across 380 bp of the 5′-LSU rDNA region and rooted the tree using *Pneumocystis carinii*. We conducted a parsimony ratchet analysis (Nixon, 1999) as implemented in PAUP* using the PAUPrat module (Sikes and Lewis, 2001) using the default module settings. The form of the best model was determined using MrModeltest 2.2 (Nylander, 2004) and Bayesian support was assessed using the settings described above in MrBayes.

3. Results

The newly designed SCGI taxon-specific primer (4c26R = 5′-CAGCGTCCTAGGAAGAAC-3′) was successfully used in a semi-nested PCR amplification to produce a 2.4 kbp region of rDNA. This strategy generated

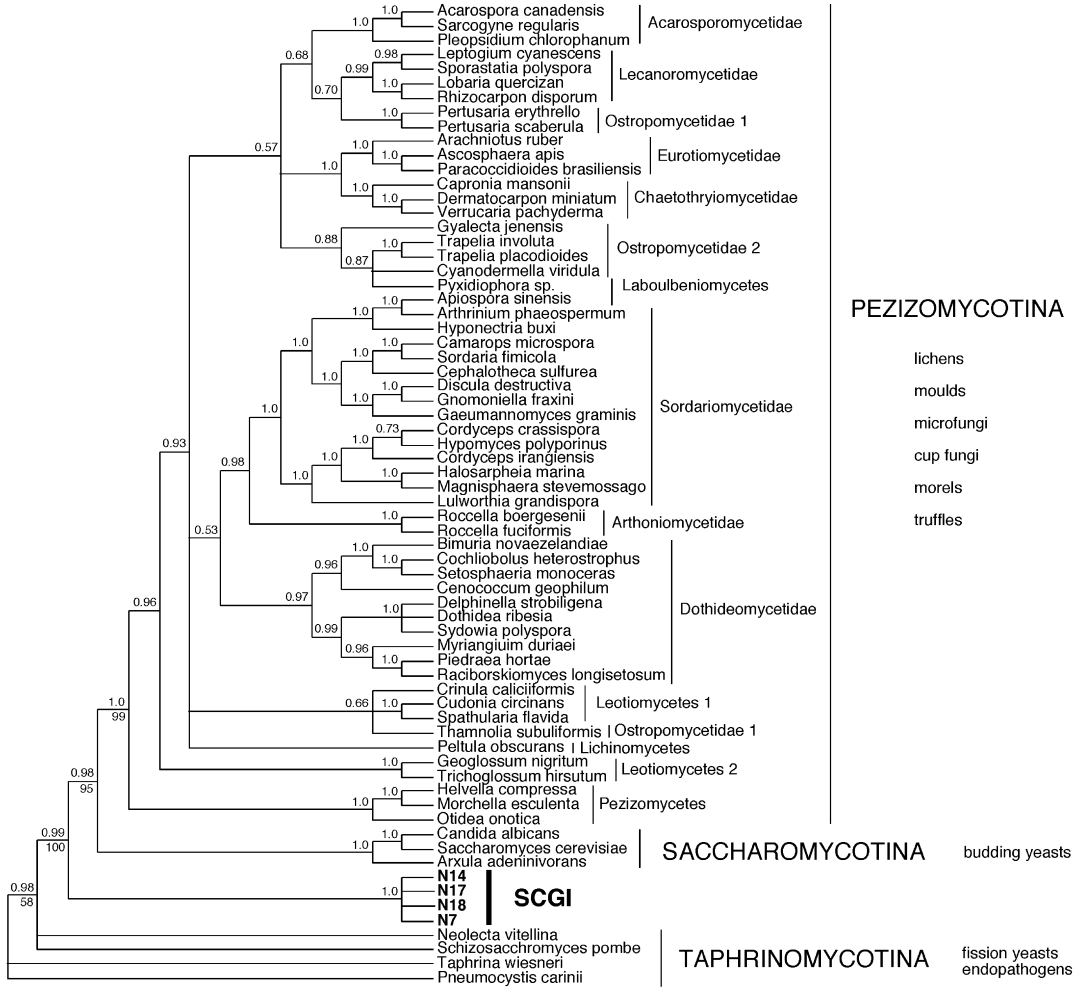


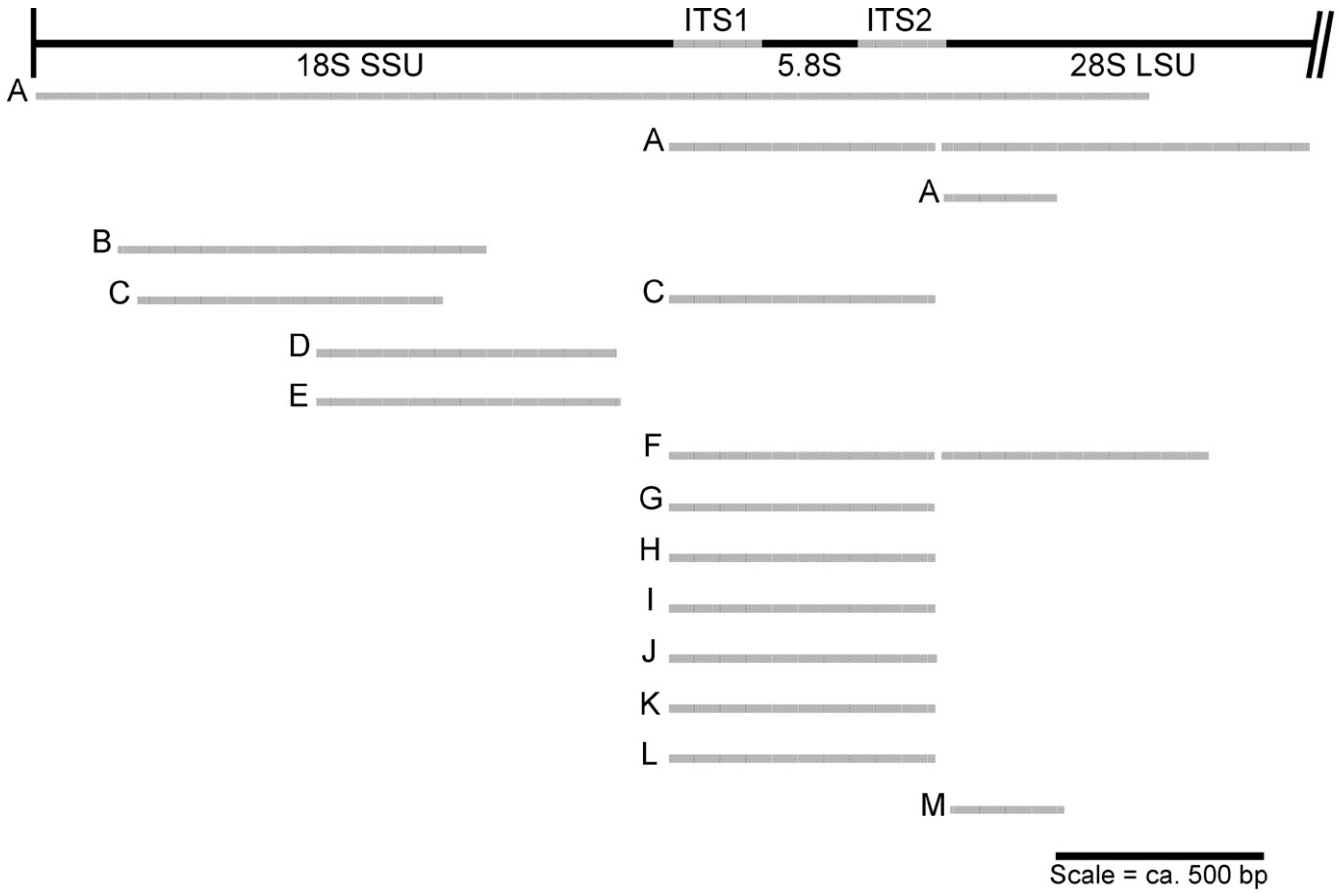
Fig. 1. Bayesian consensus tree for the SSU + LSU combined rDNA analysis. The data set includes 65 taxa and 2,364 SSU and LSU rDNA characters, and is rooted with *Pneumocystis carinii*. Class and subclass labels largely follow the convention used by Lutzoni et al. (2004). Bayesian support is shown above the branch and parsimony bootstrap support is shown below branches of interest.

fragments that span nearly the entire SSU, complete 5.8S gene and ITS spacer regions, as well as the 5′-end of LSU ribosomal genes. This allowed us to unite previously disparate data, greatly broadening this survey, and to firmly root the SCGI group within the Ascomycota. Sequences generated for this study have been submitted to GenBank with Accession Nos. EU179545–EU179599 and EU179933–EU180016.

We confirmed the relatedness of our newly generated SCGI rDNA 2.4 kbp sequences with those originally reported by Schadt et al. (2003) by conducting a maximum parsimony bootstrap analysis which indicated their monophyly with 100% statistical support (Supplementary Information). Bayesian posterior probability (BPP) is expressed on the interval 0.0–1.0. Parsimony bootstrap (PBS) is expressed as a percentage of trees in which a node is found. Analyses including the newly produced 2.4 kpb fragment indicate that the SCGI clade represents a distinct subphylum-level monophyletic group with 1.0 BPP (Fig. 1). The SSU + LSU combined alignment included 2364 characters and resolved the SCGI lineage as a clade independent of the three currently recognized subphyla. The Pezizomycotina subphylum is monophyletic (1.0 BPP/99% PBS), the Pezizomycotina–Saccharomycotina subphyla are monophyletic (0.98 BPP/95% PBS), and the Pezizomycotina–Saccharomycotina–SCGI clades are monophyletic (0.99 BPP/100% PBS). There were no strongly supported conflicts (i.e. greater than 70% PBS) with the grouping of taxa into clades recognized in the literature.

We then retrieved "unclassified" environmental fungal sequences from GenBank using BLAST and confirmed their relatedness to the SCGI clade by conducting a series of maximum parsimony bootstrap analyses using only the characters overlapping with our newly generated 2.4 kb rDNA fragments (Fig. 2 and Supplementary Information). The analyses that included only short stretches of rDNA consistently recovered a moderate to well-supported monophyletic SCGI clade with 62–100% PBS, but could not unambiguously classify these sequences in the fungal tree of life (Supplementary Information).

When we assessed the phylogenetic structure within the SCGI clade we found SCGI taxa to be widespread and com-



	Study	Location	Source
A	This study	Ontario, Canada	Temperate hemlock forest soil
		Colorado, USA	Montane pine-fir forest soil
			Treeline spruce-fir forest soil
		Costa Rica	Tropical forest oxisol
B	Vandenkoornhuyse et al. (2002)	France	*Arrhenatherum elatius* grass roots
C	O'Brien et al. (2005)	North Carolina, USA	Pine forest soil
			Hardwood forest soil
D	Anderson et al. (2003)	Scotland, UK	Grassland soil
E	Jumpponen and Johnson (2005)	Kansas, USA	Tallgrass prairie soil and rhizosphere
F	Lim et al. GenBank	British Columbia, Canada	Mycorrhizal hemlock root tips
G	Chen and Cairney (2002)	NSW, Australia	Sclerophyll forest soil
H	Izzo et al. (2005)	California, USA	Ectomycorrhizal root tips
I	Korkama et al. GenBank	Norway	Norway spruce root tips
J	Menkis et al. (2005)	Lithuania	Ectomycorrhizal *Pinus sylvestris* roots
K	Pringle et al. (2000)	North Carolina, USA	*Acaulospora colossica* fungal spore
L	Rosling et al. (2003)	Sweden	Ectomycorrhizal plant roots
M	Schadt et al. (2003)	Colorado, USA	Alpine tundra dry-meadow soil

Fig. 2. Schematic diagram of the partial rDNA regions sampled for SCGI taxa. Sequences from this study, as well as sequences from GenBank, are shown. Several sequences were from unpublished studies such as: F = Lim et al., GenBank AY394903 and AY394904; and I = Korkama et al., GenBank DQ233843 and DQ233781.

mon within soils from across North and Central America, occurring in forest and alpine tundra soil in Colorado (USA), forested soil in Ontario (Canada), as well in a tropical forest soil in Costa Rica (Fig. 3). The relative abundance of these clones within the overall fungal libraries from the surveyed localities ranged from 6.9% to 27% (Table 1).

Parsimony and Bayesian analyses of the LSU region homologous to bases 190–549 in *Saccharomyces cerevisiae* (J01355) in all SCGI taxa for which we have these sequences revealed many well-supported clades (Fig. 3). The most parsimonious tree had a tree length of 662 (CI = 0.4789, HI = 0.5211, RI = 0.8565) and is shown in Fig. 3 with Bayesian support greater than 0.90 at the nodes. The form of the best model determined by MrModeltest is as follows: equal base frequencies with a GTR model of DNA substitution with gamma distributed rate variation

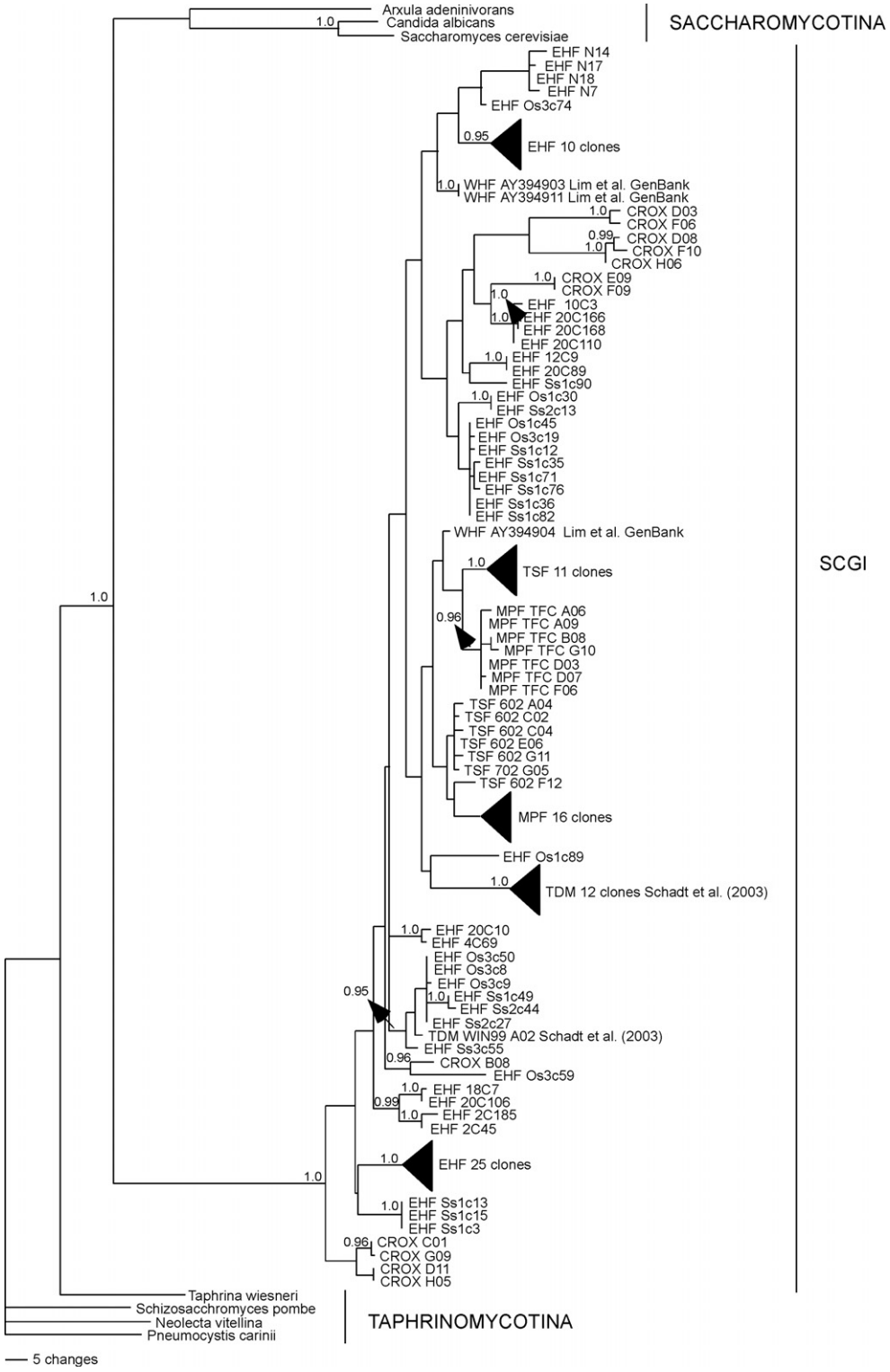


Fig. 3. Parsimony tree for the LSU region showing all SCGI taxa and several basal Ascomycota reference taxa. The data set includes 151 taxa and 352 included characters corresponding to bases 190–549 in *Saccharomyces cerevisiae* (J01355). Samples from six locations are shown: Eastern hemlock forest (EHF), Western hemlock forest (WHF), Costa Rican oxisol (CROX), Treeline spruce-fir forest (TSF), Montane pine-fir forest (MPF), and Alpine tundra dry meadow (TDM). The tree was rooted with *Pneumocystis carinii*. Bayesian support greater than 0.90 is shown above the branches at the nodes.

Table 1
Proportion of LSU SCGI clones recovered from each site

Site	Number of SCGI clones	Total clones analyzed	Percent
Tundra, CO, USA (Schadt et al., 2003)	13	125	10.4
Treeline, CO, USA	18	207	8.7
Montane, CO, USA	23	85	27
Oxisol, Costa Rica	12	50	24
Temperate, ON, Canada	75	1093	6.9

across sites. Nearly all of the tip clades within SCGI clustered by sampling location, similarly, sequences from any one sampling location are distributed across SCGI. In contrast, Bayesian analyses of the SSU and 5.8S regions of SCGI taxa resolved very few well-supported nodes (Supplementary Information).

We also compared the maximum pairwise sequence differences represented within the recognized Ascomycota subphyla and SCGI (Supplementary Information). We found that the maximum percent pairwise sequence differences in the Pezizomycotina ranged from 10.2% to 19.7%, in the Saccharomycotina from 2.2% to 12.1%, in the Taphrinomycotina from 5.3% to 17.3%, and within SCGI from 6.3% to 20% for the SSU and LSU rDNA regions analyzed in this study.

4. Discussion

4.1. SCGI phylogenetic placement and sequence divergence

The total length of rDNA analyzed in this study (2.4 kbp) far exceeds the sampling norm for this type of fungal survey (0.3–1 kbp). This greater number of characters, combined with a backbone phylogeny based on the All Fungal Tree of Life (AFTOL) phylogeny, clearly highlights the significance of the SCGI clade as a novel subphylum of Ascomycota (Fig. 1). In fact, except for the placement of the novel SCGI lineage, this tree topology largely conforms to the rDNA AFTOL phylogeny for the Ascomycota that is itself also independently supported by analyses of protein-coding genes (Lutzoni et al., 2004; Liu et al., 1999). SCGI is clearly distinct from the known Ascomycota subphyla: the Pezizomycotina, Saccharomycotina, and Taphrinomycotina. This suggests that SCGI may harbour a unique biology and ecology that is as yet entirely unstudied. To date, this group is only known from its rDNA, which is unusual at a time when most fungi have already been sorted into major evolutionary lineages (Lutzoni et al., 2004; James et al., 2006). This clearly indicates the need to update the most current fungal classification (Hibbett et al., 2007).

We found that SCGI taxa are significant soil components that may comprise up to 27% of soil fungal sequences sampled in this and other independent studies (Table 1 and Supplementary Information). Clearly, SCGI taxa comprise a substantial portion of soil fungi that was overlooked by traditional sampling methods. The repeated recovery of SCGI from so many environmental DNA-based studies emphasizes the need to recognize SCGI as an integral component of a comprehensive fungal phylogeny.

Even with the limited SCGI sampling conducted to date, the maximum pairwise sequence differences among SCGI taxa, 6.3–20%, is comparable to the maximum pairwise sequence differences found within the other recognized Ascomycota subphyla, 2.2–19.7% (Supplementary Information). This amount of sequence divergence may indicate potentially high taxon diversity within SCGI. Fig. 3 shows that many SCGI taxa cluster according to sampling site. If we were to use 99% sequence similarity to delimit operational taxonomic units (OTUs) as a proxy for species, as used in similar studies for the LSU rDNA region (such as Schadt et al., 2003), then 144 SCGI sequences in Fig. 3 represents 38 OTUs. Additionally, each OTU is comprised of clone sequences from a single site. A frequency histogram of OTUs versus sequences sampled shows that two-thirds of OTUs are rare taxa sampled only once or twice (data not shown). These results suggest that we have only just begun to sample the diversity of SCGI taxa in the soils analyzed in this study, making a comprehensive geographic assessment of SCGI diversity difficult to assess at this point.

4.2. Current theory of Ascomycota evolution

It has been hypothesized that the most basal Ascomycota are the Taphrinomycotina (Sugiyama et al., 2006; James et al., 2006; Liu et al., 1999), recognized relatively recently from molecular evidence (Nishida and Sugiyama, 1994; Liu and Hall, 2004). Yeasts and filamentous hyphae characterize the vegetative stage whereas the sexual stage is characterized by unitunicate asci and a lack of well-formed ascomata (one exception is *Neolecta*). Lifestyles range from obligate plant pathogens, animal pathogens, to saprobic fission yeasts (Berbee, 2001). Two well-known exemplars include the plant pathogen *Taphrina deformans* and *P. carinii*, a cause of pneumonia in immune-compromised patients.

The Saccharomycotina and Pezizomycotina are the next two subphyla to arise (Lutzoni et al., 2004; Liu et al., 2006). The Saccharomycotina are composed of about 1000 species with basal forms that are filamentous, others that are true budding yeasts, and more derived forms that show filamentous-yeast dimorphism (Suh et al., 2006). Most are free-living, but one genus is known to contain animal pathogens (*Candida*), and one genus contains plant pathogens (*Eremothecium*) (Berbee, 2001). The sexual stage is also characterized by the lack of ascomata and the formation of unitunicate asci. Some well-known members include the baker's and brewer's yeast, *S. cerevisiae*, as well as the opportunistic human pathogen, *Candida albicans*.

The Pezizomycotina is the largest subphylum and includes greater than 27,000 described species (Kirk

et al., 2001). A five-gene phylogeny strongly supported the monophyly of the Pezizomycotina (Spatafora et al., 2006). When the sexual stage is known, it is characterized by the formation of ascomata. Generally, the "higher" ascomycetes show more complex morphology and include most lichen-forming fungi (Gargas et al., 1995). Some well-known forms include the edible delicacy, *Morchella,* the morel mushroom and the ubiquitous *Penicillium* and *Aspergillus* moulds.

4.3. Transitional nature of SCGI

The Taphrinomycotina and the Saccharomycotina is intersected by SCGI in our phylogeny (Fig. 1). It is unlikely that SCGI taxa produce a macroscopic fruiting body, based on the observed lack of well-formed ascomata in the Taphrinomycotina and lack of ascomata in the Saccharomycotina. This would explain why SCGI taxa have not been detected in previous studies using traditional fruiting body collection methods or methods that rely on morphology for identification and classification. If SCGI taxa are obligately biotrophic, as are many of the pathogens in the Taphrinomycotina, this would explain why these taxa have been overlooked in studies using conventional culture-based isolation techniques. Although we have made no rigorous attempt to culture SCGI taxa directly from soil, if our hypothesis about an obligately symbiotic lifestyle is true, it will not be possible to find a living representative of SCGI without its host.

Based on the ancient origin of SCGI among the basal Ascomycota, this group may have diversified before the origin of the complex Ascomycete sexual reproductive structure, the ascoma, which is a defining characteristic among sexually reproducing members of the Pezizomycotina. Further study of the biology and function of SCGI would help to further develop a holistic evolutionary theory of Ascomycota fungi. This study is the first to convincingly place SCGI in the Ascomycota phylogeny with good statistical support. For now the morphological characteristics and metabolic and ecological properties associated with SCGI remain unknown and can only be speculated from comparison with their closest evolutionary relatives.

4.4. Known geographical and environmental range of SCGI

We have expanded the known range of SCGI from a single North American alpine tundra location (Schadt et al., 2003), to other sites within the Americas, Europe and Australia (Figs. 2 and 3). Members of SCGI were often very prominent among the percentage of overall fungal clones present, often comprising more than 10% of the total number of clones (Table 1 and Supplementary Information). This abundance in clone libraries suggests a numerical abundance in soils, however further experiments using in situ hybridization or other methods will be needed to confirm this (Amann et al., 1995). Using the 2.4 kbp region characterized in this study it was also possible to make comparisons across a number of studies utilizing ITS and SSU rDNA markers. These studies confirmed that SCGI members are commonly detected from soil and ectomycorrhizal root tips in North America from both western and eastern Canadian provinces, from the western, middle, and eastern states of the United States, with a southerly range that extends into Costa Rica (Schadt et al., 2003; O'Brien et al., 2005; White et al., 1990; Izzo et al., 2005; Pringle et al., 2000). SCGI members are also found in the same habitats in Europe from Scotland in the United Kingdom, with their occurrence extending eastwards and northwards to Lithuania, Sweden and Norway (Anderson et al., 2003; Menkis et al., 2005; Rosling et al., 2003). SCGI members are also found in Australia in soils from New South Wales (Chen and Cairney, 2002). The cosmopolitan distribution of these taxa on three separate continents is consistent with a hypothesis of a relatively flexible and adaptable biology capable of surviving in subsurface environments in a variety of both temperate and tropical soils.

It is also worthy to note which studies using similar PCR-based methods to study environmental samples have not detected members of the SCGI clade. Although SCGI taxa may be widespread in studies from alpine tundra, forest, and grassland soils as well as from ectomycorrhizal plant roots, despite exhaustive GenBank searches we did not detect these sequences in studies from insect guts, above-ground plant endophytes, grass roots, Sargasso Sea water, or sulphide-rich springs (Suh et al., 2005; Higgins et al., 2007; Arnold et al., 2007; Vandenkoornhuyse et al., 2002; Venter et al., 2004; Luo et al., 2005). Similarly, SCGI taxa have not been detected from soil sampled from the forefront of the Lyman glacier nor from unvegetated alpine talus soils (Meyer and S.S., unpublished; Jumpponen, 2003). These observations collectively suggest that an active rhizosphere may be an essential requirement for SCGI taxa.

4.5. Other novel fungal groups

While we were able to unify many previously uncharacterized sequences with SCGI from direct phylogenetic comparison with our 2.4 kb rDNA fragments, such as the 'Unknown Soil Fungi' clade from Jumpponen and Johnson (2005; Supplementary Information), other unknown fungal groups detected from soil have yet to be properly classified. For instance, Vandenkoornhuyse et al. (2002) sampled SSU rDNA from grass roots and they recovered two unclassified groups labelled 'IV' and 'V' in the Ascomycota. Group V clearly clusters in the Pezizomycotina in their analysis. When we tested the relationship of their Group IV with our SCGI clade, they failed to cluster together (Supplementary Information). Additionally, Schadt et al. (2003) recovered two additional unknown lineages in the Ascomycota labelled 'Group II' and 'Group III' but these are clearly nested within the Pezizomycotina in their supplementary analyses and fail to nest with SCGI in our analyses (Supplementary Information). As a result,

these unknown fungal lineages previously identified from grass roots and alpine tundra soil are more likely to represent species- to family-level novelty within the Pezizomycotina (Schadt et al., 2003; Jumpponen and Johnson, 2005).

4.6. Conclusions

Similar to the recognition of the phylum-level status of the Glomeromycota (Schußler et al., 2001) and the inclusion of the Microsporidiomycota (Keeling, 2003) in the fungi, the SCGI clade should also be recognized as a significant contribution towards reconstructing a more complete fungal phylogeny and elucidating the nature of the "missing fungi". This highlights a large gap in our knowledge of Ascomycota and emphasizes the need for further studies to characterize these fungi using fluorescent hybridization, metagenomic and targeted culture-based methods.

Acknowledgments

We thank J. Skillman and S. Margaritescu for laboratory assistance and B. Saville for helpful discussion. This work was supported by grants from the Discovery Program at the Natural Sciences and Engineering Research Council of Canada and from the Canadian Fund for Innovation (J.-M. Moncalvo), the NSF Microbial Observatories Program (S. Schmidt and A. Martin), and the Department of Energy Office of Science as part of the consortium for research into Carbon Sequestration in Terrestrial Ecosystems (CSiTE) (C. Schadt).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2007.10.002.

References

Amann, R.I., Ludwig, W., Schleifer, K.-H., 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. 59, 143–169.

Anderson, I.C., Campbell, C.D., Prosser, J.I., 2003. Potential bias of fungal 18S rDNA and internal transcribed spacer polymerase chain reaction primers for estimating fungal biodiversity in soil. Environ. Microbiol. 5, 36–47.

Arnold, A.E., Henk, D.A., Eells, R.L., Lutzoni, F., Vilgalys, R., 2007. Diversity and phylogenetic affinities of foliar fungal endophytes in loblolly pine inferred by culturing and environmental PCR. Mycologia 99, 185–206.

Bailey, C.D., Carr, T.G., Harris, S.A., Hughes, C.E., 2003. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. Mol. Phylogenet. Evol. 29, 435–455.

Berbee, M.L., 2001. The phylogeny of plant and animal pathogens in the Ascomycota. Phys. Mol. Plant Path. 59, 165–187.

Borneman, J., Triplett, E.W., 1997. Molecular microbial diversity in soils from Eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. Appl. Environ. Microbiol. 63, 2647–2653.

Bruns, T., 2006. A kingdom revised. Nature 443, 758–761.

Chen, D.M., Cairney, J.W.G., 2002. Investigation of the influence of prescribed burning on ITS profiles of ectomycorrhizal and other soil fungi at three Australian sclerophyll forest sites. Mycol. Res. 106, 532–540.

Cleveland, C.C., Townsend, A.R., Schmidt, S.K., 2002. Phosphorus limitation of microbial processes in moist tropical forests: evidence from short-term laboratory incubations and field studies. Ecosystems 5, 680–691.

Cole, J.R., Chai, B., Marsh, T.L., Farris, R.J., Wang, Q., Kulam, S.A., Chandra, S., McGarrell, D.M., Schmidt, T.M., Garrity, G.M., et al., 2003. The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. Nucleic Acids Res. 31, 442–443.

Dawson, S.C., Pace, N.R., 2002. Novel kingdom-level eukaryotic diversity in anoxic environments. Proc. Natl. Acad. Sci. 99, 8324–8329.

Egger, K.N., 1995. Molecular analysis of ectomycorrhizal fungal communities. Cdn. J. Bot. 73, S1415–S1422.

Gargas, A., DePriest, P.T., Grube, M., Tehler, A., 1995. Multiple origins of lichen symbioses in fungi suggested by SSU rDNA phylogeny. Science 268, 1492–1495.

Handelsman, J., 2004. Metagenomics: application of genomics to uncultured microorganisms. Microbiol. Mol. Biol. Rev. 68, 669–685.

Hawksworth, D.L., 1991. The fungal dimension of biodiversity: magnitude, significance, and conservation. Mycol. Res. 95, 641–655.

Hawksworth, D.L., 2001. The magnitude of fungal diversity: the 1.5 million species estimate revisited. Mycol. Res. 105, 1422–1432.

Hibbett, D.S., Binder, M., Bischoff, J.F., Blackwell, M., Cannon, P.F., Eriksson, O.E., Huhndorf, S., James, T., Kirk, P.M., Lucking, R., et al., 2007. A higher-level phylogenetic classification of the Fungi. Mycol. Res. 111, 509–547.

Higgins, K.L., Arnold, A.E., Miadlikowska, J., Sarvate, S.D., Lutzoni, F., 2007. Phylogenetic relationships, host affinity, and geographic structure of boreal and arctic endophytes from three major plant lineages. Mol. Phylogenet. Evol. 42, 543–555.

Horton, T.R., Bruns, T.D., 2001. The molecular revolution in ectomycorrhizal ecology: peeking into the black-box. Mol. Ecol. 10, 1855–1871.

Hugenholtz, P., Pace, N.R., 1996. Identifying microbial diversity in the natural environment: a molecular phylogenetic approach. Trends Biotechnol. 14, 190–197.

Izzo, A., Agbowo, J., Bruns, T.D., 2005. Detection of plot-level changes in ectomycorrhizal communities across years in an old-growth mixed-conifer forest. New Phytol. 166, 619–630.

James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J., et al., 2006. Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature 443, 818–822.

James, T.Y., Porter, D., Leander, C.A., Vilgalys, R., Longcore, J.E., 2000. Molecular phylogenetics of the Chytridiomycota supports the utility of ultrastructural data in chytrid systematics. Can. J. Bot. 78, 336–350.

Jumpponen, A., 2003. Soil fungal community assembly in a primary successional glacier forefront ecosystem as inferred from rDNA sequence analyses. New Phytol. 158, 569–578.

Jumpponen, A., Johnson, L.C., 2005. Can rDNA analyses of diverse fungal communities in soil and roots detect effects of environmental manipulations—a case study from tallgrass prairie. Mycologia 97, 1177–1194.

Keeling, P.J., 2003. Congruent evidence from α-tubulin and β-tubulin gene phylogenies for a zygomycete origin of microsporidia. Fungal. Genet. Biol. 38, 298–309.

Kirk, P.M., Cannon, P.F., David, J.C., Stalpers, J.A., 2001. Ainsworth and Bisby's Dictionary of the Fungi, ninth Ed. CABI International, UK.

Liu, Y.J., Hall, B.D., 2004. Body plan evolution of ascomycetes, as inferred from an RNA polymerase II phylogeny. Proc. Natl. Acad. Sci. 101, 4507–4512.

Liu, Y.J., Hodson, M.C., Hall, B.D., 2006. Loss of the flagellum happened only once in the fungal lineage: phylogenetic structure of Kingdom

Fungi inferred from RNA polymerase II subunit genes. BMC Evol. Biol. 6, 74.

Liu, Y.J., Whelen, S., Hall, B.D., 1999. Phylogenetic relationships among ascomycetes: evidence from an RNA polymerase II subunit. Mol. Biol. Evol. 12, 1799–1808.

Lopez-Garcia, P., Rodriguez-Valera, F., Pedros-Alio, C., Moreira, D., 2001. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. Nature 409, 603–607.

Luo, Q., Krumholz, L.R., Najar, F.Z., Peacock, A.D., Roe, B.A., White, D.C., Elshahed, M.S., 2005. Diversity of the microeukaryotic community in sulfide-rich zodletone spring (Oklahoma). Appl. Environ. Microbiol. 71, 6175–6184.

Lutzoni, F., Kauff, F., Cox, C.J., McLaughlin, D., Celio, G., Dentinger, B., Padamsee, M., Hibbett, D., James, T.Y., Baloch, E., et al., 2004. Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits. Am. J. Bot. 9, 1446–1480.

Malloch, D., 1981. Moulds: Their Isolation Cultivation and Identification. University of Toronto Press, Toronto.

Menkis, A., Vasiliauskas, R., Taylor, A.F.S., Stenlid, J., Finlay, R., 2005. Fungal communities in mycorrhizal roots of conifer seedlings in forest nurseries under different cultivation systems, assessed by morphotyping, direct sequencing and mycelial isolation. Mycorrhiza 16, 33–41.

Monson, R.K., Turnipseed, A.A., Sparks, J.P., Harley, P.C., Scott-Denton, L.E., Sparks, K., Huxman, T.E., 2002. Carbon sequestration in a high-elevation, subalpine forest. Glob. Change Biol. 8, 459–478.

Moon-van der Staay, S.Y., De Wachter, R., Vaulot, D., 2001. Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. Nature 409, 607–610.

Nishida, H., Sugiyama, J., 1994. Archiascomycetes: detection of a major new lineage within the Ascomycota. Mycoscience 35, 361–366.

Nixon, K.C., 1999. The Parsimony Ratchet, a new method for rapid parsimony analysis. Cladistics 15, 407–414.

Nylander, A.A.J., 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

O'Brien, H.E., Parrent, J.L., Jackson, J.A., Moncalvo, J.-M., Vilgalys, R., 2005. Fungal community analysis by large-scale sequencing of environmental samples. Appl. Environ. Microbiol. 71, 5544–5550.

Pace, N., 1997. A molecular view of microbial diversity and the biosphere. Science 276, 734–740.

Pennisi, E., 2005. Fungal trees grow faster with computer help. Science 309, 374.

Pringle, A., Moncalvo, J.-M., Vilgalys, R., 2000. High levels of variation in ribosomal DNA sequences within and among spores of a natural population of the arbuscular mycorrhizal fungus *Acaulospora colossica*. Mycologia 92, 259–268.

Polz, M.F., Cavanaugh, C.M., 1998. Bias in template-to-product ratios in multitemplate PCR. Appl. Environ. Microbiol. 64, 3724–3730.

Qiu, X., Wu, L., Huang, H., McDonel, P.E., Palombo, A.V., Tiedje, J.M., Zhou, J., 2001. Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. Appl. Environ. Microbiol. 67, 880–887.

Rambaut, A., 1996. Se-Al: Sequence Alignment Editor. Available from: http://evolve.zoo.ox.ac.uk/.

Rehner, S.A., Samuels, G.J., 1994. Taxonomy and phylogeny of *Gliocladium* analysed from the nuclear large subunit ribosomal DNA sequences. Mycol. Res. 98, 625–634.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Rosling, A., Landeweert, R., Lindahl, B.D., Larsson, K.-H., Kuyper, T.W., Taylor, A.F.S., Finlay, R.D., 2003. Vertical distribution of ectomycorrhizal fungal taxa in a podzol soil profile. New Phytol. 159, 775–783.

Schadt, C.W., Martin, A.P., Lipson, D.A., Schmidt, S.K., 2003. Seasonal dynamics of previously unknown fungal lineages in tundra soils. Science 301, 1359–1361.

Schußler, A., Schwarzott, D., Walker, C., 2001. A new fungal phylum, the Glomeromycota: phylogeny and evolution. Mycol. Res. 105, 1413–1421.

Scott-Denton, L.E., Sparks, K.L., Monson, R.K., 2003. Spatial and temporal controls of soil respiration rate in a high-elevation, subalpine forest. Soil Biol. Biochem. 35, 525–534.

Sikes, D.S., Lewis, P.O., 2001. Beta software, version 1. PAUPrat: PAUP implementation of the parsimony ratchet. Distributed by the authors. Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs.

Spatafora, J.W., Sung, G.H., Johnson, D., Hesse, C., O'Rourke, B., Serdani, M., Spotts, R., Lutzoni, F., Hofstetter, V., Miadlikowska, J., et al., 2006. A five-gene phylogeny of Pezizomycotina. Mycologia 98, 1018–1028.

Sugiyama, J., Hosaka, K., Suh, S.O., 2006. Early diverging Ascomycota: phylogenetic divergence and related evolutionary enigmas. Mycologia 98, 996–1005.

Suh, S.O., Blackwell, M., Kurtzman, C.P., Lachance, M.A., 2006. Phylogenetics of Saccharomycetales, the ascomycete yeasts. Mycologia 98, 1006–1017.

Suh, S.-O., McHugh, J.V., Pollock, D.D., Blackwell, M., 2005. The beetle gut: a hyperdiverse source of novel yeasts. Mycol. Res. 109, 261–265.

Suzuki, M.T., Giovannoni, S.J., 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. 62, 625–630.

Swofford, D.L., 2002. PAUP* Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Taylor, D.L., Bruns, T.D., 1999. Community structure of ectomycorrhizal fungi in a *Pinus muricata* forest: minimal overlap between the mature forest and resistant propagule communities. Mol. Ecol. 8, 1837–1850.

Townsend, A.T., Asner, G.P., Cleveland, C.C., Lefer, M.E., Bustamante, M.M., 2002. Unexpected changes in soil phosphorus dynamics along pasture chronosequences in the humid tropics. J. Geophys. Res. Atmos. 107, 8067–8076.

Vandenkoornhuyse, P., Baldauf, S.L., Leyval, C., Straczek, J., Young, J.P.W., 2002. Extensive fungal diversity in plant roots. Science 295, 2051.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., et al., 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science 304, 66–74.

Vilgalys, R., Hester, M., 1990. Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several *Cryptococcus* species. J. Bacteriol. 172, 4238–4246.

Warcup, J.H., 1965. Growth and reproduction of soil microorganisms in relation to substrate. In: Baker, K.F., Snyder, W.C. (Eds.), Ecology of Soil-Borne Plant Pathogens. University of California Press, Berkeley, pp. 52–68.

White, T.J., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. (Eds.), PCR Protocols. Academic Press, San Diego, pp. 315–322.

Woese, C.R., Pace, N.R., Olsen, G.J., 1986. Are arguments against archaebacteria valid? Nature 320, 401–402.