Tasting Soil Fungal Diversity with Earth Tongues: Phylogenetic Test of SATé Alignments for Environmental ITS Data

Zheng Wang[1]*, R. Henrik Nilsson[2,3], Francesc Lopez-Giraldez[1], Wen-ying Zhuang[4], Yu-cheng Dai[5], Peter R. Johnston[6], Jeffrey P. Townsend[1]

1 Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut, United States of America, **2** Department of Plant and Environmental Sciences, Göteborg University, Göteborg, Sweden, **3** Department of Botany, Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia, **4** Institute of Microbiology, Chinese Academy of Science, Beijing, China, **5** Institute of Microbiology, Beijing Forestry University, Beijing, China, **6** Herbarium PDD, Landcare Research, Auckland, New Zealand

Abstract

An abundance of novel fungal lineages have been indicated by DNA sequencing of the nuclear ribosomal ITS region from environmental samples such as soil and wood. Although phylogenetic analysis of these novel lineages is a key component of unveiling the structure and diversity of complex communities, such analyses are rare for environmental ITS data due to the difficulties of aligning this locus across significantly divergent taxa. One potential approach to this issue is simultaneous alignment and tree estimation. We targeted divergent ITS sequences of the earth tongue fungi (Geoglossomycetes), a basal class in the Ascomycota, to assess the performance of SATé, recent software that combines progressive alignment and tree building. We found that SATé performed well in generating high-quality alignments and in accurately estimating the phylogeny of earth tongue fungi. Drawing from a data set of 300 sequences of earth tongues and progressively more distant fungal lineages, 30 insufficiently identified ITS sequences from the public sequence databases were assigned to the Geoglossomycetes. The association between earth tongues and plants has been hypothesized for a long time, but hard evidence is yet to be collected. The ITS phylogeny showed that four ectomycorrhizal isolates shared a clade with *Geoglossum* but not with *Trichoglossum* earth tongues, pointing to the significant potential inherent to ecological data mining of environmental samples. Environmental sampling holds the key to many focal questions in mycology, and simultaneous alignment and tree estimation, as performed by SATé, can be a highly efficient companion in that pursuit.

Citation: Wang Z, Nilsson RH, Lopez-Giraldez F, Zhuang W-y, Dai Y-c, et al. (2011) Tasting Soil Fungal Diversity with Earth Tongues: Phylogenetic Test of SATé Alignments for Environmental ITS Data. PLoS ONE 6(4): e19039. doi:10.1371/journal.pone.0019039

Editor: Olivier Lespinet, Université Paris-Sud, France

Received January 21, 2011; **Accepted** March 22, 2011; **Published** April 21, 2011

Funding: ZW was supported by a Yale University internal postdoctoral fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: wang.zheng@yale.edu

Introduction

Estimates of the number of extant species of fungi range from 0.5 million to 3.5 million or more [1] [2]. While the total count is open to debate, few scientists doubt that the *c.* 100,000 species of fungi described so far [3] represent but a small portion of the global fungal diversity. Unfortunately, the rate at which new species are described is not increasing over time [4], suggesting that one or more steps in the species discovery and description processes may have to be streamlined. Traditional approaches to the systematics and taxonomy of fungi rely to a large extent on the presence of distinctive morphological characters of reproductive structures or on culture studies in the laboratory. However, many fungi do not produce detectable reproductive morphological structures during their observable life stages, and the majority of fungal lineages defy attempts at keeping them in laboratory cultures [5] [6] [7]. Furthermore, the formation of fruiting bodies of many fungi is rare under the conditions in which most described fungi are observed or collected, and the reproductive structures of many micro-fungi lack distinguishing characters altogether [8] [9] [10]. The advent of molecular (DNA sequence) data has provided a new and bountiful source of data, and many new taxa have been described using the sequence data as the chief discriminative criterion. Much of this recent work has relied on a single popular "universal" molecular marker or "DNA barcode" for the identification of fungi: the nuclear ribosomal internal transcribed spacer (ITS) region, including the three subregions the ITS1, 5.8S rDNA, and ITS2 [11] [12] [13].

Significant methodological advancements in DNA extraction and sequencing have unfortunately done little to ameliorate the challenge of bringing ITS sequences to bear upon the characterization of the diversity of environmental or otherwise cryptic fungi. Some pragmatic attempts to characterize the diversity of these "invisible fungi" from environmental samples have relied on operational taxonomic units (OTUs) that have been delimited as discrete units of sequences that share, e.g., 95–97% pairwise similarity [14] [15] [16]. Problematically, this level of similarity is within the range of intraspecific ITS sequence divergence for many fungal lineages. Furthermore, fully identified ITS sequences are available in the International Nucleotide Sequence Database

Collaboration (INSD: GenBank, EMBL, and DDBJ; [17]) for about 1% of the 1.5 million estimated number of fungal species (and some 15% of the described fungal species; [3]). As a result, many newly generated environmental ITS sequences do not have any fully identified counterparts in these databases; indeed, a full 41% of the ~145,000 fungal ITS sequences in INSD remain unidentified to species level, including more than 14,000 sequences annotated as coming from soil samples, about 10,000 sequences of mycorrhizal fungi, and some 4,000 sequences of endophytic fungi. The proportion of insufficiently identified sequences is likely to increase dramatically in the near future as high-throughput sequencing techniques become commonplace. Coupled with the very moderate pace at which the herbaria worldwide are explored for sequence data [12], this burgeoning trend toward rapid accumulation of insufficiently identified sequence data provides little hope for improving our prospects for rapid molecular identification of fungi.

Identification of every environmental sequence to species level, however, may not always be necessary for many ecological and evolutionary questions. To understand the overall fungal diversity and related ecology in some environmental samples, many ecologists would be satisfied with a tree-based genus-level estimation of fungal diversity, which would provide the likely ecology and biology of the unknown lineages. Moreover, as a complement to OTUs, a well-resolved low-level (usually generic level) phylogeny would be a very powerful tool in the characterization of fungal diversity in environmental samples, and it would also be critical for the description of new taxa.

One candidate locus for such low-level phylogenies in fungi is the well-studied ITS region. The ITS region is a popular marker for fungal DNA fingerprinting, as it is easily recovered from diverse samples due to its multicopy nature and also because of the relative ease of designing primers to match the conserved neighboring ribosomal genes [6] [18]. ITS sequences are not as widely used as other rDNA regions and protein-coding genes for higher-level phylogenetic inference due to the high rates of evolution of the region; ITS alignments are typically difficult to extend beyond the genus or subfamily level [19]. And yet multiple sequence alignments (MSAs) are a fundamental prerequisite for many downstream comparative analyses, including phylogenetic inference. Automated solutions to sequence alignment invariably call for manual inspection and adjustment [20], which makes the process intractable for very large datasets. One novel approach is simultaneous alignment and tree estimation, such as that performed by SATé, a recently released program [21] [22] that unites the inference of optimal multiple alignments and phylogenetic trees into a single process, thereby eliminating manual examination of MSAs for molecular ecological inference. The present study examines the use of SATé in exploring the diversity of earth tongue fungi, drawing from published studies as well as from environmental samples (Fig. 1). We also contrasted SATé with a simpler approach based on Clustal W alignment and Bayesian phylogenetic analysis to be able to interpret the results of SATé in light of the corresponding results from more traditional phylogenetic approaches.

Earth tongue fungi are ascomycetes that produce large pileate (mushroom-like) fruiting bodies on various substrates, and their unique dark colored and tongue- shaped fruiting bodies are common in temperate regions (Fig. 2). Although earth tongue fungi have a world-wide distribution, most species were described from North America and southwest China [23]. Their classification has changed significantly in recent times, and they are now placed in the newly erected class Geoglossomycetes [24]. A recent molecular phylogeny of the ascomycetes places this class of fungi as a deep branch around the Pezizomycetes, Orbiliomycetes, and some early lichenized ascomycetes, and this placement received moderate to strong support from different sets of genes [25]. Many earth tongues are considered as associated with bryophytes, but no hard evidence has ever been established. About 50 species [3] have been reported from four genera - *Geoglossum*, *Trichoglossum*, *Sarcoleotia*, and *Thuemenidium* - but the relationships among their species remain understudied [26] [27]. ITS sequences from these lineages are divergent in terms of sequence identity and also in sequence length. Being able to recover ITS sequences from environmental ITS using next generation sequencing techniques would be very helpful to understand the diversity, distribution, and ecology of this group of fungi. At the onset of this study, a mere five environmental ITS sequences had been classified in this group in INSD, and another 40 entries showed their highest similarity to published earth tongue sequences in the sequence search suite *emerencia* [28]. We explored these environmental sequences in the light of a reference dataset of 31 fully identified earth tongue fungi in a phylogenetic context.

Results

Earth tongue ITS data

A total of 16 ITS sequences were generated from 11 species of *Geoglossum*, *Trichoglossum*, and *Thuemenidium*, and one ITS sequence (PDD70355) generated from a New Zealand collection was subsequently identified as a species of *Microglossum*. INSD accession numbers of the newly generated ITS sequences were HQ222862 through HQ222878 (Table S1). Fifteen fully identified ITS sequences were downloaded from INSD for species of *Geoglossum*, *Trichoglossum*, and *Sarcoleotia*. As estimated with Needle of the EMBOSS suite, the similarity among earth tongues in ITS sequence was generally lower than 70% between different genera. Similarity within genera was not much higher, and the greatest distances observed in each genus ranged from 58% to 67% similarity. Twenty-three ITS sequences from 10 environmental studies were retrieved by the BLAST searches of the core ITS dataset. Reference ITS sequences from ascomycetous groups related to the Geoglossomycetes and ITS sequences identified as uncultured Pezizomycotina environmental samples were added to make a data set of 300 ITS sequences.

Alignment of ITS using Clustal W and SATé

The whole ITS alignment of all 300 ITS sequences as generated by Clustal W was deemed to be of unsatisfactory quality as the 5.8S region of many sequences was appreciably misaligned. A reduced data set of 118 taxa, including only reference sequences and all environmental ITS sequences recognized above as likely to be earth tongue related, was therefore composed and aligned with Clustal W. The alignment was manually adjusted in MacClade 4.0 (Alignment S1). For the second approach, SATé returned a best-scored ITS alignment of 1,296 positions for 300 taxa (Alignment S2). A lower-scored whole ITS alignment was also examined (Alignment S3). The anchored alignment had 1,094 positions for 296 taxa, with four taxa without ITS1 regions excluded from the whole ITS data set (Alignment S4). The 5.8S re-included ITS alignment featured 1,544 positions for 296 taxa (Alignment S5). Some 24 hours were needed to run the whole ITS data set for a 20 iterations in SATé on a regular one-core desktop PC. While the best alignment scores were generally reached before the tenth step in each SATé run, the overall best scored alignment was identified only after many more iterations on a 8-core cluster (Table S2).

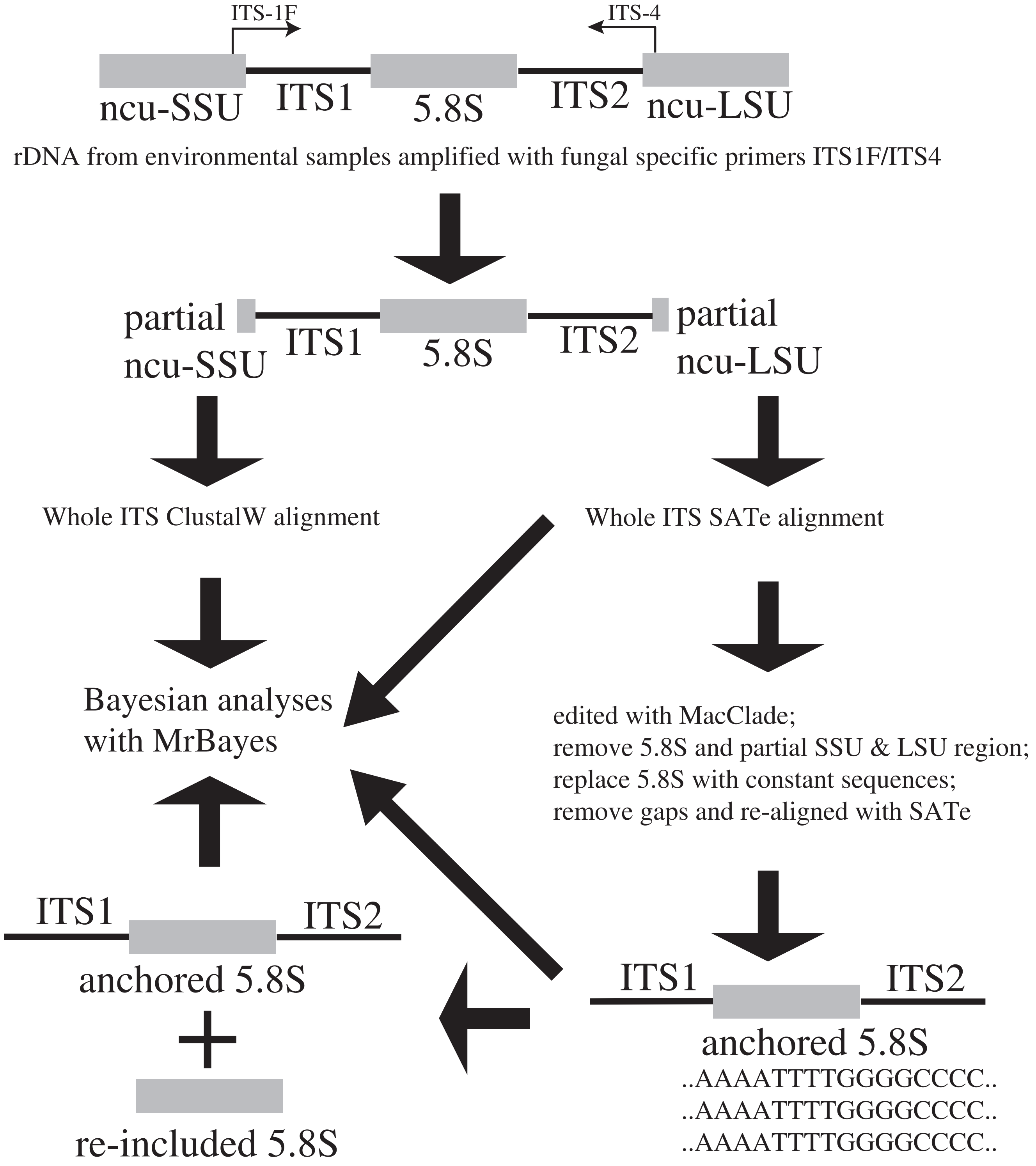

ITS-1F
ITS-4
ncu-SSU
ITS1
5.8S
ITS2
ncu-LSU
rDNA from environmental samples amplified with fungal specific primers ITS1F/ITS4
partial ncu-SSU
ITS1
5.8S
ITS2
partial ncu-LSU
Whole ITS ClustalW alignment
Whole ITS SATe alignment
Bayesian analyses with MrBayes
edited with MacClade;
remove 5.8S and partial SSU & LSU region;
replace 5.8S with constant sequences;
remove gaps and re-aligned with SATe
ITS1
ITS2
anchored 5.8S
+
re-included 5.8S
ITS1
ITS2
anchored 5.8S
..AAAATTTTGGGGCCCC..
..AAAATTTTGGGGCCCC..
..AAAATTTTGGGGCCCC..

Figure 1. Flowchart depicting how the four different alignments were prepared using Clustal W and SATé, and how they were analyzed using MrBayes.
doi:10.1371/journal.pone.0019039.g001

Phylogenetic inference based on the Clustal W whole ITS alignment

The whole ITS alignment generated by Clustal W for the reduced data set featured 118 taxa and 1,553 aligned positions, including 328 constant characters, 186 parsimony-uninformative changes, and 1,039 (67%) parsimony-informative positions. No further manual adjustment was made to this alignment. Bayesian analyses using this alignment produced a phylogeny with fairly high support for both the backbone nodes and the internal branches (Fig. 2). The Geoglossomycetes shared a clade

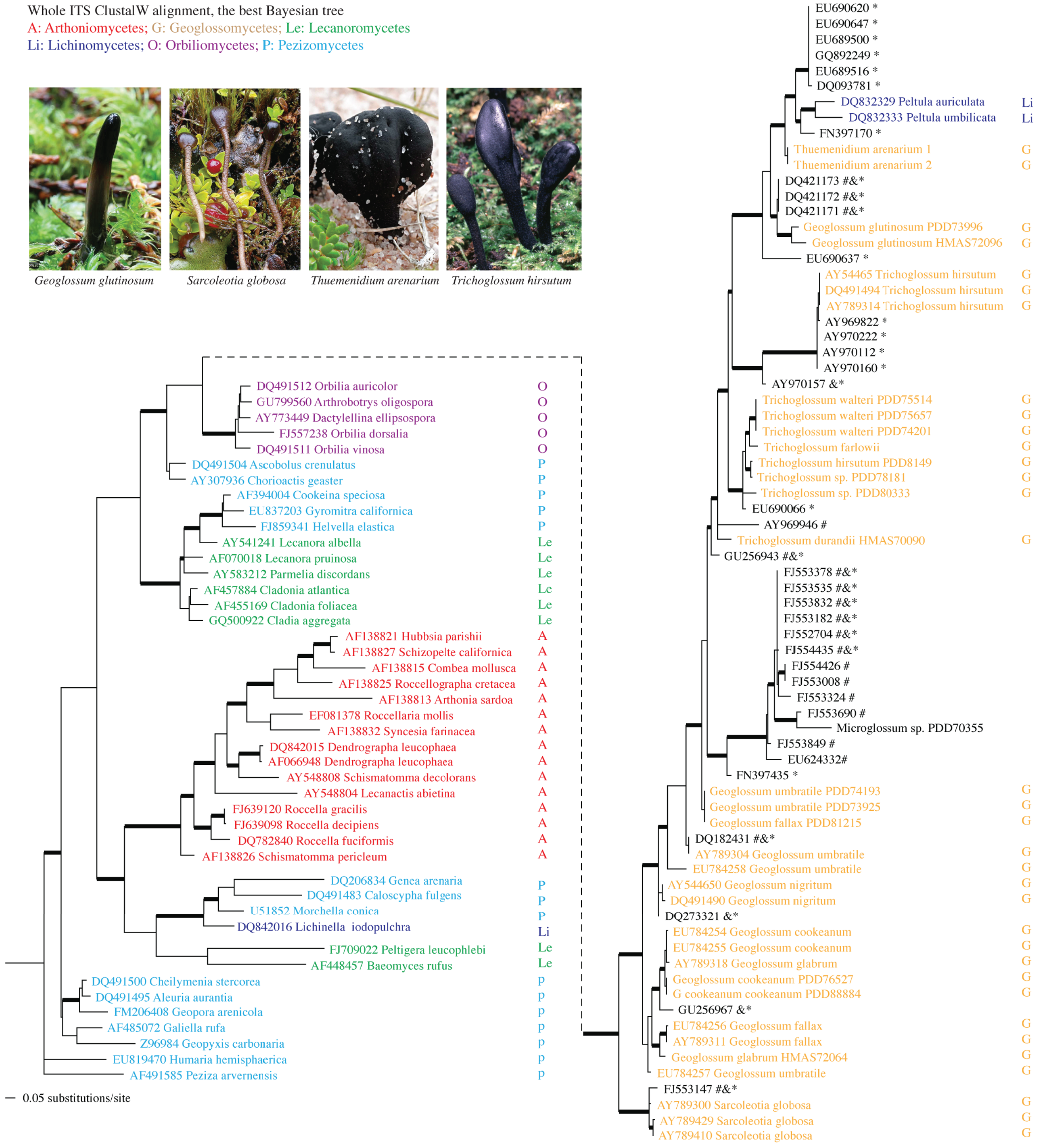

Whole ITS ClustalW alignment, the best Bayesian tree
A: Arthoniomycetes; G: Geoglossomycetes; Le: Lecanoromycetes
Li: Lichinomycetes; O: Orbiliomycetes; P: Pezizomycetes
Geoglossum glutinosum
Sarcoleotia globosa
Thuemenidium arenarium
Trichoglossum hirsutum
DQ491512 Orbilia auricolor O
GU799560 Arthrobotrys oligospora O
AY773449 Dactylellina ellipsospora O
FJ557238 Orbilia dorsalia O
DQ491511 Orbilia vinosa O
DQ491504 Ascobolus crenulatus P
AY307936 Chorioactis geaster P
AF394004 Cookeina speciosa P
EU837203 Gyromitra californica P
FJ859341 Helvella elastica P
AY541241 Lecanora albella Le
AF070018 Lecanora pruinosa Le
AY583212 Parmelia discordans Le
AF457884 Cladonia atlantica Le
AF455169 Cladonia foliacea Le
GQ500922 Cladia aggregata Le
AF138821 Hubbsia parishii A
AF138827 Schizopelte californica A
AF138815 Combea mollusca A
AF138825 Roccellographa cretacea A
AF138813 Arthonia sardoa A
EF081378 Roccellaria mollis A
AF138832 Syncesia farinacea A
DQ842015 Dendrographa leucophaea A
AF066948 Dendrographa leucophaea A
AY548808 Schismatomma decolorans A
AY548804 Lecanactis abietina A
FJ639120 Roccella gracilis A
FJ639098 Roccella decipiens A
DQ782840 Roccella fuciformis A
AF138826 Schismatomma pericleum A
DQ206834 Genea arenaria P
DQ491483 Caloscypha fulgens P
U51852 Morchella conica P
DQ842016 Lichinella iodopulchra Li
FJ709022 Peltigera leucophlebi Le
AF448457 Baeomyces rufus Le
DQ491500 Cheilymenia stercorea p
DQ491495 Aleuria aurantia p
FM206408 Geopora arenicola p
AF485072 Galiella rufa p
Z96984 Geopyxis carbonaria p
EU819470 Humaria hemisphaerica p
AF491585 Peziza arvernensis p
0.05 substitutions/site
EU690620 *
EU690647 *
EU689500 *
GQ892249 *
EU689516 *
DQ093781 *
DQ832329 Peltula auriculata Li
DQ832333 Peltula umbilicata Li
FN397170 *
Thuemenidium arenarium 1 G
Thuemenidium arenarium 2 G
DQ421173 #&*
DQ421172 #&*
DQ421171 #&*
Geoglossum glutinosum PDD73996 G
Geoglossum glutinosum HMAS72096 G
EU690637 *
AY54465 Trichoglossum hirsutum G
DQ491494 Trichoglossum hirsutum G
AY789314 Trichoglossum hirsutum G
AY969822 *
AY970222 *
AY970112 *
AY970160 *
AY970157 &*
Trichoglossum walteri PDD75514 G
Trichoglossum walteri PDD75657 G
Trichoglossum walteri PDD74201 G
Trichoglossum farlowii G
Trichoglossum hirsutum PDD8149 G
Trichoglossum sp. PDD78181 G
Trichoglossum sp. PDD80333 G
EU690066 *
AY969946 #
Trichoglossum durandii HMAS70090 G
GU256943 #&*
FJ553378 #&*
FJ553535 #&*
FJ553832 #&*
FJ553182 #&*
FJ552704 #&*
FJ554435 #&*
FJ554426 #
FJ553008 #
FJ553324 #
FJ553690 #
Microglossum sp. PDD70355
FJ553849 #
EU624332#
FN397435 *
Geoglossum umbratile PDD74193 G
Geoglossum umbratile PDD73925 G
Geoglossum fallax PDD81215 G
DQ182431 #&*
AY789304 Geoglossum umbratile G
EU784258 Geoglossum umbratile G
AY544650 Geoglossum nigritum G
DQ491490 Geoglossum nigritum G
DQ273321 &*
EU784254 Geoglossum cookeanum G
EU784255 Geoglossum cookeanum G
AY789318 Geoglossum glabrum G
Geoglossum cookeanum PDD76527 G
G cookeanum cookeanum PDD88884 G
GU256967 &*
EU784256 Geoglossum fallax G
AY789311 Geoglossum fallax G
Geoglossum glabrum HMAS72064 G
EU784257 Geoglossum umbratile G
FJ553147 #&*
AY789300 Sarcoleotia globosa G
AY789429 Sarcoleotia globosa G
AY789410 Sarcoleotia globosa G

Figure 2. The best tree generated by MrBayes for the whole ITS Clustal W alignment. Bayesian posterior probabilities higher than 0.94 are indicated with bold branches. Taxonomic groups were coded with colors and initials of the classes. Examples of ascoma morphology and habits of the earth tongues; photo credits and courtesy of: Clive Shirley (*Geoglossum glutinosum*), Thomas Læssøe (*Sarcoleotia globosa*), Andrus Votik (*Thuemenidium arenarium*), and Mark Steinmetz (*Trichoglossum hirsutum*). Environmental ITS sequences grouped in the earth tongue clades are marked with one or more annotations indicating whether the isolate would be designated an earth tongue by other approaches. Affinity of marked isolates to the earth tongues was suggested by: * the sequence search suite *emerencia*; & BLAST of known earth-tongue ITS sequence against INSD; and by # annotation in INSD as unclassified Pezizomycotina environmental samples. Newly generated sequences for this study are presented as taxa without INSD accession numbers while download sequences are listed with their INSD accession numbers.

doi:10.1371/journal.pone.0019039.g002

(BPP = 0.99) with two species of *Peltula* and with *Microglossum* (PDD70355), and within this clade, 11 sequences of environmental samples were grouped in five subclades (BPP >0.95) with different *Geoglossum*, *Sarcoleotia*, and *Trichoglossum* species. The reference lineages in Pezizomycetes, Lecanoromycetes, and Lichinomycetes were also strongly supported. The placements and sub-groupings of these clades were not consistent with previously published phylogenies using other gene sequences, however.

Phylogenetic inference based on the SATé whole ITS alignment

The best-scored whole ITS alignment generated by SATé featured 300 taxa and 1,296 aligned positions, including 191 constant characters, 210 parsimony-uninformative changes, and 895 (69%) parsimony-informative positions. Bayesian analyses using this data set without manual adjustment produced a phylogeny similar to the automatic tree simultaneously generated by SATé with some few internal branches being well supported (Fig. 3). The Geoglossomycetes was monophyletic without strong support (BPP = 0.51) with a clade of two *Peltula* species as the sister group. A total of 39 environmental ITS sequences were assigned to this clade, and 23 sequences were grouped in four subclades (BPP >0.99) with fully identified species of *Geoglossum*, *Trichoglossum*, *Thuemendidium*, and *Sarcoleotia.* All species of *Geoglossum*, except for *G. glutionosum* which is likely to be transferred to *Trichoglossum* based on other gene phylogenies, were moderately supported as a joint monophyletic lineage (BPP = 0.91). Nine environmental isolates on very long branches were retrieved within the earth tongue clade. Among the reference groups, the Orbiliomycetes was strongly supported as monophyletic with some environmental samples (BPP = 0.99), while the Arthoniomycetes and the Lecanoromycetes were not monophyletic. The overall topology of the best Bayesian tree based on the lower-scored SATé alignment was quite different from the one produced by the best-scored alignment, and all of the reference lineages, except for the Orbiliomycetes, were scattered over more than one clade each (Figure S1).

Phylogenetic inference based on the SATé anchored ITS alignment

The anchored ITS alignment of the data set using SATé featured 296 taxa and 1,094 aligned positions, including 140 constant characters, 75 parsimony-uninformative changes, and 879 (80%) parsimony-informative positions (Fig. 4). While many recent nodes were strongly supported in the best Bayesian tree from the anchored ITS alignment, few internodes along the backbone were supported, and only the Orbiliomycetes in the reference groups was recovered as monophyletic. 17 environmental isolates were grouped in the *Geoglossum glutinosum – Trichoglossum – Thuemenidium* clade (PBB = 0.95), three isolates shared a clade with *Sarcoleotia* species (BP = 1.00), and eight isolates were grouped with other *Geoglossum* species (BP = 1.00). Regarding other groups with reference taxa, a clade including Orbiliomycetes and six environmental isolates was strongly supported (BPP = 1.0), while other reference groups were not monophyletic.

Phylogenetic inference based on the SATé 5.8S re-included ITS alignment

Bayesian analysis using the 5.8S re-included ITS alignment of the large data set without manual adjustment produced a phylogeny with moderate to strong support for many internodes (Fig. 5). The whole data set included 296 taxa with 1,544 aligned but highly gapped positions, including 252 constant characters, 263 parsimony-uninformative changes, and 1,029 (67%) parsimony-informative positions. Although the earth tongue clade was not resolved as monophyletic, there was strong support (BPP>0.95) for subclades of the earth tongues: the *Sarcoleotia* clade, the *Thuemenidium-Trichoglossum-Geoglossum glutinosum* clade, and the *Geoglossum* s. str. clade. A total of 30 environmental isolates grouped in the earth tongue clade, and these isolates shared clades with identified species of *Sarcoleotia*, *Thuemenidium*, *Trichoglossum*, or *Geoglossum* with strong support (BPP>0.95). A clade including species of Orbiliomycetes and related environmental samples and its subclades received strong support (BPP >0.95) as well, while other reference groups the Arthoniomycetes, the Lecanoromycetes, and the Pezizomycetes were not recovered as monophyletic.

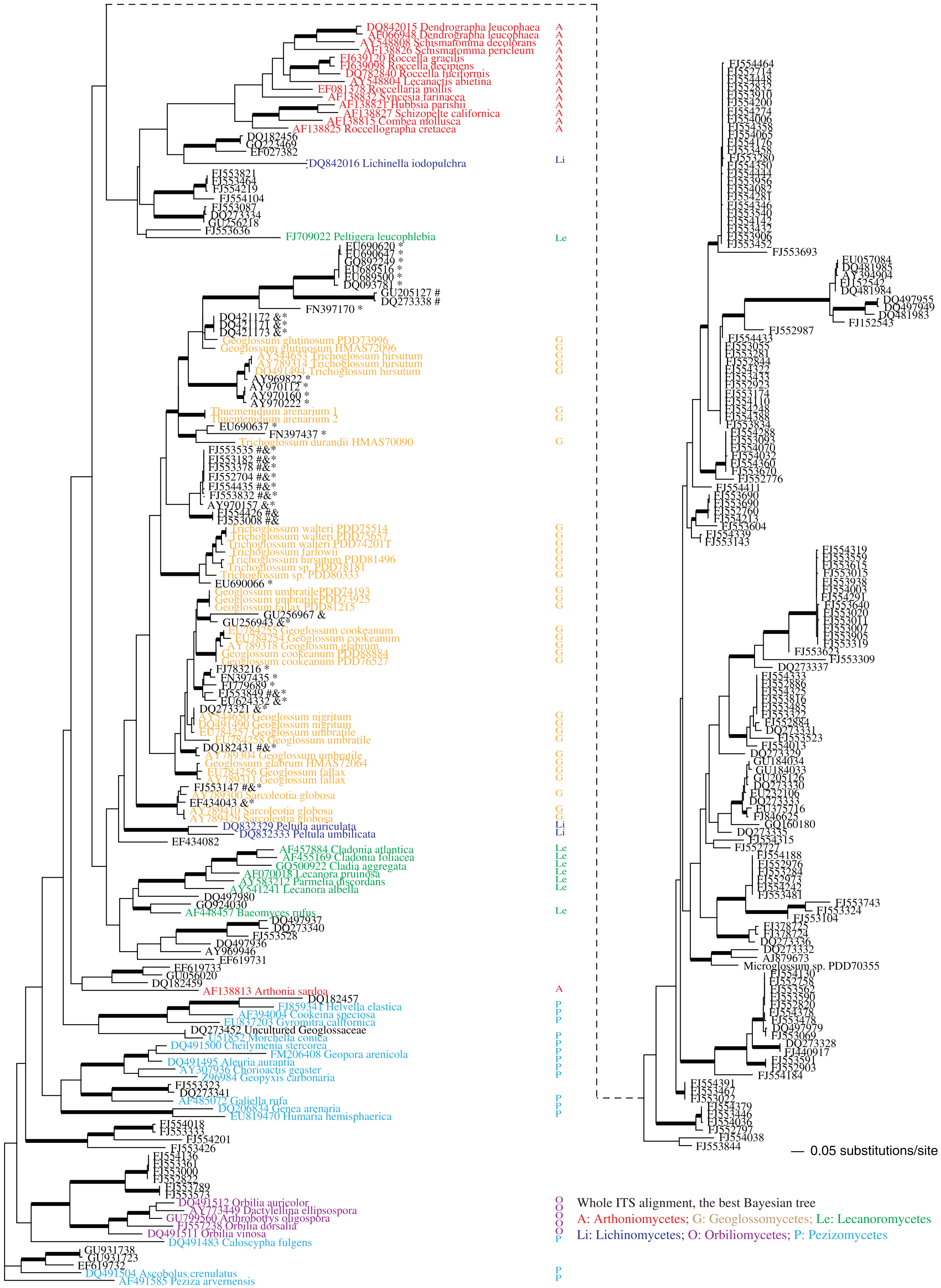
Comparison of the SATé approaches

Both the best-scored whole ITS SATé alignment and the 5.8S re-included SATé alignment produced phylogenetic trees with the highest topological similarity to published multi-gene phylogenies. The phylogenetic tree based on the 5.8S re-included SATé alignment had the largest proportion of well-supported branches and the fewest extremely long branches, although the alignment was longer than for the two above approaches. The lower scored whole ITS SATé alignment produced the most dissimilar trees, and the corresponding multiple alignment was found to be very suboptimal. With the highly conserved 5.8S region excluded, the phylogeny based on the anchored ITS SATé alignment was found to be intermediate, with several reference sequences being correctly grouped.

Discussion

The difficulties associated with aligning the ITS region across larger fungal lineages has hitherto prevented wide use of ITS data for high-level phylogeny as well as preventing use of ITS data for inference of phylogenetic structure in environmental samples. Our alignments validated these difficulties in that we found ITS sequences to be highly divergent across a wide taxon sampling, with long alignments and comparatively high percentages of parsimony-informative positions. Even so, the performance of SATé in generating large ITS alignments for an accurate estimation of a higher-level phylogeny was surprisingly good, especially since the relationships between the Pezizomycetes and the Orbiliomycetes, and among the earth tongues and the three lichenized groups, have been problematic even with multigene phylogenies (*e.g.* [25] [29–30]). Indeed, the anchored ITS SATé alignment suggests that that some degree of phylogenetic information beyond the generic level does exist in the ITS region, since the major clades recognized with multilocus data in the Ascomycota were generally also recovered in the anchored ITS phylogenies. However, no significant support for deep nodes was achieved, presumably due to large contribution of phylogenetic noise from rapidly-envolving ITS regions.

We have shown, with earth tongue fungi and related fungal classes as examples, that a progressive alignment program (SATé) outperformed traditional Clustal W in conveniently, rapidly, and accurately aligning large ITS data sets. However, a careful handling of 5.8S rDNA and residual portions of the 18S and 25S genes often found in ITS sequences is required as a few highly saturated characters can be misleading for inference of events in the initial tree [31]. In this study, the overall best scored SATé alignment of the whole ITS data set was reached after more than 15 iterations, and alignments with lower scores would produce different phylogenies. However, it is unfortunate that there is no parameter for setting the total running time for specific types of sequence data in SATé. For now, we suggest that many relevant


DQ842015 Dendrographa leucophaea A
AF066948 Dendrographa leucophaea A
AY548808 Schismatomma decolorans A
AF138826 Schismatomma pericleum A
FJ639120 Roccella gracilis A
FJ639098 Roccella decipiens A
DQ782840 Roccella fuciformis A
AY548804 Lecanactis abietina A
EF081378 Roccellaria mollis A
AF138832 Syncesia farinacea A
AF138821 Hubbsia parishii A
AF138827 Schizopelte californica A
AF138815 Combea mollusca A
AF138825 Roccellographa cretacea A
DQ182456
GQ223469
EF027382
DQ842016 Lichinella iodopulchra Li
FJ553821
FJ553464
FJ554219
FJ554104
FJ553087
DQ273334
GU256218
FJ553636
FJ709022 Peltigera leucophlebia Le
EU690620 *
EU690647 *
GQ892249 *
EU689516 *
EU689500 *
DQ093781 *
GU205127 #
DQ273338 #
FN397170 *
DQ421172 &*
DQ421171 &*
DQ421173 &*
Geoglossum glutinosum PDD73996 G
Geoglossum glutinosum HMAS72096 G
AY544653 Trichoglossum hirsutum G
AY789314 Trichoglossum hirsutum G
DQ491494 Trichoglossum hirsutum G
AY969822 *
AY970112 *
AY970160 *
AY970222 *
Thuemenidium arenarium 1 G
Thuemenidium arenarium 2 G
EU690637 *
FN397437 *
Trichoglossum durandii HMAS70090 G
FJ553535 #&*
FJ553182 #&*
FJ553378 #&*
FJ552704 #&*
FJ554435 #&*
FJ553832 #&*
AY970157 &*
FJ554426 #&
FJ553008 #&
Trichoglossum walteri PDD75514 G
Trichoglossum walteri PDD75657 G
Trichoglossum walteri PDD74201T G
Trichoglossum farlowii G
Trichoglossum hirsutum PDD81496 G
Trichoglossum sp. PDD78181 G
Trichoglossum sp. PDD80333 G
EU690066 *
Geoglossum umbratilePDD74193 G
Geoglossum umbratilePDD73925 G
Geoglossum fallax PDD81215 G
GU256967 &
GU256943 &*
EU784255 Geoglossum cookeanum G
EU784254 Geoglossum cookeanum G
AY789318 Geoglossum glabrum G
Geoglossum cookeanum PDD88884 G
Geoglossum cookeanum PDD76527 G
FJ783216 *
FN397435 *
FJ779689 *
FJ553849 #&*
EU622332 &*
DQ273321 &*
AY544650 Geoglossum nigritum G
DQ491490 Geoglossum nigritum G
EU784257 Geoglossum umbratile G
EU784258 Geoglossum umbratile G
DQ182431 #&*
AY789304 Geoglossum umbratile G
Geoglossum glabrum HMAS72064 G
EU784256 Geoglossum fallax G
AY789311 Geoglossum fallax G
FJ553147 #&*
AY789300 Sarcoleotia globosa G
EF434043 &*
AY789410 Sarcoleotia globosa G
AY789429 Sarcoleotia globosa G
DQ832329 Peltula auriculata Li
DQ832333 Peltula umbilicata Li
EF434082
AF457884 Cladonia atlantica Le
AF455169 Cladonia foliacea Le
GQ500922 Cladia aggregata Le
AF070018 Lecanora pruinosa Le
AY583212 Parmelia discordans Le
AY541241 Lecanora albella Le
DQ497980
GQ924030
AF448457 Baeomyces rufus Le
DQ497937
DQ273340
FJ553528
DQ497936
AY969946
EF619731
EF619733
GU056020
DQ182459
AF138813 Arthonia sardoa A
DQ182457
FJ859341 Helvella elastica P
AF394004 Cookeina speciosa P
EU837203 Gyromitra californica P
DQ273452 Uncultured Geoglossaceae
U51852 Morchella conica P
DQ491500 Cheilymenia stercorea P
FM206408 Geopora arenicola P
DQ491495 Aleuria aurantia P
AY307936 Chorioactis geaster P
Z96984 Geopyxis carbonaria P
FJ553323
DQ273341
AF485072 Galiella rufa P
DQ206834 Genea arenaria P
EU819470 Humaria hemisphaerica P
FJ554018
FJ553333
FJ554201
FJ553426
FJ554136
FJ553361
FJ553000
FJ552822
FJ553789
FJ553573
DQ491512 Orbilia auricolor O
AY773449 Dactylellina ellipsospora O
GU799560 Arthrobotrys oligospora O
FJ557238 Orbilia dorsalia O
DQ491511 Orbilia vinosa O
DQ491483 Caloscypha fulgens P
GU931738
GU931723
EF619732
DQ491504 Ascobolus crenulatus P
AF491585 Peziza arvernensis P
FJ554464
FJ552714
FJ554448
FJ553532
FJ553910
FJ554200
FJ554274
FJ554006
FJ554358
FJ554065
FJ554176
FJ553458
FJ553280
FJ554350
FJ554444
FJ553956
FJ554082
FJ554281
FJ554346
FJ553540
FJ554142
FJ553432
FJ553906
FJ553452
FJ553693
EU057084
DQ481985
AY394904
FJ152542
DQ481984
DQ497955
DQ497949
DQ481983
FJ152543
FJ552987
FJ554433
FJ553055
FJ553281
FJ552844
FJ554322
FJ553433
FJ552923
FJ553174
FJ554110
FJ554248
FJ554388
FJ553834
FJ554288
FJ553093
FJ554070
FJ554032
FJ554360
FJ553670
FJ552776
FJ554411
FJ553690
FJ553690
FJ552760
FJ554213
FJ553604
FJ554339
FJ553143
FJ554319
FJ553559
FJ553615
FJ553015
FJ553938
FJ554003
FJ554291
FJ553640
FJ553020
FJ553011
FJ553007
FJ553905
FJ553319
FJ553623
FJ553309
DQ273337
FJ554333
FJ552886
FJ554325
FJ553816
FJ553485
FJ553322
FJ552884
DQ273331
FJ553523
FJ554013
DQ273329
GU184034
GU184033
GU205126
DQ273330
EU232106
DQ273333
EU375716
FJ846625
GQ160180
DQ273335
FJ554315
FJ552727
FJ554188
FJ552976
FJ553284
FJ552973
FJ554242
FJ553481
FJ553743
FJ553324
FJ553104
FJ378725
FJ378724
DQ273336
DQ273332
AJ879673
Microglossum sp. PDD70355
FJ554130
FJ552758
FJ553562
FJ553590
FJ552820
FJ554378
FJ553478
DQ497979
FJ553069
DQ273328
FJ440917
FJ553591
FJ552903
FJ554184
FJ554391
FJ553467
FJ553022
FJ554379
FJ553446
FJ554036
FJ552797
FJ554038
FJ553844
0.05 substitutions/site

Whole ITS alignment, the best Bayesian tree
A: Arthoniomycetes; G: Geoglossomycetes; Le: Lecanoromycetes
Li: Lichinomycetes; O: Orbiliomycetes; P: Pezizomycetes

Figure 3. The best tree generated by MrBayes for the best-scored whole ITS SATé alignment. Bayesian posterior probabilities higher than 0.94 are indicated with bold branches. Taxonomic groups were coded with colors and initials of the classes. Environmental ITS sequences grouped in the earth tongue clades are marked with one or more annotations indicating whether the isolate would be designated an earth tongue by other approaches. Affinity of marked isolates to the earth tongues was suggested by: * the sequence search suite *emerencia*; & BLAST of known earth-tongue ITS sequence against INSD; and by # annotation in INSD as unclassified Pezizomycotina environmental samples. Newly generated sequences for this study are presented as taxa without INSD accession numbers while download sequences are listed with their INSD accession numbers.
doi:10.1371/journal.pone.0019039.g003

reference groups whose relations are well-known in addition to performing multiple SATé runs to reach the best alignment. Although it is nearly impossible to manually adjust an alignment of over hundreds divergent ITS sequences, there are segments of appreciably misaligned data in highly conserved regions in even the best-scored SATé alignment.

As SATé uses an initial guide tree to assist the alignment and tree building processes, the highly conserved 5.8S gene of the ITS region may potentially mislead alignment in other ITS regions that show lower levels of similarity [32]. In this study, we evaluated the SATé alignments using Bayesian inference of phylogeny, and we found that excluding 5.8S for the SATé run, then re-included it after, produced the least problematic (the fewest extremely long branches), the most accurate (based on the reference groups), well-resolved, and strongly supported phylogenies at all different taxonomic levels. An expanded taxon sampling would improve the phylogenetic inferences of SATé further, and with much wider and intensive sampling, the undue influence from 5.8S rDNA and other highly conserved regions is likely to be diluted.

Importantly, in this study, we demonstrated that including environmental sequences in our taxon sampling helped us to understand better about the diversity of the Geoglossomycetes. Given that environmental sequencing is becoming cheaper and easier than ever, the finding that data mining provides additional insight are relevant not only to earth tongues but to all fungal lineages. Nevertheless, while environmental sampling will extend our taxon sampling of fungi by no small measure, few of these new lineages will be matched by fully identified sequences in INSD. Brock et al. [12] showed that sequencing efforts targeted at high-quality specimens in herbaria and type collections around the world represent by far the best way to leverage extant taxonomical knowledge to inform environmental sampling and vice versa. This is consistent with our findings in this study, since a phylogenetic analysis using only sequences that lack Latin names would be difficult to root in any general standard of measure. Regarding the sampling of environmental ITS sequences from INSD to explore the diversity of certain fungal groups, we found that the genus-specific search function of *emerencia* provided the most inclusive list of relevant candidate sequences. Emerencia is very useful in searching for closely related sequences for a query sequence or genus, phylogenetic analyses, however, are vital to confirmation of the affinities of these environmental isolates with respect to known fungal groups.

Regarding the diversity of earth tongues in environmental samples, the 30 unidentified sequences that grouped with known earth tongue species in the 5.8S re-included ITS alignment were found to stem from nine scientific studies, including three unpublished ones [14] [33] [34] [35] [36] [37]. A total of 10 sequences representing four lineages that were closely related to the four genera in the earth tongue family were from British Columbia, Canada [36], and molecular isolates from this study (INSD accessions FJ550631 through FJ554464) formed the majority of uncultured Pezizomycotina entries in INSD. Although earth tongues with about 50 described species are not rare in temperate forests, we were surprised to find such a high diversity from a single sample site. Seaver [38] described 16 species of *Geoglossum* and *Trichoglossum* from North America, but recent collections were not available for this study other than five common species and some more representatives from New Zealand in the Southern Hemisphere. The second hotspot of environmental earth tongue diversity was the Duke Forest, Durham, North Carolina sampled by O'Brien et al. [14]. Six ITS sequences from this study (AY96936 through AY970290) were grouped with common North American species of *Trichoglossum* and *Geoglossum* separately. Interestingly, two ITS isolates (GU256967 and GU256943) from hair root samples of *Rhododendron* in Sichuan, China, were in a close relationship with *Geoglossum* species. The region southwest of China hosts many earth tongues species [23], and more data from earth tongues from that region would be interesting.

A close biological association between earth tongues and lower plants has long been suspected (e.g. [39]), and in this study we inferred four cases of mycorrhizal association with different type of plants based on unpublished data provided in INSD accessions. These included hair root samples of *Rhododendron* found in China, an isolate from roots of tanoak in northern California (DQ273321), root samples of oak in a tropical cloud forest (EU624332), and a terrestrial orchid *Cephalanthera longifolia* (DQ182431). All these putative associations with plants were found in the clade including fully identified species of the genus *Geoglossum*. Although the *Geoglossum* clade was strongly supported in the trees based on the 5.8S anchored and 5.8S re-included ITS alignments (Fig. 4), it was not well supported in the trees based on other alignments. Of course, BLAST searches and analysis of local alignments would also be useful for linking identified taxa to interesting environmental sequences. However, implementation of BLAST and local alignment approaches would be impossible for analyses of thousands of sequences generated from environmental studies, because knowledge of community structures of the samples is generally lacking. To gather more data on the ecology and geography of all described species of earth tongues along with more sequence data from environmental samples for phylogenetic analyses would constitute the next step toward a better understanding of this class of fungi.

We conclude that SATé is highly efficient at producing accurate multiple sequence alignments for large numbers of divergent ITS sequences from a broad taxon sampling in ascomycetes. With a careful processing of the conserved regions in the ITS data set, phylogenetic analyses based on SATé alignments can provide a reliable estimation of fungal diversity in environmental samples. Furthermore, including data from environmental samples in phylogenetic analyses will provide some insights about ecology and biology, which are understudied due to the shortcomings of traditional taxonomic methods.

Materials and Methods

Laboratory work

Eighteen recently collected specimens of *Geoglossum*, *Trichoglossum*, and *Thuemenidium* from New Zealand, China, and Canada were sampled for ITS sequences in this study. DNA extraction and polymerase chain reaction (PCR) were carried out following Wang et al [26–27], using the fungus-specific ITS1F [40] and the eukaryotic ITS4 primers [41] for amplification. PCR runs were

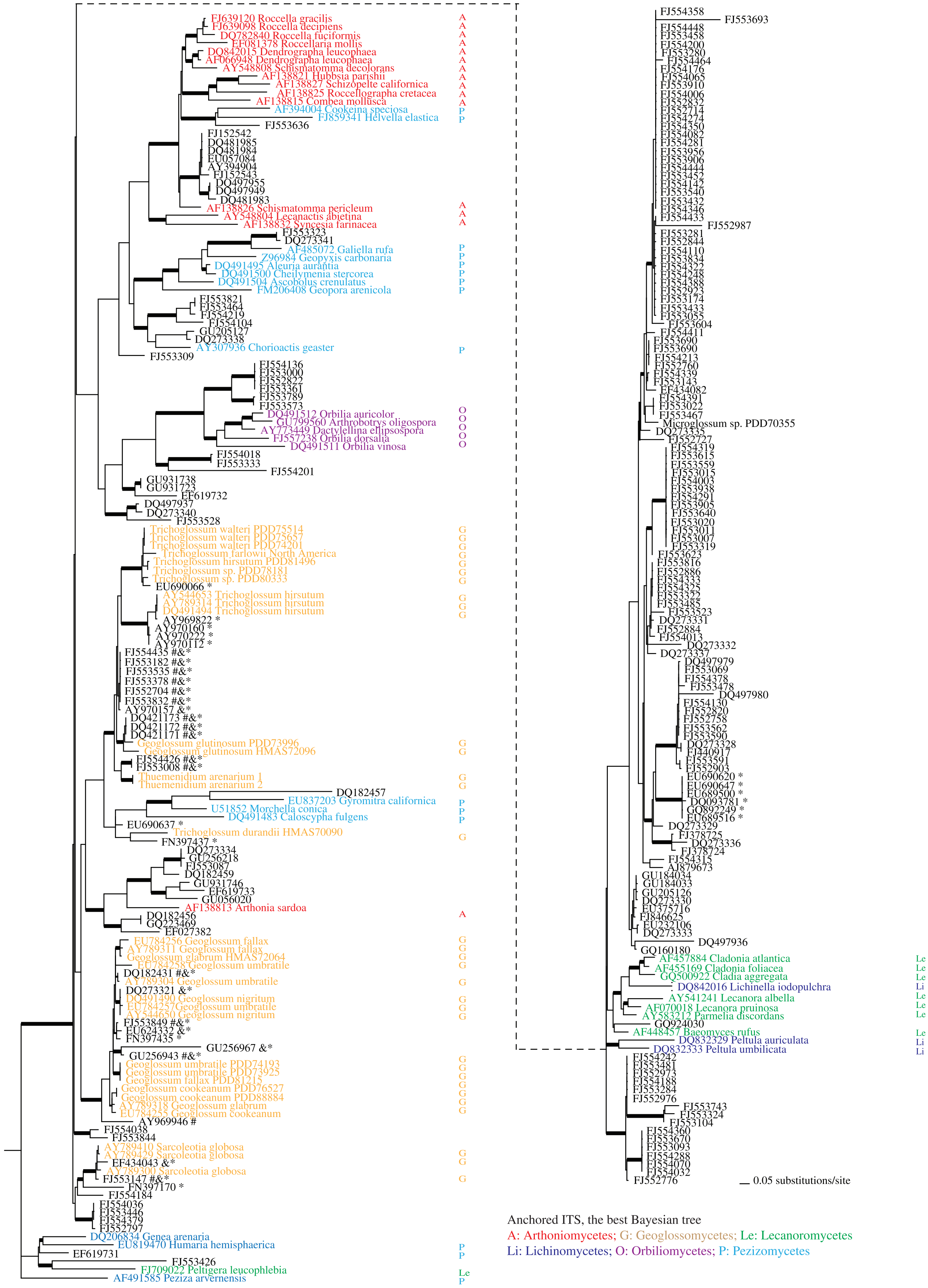

FJ639120 Roccella gracilis A
FJ639098 Roccella decipiens A
DQ782840 Roccella fuciformis A
EF081378 Roccellaria mollis A
DQ842015 Dendrographa leucophaea A
AF066948 Dendrographa leucophaea A
AY548808 Schismatomma decolorans A
AF138821 Hubbsia parishii A
AF138827 Schizopelte californica A
AF138825 Roccellographa cretacea A
AF138815 Combea mollusca A
AF394004 Cookeina speciosa P
FJ859341 Helvella elastica P
FJ553636
FJ152542
DQ481985
DQ481984
EU057084
AY394904
FJ152543
DQ497955
DQ497949
DQ481983
AF138826 Schismatomma pericleum A
AY548804 Lecanactis abietina A
AF138832 Syncesia farinacea A
FJ553323
DQ273341
AF485072 Galiella rufa P
Z96984 Geopyxis carbonaria P
DQ491495 Aleuria aurantia P
DQ491500 Cheilymenia stercorea P
DQ491504 Ascobolus crenulatus P
FM206408 Geopora arenicola P
FJ553821
FJ553464
FJ554219
FJ554104
GU205127
DQ273338
AY307936 Chorioactis geaster P
FJ553309
FJ554136
FJ553000
FJ552822
FJ553361
FJ553789
FJ553573
DQ491512 Orbilia auricolor O
GU799560 Arthrobotrys oligospora O
AY773449 Dactylellina ellipsospora O
FJ557238 Orbilia dorsalia O
DQ491511 Orbilia vinosa O
FJ554018
FJ553333
FJ554201
GU931738
GU931723
EF619732
DQ497937
DQ273340
FJ553528
Trichoglossum walteri PDD75514 G
Trichoglossum walteri PDD75657 G
Trichoglossum walteri PDD74201 G
Trichoglossum farlowii North America G
Trichoglossum hirsutum PDD81496 G
Trichoglossum sp. PDD78181 G
Trichoglossum sp. PDD80333 G
EU690066 *
AY544653 Trichoglossum hirsutum G
AY789314 Trichoglossum hirsutum G
DQ491494 Trichoglossum hirsutum G
AY969822 *
AY970160 *
AY970222 *
AY970112 *
FJ554435 #&*
FJ553182 #&*
FJ553535 #&*
FJ553378 #&*
FJ552704 #&*
FJ553832 #&*
AY970157 &*
DQ421173 #&*
DQ421172 #&*
DQ421171 #&*
Geoglossum glutinosum PDD73996 G
Geoglossum glutinosum HMAS72096 G
FJ554426 #&*
FJ553008 #&*
Thuemenidium arenarium 1 G
Thuemenidium arenarium 2 G
DQ182457
EU837203 Gyromitra californica P
U51852 Morchella conica P
DQ491483 Caloscypha fulgens P
EU690637 *
Trichoglossum durandii HMAS70090 G
FN397437 *
DQ273334
GU256218
FJ553087
DQ182459
GU931746
EF619733
GU056020
AF138813 Arthonia sardoa A
DQ182456
GQ223469
EF027382
EU784256 Geoglossum fallax G
AY789311 Geoglossum fallax G
Geoglossum glabrum HMAS72064 G
EU784258 Geoglossum umbratile G
DQ182431 #&*
AY789304 Geoglossum umbratile G
DQ273321 &*
DQ491490 Geoglossum nigritum G
EU784257Geoglossum umbratile G
AY544650 Geoglossum nigritum G
FJ553849 #&*
EU624332 &*
FN397435 *
GU256967 &*
GU256943 #&*
Geoglossum umbratile PDD74193 G
Geoglossum umbratile PDD73925 G
Geoglossum fallax PDD81215 G
Geoglossum cookeanum PDD76527 G
Geoglossum cookeanum PDD88884 G
AY789318 Geoglossum glabrum G
EU784255 Geoglossum cookeanum G
AY969946 #
FJ554038
FJ553844
AY789410 Sarcoleotia globosa G
AY789429 Sarcoleotia globosa G
EF434043 &*
AY789300 Sarcoleotia globosa G
FJ553147 #&*
FN397170 *
FJ554184
FJ554036
FJ553446
FJ554379
FJ552797
DQ206834 Genea arenaria
EU819470 Humaria hemisphaerica P
EF619731 P
FJ553426
FJ709022 Peltigera leucophlebia Le
AF491585 Peziza arvernensis P
FJ554358
FJ553693
FJ554448
FJ553458
FJ554200
FJ553280
FJ554464
FJ554176
FJ554065
FJ553910
FJ554006
FJ552832
FJ552714
FJ554274
FJ554350
FJ554082
FJ554281
FJ553956
FJ553906
FJ554444
FJ553452
FJ554142
FJ553540
FJ553432
FJ554346
FJ554433
FJ552987
FJ553281
FJ552844
FJ554110
FJ553834
FJ554322
FJ554248
FJ554388
FJ552923
FJ553174
FJ553433
FJ553055
FJ553604
FJ554411
FJ553690
FJ553690
FJ554213
FJ552760
FJ554339
FJ553143
EF434082
FJ554391
FJ553022
FJ553467
Microglossum sp. PDD70355
DQ273335
FJ552727
FJ554319
FJ553615
FJ553559
FJ553015
FJ554003
FJ553938
FJ554291
FJ553905
FJ553640
FJ553020
FJ553011
FJ553007
FJ553319
FJ553623
FJ553816
FJ552886
FJ554333
FJ554325
FJ553322
FJ553485
FJ553523
DQ273331
FJ552884
FJ554013
DQ273332
DQ273337
DQ497979
FJ553069
FJ554378
FJ553478
DQ497980
FJ554130
FJ552820
FJ552758
FJ553562
FJ553590
DQ273328
FJ440917
FJ553591
FJ552903
EU690620 *
EU690647 *
EU689500 *
DQ093781 *
GQ892249 *
EU689516 *
DQ273329
FJ378725
DQ273336
FJ378724
FJ554315
AJ879673
GU184034
GU184033
GU205126
DQ273330
EU375716
FJ846625
EU232106
DQ273333
DQ497936
GQ160180
AF457884 Cladonia atlantica Le
AF455169 Cladonia foliacea Le
GQ500922 Cladia aggregata Le
DQ842016 Lichinella iodopulchra Li
AY541241 Lecanora albella Le
AF070018 Lecanora pruinosa Le
AY583212 Parmelia discordans Le
GQ924030
AF448457 Baeomyces rufus Le
DQ832329 Peltula auriculata Li
DQ832333 Peltula umbilicata Li
FJ554242
FJ553481
FJ552973
FJ554188
FJ553284
FJ552976
FJ553743
FJ553324
FJ553104
FJ554360
FJ553670
FJ553093
FJ554288
FJ554070
FJ554032
FJ552776
0.05 substitutions/site
Anchored ITS, the best Bayesian tree
A: Arthoniomycetes; G: Geoglossomycetes; Le: Lecanoromycetes
Li: Lichinomycetes; O: Orbiliomycetes; P: Pezizomycetes

Figure 4. The best tree generated by MrBayes for the anchored ITS SATé alignment of the large data set. Bayesian posterior probabilities higher than 0.94 are indicated with bold branches. Taxonomic groups were coded with colors and initials of the classes. Environmental ITS sequences grouped in the earth tongue clades are marked with one or more annotations indicating whether the isolate would be designated an earth tongue by other approaches. Affinity of marked isolates to the earth tongues was suggested by: * the sequence search suite *emerencia*; & BLAST of known earth-tongue ITS sequence against INSD; and by # annotation in INSD as unclassified Pezizomycotina environmental samples. Newly generated sequences for this study are presented as taxa without INSD accession numbers while download sequences are listed with their INSD accession numbers.
doi:10.1371/journal.pone.0019039.g004

performed with an initial denaturation step of 2 min at 94°C, followed by 35 cycles of 30 s at 94°C, 30 s at 53°C, 90 s at 72°C, and a final step of 10 min at 72°C. PCR products were purified with QIAquick columns (Qiagen Inc., Valencia, CA, USA), sequenced at the Yale University Center for Genomics and Proteomics, and edited in SEQUENCHER 4.5 (Gene Codes, Ann Arbor, MI, USA).

Data mining

ITS sequences of European and North American isolates of *Geoglossum*, *Trichoglossum*, and *Sarcoleotia*, which were generated from voucher specimens for the Assembling the Fungal Tree of Life (AFToL) project and other studies, were retrieved from INSD and were combined with the newly generated sequences to build up a core ITS data set for earth tongue fungi. Pairwise alignments were generated and compared among ITS sequences for similarity, which was established by dividing the number of matching nucleotides by the total length of the alignment using the online version of Needle in EMBOSS [42] [43].

To investigate the unknown diversity of earth tongue fungi in the public sequence databases, insufficiently identified sequences from environmental samples were compiled for further analyses using three strategies. First, sequence similarity searches using BLAST [44] were performed using the core ITS data set against all sequences in INSD, and matching insufficiently identified sequences were retrieved starting from the top of the BLAST results progressing until no further earth tongue genera appeared. A total of 27 ITS sequences were retrieved in this way. Second, we used the genus search function of *emerencia* [45] to identify an additional 41 environmental ITS sequences whose best BLAST match was found among fully identified sequences of earth tongue genera in INSD. Third, we downloaded all ITS sequences identified as uncultured Pezizomycotina environmental samples from INSD, assuming some unidentified earth tongues might be found among them because of the position of earth tongues in the Pezizomycotina. Sequences that were submitted to INSD as reverse complementary were identified and corrected using Nilsson et al. [46]. Fourteen chimeric sequences were identified with the Fungal ITS Chimera Checker [47] and were excluded from further analysis. Similarly, all short sequences and sequences that lacked the ITS2 subregion were excluded. For phylogenetic analyses, 46 outgroups and reference ITS sequences from Pezizomycetes (15), Orbilliomycetes (5), and three lichenized classes including Arthoniomycetes (15), Lichinomycetes (3), and Lecanoromycetes (8) were added to the data set for alignment and further analyses. The accuracy of the taxonomic annotations was established by comparing these reference sequences using BLAST in INSD to verify that the top hits were sequences from the same or closely related species in the same genus as generated by different studies. The final data set included 300 unaligned ITS sequences.

Phylogenetic inference

Four alignment strategies used Clustal W [48] and SATé [21], and MrBayes [49] were applied in this study:

- the whole-ITS Clustal W. Clustal W is used for multiple alignment and MrBayes is used for phylogenetic inference. This approach mimics the one adopted by most taxonomists, and we use it as the baseline for our comparison with the more refined approaches below.
- the whole-ITS SATé. Using the same input sequences as the first approach, this alignment strategy uses SATé for joint estimation of alignment and phylogenetic trees without any prior processing of the alignment.
- the anchored ITS SATé. Using the same input sequences as the two first approaches, but with the 5.8S gene sequence removed, this approach is designed to examine the impact of the very conserved 5.8S gene sequence on the alignment and phylogenetic estimations of SATé.
- the 5.8S re-included SATé alignments. This approach relies on the final alignment of the anchored ITS SATé, but with the 5.8S gene sequence re-inserted for the phylogenetic inference. This alignment facilitates examination of the impact of the 5.8S gene sequence upon the alignment stage alone (Fig. 1).

To build these four alignments, we first used the ITS sequences of full length, including any fringe residues of 18S and 28S, to construct "whole ITS" alignments with both Clustal W and SATé. For the whole ITS analyses, the alignments created by Clustal W and SATé were used for further analyses without manual adjustment except for some trimming of highly gapped ends where necessary. SATé alignments were performed with 6 runs of 20 iterations under default settings using MAFFT for alignment and RAxML for phylogenetic inference *sans* initial guiding tree on single, double, and four-core PCs and Macs separately. One more SATé run for each data set was employed on an 8-core cluster with a 24 hour limitation as well (Table S2).

The best-scored SATé alignment of whole ITS1-5.8s rDNA-ITS2 was then edited in MacClade 4.0 [50] to remove the residues of 18S and 28S rDNA at the fringes. Due to the highly divergent rates of evolution of the ITS region, a set of 16 constant characters was added to replace the 5.8S rDNA region across all taxa, serving as an anchor to make sure that SATé kept the ITS1 and ITS2 separate for the subsequent so-called "anchored ITS" alignment (below). The regenerated data sets from MacClade were edited further to remove all gaps, then submitted to SATé to produce alignments of only ITS1-anchor-ITS2 (Table S2). The fourth alignment was produced by combining the anchored ITS data set and the removed 5.8S regions to construct a "5.8S re-included ITS" dataset.

All four alignments were analyzed with MrBayes 3.1 under the GTR + Γ +I model of nucleotide evolution by running four chains of 10 million generations. Trees were sampled every 1000th generation; trees sampled prior to convergence were discarded, and a majority-rule consensus tree was computed from the remaining trees. Bayesian posterior probabilities (BPP) were obtained from the majority-rule consensus of the remaining trees. Branches with BPP higher than 0.94 were considered as significantly supported. As there is no proper approach to directly assess the quality of alignments of such large sizes, we compared

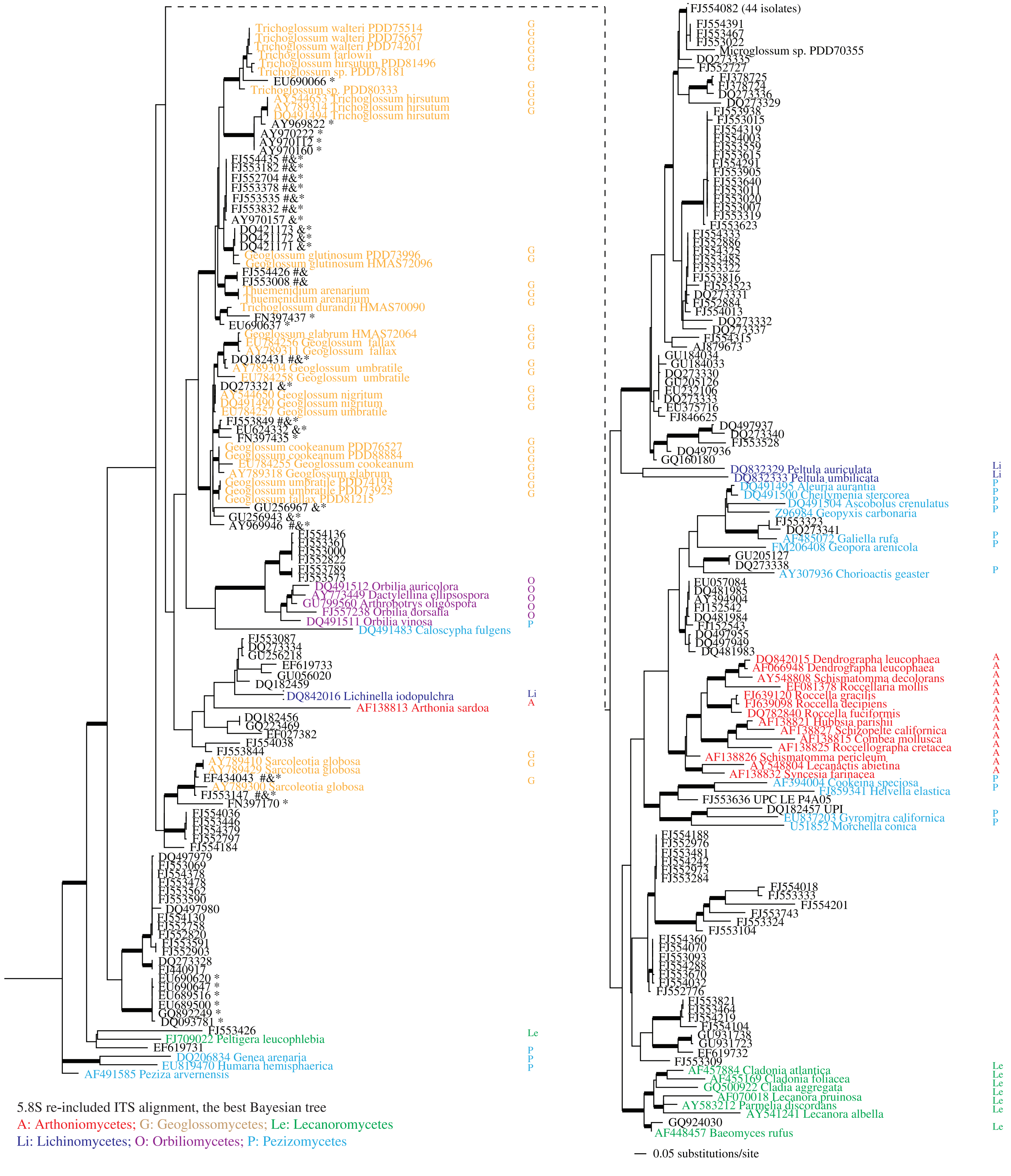

Trichoglossum walteri PDD75514 G
Trichoglossum walteri PDD75657 G
Trichoglossum walteri PDD74201 G
Trichoglossum farlowii G
Trichoglossum hirsutum PDD81496 G
Trichoglossum sp. PDD78181 G
EU690066 *
Trichoglossum sp. PDD80333 G
AY544653 Trichoglossum hirsutum G
AY789314 Trichoglossum hirsutum G
DQ491494 Trichoglossum hirsutum G
AY969822 *
AY970222 *
AY970112 *
AY970160 *
FJ554435 #&*
FJ553182 #&*
FJ552704 #&*
FJ553378 #&*
FJ553535 #&*
FJ553832 #&*
AY970157 &*
DQ421173 &*
DQ421172 &*
DQ421171 &*
Geoglossum glutinosum PDD73996 G
Geoglossum glutinosum HMAS72096 G
FJ554426 #&
FJ553008 #&
Thuemenidium arenarium G
Thuemenidium arenarium G
Trichoglossum durandii HMAS70090 G
FN397437 *
EU690637 *
Geoglossum glabrum HMAS72064 G
EU784256 Geoglossum fallax G
AY789311 Geoglossum fallax G
DQ182431 #&*
AY789304 Geoglossum umbratile G
EU784258 Geoglossum umbratile G
DQ273321 &*
AY544650 Geoglossum nigritum G
DQ491490 Geoglossum nigritum G
EU784257 Geoglossum umbratile G
FJ553849 #&*
EU624332 &*
FN397435 *
Geoglossum cookeanum PDD76527 G
Geoglossum cookeanum PDD88884 G
EU784255 Geoglossum cookeanum G
AY789318 Geoglossum glabrum G
Geoglossum umbratile PDD74193 G
Geoglossum umbratile PDD73925 G
Geoglossum fallax PDD81215 G
GU256967 &*
GU256943 &*
AY969946 #&*
FJ554136
FJ553361
FJ553000
FJ552822
FJ553789
FJ553573
DQ491512 Orbilia auricolora O
AY773449 Dactylellina ellipsospora O
GU799560 Arthrobotrys oligospora O
FJ557238 Orbilia dorsalia O
DQ491511 Orbilia vinosa O
DQ491483 Caloscypha fulgens P
FJ553087
DQ273334
GU256218
EF619733
GU056020
DQ182459
DQ842016 Lichinella iodopulchra Li
AF138813 Arthonia sardoa A
DQ182456
GQ223469
EF027382
FJ554038
FJ553844
AY789410 Sarcoleotia globosa G
AY789429 Sarcoleotia globosa G
EF434043 #&*
AY789300 Sarcoleotia globosa G
FJ553147 #&*
FN397170 *
FJ554036
FJ553446
FJ554379
FJ552797
FJ554184
DQ497979
FJ553069
FJ554378
FJ553478
FJ553562
FJ553590
DQ497980
FJ554130
FJ552758
FJ552820
FJ553591
FJ552903
DQ273328
FJ440917
EU690620 *
EU690647 *
EU689516 *
EU689500 *
GQ892249 *
DQ093781 *
FJ553426
FJ709022 Peltigera leucophlebia Le
EF619731
DQ206834 Genea arenaria P
EU819470 Humaria hemisphaerica P
AF491585 Peziza arvernensis P
5.8S re-included ITS alignment, the best Bayesian tree
A: Arthoniomycetes; G: Geoglossomycetes; Le: Lecanoromycetes
Li: Lichinomycetes; O: Orbiliomycetes; P: Pezizomycetes
FJ554082 (44 isolates)
FJ554391
FJ553467
FJ553022
Microglossum sp. PDD70355
DQ273335
FJ552727
FJ378725
FJ378724
DQ273336
DQ273329
FJ553938
FJ553015
FJ554319
FJ554003
FJ553559
FJ553615
FJ554291
FJ553905
FJ553640
FJ553011
FJ553020
FJ553007
FJ553319
FJ553623
FJ554333
FJ552886
FJ554325
FJ553485
FJ553322
FJ553816
FJ553523
DQ273331
FJ552884
FJ554013
DQ273332
DQ273337
FJ554315
AJ879673
GU184034
GU184033
DQ273330
GU205126
EU232106
DQ273333
EU375716
FJ846625
DQ497937
DQ273340
FJ553528
DQ497936
GQ160180
DQ832329 Peltula auriculata Li
DQ832333 Peltula umbilicata Li
DQ491495 Aleuria aurantia P
DQ491500 Cheilymenia stercorea P
DQ491504 Ascobolus crenulatus P
Z96984 Geopyxis carbonaria P
FJ553323
DQ273341
AF485072 Galiella rufa P
FM206408 Geopora arenicola P
GU205127
DQ273338
AY307936 Chorioactis geaster P
EU057084
DQ481985
AY394904
FJ152542
DQ481984
FJ152543
DQ497955
DQ497949
DQ481983
DQ842015 Dendrographa leucophaea A
AF066948 Dendrographa leucophaea A
AY548808 Schismatomma decolorans A
EF081378 Roccellaria mollis A
FJ639120 Roccella gracilis A
FJ639098 Roccella decipiens A
DQ782840 Roccella fuciformis A
AF138821 Hubbsia parishii A
AF138827 Schizopelte californica A
AF138815 Combea mollusca A
AF138825 Roccellographa cretacea A
AF138826 Schismatomma pericleum A
AY548804 Lecanactis abietina A
AF138832 Syncesia farinacea A
AF394004 Cookeina speciosa P
FJ859341 Helvella elastica P
FJ553636 UPC LE P4A05
DQ182457 UPI
EU837203 Gyromitra californica P
U51852 Morchella conica P
FJ554188
FJ552976
FJ553481
FJ554242
FJ552973
FJ553284
FJ554018
FJ553333
FJ554201
FJ553743
FJ553324
FJ553104
FJ554360
FJ554070
FJ553093
FJ554288
FJ553670
FJ554032
FJ552776
FJ553821
FJ553464
FJ554219
FJ554104
GU931738
GU931723
EF619732
FJ553309
AF457884 Cladonia atlantica Le
AF455169 Cladonia foliacea Le
GQ500922 Cladia aggregata Le
AF070018 Lecanora pruinosa Le
AY583212 Parmelia discordans Le
AY541241 Lecanora albella Le
GQ924030
AF448457 Baeomyces rufus Le
0.05 substitutions/site

Figure 5. The best tree generated by MrBayes for the 5.8S re-included SATé alignment of the large data set. BPP higher than 0.94 are indicated with bold branches. Taxonomic groups were coded with colors and initials of the classes. Environmental ITS sequences grouped in the earth tongue clades are marked with one or more annotations indicating whether the isolate would be designated an earth tongue by other approaches. Affinity of marked isolates to the earth tongues was suggested by: * the sequence search suite *emerencia*; & BLAST of known earth-tongue ITS sequence against INSD; and by # annotation in INSD as unclassified Pezizomycotina environmental samples. Newly generated sequences for this study are presented as taxa without INSD accession numbers while download sequences are listed with their INSD accession numbers.
doi:10.1371/journal.pone.0019039.g005

the resulting phylogenies to previously published phylogenies that were based on multilocus data of genes evolving at slower, more appropriate rates. Due to the high rates of evolution of ITS sequences, strong support for deep nodes – such as those observed in the multilocus datasets – were not expected in the ITS trees.

Supporting Information

Figure S1 The best tree generated by MrBayes for the second best-scored whole ITS SATé alignment. Bayesian posterior probabilities higher than 0.94 are indicated with bold branches. Taxonomic groups were coded with colors and initials of the classes. Environmental ITS sequences grouped in the earth tongue clades are marked with one or more annotations indicating whether the isolate would be designated an earth tongue by other approaches. Affinity of marked isolates to the earth tongues was suggested by: * the sequence search suite *emerencia*; & BLAST of known earth-tongue ITS sequence against INSD; and by # annotation in INSD as unclassified Pezizomycotina environmental samples. Newly generated sequences for this study are presented as taxa without INSD accession numbers while download sequences are listed with their INSD accession numbers.
(EPS)

Table S1 Isolate origin for earth tongue ITS sequences generated for this study.
(DOC)

Table S2 Performance of SATé on different platforms.
(DOC)

Alignment S1 ClustalW alignment of 118 taxa analyzed in this study.
(PDF)

Alignment S2 The best-scored whole ITS SATé alignment for 300 taxa.
(PDF)

Alignment S3 The second best-scored whole ITS SATé alignment.
(PDF)

Alignment S4 The best-scored anchored ITS SATé alignment for 296 taxa.
(PDF)

Alignment S5 The best-scored 5.8S re-introduced ITS SATé alignment for 296 taxa.
(PDF)

Acknowledgments

We thank the curators of PDD, HMAS, OSC, and NIFG for providing collections.

Author Contributions

Conceived and designed the experiments: ZW JPT. Performed the experiments: ZW FLG. Analyzed the data: ZW RHN FLG. Contributed reagents/materials/analysis tools: RHN FLG WZ YD PRJ. Wrote the paper: ZW RHN WZ PRJ JPT.

References

1. Hawksworth DL (2001) The magnitude of fungal diversity: the 1.5 million species estimate revisited. Mycol Res 105: 1422–1432.
2. Mueller GM, Schmit JP (2007) Fungal biodiversity: What do we know? What can we predict? Biodiversity and Conserv 16: 1–5.
3. Kirk PM, Cannon PF, Minter DW, Stalpers JA (2008) Dictionary of the Fungi, 10th ed. Wallingford, UK: CABI.
4. Hibbett DS, Ohman A, Glotzer D, Nuhn M, Kirk P, Nilsson RH (2011) Progress in molecular and morphological taxon discovery in Fungi and options for formal classification of environmental sequences. Fungal Biol Rev (DOI: 10.1016/j.fbr.2011.01.001).
5. Shenoy BD, Jeewon R, Hyde KD (2007) Impact of DNA sequence-data on the taxonomy of anamorphic fungi. Fungal Divers 26: 1–54.
6. Peay KG, Kennedy PG, Bruns TD (2008) Fungal community ecology: a hybrid beast with a molecular master. BioScience 58: 799–810.
7. Kõljalg U, Larsson K-H, Abarenkov K, Nilsson RH, Alexander IJ, et al. (2005) UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. New Phytol 166: 1063–1068.
8. Begerow D, Nilsson H, Unterseher M, Maier W (2010) Current state and perspectives of fungal DNA barcoding and rapid identification procedures. Appl Microbiol Biotechnol 87: 99–108.
9. Taylor JW, Turner E, Townsend JP, Dettman JR, Jacobson D (2006) Eukaryotic microbes, species recognition and the geographic limits of species: examples from the kingdom Fungi. Philos Trans R Soc London [Biol] 361: 1947–1963.
10. Seifert KA (2009) Progress towards DNA barcoding of fungi. Mol Ecol Resour 9: 83–89.
11. Bidartondo MI, Gardes M (2005) Fungal diversity in molecular terms: profiling, identification, and quantification in the environment. In: Dighton J, White JF, Oudemans PV, eds. The Fungal Community. Boca Raton, Florida, CRC press, Taylor & Francis Group. pp 215–239.
12. Brock PM, Döring H, Bidartondo MI (2009) How to know unknown fungi: the role of a herbarium. New Phytol 181: 719–724.
13. Eberhardt U (2010) A constructive step towards selecting a DNA barcode for fungi. New Phytol 187: 265–268.
14. O'Brien HE, Parrent JL, Jackson JA, Moncalvo J-M, Vilgalys R (2005) Fungal community analysis by large-scale sequencing of environmental samples. Appl Environ Microbiol 71: 5544–5550.
15. Jumpponen A, Jones KL (2009) Massively parallel 454-sequencing indicates hyperdiverse fungal communities in temperate *Quercus macrocarpa* phyllosphere. New Phytol 184: 438–448.
16. Buée M, Reich M, Murat C, Morin E, Nilsson RH, Uroz S, et al. (2009) 454 pyrosequencing analyses of forest soils reveal an unexpected high fungal diversity. New Phytol 184: 449–456.
17. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, et al. (2010) Database resources of the National Center for Biotechnology information. Nucleic Acids Res 38: D5–D16.
18. Coleman AW (2003) ITS2 is a double-edged tool for eukaryote evolutionary comparisons. Trends Genet 19: 370–375.
19. Thompson JD, Koehl P, Ripp R, Poch R (2005) BAliBASE 3.0: Latest developments of the multiple sequence alignment benchmark. Proteins: Struct Funct Bioinf 61: 127–136.
20. Nilsson RH, Kristiansson E, Ryberg M, Hallenberg N, Larsson K-H (2008) Intraspecific ITS variability in the kingdom Fungi as expressed in the international sequence databases and its implications for molecular species identification. Evol Bioinformatics 4: 193–201.
21. Liu K, Raghavan S, Nelesen S, Linder CR, Warrnow T (2009) Rapid and accurate large-scale coestimation of sequence alignments and phylogenetic trees. Science 324: 1561–1564.
22. Löytynoja A, Goldman N (2009) Uniting alignments and trees. Science 324: 1528–1529.
23. Zhuang WY (1998) Flora Fungorum Sinicorum. Vol. 8. Sclerotiniaceae et Geoglossaceae. Beijing: Science Press.
24. Schoch CL, Wang Z, Townsend JP, Spatafora JW (2009) Geoglossomycetes *cl. nov.*, Geoglosssales *ord. nov.* and taxa above class rank in the Ascomycota tree of life. Persoonia 22: 129–138.
25. Schoch CL, Sung G-H, López-Giráldez F, Townsend JP, Miadlikowska J, et al. (2009) The Ascomycota Tree of Life: A phylum wide phylogeny clarifies the origin and evolution of fundamental reproductive and ecological traits. Syst Biol 58: 224–239.
26. Wang Z, Binder M, Schoch CL, Johnston PR, Spatafora JW, et al. (2006) Evolution of helotialean fungi (Leotiomycetes, Pezizomycotina): a nuclear rDNA phylogeny. Mol Phylogenet Evol 41: 295–312.
27. Wang Z, Johnston PR, Takamatsu S, Spatafora JW, Hibbett DS (2006) Toward a phylogenetic classification of the Leotiomycetes based on rDNA data. Mycologia 98: 1065–1075.
28. Nilsson RH, Kristiansson E, Ryberg M, Hallenberg N, Larsson K-H (2005) Approaching the taxonomic affiliation of unidentified sequences in public databases – an example from the mycorrhizal fungi. BMC Bioinformatics 6: 178.
29. Lutzoni F, Kauff F, Cox CJ, McLaughlin D, Celio G, et al. (2004) Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits. Am J Bot 91: 1446–1480.
30. James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, et al. (2006) Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature 443: 818–822.
31. Penn O, Privman E, Landan G, Graur D, Pupko T (2010) An alignment confidence score capturing robustness to guide tree uncertainty. Mol Biol Evol 27: 1759–1767.

32. Blair C, Murphy RW (2011) Recent trends in molecular phylogenetic analysis: where to next? J Heredity 102: 130–138.
33. Waldrop MP, Zak DR, Blackwood CB, Curtis CD, Tilman D (2006) Resource availability controls fungal diversity across a plant diversity gradient. Ecol Lett 9: 1127–1135.
34. Taylor DL, Herriott IC, Long J, O'Neill K (2007) TOPO TA is a-ok: a test of phylogenetic bias in fungal environmental clone library construction. Environ Microbiol 9: 1329–1334.
35. Morris MH, Pérez-Pérez MA, Smith ME, Bledsoe CS (2008) Multiple species of ectomycorrhizal fungi are frequently detected on individual oak root tips in a tropical cloud forest. Mycorrhiza 18: 375–383.
36. Hartmann M, Lee S, Hallam SJ, Mohn WW (2009) Bacterial, archaeal and eukaryal community structures throughout soil horizons of harvested and naturally disturbed forest stands. Environ Microbiol 11: 3045–3062.
37. Napoli C, Mello A, Borra A, Vizzini A, Sourzat P, Bonfante P (2010) *Tuber melanosporum*, when dominant, affects fungal dynamics in truffle grounds. New Phytol 185: 237–247.
38. Seaver FJ (1951) The North American cup-fungi (in-operculate). Seaver, New York.
39. Ohenoja E, Wang Z, Townsend JP, Mitchel D, Voitk A (2010) Northern species of earth tongue genus *Thuemenidium* revisited, considering morphology, ecology and molecular phylogeny. Mycologia 102: 1089–1095.
40. Gardes M, Bruns TD (1993) ITS primers with enhanced specificity for basidiomycetes: applicatioin to the identification of mycorrhiza and rusts. Mol Ecol 2: 113–118.
41. White TJ, Brun TD, Lee S, Taylor J (1990) Analysis of phylogenetic relationships by amplification and direct sequencing of ribosomal RNA genes. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds. PCR protocols: a guide to methods and applications. New York: Academic Press. pp 315–322.
42. Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48: 443–453.
43. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16: 276–277.
44. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.
45. Ryberg M, Kristiansson E, Sjökvist E, Nilsson RH (2009) An outlook on the fungal internal transcribed spacer sequences in GenBank and the introduction of a web-based tool for the exploration of fungal diversity. New Phytol 181: 471–477.
46. Nilsson RH, Veldre V, Wang Z, Eckart M, Branco S, et al. (2011) A note on the incidence of reverse complementary fungal ITS sequences in the public sequence databases and a software tool for their detection and reorientation. Mycoscience (DOI: 10.1007/s10267-010-0086-z).
47. Thompson JD, Gibson TJ, Jeanmougin F, Higgins DG (1997) The Clustal-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24: 4876–4882.
48. Nilsson RH, Abarenkov K, Veldre V, Nylinder S, De Wit P, et al. (2010) An open source chimera checker for the fungal ITS region. Mol Ecol Resour 10: 1076–1081.
49. Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754–755.
50. Maddison DR, Maddison WP (2000) MacClade version 4: Analysis of phylogeny and character evolution. Sunderland Massachusetts: Sinauer Associates.