Mycologia, 99(2), 2007, pp. 185–206.

Diversity and phylogenetic affinities of foliar fungal endophytes in loblolly pine inferred by culturing and environmental PCR

A. Elizabeth Arnold[1]
Daniel A. Henk
Rebecca L. Eells
François Lutzoni
Rytas Vilgalys

Department of Biology, Duke University, Durham, North Carolina 27708

Abstract: We examined endophytic fungi in asymptomatic foliage of loblolly pine (*Pinus taeda*) in North Carolina, USA, with four goals: (i) to evaluate morphotaxa, BLAST matches and groups based on sequence similarity as functional taxonomic units; (ii) to explore methods to maximize phylogenetic signal for environmental datasets, which typically contain many taxa but few characters; (iii) to compare culturing vs. culture-free methods (environmental PCR of surface sterilized foliage) for estimating endophyte diversity and species composition; and (iv) to investigate the relationships between traditional ecological indices (e.g. Shannon index) and phylogenetic diversity (PD) in estimating endophyte diversity and spatial heterogeneity. Endophytes were recovered in culture from 87 of 90 *P. taeda* leaves sampled, yielding 439 isolates that represented 24 morphotaxa. Sequence data from the nuclear ribosomal internal transcribed spacer (ITS) for 150 isolates revealed 59 distinct ITS genotypes that represented 24 and 37 unique groups based on 90% and 95% sequence similarity, respectively. By recoding ambiguously aligned regions to extract phylogenetic signal and implementing a conservative phylogenetic backbone constraint, we recovered well supported phylogenies based on ca. 600 bp of the nuclear ribosomal large subunit (LSUrDNA) for 72 Ascomycota and Basidiomycota, 145 cultured endophytes and 33 environmental PCR samples. Comparisons with LSUrDNA-delimited species showed that morphotaxa adequately estimated total species richness but rarely corresponded to biologically meaningful groups. ITS BLAST results were variable in their utility, but ITS genotype groups based on 90% sequence similarity were concordant with LSUrDNA-delimited species. Environmental PCR yielded more genotypes per sampling effort and recovered several distinct clades relative to culturing, but some commonly cultured clades were never found (Sordariomycetes) or were rare relative to their high frequency among cultures (Leotiomycetes). In contrast to traditional indices, PD demonstrated spatial heterogeneity in endophyte assemblages among *P. taeda* trees and study plots. Our results highlight the need for caution in designating taxonomic units based on gross cultural morphology or ITS BLAST matches, the utility of phylogenetic tools for extracting robust phylogenies from environmental samples, the complementarity of culturing and environmental PCR, the utility of PD relative to traditional ecological indices, and the remarkably high diversity of foliar fungal endophytes in this simplified temperate ecosystem.

Key words: Ascomycota, backbone constraint, Basidiomycota, diversity, environmental PCR, fungal endophytes, ITS, LSUrDNA, phylogenetic diversity, symbiosis

INTRODUCTION

Microfungal pathogens, parasites, decomposers and mutualists influence ecosystem productivity and nutrient cycling, plant community structure and species-specific responses of plants to stressors (see Alexopoulos et al 1996). Yet the diversity of plant-associated microfungi, their identity and phylogenetic positions, the number and phylogenetic affinities of unculturable species and the utility of species concepts in delimiting taxonomic boundaries remain largely unknown. These issues are especially challenging in the case of cryptic symbionts whose presence is not readily observed and which frequently do not fruit in culture (e.g. fungal endophytes of foliage).

Fungal endophytes—microfungi that colonize and live within healthy plant tissues without inducing symptoms of disease (Petrini 1991)—comprise a large but little explored portion of fungal diversity (Fröhlich and Hyde 1999, Hawksworth 2001). Endophytes are thought to be ubiquitous among terrestrial plants, having been recovered from all major plant lineages in a wide range of terrestrial communities (see Stone et al 2001). Numerous investigations in the past three decades have shown that endophytes are abundant in asymptomatic leaves of many conifers (e.g. Carroll

Accepted for publication 12 Dec 2006.

[1] Corresponding author. Current address: Division of Plant Pathology and Microbiology, Department of Plant Sciences, University of Arizona. Tel: 520-621-7212; Fax: 520-621-9290. E-mail: Arnold@ag.arizona.edu

and Carroll 1978, Stone 1985, Sherwood-Pike et al 1986, Legault et al 1989, Petrini 1991, Rollinger and Langenheim 1993, Dobranic et al 1995, Stanosz et al 1997, Deckert and Peterson 2000, Kriel et al 2001, Müller et al 2001). Most endophytes of conifer leaves are filamentous Ascomycota (e.g. Petrini 1986). Unlike the clavicipitaceous endophytes of grasses, conifer endophytes generally are transmitted among hosts via contagious spread (i.e. horizontal transmission) and occur as numerous, localized infections that increase in number, density and species diversity as leaves age (Stone 1987). Up to 110 species of endophytic fungi have been found in leaves of well surveyed species such as *Pseudotsuga menziesii* (Carroll and Carroll 1978), highlighting the ability of conifers to engage in symbioses with a tremendous diversity of fungi in their above-ground tissues.

Several conifer endophytes augment host defense against natural enemies (e.g. *Rhabdocline parkeri* in *P. menziesii*; Carroll and Carroll 1978, Carroll 1986), a finding echoed by recent work on some angiosperm-associated endophytes (Redman et al 2002, Arnold et al 2003, Arnold and Lewis 2005). In most cases however the ecological roles of horizontally transmitted endophytes are not known, reflecting both a lack of study of many hosts and methodological issues that impede ecological studies. These methodological issues are diverse but center on three main challenges: identifying sterile endophytes, delimiting functional taxonomic units that are biologically meaningful, and understanding the biases imposed by culturing.

Because sterile cultures lack the taxonomic characters needed for identification, morphotaxa, based on gross colony features, are used frequently as functional taxonomic units (e.g. Guo et al 2000, 2003; Arnold et al 2000, 2003). In some cases molecular sequence data from the nuclear ribosomal internal transcribed spacer region (ITS) have been used to identify sterile cultures and to evaluate morphotaxon boundaries (Arnold 2002, Lacap et al 2003). ITS data are considered useful for these purposes due to the rapid rate of evolution in the spacer regions, the relative ease with which ITS data can be recovered and the abundance of ITS data in GenBank (21 075 fungal ITS sequences as of early 2004, Lutzoni et al 2004). Taxonomic boundaries are estimated on the basis of ITS BLAST matches to different taxa, comparisons of ITS sequence divergence, or phylogenetic analyses of endophytes and closely related species (for which ITS data can be aligned). In turn identifications frequently are based on ITS BLAST affinity for named taxa in GenBank. However most fungi are not represented in GenBank and some GenBank records are misidentified or lack taxonomic information. The accuracy of matches based on the nonphylogenetic BLAST algorithm is not clear, and the degree to which different BLAST matches precisely and accurately distinguish species has not been established. The correspondence of ITS sequence divergence with true species boundaries is not known for many groups of fungi and is likely to vary among clades (see Jacobs and Rehner 1998, Lieckfeldt and Seifert 2000, Kim and Breuil 2001). Some studies have used phylogenetic analyses of ITS data to confirm BLAST identifications, but in those cases explicit criteria have not been provided for delimiting species (e.g. Guo et al 2003). To our knowledge no study has used another locus to critically evaluate endophyte species boundaries and identifications based on ITS data nor implemented a formal phylogenetic criterion for delimiting endophyte species.

Similarly little is known regarding the abundance and diversity of unculturable endophytes, limiting our understanding of endophyte infection frequencies, taxonomic composition and diversity. Culture-free methods such as environmental PCR of surface sterilized leaves may be especially key for uncovering endophytes with obligate host associations, species that grow slowly or that do not grow readily on standard media, and species that lose in competitive interactions during the culturing process. However to our knowledge conifer foliage has not been examined for unculturable endophytic symbionts. Previous environmental PCR of asymptomatic, living foliage from other hosts has used only ITS data without direct comparisons to concurrently cultured endophytes (Guo et al 2001). As a result the diversity and taxonomic affinities of unculturable foliar endophytes and their relationships to cultured endophytes have not been established in a broad phylogenetic context.

Finally, traditional measures of diversity (e.g. Shannon index) and community similarity (e.g. Sorensen's index) have been used in endophyte studies (e.g. Gamboa and Bayman 2001), but their interpretation is hindered by uncertain species definitions and the lack of a statistical framework for comparing values. In contrast to traditional measures of diversity and similarity, phylogenetic diversity (PD) takes into account the taxonomic breadth of samples without relying on morphotaxa, species or sequence-type designations (Vane-Wright et al 1991, Faith 1992, Shaw et al 2003). PD provides information regarding the distribution and diversity of samples at higher taxonomic levels, which complements the study of diversity at the species level and can be compared directly using standard statistical methods. However PD measures require inclusive phylogenies, which to date are lacking in endophyte studies.

Resolution of these issues requires that endophytes be integrated into a phylogenetic context using loci that are informative across diverse fungal lineages. However the high abundance and diversity of endophytes typically leads to a tradeoff between the number of isolates that can be sequenced and the number of loci that can be obtained. One solution is to choose a locus with a high degree of phylogenetic signal at high and low taxonomic levels, high recovery rate, low sequencing cost (defined by number of reactions per locus), and relatively large number of sequenced taxa available for comparison. Data from the nuclear ribosomal large subunit (LSUrDNA) have been used in more than 150 phylogenetic analyses for fungi (1991–2003, Lutzoni et al 2004), and in some cases served as the primary source of phylogenetic signal (e.g. Allen et al 2003, McLenon and Moncalvo 2004). Although individual loci are of limited use in reconstructing deep nodes in the fungal and Ascomycota trees of life (e.g. Tehler et al 2000, Berbee 2001, Tehler 2003, Liu and Hall 2004, Reeb et al 2004, Lutzoni et al 2004), numerous tools exist in phylogenetic biology that have not been applied yet in studies of endophyte diversity and which can be used to augment the phylogenetic signal from single-locus datasets. These include gathering phylogenetic signal from ambiguously aligned regions, and the implementation of backbone constraints—phylogenetic constraints that establish deep relationships for a subset of named taxa while permitting all other taxonomic units attach to any point on trees during the search procedure.

Here we present results of a culture- and environmental PCR-based survey of fungal endophytes inhabiting healthy foliage of an ecologically and economically important conifer species, loblolly pine (*Pinus taeda*, Pinaceae) in central North Carolina (USA). The goals of our study were (i) to develop a rigorous phylogenetic approach for assessing the biological relevance of morphotaxa, ITS genotype groups and ITS BLAST matches as functional taxonomic units, (ii) to explore methods to maximize phylogenetic signal for highly diverse environmental samples, (iii) to compare endophyte diversity and species composition given a culture-based vs. environmental sampling approach, and (iv) to examine the utility of phylogenetic diversity in assessing endophyte community structure.

MATERIALS AND METHODS

Host species and study site.—*Pinus taeda* (loblolly pine) is an economically important timber species native to the southeastern United States. Entomopathogenic, mycorrhizal and pathogenic fungi associated with *P. taeda* have been studied previously (Sluder 1993, Carey and Kelley 1994, Sung et al 1995, West and Jones 2000, Mihelcic et al 2003). To our knowledge the foliar endophytes associated with *P. taeda* have not been examined within its native range.

Field sampling was conducted in Jun 2003 in the Blackwood Division of Duke Forest (Orange County, NC; 35°58′N, 79°05′W). Duke Forest comprises 7900 acres of mixed hardwood and pine stands spanning four counties (Durham, Orange, Chatham and Alamance) in the eastern piedmont of North Carolina. The area receives 1140 mm of precipitation annually and has a mean annual temperature of 15.5 C. Our study site is part of a 200-acre, planted cohort of *P. taeda* and a naturally regenerating understory of hardwood saplings (e.g. *Liquidambar styraciflua*, *Acer rubrum* and *Cornus florida*). Samples were collected in three 707 m^2 plots in this contiguous forest, which serve as the control (unmanipulated) plots at the Free Air Carbon Exchange (FACE) site in Duke Forest. Comparisons of endophyte communities under ambient and elevated CO_2 will be presented in a forthcoming paper (Henk et al in preparation).

Endophyte isolations.—From three haphazardly chosen trees at the center of each plot, we collected 10 fascicles > 1 y old from branches 10–12 m above ground. Within 24 h of harvesting, one asymptomatic needle per fascicle (n = 90 needles) was washed thoroughly in running tap water, cut into five segments of uniform length and surface sterilized by sequential immersion in 95% ethanol (5 s), 0.5% NaOCl (2 min) and 70% ethanol (2 min) (Arnold 2002). Segments were surface-dried under sterile conditions before plating on 2% malt-extract agar (MEA), which supports growth by diverse endophytes (Fröhlich and Hyde 1999). Plates were sealed, checked daily for hyphal growth and incubated at room temperature for up to 8 wk. One colony per segment was randomly selected for isolation into axenic culture on 2% MEA. When present additional colonies with apparently unique morphology also were isolated. All isolates were digitally photographed and vouchers have been deposited at the Robert L. Gilbertson Mycological Herbarium (University of Arizona).

Seven days after isolation, pure cultures in 60 mm Petri plates were grouped into morphotaxa based on colony shape, height and color of aerial hyphae, base color, growth rate, margin characteristics, surface texture and depth of growth into medium (Arnold 2002). Cultures differing in two or more characteristics were placed in different morphotaxa. These characters conservatively reconstruct taxonomic boundaries among some clades of endophytic fungi (see Arnold 2002, Lacap et al 2003, Guo et al 2003), but their utility in large endophyte surveys has not been established.

DNA extraction and PCR.—Representatives of all morphotaxa (439 isolates representing 24 morphotaxa) were randomized with regard to tree and plot of origin and subsampled for analysis of the nuclear ribosomal internal transcribed spacer regions (ITS1 and ITS2) and 5.8s gene (ITS) and a ca. 600 base pair (bp) portion at the 5′ end of nuclear ribosomal large subunit (LSUrDNA). Numbers of isolates sequenced per morphotaxon were proportional to

the relative abundance of each morphotaxon. All trees and plots were represented with approximately equal frequency among sequenced isolates.

DNA extraction followed Arnold and Lutzoni (2007). ITS and ca. 600 bp of LSUrDNA were amplified as a single fragment using primers ITS1F and LR3 (Gardes and Bruns 1993, Vilgalys and Sun 1994) or as two fragments using the additional primers ITS4 and LROR (Vilgalys and Hester 1990, White et al 1990) using 25 μL reactions (2.5 μL dNTPs, 2.5 μL BSA, 2.5 μL 10× buffer, 1.25 μL of each primer, 0.125 μL Taq, 13.875 μL water, and 1.0 μL DNA template). Amplification was performed on MWG Ag Biotech Primus 96 plus HPL and MJ Research PTC-200 thermocyclers following Arnold and Lutzoni (2007). Products that yielded single bands when viewed with EtBr or SYBR Green were purified with QIAGEN or Microcon columns and sequenced in two directions using primers ITS1F, ITS4, LROR and LR3.

Environmental PCR.—Nine asymptomatic leaves per tree were surface sterilized as described above and stored at −80 C for 7 d before DNA extraction. An additional set of nine leaves were collected from three individuals of *P. taeda* in three additional study plots, which were surveyed concurrently as part of a separate project (Henk et al in preparation). These plots were fumigated with elevated CO_2 (ambient + 200 μmol/mol) and are paired at small spatial scales with the plots from which we obtained cultures. Clones from these trees are included here because we observed no significant difference in endophyte species composition between the control and elevated CO_2 treatments (Henk et al in preparation).

Leaf samples were bulked by tree and ground under liquid nitrogen with a sterile mortar and pestle. Approximately 0.2 g of homogenized material per sample was added to 750 μL of CTAB extraction buffer, treated with 750 μL phenol : chloroform : isoamyl alcohol and extracted as described above. Pellets were washed in 80% ethanol (−20 C), dried 20 min in a rotary vacufuge and eluted in 100 μL of sterile water. DNA extractions were used directly in PCR to amplify ITS and ca. 600 bp of LSUrDNA as a single fragment with primers ITS1F and LR3. PCR products were ligated to a cloning vector with the TOPO cloning kit (Invitrogen) following the manufacturer's instructions. After transformation up to five positive clones per tree were selected randomly for secondary PCR and sequencing as described above.

Preparation of sequence data.—ITS data were obtained from 150 cultured isolates (APPENDIX 1). LSUrDNA data were obtained for 145 isolates (APPENDIX 2). ITS and LSUrDNA data were recovered from 42 clones (APPENDIX 3). Sequence data were screened by visual inspection of chromatograms and automatically using quality assignments for individual bases by the program Phred (Ewing et al 1998). Contigs were assembled with Phrap (Ewing et al 1998), followed by manual inspection and editing in Sequencher 4.1 or 4.2 (Gene Codes, Ann Arbor, Michigan). Phred and Phrap analyses were aided by a BioPython script (written by F. Kauff, Duke University). ITS consensus sequences were trimmed to homologous starting and ending points, such that length variation reflected only insertion or deletion events and partial sequences would not influence delimitations of ITS genotype groups or BLAST searches. BLAST searches of ITS sequence data in the NCBI GenBank database were used to guide taxon sampling for subsequent phylogenetic analysis of LSUrDNA.

Delimitation of ITS genotype groups.—Sequencher 4.1 was used to assemble consensus ITS sequences into contigs defined by 90%, 95% and 99% similarity, with an expectation of at least 40% sequence overlap (''global'' delimitations). Because Sequencher was not designed for assembling multitaxon contigs, we also used *needle*, available through EMBOSS (Rice et al 2000), for pairwise comparisons of all ITS sequences. A custom BioPERL script (written by J. Stajich, Duke University) was used to assemble sequences into groups defined by 90%, 95%, and 99% similarity (''pairwise'' delimitations). Global and pairwise groups at each level of similarity are shown (APPENDIX 1).

Ecological analyses.—Species accumulation curves and bootstrap estimates of total richness were inferred with EstimateS (Colwell RK. 1996. EstimateS: statistical estimation of species richness and shared species from samples. Freeware, online at http://viceroy.eeb.uconn.edu/EstimateS). Based on the observed concordance of species boundaries inferred from the LSUrDNA phylogeny and ITS genotype groups delimited by 90% sequence similarity (see below), and the ease with which ITS genotype groups were designated in Sequencher, we used 90% ITS genotype groups based on global comparisons as functional taxonomic units for further ecological analyses.

Diversity of cultured endophytes was measured by the Shannon diversity index (H'). Endophyte community similarity among trees and plots was assessed with Soerensen's index of similarity (Arnold et al 2001). Mean similarity values were compared with nonparametric Wilcoxon tests implemented in JMP (Sall and Lehman 1996).

To compare the utility of culturing vs. environmental sampling, ITS sequence groups obtained by culturing were randomized and subsampled with replacement to yield partitions of 42 sequences. Each partition was then assessed for (i) species richness and (ii) the proportion of singleton species ($N = 1000$ replicates, comparisons based on global ITS comparisons using 90% sequence similarity). The resulting distributions were compared against observed values obtained by environmental PCR, with P-values indicating the probability of recovering the same, greater or lesser values based on culturing alone.

LSUrDNA alignment and phylogenetic analyses.—We first inferred phylogenetic placement of 145 cultured endophytes. LSUrDNA data were aligned with 72 representative Ascomycota and Basidiomycota to generate a matrix of 217 OTU (see APPENDIX 4 for GenBank accession numbers). A core alignment for 46 named taxa (following Lutzoni et al 2001) was generated manually in MacClade 4.06 (Maddison and Maddison 2003) with the LSUrDNA secondary structure model for *Saccharomyces cerevisiae* (Cannone et al 2002). All remaining sequences were aligned to this core with Clustal X (Thompson et al 1997) (gap opening cost =

15, gap extension cost = 6) followed by manual adjustment. Seven ambiguously aligned regions comprising > 200 characters (ca. 40% of available data) were excluded, leaving 323 included characters.

Preliminary analyses failed to recover a well supported and well resolved topology (results not shown). Unambiguously aligned portions of the data matrix then were subjected to a symmetric step matrix containing costs inversely proportional to estimated frequency of changes, estimated by summarizing character states (four nucleotides and gaps as a fifth character state) at each position of the unambiguously aligned regions using the FULL DETAIL character status option in PAUP* 4.0b8a (Swofford 2001). STMatrix 2.1 (written by S. Zoller, available at http://www.lutzonilab.net) was used to calculate proportional frequency of changes, which were converted into cost of changes with the negative natural logarithm (Felsenstein 1981, Wheeler 1990). All ambiguous regions were then recoded into 23 characters with the ''nucleotide option'' of the program arc v1.5 (written by F. Kauff, available at http://www.lutzonilab.net), which takes into account the length of a given ambiguous region, the relative frequencies of bases and base pairs, and their distribution among sequences (see Miadlikowska et al 2003). Each of the 23 characters was subjected to a specific weight inversely proportional to the number of linked characters (see arc manual), generating a total of 484 characters (161 arc characters and 323 unambiguously aligned sites). Congruence between arc characters and unambiguously aligned characters was assessed with a reciprocal 70% bootstrap proportion (Reeb et al 2004) with bootstrap values calculated by neighbor joining with average mean distances derived from 1000 bootstrapped datasets for each data partition. No significant conflicts were detected between the down-weighted arc and unambiguously aligned characters. However these 484 characters were insufficient for resolving relationships among the 217 OTU (results not shown).

For this reason we decoupled the placement of unknown endophytes within a phylogenetic tree from the recovery of phylogenetic relationships among identified taxa. Over the past decade fungal systematists have begun to elucidate the relationships among major lineages of Ascomycota, and several published hypotheses exist based on multilocus phylogenies (e.g. Lutzoni et al 2001, Lumbsch et al 2002, Liu and Hall 2004, Lutzoni et al 2004, Reeb et al 2004). Based on consensuses among these studies and on additional highly supported internodes revealed in multilocus studies by Lutzoni et al (2004) and Reeb et al (2004), we constructed a conservative topological constraint tree for a subset of the deep nodes uniting the named taxa in our analysis (i.e. a backbone constraint). Nodes were constrained if they contained conspecific taxa relative to Reeb et al (2004) and Lutzoni et al (2004) and if these studies indicated high support values (>70%) based on neighbor joining, maximum parsimony, or maximum likelihood bootstrap in the two, three, or four-gene trees (Reeb et al 2004: LSUrDNA + SSUrDNA + RPB2 for Ascomycota; Lutzoni et al 2004: LSUrDNA + SSUrDNA for Fungi; LSUrDNA + SSUrDNA + RPB2 and LSUrDNA + SSUrDNA + RPB2 + mitSSU for Ascomycota + Basidiomycota). Our resulting constraint tree was conservative, including 35 representative taxa relative to the 217 OTU present in our analysis (16.1%). The 182 OTU not included in the constraint, including named taxa and fungal endophyte sequences, were free to attach to any point on the trees during the search procedure.

Subsequent parsimony analyses were conducted with a step-matrix for the unambiguously aligned sites, down-weighted arc characters, and the constraint tree, which was implemented as a backbone constraint in PAUP* 4.0b8a or b10 (Swofford 2001) with TBR branch swapping and MULTREES selected. Because of the high ratio of taxa to characters our searches were unable to recover all equally most parsimonious trees. To determine the length of the most parsimonious trees for the 217-OTU dataset, we first performed a search with 1000 random addition sequences (RAS), saving only one tree per RAS. The shortest tree length (3807.39 steps) was encountered 194 times, yielding 194 nonidentical trees. A second search then was initiated in which only one tree ≥ 3809.00 steps was saved per RAS and all trees shorter than this threshold length were saved. This second search was stopped after more than 1 wk of TBR swapping, yielding 4848 trees of 3807.39 steps from a single island of equally most parsimonious trees (EMPT). To determine whether sampling more EMPT would change the strict consensus tree, five strict consensus trees were generated summarizing the first 1000, 2000, 3000, 4000, 4848 and 4848 + 194 trees, respectively. Using the default settings of the tree-to-tree distances in PAUP* 4.0b8a, the five consensus trees were found to be identical. The strict consensus tree for these 5042 trees is shown (FIG. 3), with MP branch lengths obtained with the tree score option in PAUP* 4.0b8a under the settings described above.

To compare branch length estimations derived from different optimization criteria and to generate conservative branch lengths for estimating phylogenetic diversity, we generated branch lengths for the same strict consensus using mean character difference (i.e. distance, APPENDIX 5), and used these values for phylogenetic diversity measurements as described below. To assess support for branches, we implemented nonparametric bootstrapping with the same step-matrix and backbone constraint as for the original dataset, except that 10 random addition sequences were performed for each of the 1212 bootstrapped datasets and only one tree was saved for each RAS followed by TBR. As a separate estimate of phylogenetic confidence, bootstrap proportions were estimated with neighbor joining (NJ) with mean character differences (including arc characters) and the same backbone constraint.

A similar procedure was used to infer relationships among cultured endophytes, cloned samples from environmental PCR, and named taxa. The REDUNDANT TAXA option in MacClade 4.06 (Maddison and Maddison 2003), which is not sensitive to sequencing artifacts such as total sequence length, was used first to identify and merge taxa with identical sequences. The resulting dataset included 72 exemplar taxa, 66 unique genotypes of cultured endophytes and 33 cloned sequences (171 OTU). Nine of 42 clones

were excluded from the analysis on the basis of BLAST results and preliminary phylogenetic analyses (see Results). Inclusion of cloned sequences did not alter the number or size of the ambiguous regions of the alignment. All other methods, including recoding of ambiguous regions with arc, down-weighting of arc characters, generation of a step-matrix, implementation of the backbone constraint and use of search parameters were as described above.

A first search of 1000 RAS, in which one tree per RAS was saved, recovered 10 EMPT (4178.39 steps). A second search, in which all trees ≤ 4179 steps were saved, was terminated when 24 287 trees exceeded available memory. Two additional searches of 1000 RAS, in which one tree was saved per RAS, each found 15 trees of 4178.39 steps. Four strict consensus trees were generated, summarizing 24 287 trees, 24 287 + 10 trees, 24 287 + 10 + 15 trees, and 24 287 + 10 + 15 + 15 trees, respectively. The four consensus trees were identical on the basis of tree-to-tree distance. Nonparametric bootstrapping was implemented as described above, except that 300 random addition sequences were performed for each of the 1000 bootstrapped datasets and only one tree was saved for each random addition sequence followed by TBR. NJ bootstrap proportions were calculated as described above. The strict consensus tree resulting from these analyses is provided (FIG. 4).

Phylogenetic circumscription of endophyte taxa.—A ''distal criterion'' (Arnold 2002) was used to infer taxon boundaries among endophytes based on the strict consensus tree from the LSUrDNA analyses of cultured endophytes and core taxa (FIG. 3). Using this approach we first designated species, genus or family-level boundaries for named taxa only on the strict consensus trees. We then assessed the position of each endophyte, considering endophytes to be distinct from one another if their placement was distal to boundaries delimiting known taxonomic groups based on exemplar taxa alone.

Phylogenetic diversity.—Phylogenetic diversity (PD) accounts directly for the phylogenetic breadth of taxa without depending on species, morphotaxa or sequence-type definitions (Vane-Wright et al 1991, Faith 1992, Shaw et al 2003). PD can be calculated in numerous ways (see Vane-Wright et al 1991, Faith 1992, Williams et al 1994, Humphries et al 1995, Moritz and Faith 1998). We calculated PD as the sum of branch lengths from a given data partition that included all and only those taxa within that focal partition (Faith 1992). Here partitions were defined as study plots and individual pine trees. PAUP* 4.0b10 (Swofford 2003) was used to prune taxa from the initial, strict consensus LSUrDNA tree, leaving only the isolates from each focal partition. Branch lengths, estimated with neighbor joining on the strict consensus tree inferred from parsimony searches (APPENDIX 5), were summed for remaining subtrees. To account for isolation frequency we used a bootstrapping approach, subsampling 14 taxa (the minimum number isolated from any partition) from each partition for 100 bootstrap replicates. Bootstrapping and PD calculations were aided by Python scripts (DAH and Cymon Cox, unpublished data, available on request from DAH). For each partition we calculated mean PD and standard deviations, which were compared using nonparametric tests due to a significant deviation of PD distributions from normality (Shapiro-Wilk W, $P > 0.05$). PD values were compared with traditional indices of diversity and community similarity based on ITS genotype groups for cultured endophytes using linear regression.

RESULTS

Fungal endophytes were abundant and diverse in healthy, mature foliage of *Pinus taeda*. Culturable endophytes were recovered from 96.7% of sampled leaves, yielding a total of 439 isolates. Infection rates were consistent among plots, trees, and leaves (TABLE I).

Twenty-four morphotaxa were recovered. Morphotaxa abundances conformed to a log-normal distribution (Shapiro-Wilk $W = 0.9328$; $P = 0.1156$). Three morphotaxa were recovered only once; all other morphotaxa were recovered from multiple leaves, with the most common morphotaxon accounting for 181 isolates. When adjusted for the number of isolates recovered, morphotaxon richness did not differ among trees or plots (TABLE I). Eighteen to 20 morphotaxa were recovered per plot ($N = 30$ leaves/plot), and 9–16 morphotaxa were recovered per tree (mean ± SE: 12.00 ± 0.87 morphotaxa per 10 leaves sampled) (TABLE I).

Morphotaxa accumulated rapidly with each additional tree and needle sampled (FIG. 1). Bootstrap estimates indicate that ca. 93% of expected richness was recovered by our sampling, such that surveys of nine trees were sufficient to recover the majority of morphological richness. In contrast ITS genotypes continued to accumulate with increased sampling (FIG. 1). As sample sizes increased beyond four *P. taeda* trees (ca. 36 *P. taeda* leaves), richness based on morphotaxa markedly underestimated ITS genotype richness (FIG. 1).

Overall > 54% of morphotaxon- and ITS-genotype comparisons resulted in one-to-one correspondence, wherein each morphotaxon corresponded to a single ITS genotype and vice versa (FIG. 2). However many ITS genotypes contained multiple morphotaxa, reflecting phenotypic plasticity or misleading mycelial characteristics. Conversely many morphotaxa contained multiple ITS genotypes (cryptic taxa). Some morphotaxa contained 2–7 ITS genotypes based on global comparisons (FIG. 2) and up to 10 ITS genotypes based on pairwise comparisons (data not shown). Genotype groups based on 90% ITS sequence similarity, inferred with global comparisons, showed high congruence with phylogenetically delimited taxa based on LSUrDNA analyses (see below).

TABLE I. Frequency of endophyte infection among surveyed leaves, numbers of isolates recovered, numbers of morphotaxa recorded, and ratio of morphotaxa to isolates obtained in a culture-based survey of fungal endophytes from asymptomatic, >1 y old foliage of mature *Pinus taeda* in Duke Forest (NC, USA) (N = 10 leaves/tree). Plot numbers correspond to pre-existing study plots in Duke Forest.

Plot	Tree	Infection frequency	Isolates recovered	Morphotaxa	Morphotaxa per isolate
1	14004	0.80	39	11	0.28
	15092	1.00	52	11	0.21
	17045	1.00	42	15	0.36
	Total	—	133	18	—
	Mean	0.93	44.3 ± 3.93	12.3 ± 1.33	0.28 ± 0.04
5	51003	1.00	50	16	0.32
	52003	1.00	51	9	0.18
	54003	1.00	53	10	0.19
	Total	—	154	20	—
	Mean	1.00	51.3 ± 0.88	11.7 ± 2.19	0.23 ± 0.05
6	61007	1.00	54	14	0.26
	65003	1.00	50	13	0.26
	67004	0.90	48	9	0.16
	Total	—	152	19	—
	Mean	0.97	54.0 ± 2.31	12.0 ± 1.53	0.22 ± 0.04
	Total	—	439	24	—
	Mean/tree	0.97 ± 0.02	49.9 ± 1.97	12.0 ± 0.87	0.25 ± 0.02

Spatial heterogeneity of cultured endophytes.—Among nonsingleton ITS genotype groups recovered by culturing, 100%, 85.7% and 85.7% were recovered from at least two study plots (based on 90%, 95% and 99% similarity, respectively). The homogenous distribution of ITS genotypes is reflected by high similarity values and by the lack of significant differences in endophyte assemblages within vs. between plots (TABLE II).

Phylogenetic affinities of cultured endophytes.—By arc-recoding ambiguous regions, implementing a conservative backbone constraint for a subset of taxa and comparing tree-to-tree distance to assess search completeness, we recovered a well supported phylogeny based on 600 bp of LSUrDNA data for 145 cultured endophytes and 72 named exemplar taxa (FIG. 3). Major lineages of fungi were resolved as monophyletic and with high bootstrap support (≥70%, inferred with maximum parsimony [MP] or neighbor joining [NJ]). Resolution and support at deeper nodes largely reflects the backbone constraint used in our analyses (see insets), but the presence of unconstrained, named taxa in their expected clades indicates phylogenetic signal that was not guided by our backbone constraints alone. Based on the placement of named taxa, the resulting topology is consistent with previously published phylogenetic hypotheses based on multiple loci (e.g. Reeb et al 2004, Lutzoni et al 2004).

Twenty-two morphotaxa representing 20 unique ITS genotypes (90% sequence similarity, global comparisons) were represented among the cultured endophytes for which LSUrDNA data were examined (FIG. 3). Based on our conservative estimation of species boundaries on the LSUrDNA tree, we infer that 20–23 unique species are present in this sample. Morphotaxa were generally not congruent with LSUrDNA-delimited species, but ITS genotype groups (90% similarity, global criterion) were highly concordant with the LSUrDNA topology (FIG. 3). ITS genotype groups based on 90% sequence similarity did not conform to the expected log-normal distribution (Shapiro-Wilk $W = 0.7758$, $P < 0.0001$) despite the evidence for their utility as biologically relevant taxonomic units.

Cultured endophytes represented both a single clade within the Basidiomycota (seven isolates) and numerous lineages of Ascomycota (FIG. 3). Together the basidiomycetous endophytes showed highest affinity for three different taxa of Basidiomycota when subjected to BLAST searches in Sep 2003: *Lentinellus vulpinus* (Aphyllophorales: endophytes 1980, 2313, 2072A), *Aleurocystidiellum subcruentatum* (Aphyllophorales: endophyte 2361, 2162) and *Thelephora americana* (Thelephorales: endophytes 2617 and 2618). When searches were repeated in 2004 all showed highest affinity for *Fomitiporia hesleri* (Hymenochaetales, conordinal with *Phellinus gilvus*), albeit with low quality scores (APPENDIX 3). Together these endophytes form a single, well supported clade nested within the Homobasidiomycetes (with low

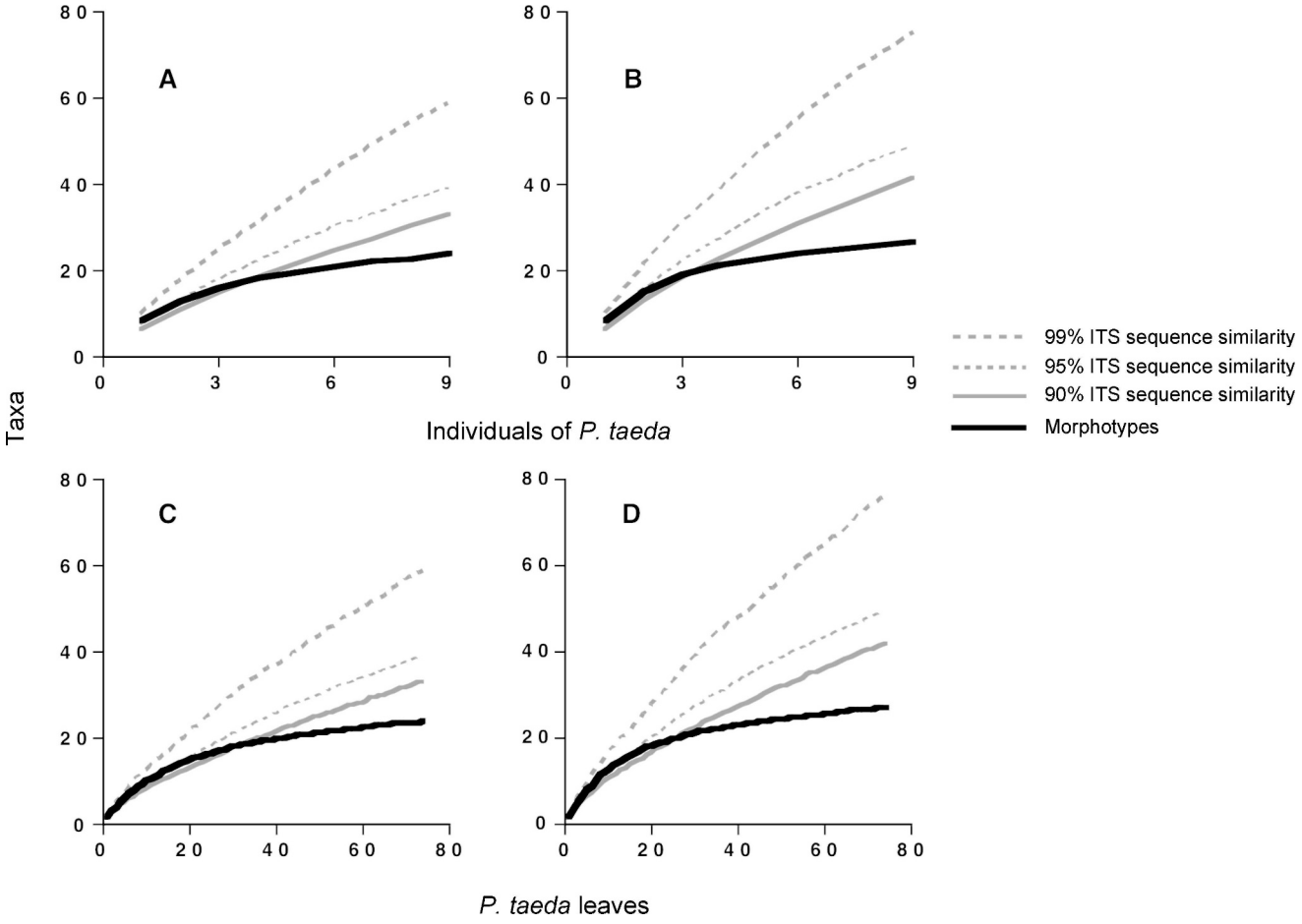

FIG. 1. Relationship of sampling effort and taxonomic richness for cultured endophytes from asymptomatic, mature foliage of *Pinus taeda* in Duke Forest (NC, USA). A, B indicate the accumulation of taxa as a function of *P. taeda* trees sampled. A. Accumulation of morphotaxa and ITS genotypes (based on global comparisons); B. Estimated richness based on bootstrap analyses. C, D. indicate the accumulation of taxa as a function of leaves sampled. C. Accumulation of morphotaxa and ITS genotypes; D. Bootstrap estimates. Curves are based on 50 randomization of sample order.

bootstrap support) and sister of the well supported clade containing *Thelephora americana*, *Boletus pallidus* and *Coprinopsis cinerea* (FIG. 3A). The exact placement of this clade of endophytes (the ''basidiomycete clade'') remains to be determined, in part due its placement on a long branch relative to other basidiomycete taxa and its variable placement in our two analyses (see below). Together these seven isolates represented two distinct morphotaxa; however, all endophytes in this clade were members of a single ITS genotype group (defined by 90% sequence similarity; FIG. 3A, APPENDIX 2).

Within the Ascomycota cultured endophytes were distributed among the Sordariomycetes, Dothideomycetes, Eurotiomycetes and Leotiomycetes (FIG. 3). Superordinal matches based on ITS BLAST results, when available, were consistent with LSUrDNA results for all ascomycetous endophytes. Placement of endophytes at the ordinal and family levels with ITS BLAST results was confirmed largely by the LSUrDNA analyses (FIG. 3, APPENDIX 2), although in some cases orders or families were not monophyletic in our analyses.

Among the Sordariomycetes, morphotaxa were incongruent with taxonomic boundaries inferred with LSUrDNA data (FIG. 3A). In contrast ITS genotypes were highly congruent with LSUrDNA species boundaries. Thirteen cultured endophytes were associated with the Xylariales and represent 4–6 unique species. Seven endophyte isolates representing four species were placed within the remainder of the Sordariomycetes. Eleven endophytes that represented seven or eight species were placed within various lineages of Dothideomycetes (FIG. 3A).

The most frequently encountered endophytes of *P. taeda* in our culture-based surveys form two well supported clades within the Leotiomycetes, corresponding to two unique species (FIG. 3B). The

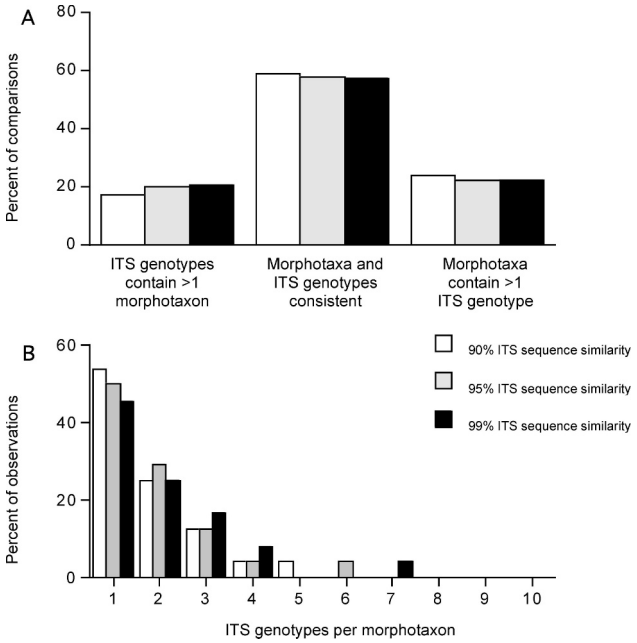

FIG. 2. Correspondence of endophyte morphotaxa and ITS genotypes based on global groupings at 90%, 95% and 99% ITS sequence similarity. A. Percent of observed cases in which ITS genotypes comprised multiple morphotaxa; morphotaxa and genotypes showed one-to-one correspondence; and morphotaxa comprised multiple ITS genotypes (global comparisons). B. Frequency distribution indicating the occurrence of multiple ITS genotypes within morphotaxa (global comparisons).

leotiomycetous endophytes were highly problematic in terms of morphotaxon designations and accounted for the majority of splitting of ITS groups among morphotaxa (see FIG. 2). The larger clade, comprising 84 endophytes in 10 morphotaxa but only one ITS genotype group, was recovered from all *P. taeda* individuals surveyed, and is well supported as sister of *Lophodermium pinastri* (Rhytismataceae) (FIG. 3B). This clade of endophytes (the ''*Lophodermium* clade'') contains several well supported branches on the LSUrDNA tree, indicating the presence of more than a single clone throughout our study plots. An additional group of 21 endophytes, which had highest ITS BLAST affinity for *Bisporella citrina* (Sep 2003) and an unnamed fungus from an ectomycorrhizal root tip (Sep 2004), forms a second well supported clade. Endophyte 1990 appears as the sister of the rest of the Leotiomycetes and has highest ITS BLAST affinity for *Bisporella citrina* (Sep 2004); however, without denser taxon sampling, its ordinal affinity is not clear.

Finally endophyte 2072 was strongly supported as sister of *Phaeoacremonium chlamydosporum*, and together these taxa were resolved as members of the well supported Chaetothyriales (Chaetothyriomycetidae, Eurotiomycetes) (FIG. 3B). This endophyte was placed in a morphotaxon group with various Sordariomycetes (e.g. endophyte 1829) and Leotiomycetes (e.g. endophytes 1723, 1758); however it represented a unique ITS genotype among our sequenced isolates and had highest ITS BLAST affinity for *Cladophialophora hachijoensis* (Herpotrichiellaceae, Chaetothyriales).

Recovery of endophyte taxa via environmental PCR.—LSUrDNA and ITS data were obtained for 42 clones representing surface sterilized needle samples from 12 *P. taeda* trees (Appendix 3). Nine clones were excluded from analysis, including (i) five invariable clones from a single *P. taeda* individual that had highest BLAST affinities for *Spizellomyces punctatus* (Chytridiomycota) and (ii) four invariable clones from a single *P. taeda* individual with highest ITS BLAST affinities for the lichenized taxon *Acarospora complanata* (Acarosporomycetidae). The former were excluded because of the small number of phylogenetically informative characters relative to the taxonomic breadth of our sample. The latter were excluded because studies of *Acarospora* were under way concurrently in our laboratories, suggesting possible contamination given the unexpected re-

TABLE II. Similarity of endophyte assemblages with regard to spatial aggregation of *P. taeda* trees in Duke Forest based on ITS genotype groups for nonsingleton taxa. Similarity values were defined with presence/absence (Soerensen's index) and abundance data (Morisita-Horn index) and were calculated with *EstimateS* (Colwell 1996). Chi-square and P-values reflect nonparametric comparisons of similarity for trees in the same study plots vs. trees in different study plots (three comparisons chosen randomly and without replacement from all possible pairwise comparisons of trees in different plots). Only one comparison was significant, but in this case similarity was higher between plots than within.

ITS genotype groups	Soerensen's index				Morisita-Horn index			
	Within plots	Between plots	χ^2_1	P	Within plots	Between plots	χ^2_1	P
90%	0.27 ± 0.03	0.29 ± 0.02	1.33	0.2485	0.75 ± 0.05	0.75 ± 0.16	0.02	0.8945
95%	0.20 ± 0.03	0.29 ± 0.03	3.98	0.0460	0.74 ± 0.05	0.74 ± 0.05	0.00	0.9647
99%	0.20 ± 0.07	0.22 ± 0.02	1.04	0.3086	0.66 ± 0.03	0.57 ± 0.04	0.38	0.5359

covery of these primarily saxicolous, lichenized fungi. In sum clones from 11 *P. taeda* trees were included among the 33 clones in our analyses.

Results of phylogenetic analyses of LSUrDNA for core taxa, unique LSUrDNA genotypes of cultured endophytes and clones from surface sterilized foliage are shown (FIG. 4). Cloned endophytes associated with *P. taeda* are distributed across multiple lineages of Ascomycota and Basidiomycota. Up to five unique species were recovered per study plot, and up to four species were recovered per five clones harvested for each *P. taeda* individual.

Together the 33 clones included in this analysis represented 17 unique BLAST matches and 15 distinct ITS genotypes. Randomization analyses indicated that the richness recovered by environmental PCR was significantly higher than expected ($P = 0.03$) relative to a culturing approach alone (FIG. 5). When adjusted to reflect only those plots from which cultures and clones both were obtained, environmental sampling still recovered a higher richness per sampling effort (nine genotypes per 18 clones) than culturing alone (mean = 6.6 ± 0.40 species per 18 cultures; $P < 0.0002$). A lower proportion of singletons was recovered by cloning than expected based on culturing alone, although the results were not significant ($P = 0.08$, FIG. 5). Based on species designations on the LSUrDNA tree for clones and cultures, similarity between the communities inferred by cloning vs. culturing was low (Soerensen's index = 0.312 based on 90% ITS genotype groups).

Within the Basidiomycota at least four unique species of fungi were recovered only through environmental PCR (FIG. 4A). Basidiomyceteous clones were recovered from three of 11 *P. taeda* individuals for which direct PCR data were analyzed and were found in three different study plots. This analysis resulted in a different placement of the basidiomycete clade of cultured endophytes relative to FIG. 3, although the position of this branch is not well supported in either analysis.

Ascomycota were recovered from 10 of 11 *P. taeda* trees for which environmental samples were assessed (FIG. 4B, C). Within the Ascomycota at least nine species were recovered among 26 clones. Based on our conservative species delimitations six of these species were recovered among sequenced cultures, but two were novel (clones 0253 and 0251, clone 0285). Two additional clones (0265 and 0261) and clades of clones (0279–0278 and 0246–0286) likely represent unique species but could not be distinguished given limited taxon sampling of exemplar taxa.

Ascomycetous endophytes recovered by direct PCR were primarily Dothideomycetes and Leotiomycetes (FIG. 4B, C), which were represented commonly among cultured endophytes. In addition one clone was supported strongly as sister of *Dendrographa minor* (Roccellaceae), a member of the Arthoniomycetes. Endophytic fungi were not known previously from this lineage, which includes both lichenized and nonlichenized species. The leotiomycetous endophytes commonly isolated as cultures were represented among clones (FIG. 4C), although at a lower frequency than expected based on the culturing results. No Sordariomycetes were recovered by environmental PCR, despite the prevalence of this lineage among cultured endophytes.

Phylogenetic diversity vs. traditional diversity measures.—In contrast to traditional measures of similarity (TABLE II) and diversity (TABLE III), phylogenetic diversity (PD) differed significantly among cultured endophytes at the scales of study plots and *P. taeda* trees (TABLE III).

When different methods for estimating endophyte richness and diversity were compared, richness based on morphotaxa was positively associated with richness based on ITS genotypes (90% sequence similarity and global delimitation), albeit with a low R^2 value (FIG. 6). As expected, richness of ITS genotype groups was strongly associated with diversity (H') inferred with ITS genotype groups. Although positively correlated, PD was not significantly associated with species richness or diversity estimated with ITS genotypes (FIG. 6).

DISCUSSION

The near ubiquity of endophytes in leaves of *P. taeda* and their phylogenetic breadth indicate that foliar endophytes are a cryptic and remarkably diverse component of this simplified temperate forest. New genotypes accumulated rapidly with each additional leaf and *P. taeda* tree sampled, rapidly outpacing the morphotaxon accumulation curve. Based on the lack of an asymptote in the ITS genotype accumulation curves and the concordance of ITS genotypes (90% similarity) with LSUrDNA-delimited species, it appears that many endophyte species await recovery in this forest ecosystem.

LSUrDNA analyses indicate that one phylogenetically unique taxon was recovered per every seven isolates examined, despite the high abundance and ubiquity of the *Lophodermium* clade. Representatives of at least five classes of dicaryomycota (Eurotiomycetes, Leotiomycetes, Sordariomycetes, Dothideomycetes and a Basidiomycota lineage) were identified based on phylogenetic analyses of only 22 morphologically distinct samples. Inclusion of 33 environ-

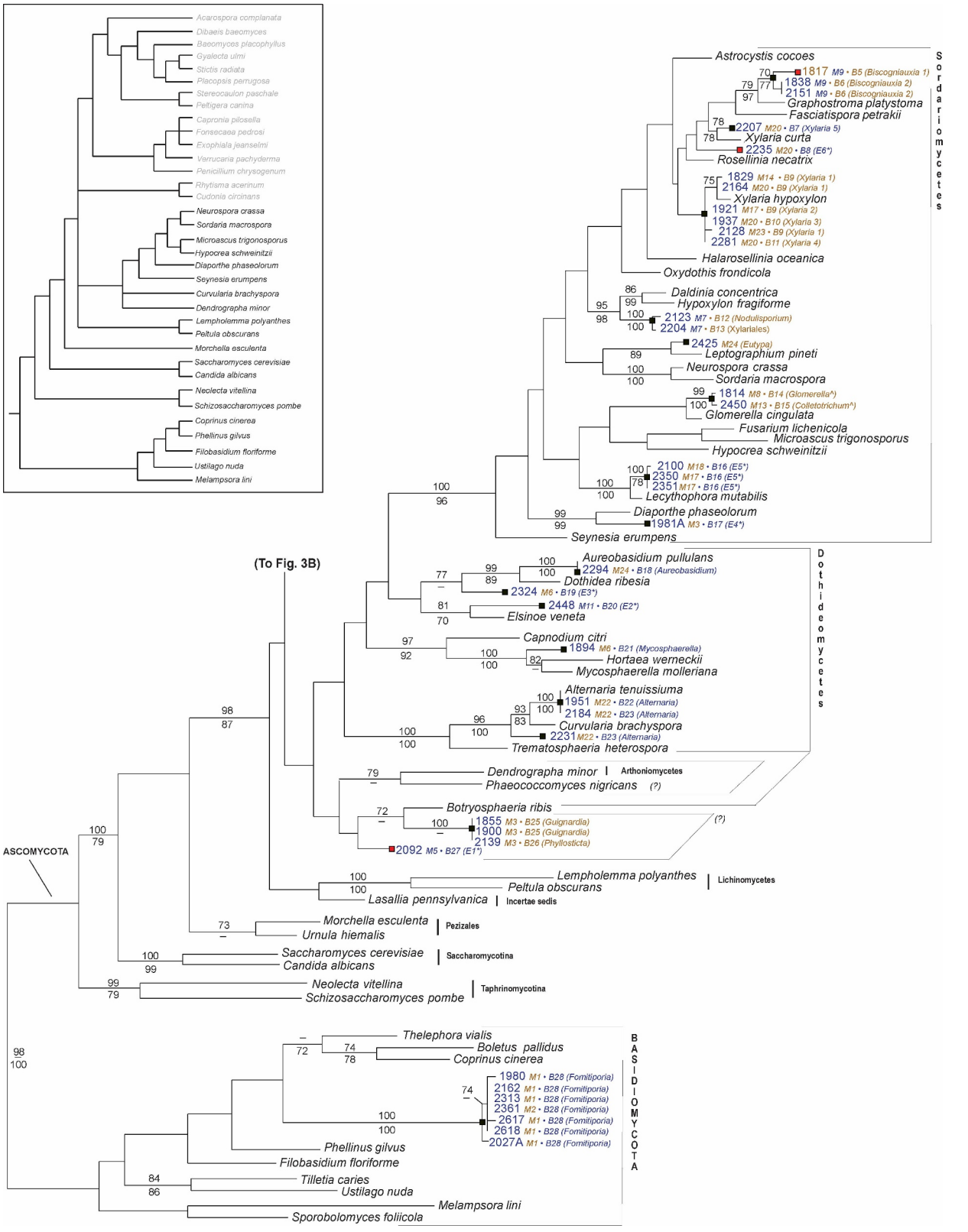
Acarospora complanata
Dibaeis baeomyces
Baeomyces placophyllus
Gyalecta ulmi
Stictis radiata
Placopsis perrugosa
Stereocaulon paschale
Peltigera canina
Capronia pilosella
Fonsecaea pedrosi
Exophiala jeanselmi
Verrucaria pachyderma
Penicillium chrysogenum
Rhytisma acerinum
Cudonia circinans
Neurospora crassa
Sordaria macrospora
Microascus trigonosporus
Hypocrea schweinitzii
Diaporthe phaseolorum
Seynesia erumpens
Curvularia brachyspora
Dendrographa minor
Lempholemma polyanthes
Peltula obscurans
Morchella esculenta
Saccharomyces cerevisiae
Candida albicans
Neolecta vitellina
Schizosaccharomyces pombe
Coprinus cinerea
Phellinus gilvus
Filobasidium floriforme
Ustilago nuda
Melampsora lini
Astrocystis cocoes
1817 M9 • B5 (Biscogniauxia 1)
1838 M9 • B6 (Biscogniauxia 2)
2151 M9 • B6 (Biscogniauxia 2)
Graphostroma platystoma
Fasciatispora petrakii
2207 M20 • B7 (Xylaria 5)
Xylaria curta
2235 M20 • B8 (E6*)
Rosellinia necatrix
1829 M14 • B9 (Xylaria 1)
2164 M20 • B9 (Xylaria 1)
Xylaria hypoxylon
1921 M17 • B9 (Xylaria 2)
1937 M20 • B10 (Xylaria 3)
2128 M23 • B9 (Xylaria 1)
2281 M20 • B11 (Xylaria 4)
Halorosellinia oceanica
Oxydothis frondicola
Daldinia concentrica
Hypoxylon fragiforme
2123 M7 • B12 (Nodulisporium)
2204 M7 • B13 (Xylariales)
2425 M24 (Eutypa)
Leptographium pineti
Neurospora crassa
Sordaria macrospora
1814 M8 • B14 (Glomerella^)
2450 M13 • B15 (Colletotrichum^)
Glomerella cingulata
Fusarium lichenicola
Microascus trigonosporus
Hypocrea schweinitzii
2100 M18 • B16 (E5*)
2350 M17 • B16 (E5*)
2351 M17 • B16 (E5*)
Lecythophora mutabilis
Diaporthe phaseolorum
1981A M3 • B17 (E4*)
Seynesia erumpens
Sordariomycetes
(To Fig. 3B)
Aureobasidium pullulans
2294 M24 • B18 (Aureobasidium)
Dothidea ribesia
2324 M6 • B19 (E3*)
2448 M11 • B20 (E2*)
Elsinoe veneta
Capnodium citri
1894 M6 • B21 (Mycosphaerella)
Hortaea werneckii
Mycosphaerella molleriana
Alternaria tenuissiuma
1951 M22 • B22 (Alternaria)
2184 M22 • B23 (Alternaria)
Curvularia brachyspora
2231 M22 • B23 (Alternaria)
Trematosphaeria heterospora
Dothideomycetes
Dendrographa minor
Phaeococcomyces nigricans (?)
Arthoniomycetes
Botryosphaeria ribis
1855 M3 • B25 (Guignardia)
1900 M3 • B25 (Guignardia)
2139 M3 • B26 (Phyllosticta)
2092 M5 • B27 (E1*)
(?)
ASCOMYCOTA
Lempholemma polyanthes
Peltula obscurans
Lasallia pennsylvanica
Incertae sedis
Lichinomycetes
Morchella esculenta
Urnula hiemalis
Pezizales
Saccharomyces cerevisiae
Candida albicans
Saccharomycotina
Neolecta vitellina
Schizosaccharomyces pombe
Taphrinomycotina
Thelephora vialis
Boletus pallidus
Coprinus cinerea
1980 M1 • B28 (Fomitiporia)
2162 M1 • B28 (Fomitiporia)
2313 M1 • B28 (Fomitiporia)
2361 M2 • B28 (Fomitiporia)
2617 M1 • B28 (Fomitiporia)
2618 M1 • B28 (Fomitiporia)
2027A M1 • B28 (Fomitiporia)
Phellinus gilvus
Filobasidium floriforme
Tilletia caries
Ustilago nuda
Melampsora lini
Sporobolomyces foliicola
BASIDIOMYCOTA

FIG. 3A

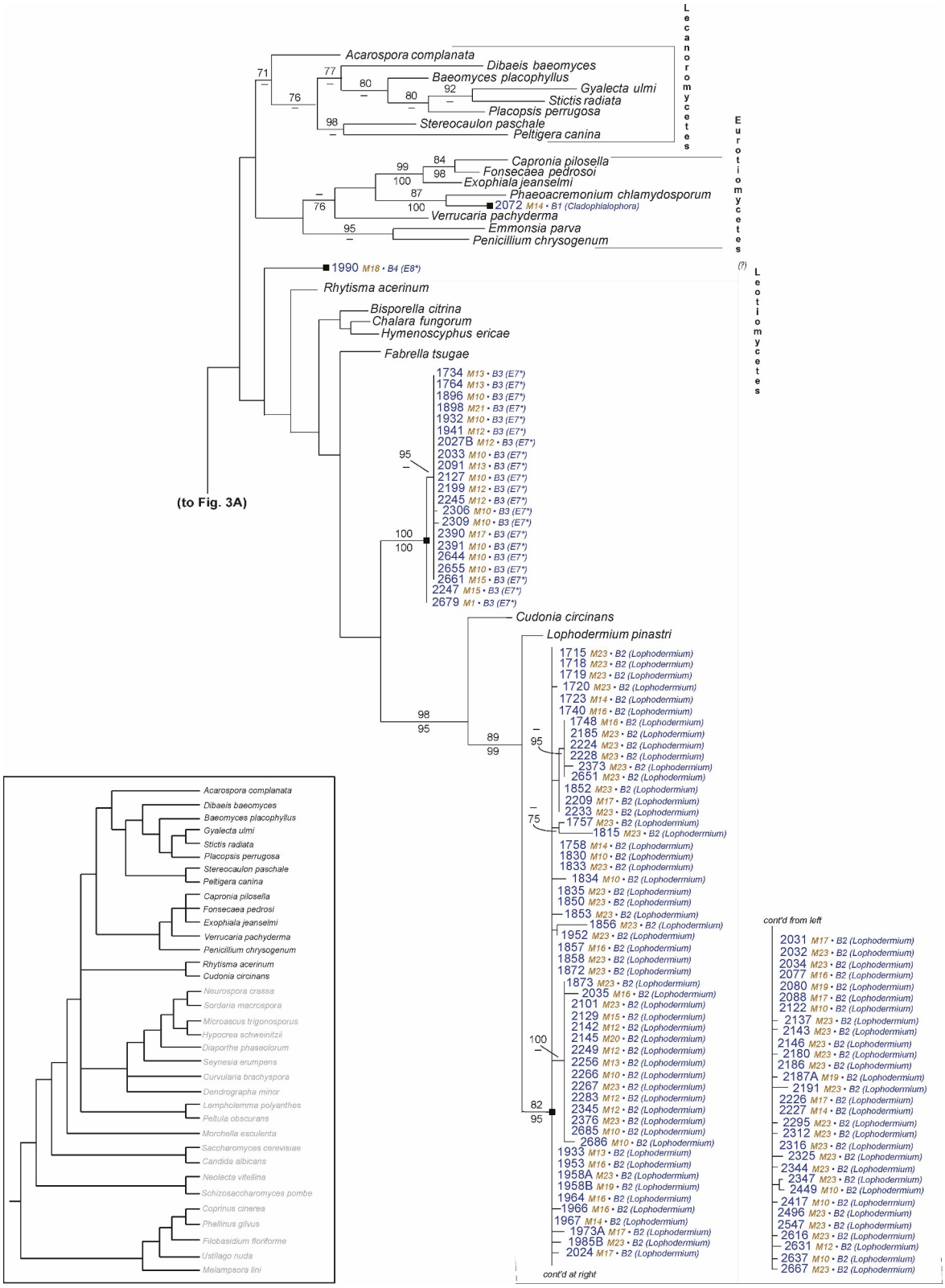
Lecanoromycetes
Eurotiomycetes
Leotiomycetes
(?)
Acarospora complanata
Dibaeis baeomyces
Baeomyces placophyllus
Gyalecta ulmi
Stictis radiata
Placopsis perrugosa
Stereocaulon paschale
Peltigera canina
Capronia pilosella
Fonsecaea pedrosoi
Exophiala jeanselmi
Phaeoacremonium chlamydosporum
2072 M14 • B1 (Cladophialophora)
Verrucaria pachyderma
Emmonsia parva
Penicillium chrysogenum
1990 M18 • B4 (E8*)
Rhytisma acerinum
Bisporella citrina
Chalara fungorum
Hymenoscyphus ericae
Fabrella tsugae
1734 M13 • B3 (E7*)
1764 M13 • B3 (E7*)
1896 M10 • B3 (E7*)
1898 M21 • B3 (E7*)
1932 M10 • B3 (E7*)
1941 M12 • B3 (E7*)
2027B M12 • B3 (E7*)
2033 M10 • B3 (E7*)
2091 M13 • B3 (E7*)
2127 M10 • B3 (E7*)
2199 M12 • B3 (E7*)
2245 M12 • B3 (E7*)
2306 M10 • B3 (E7*)
2309 M10 • B3 (E7*)
2390 M17 • B3 (E7*)
2391 M10 • B3 (E7*)
2644 M10 • B3 (E7*)
2655 M10 • B3 (E7*)
2661 M15 • B3 (E7*)
2247 M15 • B3 (E7*)
2679 M1 • B3 (E7*)
(to Fig. 3A)
Cudonia circinans
Lophodermium pinastri
1715 M23 • B2 (Lophodermium)
1718 M23 • B2 (Lophodermium)
1719 M23 • B2 (Lophodermium)
1720 M23 • B2 (Lophodermium)
1723 M14 • B2 (Lophodermium)
1740 M16 • B2 (Lophodermium)
1748 M16 • B2 (Lophodermium)
2185 M23 • B2 (Lophodermium)
2224 M23 • B2 (Lophodermium)
2228 M23 • B2 (Lophodermium)
2373 M23 • B2 (Lophodermium)
2651 M23 • B2 (Lophodermium)
1852 M23 • B2 (Lophodermium)
2209 M17 • B2 (Lophodermium)
2233 M23 • B2 (Lophodermium)
1757 M23 • B2 (Lophodermium)
1815 M23 • B2 (Lophodermium)
1758 M14 • B2 (Lophodermium)
1830 M10 • B2 (Lophodermium)
1833 M23 • B2 (Lophodermium)
1834 M10 • B2 (Lophodermium)
1835 M23 • B2 (Lophodermium)
1850 M23 • B2 (Lophodermium)
1853 M23 • B2 (Lophodermium)
1856 M23 • B2 (Lophodermium)
1952 M23 • B2 (Lophodermium)
1857 M16 • B2 (Lophodermium)
1858 M23 • B2 (Lophodermium)
1872 M23 • B2 (Lophodermium)
1873 M23 • B2 (Lophodermium)
2035 M16 • B2 (Lophodermium)
2101 M23 • B2 (Lophodermium)
2129 M15 • B2 (Lophodermium)
2142 M12 • B2 (Lophodermium)
2145 M20 • B2 (Lophodermium)
2249 M12 • B2 (Lophodermium)
2256 M13 • B2 (Lophodermium)
2266 M10 • B2 (Lophodermium)
2267 M23 • B2 (Lophodermium)
2283 M12 • B2 (Lophodermium)
2345 M12 • B2 (Lophodermium)
2376 M23 • B2 (Lophodermium)
2685 M10 • B2 (Lophodermium)
2686 M10 • B2 (Lophodermium)
1933 M13 • B2 (Lophodermium)
1953 M16 • B2 (Lophodermium)
1958A M23 • B2 (Lophodermium)
1958B M19 • B2 (Lophodermium)
1964 M16 • B2 (Lophodermium)
1966 M16 • B2 (Lophodermium)
1967 M14 • B2 (Lophodermium)
1973A M17 • B2 (Lophodermium)
1985B M23 • B2 (Lophodermium)
2024 M17 • B2 (Lophodermium)
cont'd at right
cont'd from left
2031 M17 • B2 (Lophodermium)
2032 M23 • B2 (Lophodermium)
2034 M23 • B2 (Lophodermium)
2077 M16 • B2 (Lophodermium)
2080 M19 • B2 (Lophodermium)
2088 M17 • B2 (Lophodermium)
2122 M10 • B2 (Lophodermium)
2137 M23 • B2 (Lophodermium)
2143 M23 • B2 (Lophodermium)
2146 M23 • B2 (Lophodermium)
2180 M23 • B2 (Lophodermium)
2186 M23 • B2 (Lophodermium)
2187A M19 • B2 (Lophodermium)
2191 M23 • B2 (Lophodermium)
2226 M17 • B2 (Lophodermium)
2227 M14 • B2 (Lophodermium)
2295 M23 • B2 (Lophodermium)
2312 M23 • B2 (Lophodermium)
2316 M23 • B2 (Lophodermium)
2325 M23 • B2 (Lophodermium)
2344 M23 • B2 (Lophodermium)
2347 M23 • B2 (Lophodermium)
2449 M10 • B2 (Lophodermium)
2417 M10 • B2 (Lophodermium)
2496 M23 • B2 (Lophodermium)
2547 M23 • B2 (Lophodermium)
2616 M23 • B2 (Lophodermium)
2631 M12 • B2 (Lophodermium)
2637 M10 • B2 (Lophodermium)
2667 M23 • B2 (Lophodermium)
Acarospora complanata
Dibaeis baeomyces
Baeomyces placophyllus
Gyalecta ulmi
Stictis radiata
Placopsis perrugosa
Stereocaulon paschale
Peltigera canina
Capronia pilosella
Fonsecaea pedrosi
Exophiala jeanselmi
Verrucaria pachyderma
Penicillium chrysogenum
Rhytisma acerinum
Cudonia circinans
Neurospora crassa
Sordaria macrospora
Microascus trigonosporus
Hypocrea schweinitzii
Diaporthe phaseolorum
Seynesia erumpens
Curvularia brachyspora
Dendrographa minor
Lempholemma polyanthes
Peltula obscurans
Morchella esculenta
Saccharomyces cerevisiae
Candida albicans
Neolecta vitellina
Schizosaccharomyces pombe
Coprinus cinerea
Phellinus gilvus
Filobasidium floriforme
Ustilago nuda
Melampsora lini

FIG. 3B

mental samples added at least six species to the estimate of endophyte richness and altered our conclusions regarding the representation of major fungal lineages among endophytes of *P. taeda*.

Traditional diversity and similarity indices failed to recover evidence for spatial structure in endophyte communities with regard to host individuals and study plots. In part this reflects the wide distribution of common species among hosts and sites (e.g. the *Lophodermium* clade). The occurrence of LSUrDNA species and ITS genotypes in many host individuals and plots might reflect the relative uniformity of this experimental forest, which has a monotypic canopy of consistent age. However novel ITS genotypes accumulated rapidly as additional hosts and plots were sampled, especially once sampling exceeded four trees (36 leaves). In many cases these novel ITS genotypes represented taxa that were phylogenetically distant from the more commonly recovered endophytes. Accordingly phylogenetic diversity (PD) provided strong evidence for small scale differences in endophyte communities. PD complements traditional measures of diversity, as its sensitivity to the phylogenetic breath of samples provides insight into the structure and variability of communities with regard to higher order relationships. Given recent interest in endophytes as sources of novel bioactive compounds, biological control agents and sources of genotypic diversity, PD provides a useful guide for structuring endophyte surveys in forest ecosystems.

PD measures require inclusive phylogenies and are useful when based on well resolved and well supported trees. By arc-recoding ambiguously aligned regions to extract phylogenetic signal, implementing a conservative backbone constraint tree for deep relationships among a subset of named taxa and comparing tree-to-tree distance to reduce search times, we recovered robust phylogenies for endophytes of *P. taeda* and representative Ascomycota and Basidiomycota with the LROR-LR3 region of LSUrDNA. This suite of methods, when used together, provided a rapid and comprehensive assessment of diversity and taxonomic composition of this cryptic microbial community.

Morphotaxa, ITS genotypes, and BLAST results.—Hata and Futai (1996) found that endophytes were common in foliage of *P. taeda* when sampled in an arboretum far from its native range (Japan). Based on morphotaxa, ITS genotypes and LSUrDNA-delimited species, our surveys recovered nearly twice the richness of endophytes observed in that study and provide the first characterization of the endophyte

←

FIG. 3. Phylogenetic relationships among 145 isolates of endophytic fungi obtained in culture from asymptomatic foliage of *Pinus taeda* and 72 representative species of Ascomycota and Basidiomycota. Tree represents the strict consensus of 5042 most parsimonious trees (tree length = 3807.39) inferred from ca. 600 bp at the 5′ end of the nuclear large subunit (LSUrDNA) using arc-recoded characters and a conservative backbone constraint (inset). A. Basidiomycota, early diverging lineages of Ascomycota (Taphrinomycotina, Saccharomycotina and Pezizales), Lichinomycetes, Sordariomycetes, Dothideomycetes and Arthoniomycetes, with the relevant portion of the constraint tree presented with black text in the upper left inset. B. Leotiomycetes, Eurotiomycetes and Lecanoromycetes, with the relevant portion of the constraint tree presented with black text in the lower left inset. OTU that cannot be placed with confidence in particular lineages are designated with a question mark. Bootstrap values ≥70% are shown above branches (maximum parsimony bootstrap, MP) and below branches (neighbor joining bootstrap, NJ). Branch lengths were inferred with maximum parsimony; branch lengths estimated by distance for phylogenetic diversity measurements are shown (Appendix 5). Black boxes on branches indicate conservatively placed taxonomic boundaries for endophytes; all endophytes distal to a black box are considered part of unique species relative to those distal to other black boxes. Red boxes on branches indicate cases in which the designation is uncertain; endophytes distal to these boxes likely represent unique species relative to other closely related endophytes. Endophyte numbers (given as four-digit identification numbers) are color-coded to indicate the quality of their ITS genotype group, morphotaxon (given as M followed by a one- or two-digit value) and BLAST group designations (given as B, followed by a one- or two-digit value), relative to the designation of species boundaries among endophytes inferred from the LSUrDNA tree. Endophyte numbers are blue if LSUrDNA taxon boundaries are consistent with ITS genotype groups defined by 90% sequence similarity using global comparisons; numbers are brown if inconsistent with these ITS groups (see Appendix 2 for ITS groups). Morphotaxon numbers are blue if consistent with taxonomic boundaries inferred with the LSUrDNA tree but brown if inconsistent (see Appendix 2 for morphotaxon numbers). BLAST groups are blue if consistent with LSUrDNA taxa (see Appendix 2 for BLAST matches). The top BLAST match for each isolate is indicated; where possible the genus of the top match is given. Endophytes whose ITS data showed highest BLAST affinity to environmental samples are listed as *E*, with an asterisk indicating a match coherent with the LSUrDNA taxonomy. Cases in which isolates had highest BLAST affinity for members of different genera that are known to be anamorphs/teleomorphs are indicated with ^ after the genus name.

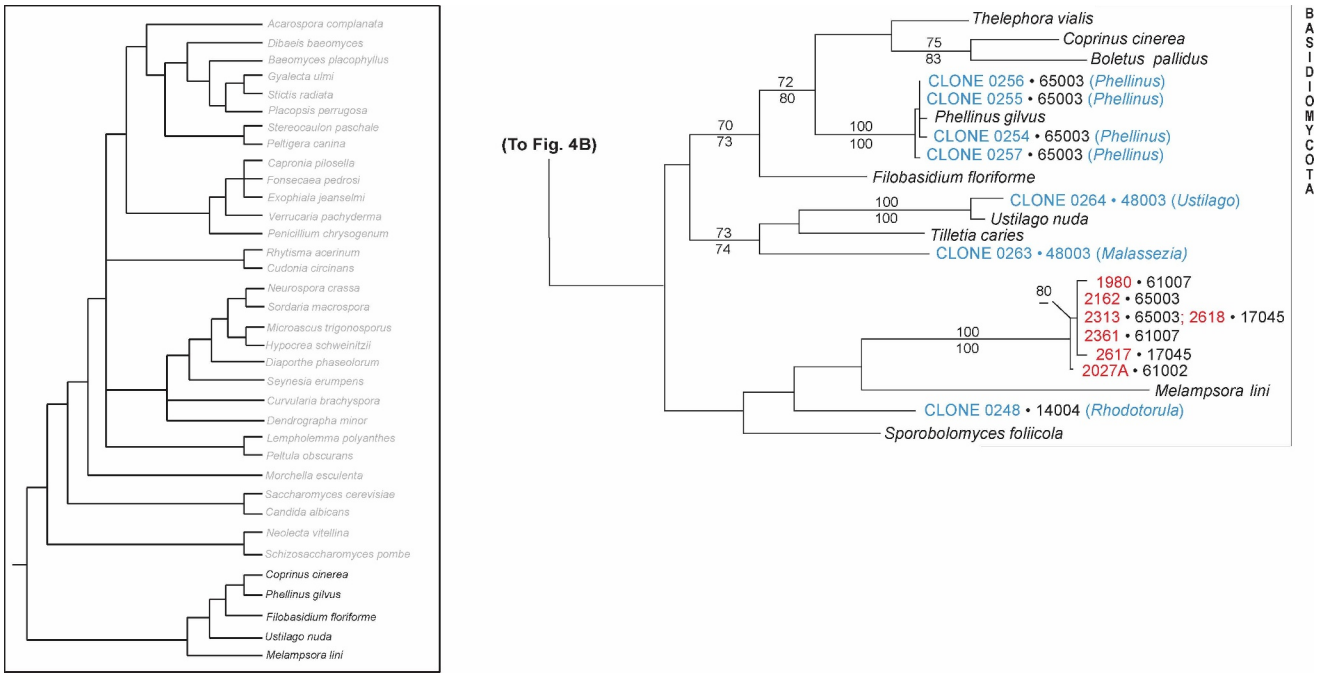

FIG. 4A

community associated with this ecologically and economically important host within its natural range.

Although morphotaxa were adequate for estimating species richness, they were generally incongruent with LSUrDNA-delimited species. In contrast the concordance between ITS genotypes (90% similarity, global delimitation) and LSUrDNA-delimited taxa indicates that ITS genotype groups can serve as useful, if highly conservative, designators of species boundaries. Grouping ITS sequences into genotype groups, especially under the global criterion (as implemented in Sequencher) is straightforward and convenient for estimating species richness. However it should be stressed that there is no threshold value of sequence similarity that is universally useful for distinguishing species of fungi, such that ITS genotypes are best employed as a proxy for estimating species boundaries. Different biological species of fungi frequently have identical ITS sequences (Lieckfeldt and Seifert 2000) and cryptic species sometimes may be revealed only through a phylogenetic approach based on other loci (Taylor et al 2000). Relatively high ITS variability also has been observed within many recognized species of fungi (see Vilgalys and Sun 1994). Interestingly neither ITS genotypes nor LSUrDNA-delimited species conformed to the expected log-normal distribution characteristic of endophyte studies based on traditional species concepts (Lodge et al 1996, Gamboa and Bayman 2001), in part due to skew induced by rare genotypes. It will be of interest to assess whether log-normal distributions hold for other symbiotic microbes and/or persist when molecular data are included.

By coupling ITS BLAST results with a phylogenetic approach based on LSUrDNA data, we explicitly assessed the phylogenetic and taxonomic accuracy of BLAST matches, which are used often to identify sterile fungal isolates or environmental samples. The 145 cultured isolates included in the LSUrDNA analysis represented 29 unique BLAST matches, corresponding to ≥20 different species in GenBank. Our results indicated 20–23 unique species of cultured endophytes based on the LSUrDNA tree, and ITS BLAST-based taxonomy was highly concordant with LSUrDNA phylotypes. These data suggest that GenBank matches, if based on well identified taxa, can be sufficient for estimating species richness and upper level taxonomic placement. However unique matches do not necessarily translate into unique species. The prevalence of unnamed samples in GenBank, the occurrence of misidentified taxa and the rapid growth of the database, which yielded highly divergent matches at the genus and family levels when BLAST results from 2003 and 2004 were compared, underscore the need for caution when estimating taxonomic composition based on BLAST results alone.

Many endophytes recovered from *P. taeda* had highest ITS BLAST affinities for unnamed species in GenBank (coded as environmental isolates on FIG. 3) and were reconstructed as distinct lineages relative to

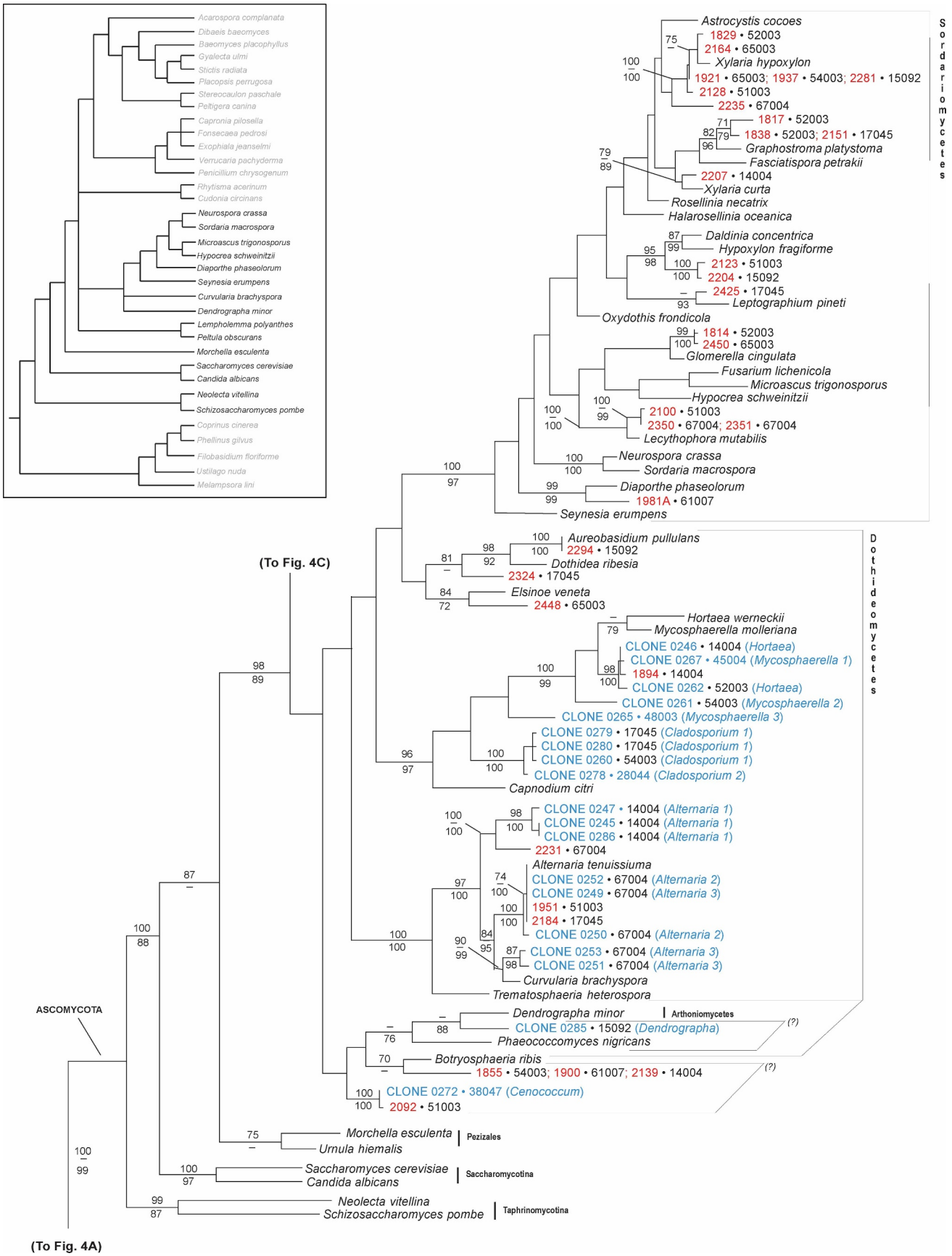
Acarospora complanata
Dibaeis baeomyces
Baeomyces placophyllus
Gyalecta ulmi
Stictis radiata
Placopsis perrugosa
Stereocaulon paschale
Peltigera canina
Capronia pilosella
Fonsecaea pedrosi
Exophiala jeanselmi
Verrucaria pachyderma
Penicillium chrysogenum
Rhytisma acerinum
Cudonia circinans
Neurospora crassa
Sordaria macrospora
Microascus trigonosporus
Hypocrea schweinitzii
Diaporthe phaseolorum
Seynesia erumpens
Curvularia brachyspora
Dendrographa minor
Lemphoiemma polyanthes
Peltula obscurans
Morchella esculenta
Saccharomyces cerevisiae
Candida albicans
Neolecta vitellina
Schizosaccharomyces pombe
Coprinus cinerea
Phellinus gilvus
Filobasidium floriforme
Ustilago nuda
Melampsora lini
Sordariomycetes
Astrocystis cocoes
1829 • 52003
2164 • 65003
Xylaria hypoxylon
1921 • 65003; 1937 • 54003; 2281 • 15092
2128 • 51003
2235 • 67004
1817 • 52003
1838 • 52003; 2151 • 17045
Graphostroma platystoma
Fasciatispora petrakii
2207 • 14004
Xylaria curta
Rosellinia necatrix
Halarosellinia oceanica
Daldinia concentrica
Hypoxylon fragiforme
2123 • 51003
2204 • 15092
2425 • 17045
Leptographium pineti
Oxydothis frondicola
1814 • 52003
2450 • 65003
Glomerella cingulata
Fusarium lichenicola
Microascus trigonosporus
Hypocrea schweinitzii
2100 • 51003
2350 • 67004; 2351 • 67004
Lecythophora mutabilis
Neurospora crassa
Sordaria macrospora
Diaporthe phaseolorum
1981A • 61007
Seynesia erumpens
(To Fig. 4C)
Dothideomycetes
Aureobasidium pullulans
2294 • 15092
Dothidea ribesia
2324 • 17045
Elsinoe veneta
2448 • 65003
Hortaea werneckii
Mycosphaerella molleriana
CLONE 0246 • 14004 (Hortaea)
CLONE 0267 • 45004 (Mycosphaerella 1)
1894 • 14004
CLONE 0262 • 52003 (Hortaea)
CLONE 0261 • 54003 (Mycosphaerella 2)
CLONE 0265 • 48003 (Mycosphaerella 3)
CLONE 0279 • 17045 (Cladosporium 1)
CLONE 0280 • 17045 (Cladosporium 1)
CLONE 0260 • 54003 (Cladosporium 1)
CLONE 0278 • 28044 (Cladosporium 2)
Capnodium citri
CLONE 0247 • 14004 (Alternaria 1)
CLONE 0245 • 14004 (Alternaria 1)
CLONE 0286 • 14004 (Alternaria 1)
2231 • 67004
Alternaria tenuissiuma
CLONE 0252 • 67004 (Alternaria 2)
CLONE 0249 • 67004 (Alternaria 3)
1951 • 51003
2184 • 17045
CLONE 0250 • 67004 (Alternaria 2)
CLONE 0253 • 67004 (Alternaria 3)
CLONE 0251 • 67004 (Alternaria 3)
Curvularia brachyspora
Trematosphaeria heterospora
ASCOMYCOTA
Dendrographa minor
Arthoniomycetes
CLONE 0285 • 15092 (Dendrographa)
(?)
Phaeococcomyces nigricans
Botryosphaeria ribis
1855 • 54003; 1900 • 61007; 2139 • 14004
(?)
CLONE 0272 • 38047 (Cenococcum)
2092 • 51003
Morchella esculenta
Pezizales
Urnula hiemalis
Saccharomyces cerevisiae
Saccharomycotina
Candida albicans
Neolecta vitellina
Taphrinomycotina
Schizosaccharomyces pombe
(To Fig. 4A)

FIG. 4B

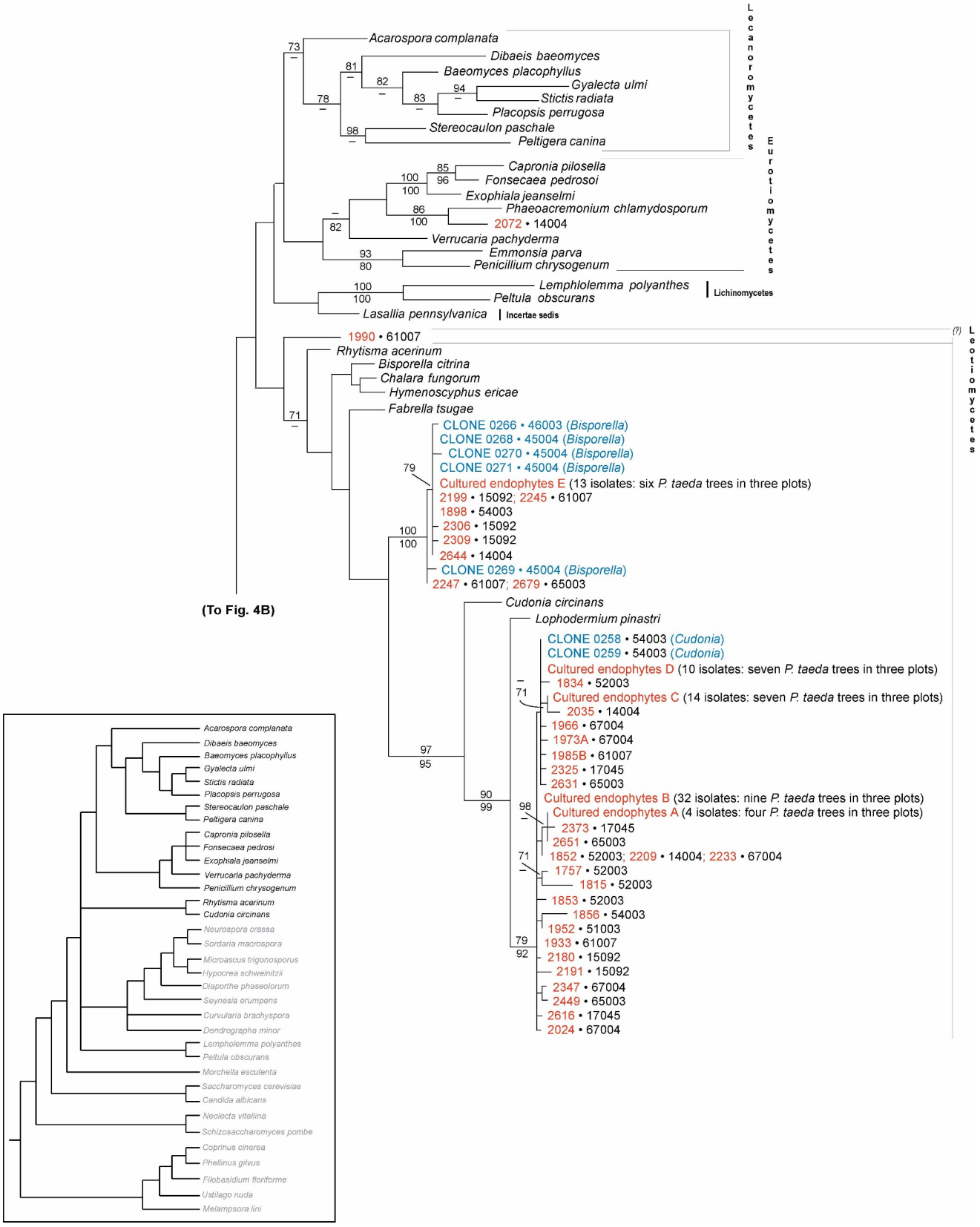
Acarospora complanata
Dibaeis baeomyces
Baeomyces placophyllus
Gyalecta ulmi
Stictis radiata
Placopsis perrugosa
Stereocaulon paschale
Peltigera canina
Lecanoromycetes
Capronia pilosella
Fonsecaea pedrosoi
Exophiala jeanselmi
Phaeoacremonium chlamydosporum
2072 • 14004
Verrucaria pachyderma
Emmonsia parva
Penicillium chrysogenum
Eurotiomycetes
Lemphlolemma polyanthes
Peltula obscurans
Lichinomycetes
Lasallia pennsylvanica
Incertae sedis
1990 • 61007
(?)
Leotiomycetes
Rhytisma acerinum
Bisporella citrina
Chalara fungorum
Hymenoscyphus ericae
Fabrella tsugae
CLONE 0266 • 46003 (Bisporella)
CLONE 0268 • 45004 (Bisporella)
CLONE 0270 • 45004 (Bisporella)
CLONE 0271 • 45004 (Bisporella)
Cultured endophytes E (13 isolates: six P. taeda trees in three plots)
2199 • 15092; 2245 • 61007
1898 • 54003
2306 • 15092
2309 • 15092
2644 • 14004
CLONE 0269 • 45004 (Bisporella)
2247 • 61007; 2679 • 65003
(To Fig. 4B)
Cudonia circinans
Lophodermium pinastri
CLONE 0258 • 54003 (Cudonia)
CLONE 0259 • 54003 (Cudonia)
Cultured endophytes D (10 isolates: seven P. taeda trees in three plots)
1834 • 52003
Cultured endophytes C (14 isolates: seven P. taeda trees in three plots)
2035 • 14004
1966 • 67004
1973A • 67004
1985B • 61007
2325 • 17045
2631 • 65003
Cultured endophytes B (32 isolates: nine P. taeda trees in three plots)
Cultured endophytes A (4 isolates: four P. taeda trees in three plots)
2373 • 17045
2651 • 65003
1852 • 52003; 2209 • 14004; 2233 • 67004
1757 • 52003
1815 • 52003
1853 • 52003
1856 • 54003
1952 • 51003
1933 • 61007
2180 • 15092
2191 • 15092
2347 • 67004
2449 • 65003
2616 • 17045
2024 • 67004
Acarospora complanata
Dibaeis baeomyces
Baeomyces placophyllus
Gyalecta ulmi
Stictis radiata
Placopsis perrugosa
Stereocaulon paschale
Peltigera canina
Capronia pilosella
Fonsecaea pedrosi
Exophiala jeanselmi
Verrucaria pachyderma
Penicillium chrysogenum
Rhytisma acerinum
Cudonia circinans
Neurospora crassa
Sordaria macrospora
Microascus trigonosporus
Hypocrea schweinitzii
Diaporthe phaseolorum
Seynesia erumpens
Curvularia brachyspora
Dendrographa minor
Lempholemma polyanthes
Peltula obscurans
Morchella esculenta
Saccharomyces cerevisiae
Candida albicans
Neolecta vitellina
Schizosaccharomyces pombe
Coprinus cinerea
Phellinus gilvus
Filobasidium floriforme
Ustilago nuda
Melampsora lini

FIG. 4C

TABLE III. Taxonomic richness (S_{obs}), Shannon diversity (H'), and phylogenetic diversity (PD) for endophytes inhabiting mature, asymptomatic foliage for study plots and *P. taeda* trees in Duke Forest. Data reflect numbers of isolates per partition for which ITS sequence data were gathered (N), number of nonsingleton morphotaxa (morphotaxa that were isolated more than once, M) and numbers of operational taxonomic units inferred with 90%, 95% and 99% sequence similarity values from global and pairwise comparisons (i.e. ITS genotypes; $N = 150$ isolates). PD values were calculated for individual *P. taeda* trees and plots by summing branch lengths for each partition inferred with neighbor joining distance on the strict consensus tree obtained from parsimony searches of the 217 OTU dataset (72 core taxa and 145 cultured endophytes; topology as in Fig. 3 and Appendix 5, branch lengths as in Appendix 5). Significantly different PD values, determined using nonparametric Tukey-Kramer tests with a Bonferroni correction for multiple comparisons, are indicated with differing superscripts. Different numerical superscripts indicate significant differences between trees; different letters indicate significant differences between plots

			S_{obs}							Diversity (H')							
				Global			Pairwise				Global			Pairwise			
Plot	Tree	N	M	90%	95%	99%	90%	95%	99%	M	90%	95%	99%	90%	95%	99%	PD
1	14004	14	8	7	7	7	7	7	9	1.86	1.45	1.45	1.45	1.45	1.45	1.96	0.340 ± 0.009^{1}
	15092	12	9	6	6	6	6	6	7	2.09	1.47	1.47	1.47	1.47	1.47	1.69	0.587 ± 0.010^{2}
	17045	19	9	9	9	10	9	10	11	1.90	1.82	1.82	1.82	1.82	1.89	2.05	0.794 ± 0.014^{3}
	All	45	16	18	18	19	18	20	23	2.33	2.06	2.06	2.15	2.06	2.06	2.62	0.574 ± 0.012^{a}
5	51003	18	13	9	9	9	10	10	11	2.42	1.82	1.82	1.95	1.82	1.82	2.11	0.489 ± 0.011^{4}
	52003	16	6	5	5	6	5	5	12	1.36	1.03	1.03	1.24	1.24	1.24	2.39	0.400 ± 0.008^{2}
	54003	20	10	6	6	6	8	10	13	2.01	1.40	1.40	1.40	1.67	1.86	2.31	0.417 ± 0.008^{5}
	All	54	20	16	16	18	19	22	29	2.44	1.83	1.83	1.96	2.01	2.14	2.89	0.436 ± 0.006^{b}
6	61007	18	9	6	6	7	6	7	10	2.23	1.58	1.58	1.73	1.39	1.55	2.03	0.638 ± 0.013^{6}
	65003	16	10	6	6	9	6	8	12	2.18	1.60	1.60	2.01	1.60	1.98	2.39	0.766 ± 0.012^{7}
	67004	17	6	3	3	4	3	3	8	0.56	0.55	0.55	0.76	0.55	0.55	1.79	0.395 ± 0.009^{2}
	All	51	16	10	10	14	11	14	25	2.40	1.57	1.57	1.85	1.62	1.84	2.72	$0.600 \pm 0.011^{a, c}$

named species in our LSUrDNA analyses. These results argue for the continued integration of environmental isolates and little known groups of symbiotic microfungi into a broad phylogenetic context, both as a means to identify these unknown fungi, and for adequately inferring the fungal tree of life. Equally important is the integration of sequence data for reliably identified fungi into existing sequence databases. Perhaps most important, a critical step remains for future studies: to reconcile traditional taxonomic approaches for those fungi that do fruit in culture with the molecular approaches used here.

←

FIG. 4. Phylogenetic relationships among 66 unique genotypes of cultured endophytic fungi (blue), 33 endophyte sequences obtained via environmental PCR of surface sterilized foliage (red) and 72 representative species of Ascomycota and Basidiomycota (black). Tree represents strict consensus of 24 327 most parsimonious trees (tree length = 4178.39) inferred with arc-recoded characters and a conservative backbone constraint (inset). A. Basidiomycota, with the relevant portion of the constraint tree used in analyses depicted with black text in the inset. B. Early diverging lineages of Ascomycota and Sordariomycetes, Dothideomycetes and Arthoniomycetes, with the relevant portion of the constraint tree (inset). C. Leotiomycetes, Eurotiomycetes and Lecanoromycetes, with the relevant portion of the constraint tree. Cases in which numerous genotypes obtained by culturing were collapsed into a single OTU are shown in red with notes indicating the number of isolates, host trees, and plots represented within the OTU. Bootstrap values ≥70% are shown above branches (parsimony) or below branches (NJ). Both cultured and cloned endophytes are annotated to indicate the plot and host individual from which they were obtained; five-digit numbers indicate host trees, with numbers in black text indicating plots from which cultures and clones were obtained, and numbers in blue text indicating plots in which only clones were analyzed. Tree IDs correspond to tag numbers for *P. taeda* individuals at the Duke Forest FACE site, with plots designated by the first digit in each tree ID. Genera of top BLAST matches are provided for all cloned endophytes, with numbers after genus names indicating different top-matching species within the same given genus (see Appendix 3). C. Genotype groups A–E include these isolates: A. 1748, 2185, 2224, 2228; B. 1715, 1718, 1719, 1723, 1740, 1758, 1830, 1833, 1835, 1850, 1857, 1858, 1872, 1953, 1958A, 1958B, 1964, 1967, 2031, 2032, 2034, 2077, 2080, 2088, 2122, 2146, 2186, 2226, 2227, 2316, 2496, 2547; C. 1873, 2101, 2129, 2142, 2145, 2249, 2256, 2266, 2267, 2283, 2345, 2376, 2685, 2686; D. 1720, 2137, 2143, 2187A, 2295, 2312, 2344, 2417, 2637, 2667; E. 1734, 1764, 1896, 1932, 1941, 2027B, 2033, 2091, 2127, 2390, 2391, 2655, 2661.

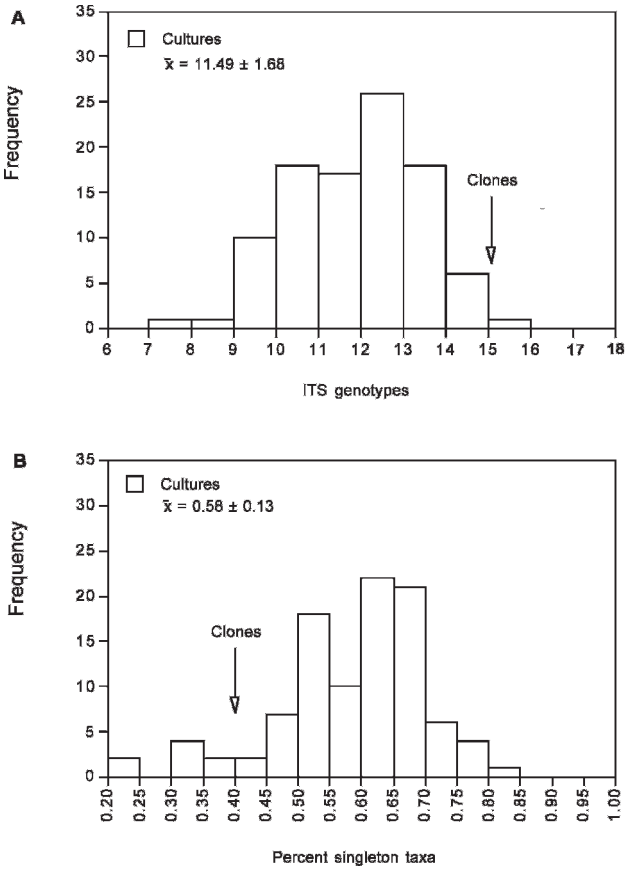

FIG. 5. Results of randomization analyses for partitions of 42 cultured isolates sampled randomly and with replacement from all cultured isolates for which ITS sequence data were obtained. A. Frequency distribution of the number of ITS genotypes, defined with global comparisons at 90% sequence similarity, encountered among 42 isolates (mean = 11.49 ± 1.68 genotypes, generated from 1000 replicate randomizations), compared with observed richness among endophytes amplified directly by environmental sampling of asymptomatic foliage (N = 42 clones) (P = 0.03). B. Frequency distribution of the proportion of singleton genotypes recovered among partitions of 42 cultures (mean = 0.58 ± 0.13 singletons from 1000 replicate randomizations), compared with the observed proportion of singletons among cloned endophytes (N = 42 clones) (P = 0.08).

Inferring phylogenetic relationships and taxonomic boundaries of endophytes.—Large studies assessing community structure of endophytes or other microfungi generally require that sequence data be generated for a large number of taxa. Because identification of environmental samples, including unnamed cultures or clones obtained by environmental PCR, often requires quickly evolving loci, many studies will be constrained by the degree to which deeper relationships and thus taxonomic placement of samples can be resolved. The backbone constraint method is useful in this regard. However should an incorrect topology be used to build the constraint, the position of unknown taxa will be biased in ways that might be difficult to detect. We implemented a constraint that was (i) supported by previous, multilocus studies, (ii) conservative, in that only a small subset of named taxa was included, and (iii) conservative in that some relationships, although constrained at deeper levels, were allowed to exist in the constraint tree as polytomies. These cautionary steps, coupled with supporting evidence from ITS BLAST results, lead us to view the placement of endophytes as presented (FIG. 3, 4) with confidence.

Many endophytes recovered from *P. taeda* represent orders of fungi known among endophytes of other conifers, including Chaetothyriales, Leotiales and Xylariales (e.g. Carroll and Carroll 1978, Petrini and Carroll 1981). In addition endophytic Phyllachorales, Sordariales, Pleosporales and Dothideales have been reported frequently from other hosts (see Alexopoulos et al 1996) and have been recovered from angiosperm hosts in Duke Forest (Arnold unpublished data). Basidiomycota are relatively rare as cultured endophytes but have been recorded previously from coniferous hosts such as *Juniperus communis* and *Pinus cembra* (Petrini and Müller 1979, Petrini 1986). Basidiomycetous endophytes obtained in our study lacked clamp connections and could not have been identified to phylum based on cultural characteristics alone.

Survey data for poorly known symbionts are usually insufficient for developing a fully phylogenetic approach for delimiting species boundaries because such datasets typically lack thorough geographic and taxon sampling. Our approach to circumscribing species boundaries on the LSUrDNA phylogeny is conservative, in that it does not necessarily allow endophytes on long branches to be distinguished as unique species relative to one another. In addition this method is sensitive to taxon sampling, with richness estimates likely to increase as sampling of named taxa becomes denser. This approach provides a minimum estimate of the number of phylogenetically distinct endophytes and permits comparisons with morphotaxa, ITS genotype groups and ITS BLAST matches.

While the approaches presented here are useful for rapid surveys, they are limited in terms of their utility in integrating little known symbionts into the broader fungal tree of life. In particular loci that are phylogenetically informative across the dicaryomycota will be necessary if different optimization criteria are to be used to infer the placement of little known symbionts (e.g. likelihood, Bayesian methods): given

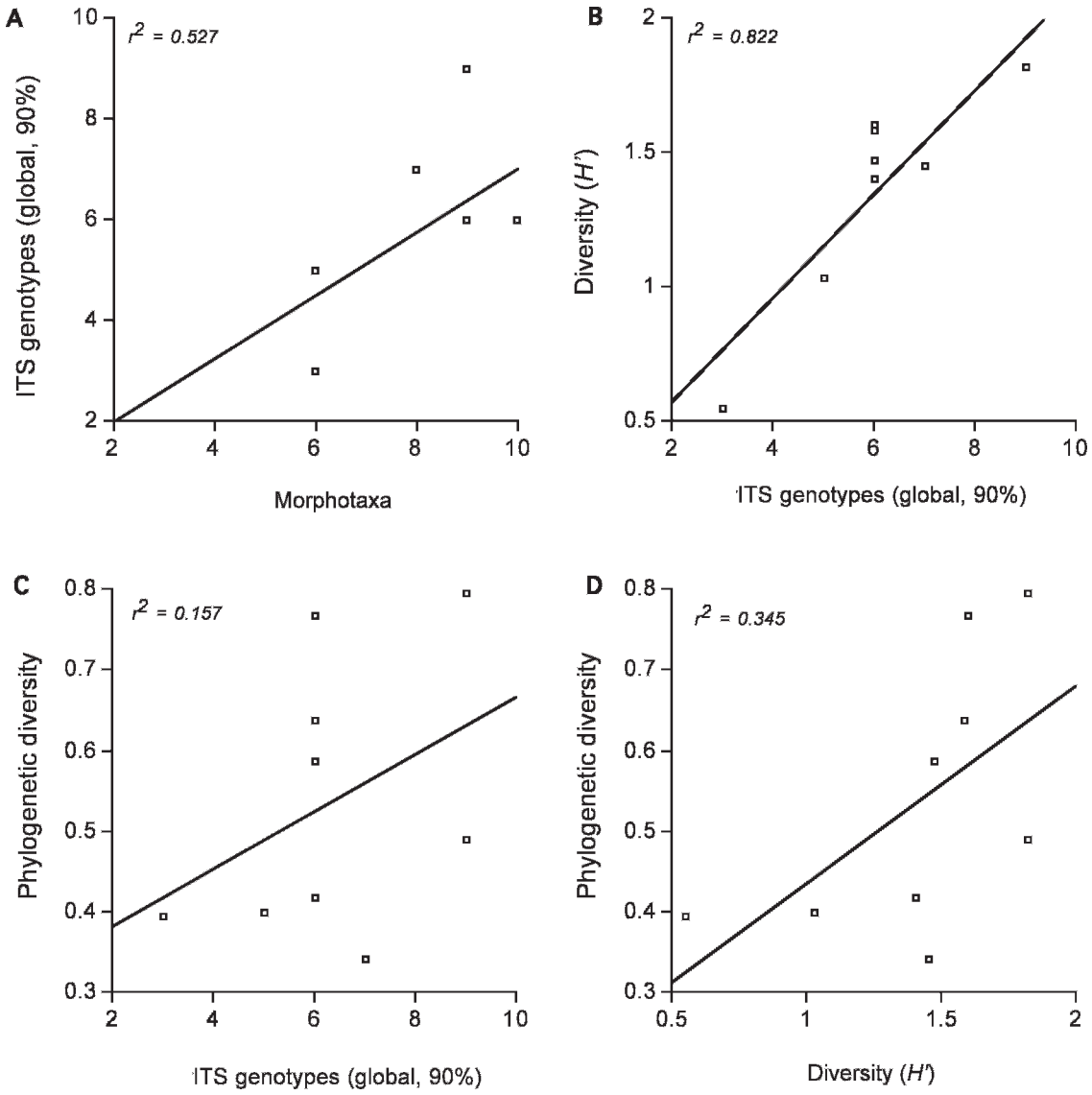

FIG. 6. Relationships of methods for assessing endophyte richness and/or diversity. Although morphotaxa, ITS genotypes, Shannon diversity (*H'*) and phylogenetic diversity are positively correlated, each measure is a relatively poor predictor of the others. A. Relationship of nonsingleton morphotaxa to ITS genotype designations (90% sequence similarity using global comparisons) for common isolates obtained from *P. taeda* trees. B. Relationship between ITS genotype groups and diversity (Shannon index, *H'*) based on ITS groups. C. Relationship between ITS genotype groups and phylogenetic diversity inferred from LSUrDNA data. D. Relationship between diversity (Shannon index, based on ITS genotypes groups using global comparisons and 90% sequence similarity) and phylogenetic diversity (PD) inferred from LSUrDNA analyses.

current technology, arc-recoding and backbone constraints can only be implemented together under parsimony. In particular Bayesian methods will allow phylogenetic uncertainty to be incorporated into PD analyses, increasing the statistical power of this approach to assessing community structure (Henk et al., in prep). We recommend that a minimum of 1.4 kb at the 5′ end of the LSUrDNA (LROR-LR7), in addition to ITS, be sequenced in future endophyte studies.

Recovery of taxa by direct PCR.—Environmental sampling has gained popularity as a means to assess the diversity and composition of microbial communities. Allen et al (2003) showed that inferences regarding the composition of rhizosphere fungal communities are complementary when inferred with both an environmental-PCR and culturing approach. Our data suggest that culturing alone will underestimate endophyte diversity and misrepresent the taxonomic composition of foliar endophyte commu-

nities. In particular the increased prevalence of basidiomycetous endophytes recovered by environmental sampling suggests that these fungi might occur more frequently as endophytes than has been inferred previously. Conversely the lack of Sordariomycetes among our cloned isolates, despite their prevalence among cultured endophytes, warrants further investigation. Whether unsuccessful recovery by environmental PCR indicates PCR bias during cloning or that some fungi have relatively low biomass but then grow quickly in culture (e.g. various Sordariomycetes) remains to be resolved. Such questions are key to understanding the diversity and ecology of plant-endophyte symbioses and will be critical as a basis for assessing the evolution of fungal symbiotic lifestyles.

ACKNOWLEDGMENTS

We thank A. Paulick, L. Bukovnik, S. Sarvate, S. Hill, K.L. Higgins and M. McGregor for technical assistance, J. Stajich for sharing his BioPERL script, C. Cox for developing tools for assessing PD, F. Kauff for scriptwriting to automate aspects of our analyses, J. Parrent, H. O'Brien, D. R. Maddison and the Endophyte Discussion Group at Duke University (E.C. Davis, K. Ryall, J. Miadlikowska, M. Hersch, T.S. Feldman and K.L. Higgins) for helpful discussion, and J. Miadlikowska and V. Reeb for assistance with sequence alignment. For financial support we thank the National Science Foundation: NSF Postdoctoral Fellowship in Microbial Biology DEB-0200413 (AEA); DEB-0342925 (AEA); NSF Career Grant DEB-0133891 (FL); NSF Assembling the Tree of Life grant DEB-0228668 (FL and RV); and NSF Microbial Observatories grant DEB-0084207 (RV). We gratefully acknowledge the Howard Hughes Summer Fellowship Program at Duke University for supporting undergraduate research (RLE).

LITERATURE CITED

Alexopoulos CJ, Mims CW, Blackwell M. 1996. Introductory Mycology. New York: Wiley & Sons.

Allen TR, Millar T, Berch SM, Berbee ML. 2003. Culturing and direct DNA extraction find different fungi from the same ericoid mycorrhizal roots. New Phytol 160: 255–272.

Arnold AE. 2002. Neotropical fungal endophytes: diversity and ecology [Doctoral dissertation]. Tucson: University of Arizona. 337 p.

———, Maynard Z, Gilbert GS, Coley PD, Kursar TA. 2000. Are tropical fungal endophytes hyperdiverse? Ecol Lett 3:267–274.

———, ———, ———. 2001. Fungal endophytes in dicotyledonous neotropical trees: patterns of abundance and diversity. Mycol Res 105:1502–1507.

———, Mejía LC, Kyllo D, Rojas E, Maynard Z, Robbins NA, Herre EA. 2003. Fungal endophytes limit pathogen damage in a tropical tree. Proc Nat Acad Sci USA 100: 15649–15654.

———, Lewis LC. 2005. Ecology and evolution of fungal endophytes, and their role against insects. In: Vega R, Blackwell M, eds. Ecological and Evolutionary Advances in Insect-Fungus Associations. Oxford, UK: Oxford University Press. p 74–96.

Berbee ML. 2001. The phylogeny of plant and animal pathogens in the Ascomycota. Physiol Mol Plant Path 59:165–187.

Cannone JJ, Subramanian S, Schnare MN, Collett JR, R'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Müller KM, Pande N, Shang Z, Yu N, Guttell RR. 2002. The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BioMed Cent Bioinformat 3:15.

Carey WA, Kelly WD. 1994. Interaction of ozone exposure and *Fusarium subglutinans* inoculation on growth and disease development of loblolly pine seedlings. Environ Pollut 84:35–43.

Carroll GC. 1986. The biology of endophytism in plants with particular reference to woody perennials. In: Fokkema NJ, van den Huevel J, eds. Microbiology of the Phyllosphere. Cambridge, UK: Cambridge University Press. p 205–222.

———. 1995. Forest endophytes—pattern and process. Can J Bot 73:S1316–S1324.

———, Carroll FE. 1978. Studies on the incidence of coniferous needle endophytes in the Pacific Northwest. Can J Bot 56:3034–3043.

Deckert RJ, Peterson RL. 2000. Distribution of foliar fungal endophytes of *Pinus strobus* between and within host trees. Can J For Res 30:1436–1442.

Dobranic JK, Johnson JA, Alikhan QR. 1995. Isolation of endophytic fungi from eastern larch (*Larix laricina*) leaves from New Brunswick, Canada. Can J Microbiol 41:194–198.

Ewing B, Hillier L, Wendl LC, Green P. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res 8:186–194.

Faith DP. 1992. Conservation evaluation and phylogenetic diversity. Biolog Conserv 61:1–10.

Felsenstein J. 1981. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. Biolog J Linnaean Soc 16:183–196.

Fröhlich J, Hyde KD. 1999. Biodiversity of palm fungi in the tropics: are global fungal diversity estimates realistic? Biodivers Conser 8:977–1004.

Gamboa MA, Bayman P. 2001. Communities of endophytic fungi in leaves of a tropical timber tree (*Guarea guidonia*: Meliaceae). Biotropica 33:352–360.

Gardes M, Bruns TD. 1993. ITS primers with enhanced specificity of basidiomycetes: application to the identification of mycorrhizae and rusts. Mol Ecol 2:113–118.

Gilbert GS. 2003. Evolutionary ecology of plant diseases in natural systems. Ann Rev Phytopathol 40:13–43.

Guo LD, Hyde KD, Liew ECY. 2000. Identification of endophytic fungi from *Livistona chinensis* based on morphology and rDNA sequences. New Phytolog 147: 617–630.

———, ———, ———. 2001. Detection and taxonomic

placement of endophytic fungi within frond tissues of *Livistona chinensis* based on rDNA sequences. Mol Phylogenet Evol 20:1–13.

———, Huang GR, Wang Y, He WH, Zheng WH, Hyde KD. 2003. Molecular identification of white morphotype strains of endophytic fungi from *Pinus tabulaeformis*. Mycol Res 107:680–688.

Hamamoto M, Nagahama T, Tamura M. 2002. Systematic study of basidiomycetous yeasts —evaluation of the ITS regions of rDNA to delimit species of the genus *Rhodosporidium*. FEMS Yeast Res 2:409–413.

Hata K, Futai K. 1996. Variation in fungal endophyte populations in needles of the genus *Pinus*. Can J Bot 74:103–114.

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

———. 1994. Ascomycete systematics: problems and perspectives in the nineties. New York: Plenum Press.

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

Humphries CJ, Williams PH, Vane-Wright RI. 1995. Measuring biodiversity value for conservation. Ann Rev Ecol Syst 26:93–111.

Jacobs KA, Rehner SA. 1998. Comparison of cultural and morphological characters and ITS sequences in anamorphs of *Botryosphaeria* and related taxa. Mycologia 90:601–610.

Kim SH, Breuil K. 2002. Common nuclear ribosomal internal transcribed spacer sequences occur in the sibling species *Ophiostoma piceae* and *O. quercus*. Mycol Res 105:331–337.

Kirk PM, Cannon PF, David JC, Stalpers JA. 2001. Dictionary of the Fungi. 9th ed. Wallingford, UK: CABI Publishing.

Kriel WM, Swart WJ, Crous PW. 2000. Foliar endophytes and their interactions with host plants, with specific reference to the gymnospermae. Advance Bot Res 33: 1–34.

Lacap DC, Hyde KD, Liew ECY. 2003. An evaluation of the fungal 'morphotype' concept based on ribosomal DNA sequences. Fung Divers 12:53–66.

Legault D, Dessureault M, Laflamme G. 1989. Mycoflora of the needles of *Pinus banksiana* and *Pinus resinosa* 1. Endophytic fungi. Can J Bot 67:2052–2060.

Lieckfeldt E, Seifert KA. 2000. An evaluation of the use of ITS sequences in the taxonomy of the Hypocreales. Stud Mycol 45:35–44.

Liu YJ, Hall BD. 2004. Body plan evolution of ascomycetes, as inferred from an RNA polymerase II subunit. Mol Biol Evol 16:1799–1808.

Lodge DJ, Fisher BJ, Sutton BC. 1996. Endophytic fungi of *Manilkara bidentata* leaves in Puerto Rico. Mycologia 88:733–738.

Lumbsch HT, Wirtz N, Lindemuth R, Schmitt I. 2002. Higher level phylogenetic relationships of euascomycetes (Pezizomycotina) inferred from a combined analysis of nuclear and mitochondrial sequence data. Mycol Prog 1:57–70.

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

———, Kauff F, Cox C, McLaughlin D, Celio G, Dentiger B, PAdamsee M, Hibbett D, James TY, Baloch E, Grube M, Reeb V, Hofstetter V, Schoch C, Arnold AE, Miadlikowska J, Spatafora J, Johnson D, Hambleton S, Crockett M, Shoemaker R, Sung G-H, Lucking R, Lumbsch T, O'Donnell K, Binder M, Diederich P, Ertz D, Gueidan C, Hansen K, Harris RC, Hosaka K, Lim Y-W, Matheny B, Nishida H, Pfister D, Rogers J, Rossman A, Schmitt I, Sipman H, Stone J, Sugiyama J, Yahr R, Vilgalys R. 2004. Assembling the Fungal Tree of Life: progress, classification and the evolution of subcellular traits. Am J Bot 91:1446–1480.

Maddison DR, Maddison WP. 2003. MacClade: analysis of phylogeny and character evolution, Version 4.06. Sunderland, MA: Sinauer Associates.

McLenon T, Moncalvo J-M. 2004. The basidiomycete community in a hemlock-dominated plot in a southern Ontario temperate forest. Inoculum 54:35.

Miadlikowska J, Lutzoni F, Goward T, Zoller S, Posada D. 2003. New approach to an old problem: incorporating signal from gap-rich regions of ITS and nrDNA large subunit into phylogenetic analyses to resolve the *Peltigera canina* complex. Mycologia 95:1181–1203.

Mihelcic N, Hanula JL, Debarr GL. 2003. Pathogenic microorganisms associated with the southern pine coneworm (Lepidoptera: *Pyralidae*) attacking loblolly pine. Florida Entomol 86:1–7.

Moritz C, Faith DP. 1998. Comparative phylogeography and the identification of genetically divergent areas for conservation. Mol Ecol 7:419–430.

Müller MM, Valjakka R, Suokko A, Hantula J. 2001. Diversity of endophytic fungi of single Norway spruce needles and their role as pioneer decomposers. Mol Ecol 10: 1801–1810.

Petrini O. 1986. Taxonomy of endophytic fungi of aerial plant tissues. In: Fokkema NJ, van den Huevel J, eds. Microbiology of the Phyllosphere. Cambridge, UK: Cambridge University Press. p 175–187.

———. 1991. Fungal endophytes of tree leaves. In: Andrews JH, Hirano SS, eds. Microbial Ecology of Leaves. New York: Springer-Verlag. p 179–197.

———, Carroll G. 1981. Endophytic fungi in foliage of some Cupressaceae in Oregon. Can J Bot 59:629–636.

———, Müller E. 1979. Pilzliche Endophyten am Beispiel von *Juniperus communis* L. Sydowia 32:224–251.

Reeb V, Lutzoni F, Roux C. 2004. Contribution of *RPB*2 to multilocus phylogenetic studies of the Pezizomycotina (euascomycetes, Fungi) with special emphasis on the lichen-forming Acarosporaceae and evolution of polyspory. Mol Phylogenet Evol 32:1036–1060.

Rice P, Longden I, Bleasby O. 2000. EMBOSS: the European molecular biology open software suite. Trend Genet 16:276–277.

Rollinger JL, Langenheim JH. 1993. Geographic survey of fungal endophyte community composition in leaves of coastal redwood. Mycologia 85:149–156.

Rossman AY. 1994. A strategy for an all-taxa inventory of

fungal biodiversity. In: Peng CI, ed. Biodiversity and Terrestrial Ecosystems. Bulletin of the Academy Sinica Institute of Botany; online, at http://www.ars.usda.gov/services/docs.htm?docid=9426.

Sall J, Lehman A. 1996. JMP Start Statistics. Cary, NC: SAS Institute.

Shaw AJ, Cox CJ, Boles SB. 2003. Global patterns in peat moss biodiversity. Mol Ecol 12:2553–2570.

Sherwood-Pike M, Stone J, Carroll GC. 1986. *Rhabdocline parkeri*, a ubiquitous foliar endophyte of Douglas-fir. Can J Bot 64:1849–1855.

Sluder ER. 1993. Results at age 15 years from a half-diallel cross among 10 loblolly pines selected for resistance to fusiform rust (*Cronartium quercuum* f-sp. *fusiforme*). Sylvae Genet 42:223–230.

Stanosz GR, Smith DR, Guthmiller MA, Stanosz JC. 1997. Persistence of *Sphaeropsis sapinea* on or in asymptomatic shoots of red and jack pines. Mycologia 89:525–530.

Stone JK. 1985. Foliar endophytes of *Pseudotsuga menziesii* (Mirb.) Franco. Cytology and physiology of the host-endophyte relationship [Doctoral dissertation]. Eugene: University of Oregon.

———. 1987. Initiation and development of latent infections by *Rhabdocline parkeri* on Douglas-fir. Can J Bot 65:2614–2621.

———, Bacon CW, White JR. 2000. An overview of endophytic microbes: endophytism defined. In: Bacon CW, White JF, eds. Microbial Endophytes. New York: Marcel Dekker. p 3–29.

Sung SJS, White LM, Marx DH, Otrosina WJ. 1995. Seasonal ectomycorrhizal fungal biomass development on loblolly pine (*Pinus taeda* L.) seedlings. Mycorrhiza 5:439–447.

Swofford DL. 2001. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4b8. Sunderland, MA: Sinauer Associates.

———. 2004. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4b10. Sunderland, MA: Sinauer Associates.

Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS, Fisher MC. 2000. Phylogenetic species recognition and species concepts in fungi. Fung Genet Biol 31:21–32.

Tehler A, Little DP, Farris JS. 2003. The full-length phylogenetic tree from 1551 ribosomal sequences of chitinous fungi, *Fungi*. Mycol Res 107:901–916.

———, Farris JS, Lipscomb DL, Källersjo M. 2000. Phylogenetic analyses of the fungi based on large rDNA datasets. Mycologia 92:459–474.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The Clustal X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl Acid Res 24:4876–4882.

Vane-Wright RI, Humphries CJ, Williams PH. 1991. What to protect? systematics and the agony of choice. Biol Conserv 55:235–254.

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

———, Sun BL. 1994. Ancient and recent patterns of geographic speciation in the oyster mushroom *Pleurotus* revealed by phylogenetic analysis of ribosomal DNA sequences. Proc Nat Acad Sci USA 91:4599–4603.

West L, Jones RH. 2000. Responses of understory tree seedlings to alteration of the soil fungal community in mid- and late-successional forests. For Ecol Manage 134:125–135.

Wheeler WC. 1990. Combinatorial weights in phylogenetic analysis: a statistical parsimony procedure. Cladistics 6: 269–275.

White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Chapter 38. In: Innis M, Gelfand D, Sninsky J, White T, eds. PCR Protocols: a guide to methods and applications. Orlando, Florida: Academic Press. p 315–322.

Williams PH, Gaston KJ, Humphries CJ. 1994. Do conservationists and molecular biologists value differences between organisms in the same way? Biodivers Lett 2: 67–78.