MOLECULAR ECOLOGY

Molecular Ecology (2011) **20**, 3001–3013 doi: 10.1111/j.1365-294X.2011.05110.x

Species delimitation in fungal endophyte diversity studies and its implications in ecological and biogeographic inferences

ROMINA GAZIS,* STEPHEN REHNER† and PRISCILA CHAVERRI*

*Department of Plant Sciences and Landscape Architecture, University of Maryland, 2112 Plant Sciences Building, College Park, MD 20742, USA, †United States Department of Agriculture, Agricultural Research Service, Systematic Mycology and Microbiology Laboratory, 10300 Baltimore Avenue, Building 011A, Room 304, Beltsville, MD 20705, USA

Abstract

The estimation of species diversity in fungal endophyte communities is based either on species counts or on the assignment of operational taxonomic units (OTUs). Consequently, the application of different species recognition criteria affects not only diversity estimates but also the ecological hypotheses that arise from those observations. The main objective of the study was to examine how the choice and number of genetic markers and species delimitation criteria influence biodiversity estimates. Here, we compare approaches to defining species boundaries in three dominant species complexes of tropical endophytes, specially *Colletotrichum gloeosporioides* agg., *Pestalotiopsis microspora* agg. and *Trichoderma harzianum* agg., from two Amazonian trees: *Hevea brasiliensis* and *H. guianensis*. Molecular tools were used to describe and compare the diversity of the different assemblages. Multilocus phylogenetic analyses [*gpd*, internal transcribed spacer (ITS) and *tef1*] and modern techniques for phylogenetic species delimitation were overlaid with ecological data to recognize putative species or OTUs. The results demonstrate that ITS alone generally underestimates the number of species predicted by other nuclear loci. These results question the use of ITS and arbitrary divergence thresholds for species delimitation.

Keywords: barcode of fungi, fungal biodiversity, genealogical sorting index, genetic markers, species complex, species identification

Received 6 December 2010; revision received 21 February 2011; accepted 3 March 2011

Introduction

Endophytes are microorganisms that live for all, or part of their life cycle, within aboveground plant tissues without causing visible signs of infection. Studies have shown that individual plants may harbour dozens of endophytic fungal species (Arnold & Lutzoni 2007). Fungal endophytes contribute to the hyperdiversity of fungi (Hawksworth 2001; Arnold 2008), and surveys in tropical moist forests suggest that the majority of the 'undiscovered' endophyte diversity occurs in tropical trees (Frohlich & Hyde 1999; Arnold *et al.* 2000; Arnold & Lutzoni 2007; Arnold 2008).

Correspondence: Romina Gazis, Fax: 301 314 9308;
E-mail: rgazis@umd.edu

Comprehensive surveys of fungal endophyte diversity are challenging because taxonomic literature for many fungal genera does not exist, particularly for tropical taxa. This lack of taxonomic resources reflects in part a prior lack of taxonomic investigation and awareness of these fungi. At a more fundamental level, taxonomic progress in these groups has been hampered by difficulties in formulating effective species recognition criteria (Arnold 2008). Morphological species recognition (MSR) (Burnett 2003; Lacap *et al.* 2003) criteria have been difficult to develop for many endophyte taxa because they are phenotypically simple and often do not develop taxonomically informative vegetative, sexual (Wang & Guo 2007; Thomas *et al.* 2008) or asexual (Reynolds 1993; Taylor *et al.* 1999) reproductive structures *in vitro*. The lack of morphological

synapomorphies may be a consequence of genetic isolation preceding the appearance of phenotypically diagnostic character states (Taylor *et al.* 2000). Biological species recognition (BSR) (Mayr 1942, 1963) criteria are also difficult to frame because mating systems are often unknown and most species cannot be induced to produce their sexual states in culture. In addition, closely related species can retain interspecific interbreeding as ancestral character (Zervakis *et al.* 2004; Dettman *et al.* 2008).

In contrast to both BSR and MSR, the 'phylogenetic species recognition' (PSR) criterion uses nucleic acid variation to circumscribe species of phenotypically uniform, apparently asexual and even unculturable fungal lineages (O'Donnell *et al.* 1998; Arnold & Lutzoni 2007). According to this concept, species are considered to be an 'irreducible cluster of organisms diagnosably different from other such clusters and within which there is a parental pattern of ancestry and descent' (Cracraft 1983, 1989). However, in single-gene genealogies, deciding where to place the species boundaries is subjective, creating uncertainty on species' limits (Taylor *et al.* 2000). In spite of clear evidence that a single-gene genealogy does not necessarily reflect the organism's phylogeny (Rosenberg 2002), most published endophyte diversity studies based their species delimitation on the PSR inferred from a single locus (Guo *et al.* 2003; Murali *et al.* 2006; Promputtha *et al.* 2007).

The internal transcribed spacers (ITS1 and ITS2) and 5.8S region of the nuclear ribosomal repeat unit (ITS) are the most widely used molecular marker in endophyte diversity studies (Guo *et al.* 2003; Murali *et al.* 2006; Promputtha *et al.* 2007; U'ren *et al.* 2009). The use of ITS has many indisputable advantages, the main advantage being the ease by which it is amplified among all lineages of fungi using universal primers (Nilsson *et al.* 2008) and the large size of the available database (Vilgalys 2003; Lutzoni *et al.* 2004). The use of ITS as a species delimiter has several disadvantages, one being the range of intraspecific variation reported in the literature (Lieckfeldt & Seifert 2000; Lacap *et al.* 2003; Nilsson *et al.* 2008). In an effort to standardize the delimitation of species, investigators have proposed the use of a sequence similarity percentage as species proxy (Arnold & Lutzoni 2007; Higgins *et al.* 2007; Hoffman & Arnold 2008). However, many studies have demonstrated that ITS is not sufficient for species delimitation, especially in rapidly evolving or highly diverse genera or species complexes (Lacap *et al.* 2003; Hoffman & Arnold 2008).

To avoid subjectivity in delimiting species boundaries, Avise & Ball (1990) and later Baum & Shaw (1995) proposed using more than one gene genealogy and to rely on their concordance to recognize reproductive isolated units [genealogical species concept (GSC)]. The GSC defines a species as a 'basal group of organisms all of whose genes coalesce more recently with each other than with those of any organism outside the group'. Species genealogies should be concordant because of the effects of genetic isolation and drift with associated lineage sorting and coalescence (Fisher *et al.* 2002). Recombination between individuals of the same species can create conflict among gene trees, and the transition between concordance to conflict determines the limits of species (Taylor *et al.* 2000). The use of GSC to determine species limits is now commonly used in fungal systematics (Miller & Huhndorf 2004; Schoch *et al.* 2006); however, it is not widespread in endophyte research.

Diversity estimates are currently based on species counts or on the assignment of operational taxonomic units (OTUs). Consequently, any variation in this number may not only affect the diversity estimates but also the ecological hypotheses that may arise from those observations (Gotelli & Colwell 2001; Agapow *et al.* 2004; Magurran 2008). For instance, if two communities (e.g. different hosts or same host distributed in different areas) are compared, the similarity index will be affected by how accurately it can be determined whether two individuals belong to the same species unit or if they represent unique entities.

The main objective of the present study was to determine how reliance on a single genetic marker (i.e. ITS) can influence biodiversity estimates and the understanding of a community's ecology (species diversity, abundance, composition and distinctiveness) and biogeography. To test the hypothesis that the most common groups of endophytes contain more than one cryptic species, PSR criteria were applied by exploring the congruence of gene genealogies for three independent nuclear loci: ITS nrDNA, translation elongation factor 1α (*tef1*) and the glyceraldehyde-3-phosphate dehydrogenase (*gpd*). The second objective was to determine whether there is concordance between ecological, biogeographic and phylogenetic data. To accomplish this objective, other lines of evidence such as geographic location or management type were overlaid on the multilocus phylogeny to determine whether there is a geographic structure within the studied groups. We hypothesized that delimitation of cryptic species would correlate with host, distribution and land management type. To achieve our objectives, we studied three groups of endophytic taxa: *Colletotrichum gloeosporioides* species complex or 'aggregate' (hereafter, *C. gloeosporioides* agg.), *Pestalotiopsis microspora* species complex or 'aggregate' (hereafter, *P. microspora* agg.) and *Trichoderma harzianum* species complex or 'aggregate' (hereafter, *T. harzianum* agg.).

Materials and methods

Study taxa

Colletotrichum gloeosporioides (Ascomycota, Sordariomycetes, Glomerellaceae), *Pestalotiopsis microspora* (Ascomycota, Sordariomycetes, Xylariales, Amphisphaeriaceae) and *Trichoderma harzianum* (Ascomycota, Sordariomycetes, Hypocreales, Hypocreaceae) species complexes were chosen as model organisms for this study because of their common occurrence and abundance as endophytes (Evans *et al.* 2003; Jeewon *et al.* 2004; Lu *et al.* 2004; Rubini *et al.* 2005; Rojas *et al.* 2010). Each of these species complexes is depauperate in morphological characters useful in species recognition, and little is known about their reproductive biology; thus, neither MSR nor BSC are effective for species identification in these taxa. In addition, low innate ITS nucleotide variability within these complexes hampers attempts to delimit cryptic species (Lieckfeldt & Seifert 2000; Jeewon *et al.* 2004; Lu *et al.* 2004; Rojas *et al.* 2010).

Source of endophytic isolates

Endophytic isolates were collected from leaves and sapwood of two *Hevea* species (*H. brasiliensis* and *H. guianensis,* Euphorbiaceae). Trees were located in Peru (Iberia, Iquitos, Los Amigos and Tambopata) and Cameroon (Ekona). *Hevea brasiliensis* trees were distributed under two types of management: wild (Iquitos and Tambopata) and plantations (Iberia and Ekona); whereas the *H. guianensis* population was only sampled from the wild (Los Amigos). At each locality, 15 trees were sampled for endophytes. Collection techniques and sampling protocols are described fully in a previously published study (Gazis & Chaverri 2010). All strains were initially identified to genus using morphological characters and by sequencing the ITS locus and referencing it to GenBank database. The abundance of the three taxa varied among sampling localities. Table S1 (Supporting information) indicates the number of isolates, collection locality, host and GenBank accession numbers.

DNA extraction, amplification and sequencing

Genomic DNA was extracted from the mycelial mat using Power Plant™ DNA isolation kit (MO BIO Laboratories Inc., Solana Beach, CA, USA). ITS region was amplified using ITS5 and ITS4 primers (White *et al.* 1990); *tef1* was amplified using the primers EF-728 (Carbone & Kohn 1999) and EF2 (Jacobs *et al.* 2004) and the *gpd* region was amplified using the primers GPD1 and GPD2 (Berbee *et al.* 1999). The PCR conditions for the ITS, *tef1* and *gpd* amplification are described in previous publications (Berbee *et al.* 1999; Gazis & Chaverri 2010; Rojas *et al.* 2010). PCR products were purified using ExoSAP-IT® (USB Corporation, Cleveland, OH, USA) and sequenced at the University of Maryland Sequencing Facility.

Diversity and phylogenetic analyses

Bidirectional sequences were assembled and edited with Sequencher™ 4.9 (Gene Codes Corporation, MI, USA). Sequence for each taxon and locus was aligned with MAFFT version 6 using the E-INS-i strategy (Katoh *et al.* 2005) and refined manually using MESQUITE version 7.2 (Maddison & Maddison 2009). The software DOTUR (Schloss & Handelsman 2005) and MEGA version 4 (Tamura *et al.* 2007) were used to assess the amount of genetic variation (distance) within and among putative species ('phylogenetic species' or OTUs recognized in this study). MEGA was also used to calculate the number of nucleotide differences and the p-distance between the established OTUs. P-distance is the proportion of nucleotide sites at which two sequences being compared are different. It is obtained by dividing the number of nucleotide differences by the total number of nucleotides compared. In MEGA, distances were calculated directly from the nexus file, but for DOTUR, a distance matrix generated in PHYLIP version 3.68 (Felsenstein 1989, 2008) was required. The furthest neighbour algorithm was used for the clustering of OTUs.

Maximum likelihood (ML), maximum parsimony (MP) and Bayesian inference (BI) analyses were performed on separate and concatenated data. ML phylogenetic trees with bootstrap analysis were constructed with RAxML version 7.0.4 (Stamatakis *et al.* 2008) using the general time-reversible (GTR) evolutionary model and the gamma model of rate heterogeneity settings. Unordered characters, random taxon addition sequences, gaps treated as missing data and the tree bisection-reconnection (TBR) branch swapping were used in the analyses. Maximum parsimony analysis was conducted by heuristic search in PAUP* 4.0 (Swofford 2002) with the following settings: all characters were equally weighted, gaps were treated as missing characters, starting trees obtained by random addition with 1000 replicates, and TBR branch swapping algorithm. Nodal support for MP and ML was determined by nonparametric bootstrapping, performing 1000 replicates with a heuristic search consisting of 100 stepwise random addition replicates and TBR branch swapping for each bootstrap replicate. MrBayes version 3.1 (Ronquist & Huelsenbeck 2003) was used to construct phylogenies under BI. All searches were performed using four chains for a total of 10 000 000 generations with

trees sampled every 100 generations. Convergence of log likelihoods (−Ln) was assessed with TRACER version 1.4 (Rambaud & Drummond 2007). *Trichoderma* spp. (Table S1, Supporting information) sequences were used as outgroup in each data set.

jModelTest was used to select the models of nucleotide substitution for the ML and BI analyses (Posada 2008). To determine whether the data sets could be combined, highly supported clades were compared among trees (reciprocal bootstrap support) generated from different data sets to detect conflicts (Mason-Gamer & Kellogg 1996). High support refers to bootstrap support values ≥70%. If no conflict exists between the highly supported clades, this suggests that the genes sequenced share similar phylogenetic history and resolution, and combining the data sets can ultimately increase support.

Species recognition/delimitation

Three methods were used for species delimitation: (i) genealogical concordance phylogenetic species recognition (GCPSR) (Taylor *et al.* 2000), (ii) the genealogical sorting index (*gsi*) (Cummings *et al.* 2008) and (iii) reticulate networks (Huson & Bryant 2006, 2010). The GCPSR assumes a complete sorting of lineages and thus reciprocal monophyly. However, studies suggest that even if speciation is effective immediately, the time required might not be enough for evolutionary changes to appear allowing two distinct lineages to be recognized (Knowles & Carstens 2007; Cummings *et al.* 2008; O'Meara 2010). The genealogical sorting index (*gsi*) is a statistic that measures the degree of genealogical divergence in a specified group of taxa. The *gsi* supplements the bootstrap support and the posterior probability of monophyly, calculated from the multilocus analysis, by providing an independent measure of monophyly on a scale between zero and one. Interpretation of monophyly based on the *gsi* statistic follows tests using the posterior probability of monophyly through permutation testing (Cummings *et al.* 2008). Because uneven sample sizes among species can shift *P*-values downward for smaller groups, significance of the *gsi* was inferred at $P < 0.001$ (Polihronakis 2009). The *gsi* was implemented using the web interface (http://www.genealogicalsorting.org/). Rooted, weighted trees were used in the analysis. *Gsi* was calculated for nodes that contain putative or hypothetical species, even if their bootstrap support was low (e.g. 50–70%). Reticulate networks based on the combined data set were developed to detect potential recombination events between groups (Holmes *et al.* 1999). This analysis was undertaken using the algorithms implemented in SplitsTree program version 4.11.3 (Huson & Bryant 2010). The GTR character substitution model and the reticulation network split transformation under the recombination 2007 method were used. Bootstrap support for each split was estimated for 1000 replicates.

Terminal clades that were highly supported in the combined analyses (ITS + *tef1* + *gpd*) were considered as species and recognized as unique reproductively isolated units. Individual nodes were considered well supported by the data and analyses when both MP and ML bootstrap presented values ≥70% and when Bayesian posterior probability (BPP) was ≥0.90. Monophyly was supported by the *gsi* when values were ≥0.90 and were significant at $P < 0.001$ (Cummings *et al.* 2008; Polihronakis 2009). Some lineages were not considered in the analyses because of limited sample size and the inability to determine statistical support of their monophyly (e.g. LA11 and T7 in Fig. 2c), although it is suspected that these lineages likely represent additional species. In ecological surveys, these lineages are regarded as 'singletons' and are generally excluded from diversity estimates (Arnold & Lutzoni 2007; Davis & Shaw 2008).

Effects of species delimitation on endophyte diversity measures: a case study

After defining species limits, the ITS alignment was reviewed for each taxonomic group and the sequence divergence was determined between and among identified OTUs. Furthermore, a selected data set was used as a case study to investigate the effects of varying the ITS similarity threshold (90–100%). This data set is composed of 106 ITS sequences belonging to endophytic Ascomycota and two Basidiomycota collected from leaves and stems of wild *H. guianensis* (Table S2, Supporting information).

Results

Phylogenetic analyses: single locus

ITS, *tef1* and *gpd* each produced discrete amplicons and sequence chromatograms and thus behave as single-copy genes. Overall, this study yielded 484 new sequences; all of them have been deposited in GenBank (Table S1, Supporting information). All methods used in phylogenetic inference (MP, ML, BI) identified the same lineages within each group; the MP trees are presented with ML and MP bootstrap values and BI posterior probabilities indicated at each node (Figs 1 and 2; Table S3, Supporting information summarizes the characteristics of the data and results of phylogenetic analyses.). Results from the comparison of 70% bootstrap trees from the individual gene analyses did not reveal

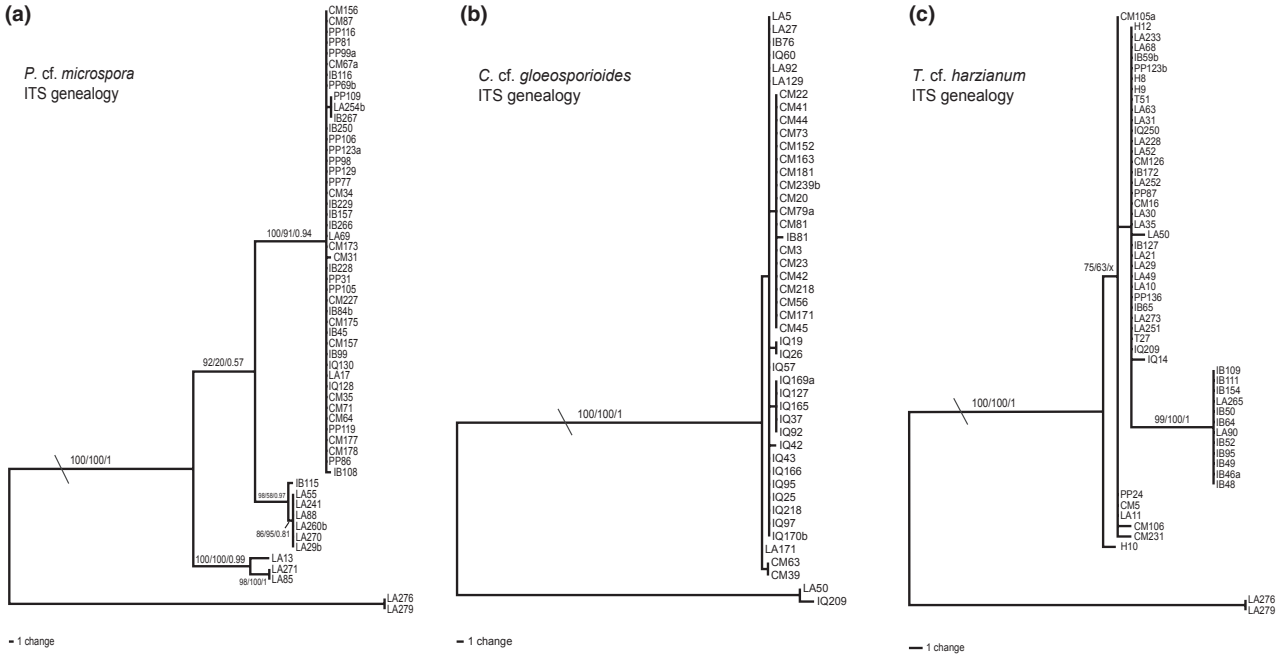

Fig. 1 Phylogenies resulting from maximum parsimony analyses of the internal transcribed spacer (ITS) nrDNA (a) *Pestalotiopsis microspora* agg., (b) *Colletotrichum gloeosporioides* agg. and (c) *Trichoderma harzianum* agg. Nuclear locus included ITS1, ITS2 and 5.8S. Node support values are indicated as follows: bootstrap support in maximum parsimony/bootstrap support in maximum likelihood/Bayesian posterior probability = (MP/ML/BPP). Branches that have been shortened are denoted by (\). Terminal labels denote the following geographic origin, management type and host species: LA (Los Amigos—Peru, wild, *H. guianensis*), IB (Iberia—Peru, plantation, *H. brasiliensis*), IQ (Iquitos—Peru, wild, *H. brasiliensis*), CM (Cameroon, plantation, *H. brasiliensis*). Analyses' parameters are indicated in Table S3 (Supporting information).

any major conflicts among phylogenies (ITS, *tef1* and *gpd* trees are included in Fig. 1, Figs S1 and S2, Supporting information). Phylogenetic analyses based on ITS suggest that the *P. microspora* agg. data set is composed of three distinct putative species or OTUs, *T. harzianum* agg. by two OTUs, and that all the isolates from *C. gloeosporioides* agg. data set belong to a single OTU (Fig. 1a–c). Table S4 (Supporting information) summarizes the number of OTUs (clades) inferred by each method and locus and the values that accompanied each analysis.

In the distance analyses conducted with DOTUR and MEGA, the number of OTUs varied depending on the locus used and on the similarity threshold applied to the data set (Table S5, Supporting information). For the *C. gloeosporioides* agg. and *T. harzianum* agg. data sets, the ITS region was the least informative of the three loci. For *C. gloeosporioides* agg., ITS grouped all isolates in one clade, even when the sequence similarity was increased to 99%. The ITS region was also invariable in the *T. harzianum* agg. data set, grouping all isolates into a single OTU at 98% similarity. The number of OTUs in *T. harzianum* agg. increased to three when the similarity threshold was raised to 99%. By contrast, the ITS was a relatively variable region for the *P. microspora* agg. group for which the use of a 95% threshold revealed three OTUs. Of the three loci examined, the *tef1* region was the most informative and variable. The *gpd* locus was more variable than ITS, but less variable than *tef1*, especially for *C. gloeosporioides* agg.

Phylogenetic analysis: multilocus approach and species delimitation

No incongruence was observed among gene trees for the three taxa; therefore, the three loci were concatenated and multilocus phylogenies were inferred using ML, MP and BI. These and the results from the genealogical sorting index (*gsi*) are presented in Fig. 2. The combination of phylogenetic analyses and the *gsi* values suggests that the *P. microspora* agg. data set included four putative species, the *C. gloeosporioides* agg. data set included ten species and the *T.* cf. *harzianum* data set included at least six species (Fig. 2). The *T. harzianum* agg. data set presented several terminal clades of unknown taxonomic position or clade affiliation. This pattern of diversification within the *T. harzianum* complex has been reported in another study (Druzhinina *et al.* 2010). One working hypothesis for the occurrence of these isolated groups is that they are 'relict lineages' (Druzhinina *et al.* 2010) but other factors such as sampling bias or chance inclusion of a geographic or

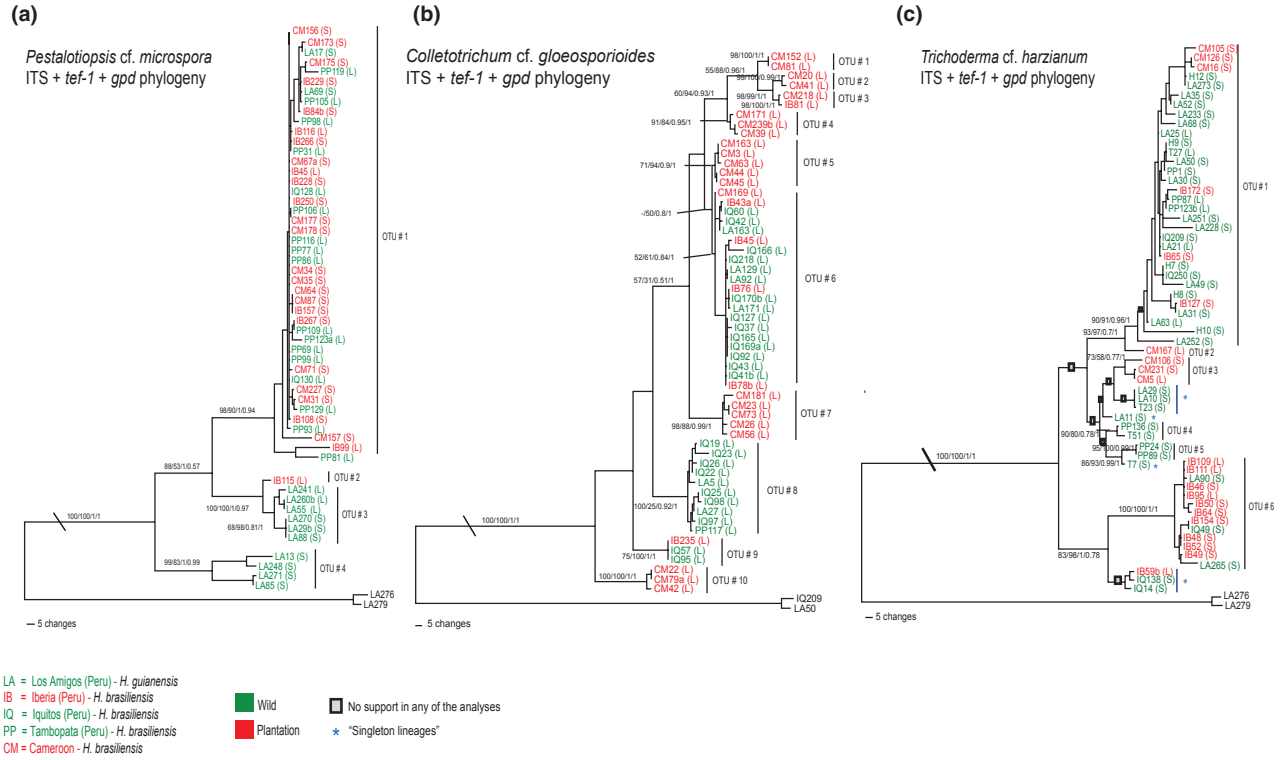

Fig. 2 Phylogenies of 55 *Pestalotiopsis microspora* agg. (a), 56 *Colletotrichum gloeosporioides* agg. (b) and 61 *Trichoderma harzianum* agg. (c) isolates resulting from maximum parsimony analysis of the concatenated three nuclear regions [complete *gpd*, internal transcribed spacer (ITS) and *tef1*]. Node support values are indicated as follows: bootstrap support in maximum parsimony/bootstrap support in maximum likelihood/Bayesian posterior probability/genealogical concordance value = (MP/ML/BPP/gsi). Branches that have been shortened are denoted by (\). Distribution within the tree is indicated by (S) = sapwood and (L) = leaves. Analyses' parameters are indicated in Table S3 (Supporting information).

ecological migrant can also create this phylogenetic pattern (Heath *et al*. 2008). Therefore, for purposes of this study, these terminal unsupported clades are considered as singletons and excluded from diversity analyses.

Results from the reticulate network analysis supported the multilocus grouping. Reticulation events were observed in clades that showed low bootstrap support for the combined analysis as well as low *gsi* values. Phylogenetic networks for the three data sets are shown in Fig. 3a–c.

Estimation of intra- and interspecific divergence

After the isolates of each group were assigned to OTUs, sequence divergence for the ITS region was calculated within and between species (Table S3, Supporting information). Tables S6–S8 (Supporting information) present ITS p-distances for each data set. Overall, ITS was the least variable of all three genes, with a mean sequence divergence of 0.0042% for sister taxa and 0.0177% for nonsister taxa. Interspecific divergence was calculated as the average distance between the sister and nonsister clades. *Pestalotiopsis microspora* agg. had the highest interspecific ITS sequence divergence (0.037%), whereas *C.* cf. *gloeosporioides* and *T. harzianum* agg. had the lowest interspecific sequence divergence (0.00191% and 0.0045%). Intraspecific sequence divergence was significantly lower for all groups with a mean of 0.001%. The locus with highest variability was *tef1*, with the highest level of polymorphism in *C. gloeosporioides* agg. *Gpd* provided better resolution and higher sequence divergence than ITS but much lower than *tef1*.

Effects of species delimitation on ecological and biogeographic inferences

Overall, the combined analyses for all three taxa revealed correspondence between resolved clades and plant host association or geographic origin. Host specialization can be clearly observed in the *P. microspora* agg. (Fig. 2a). In this group, the combined analyses revealed four putative species or OTUs. Two of the OTUs, OTUs 3 and 4 are restricted to *H. guianensis*, whereas with two exceptions, OTU1 includes isolates from *H. brasiliensis*. Biogeographic structure among

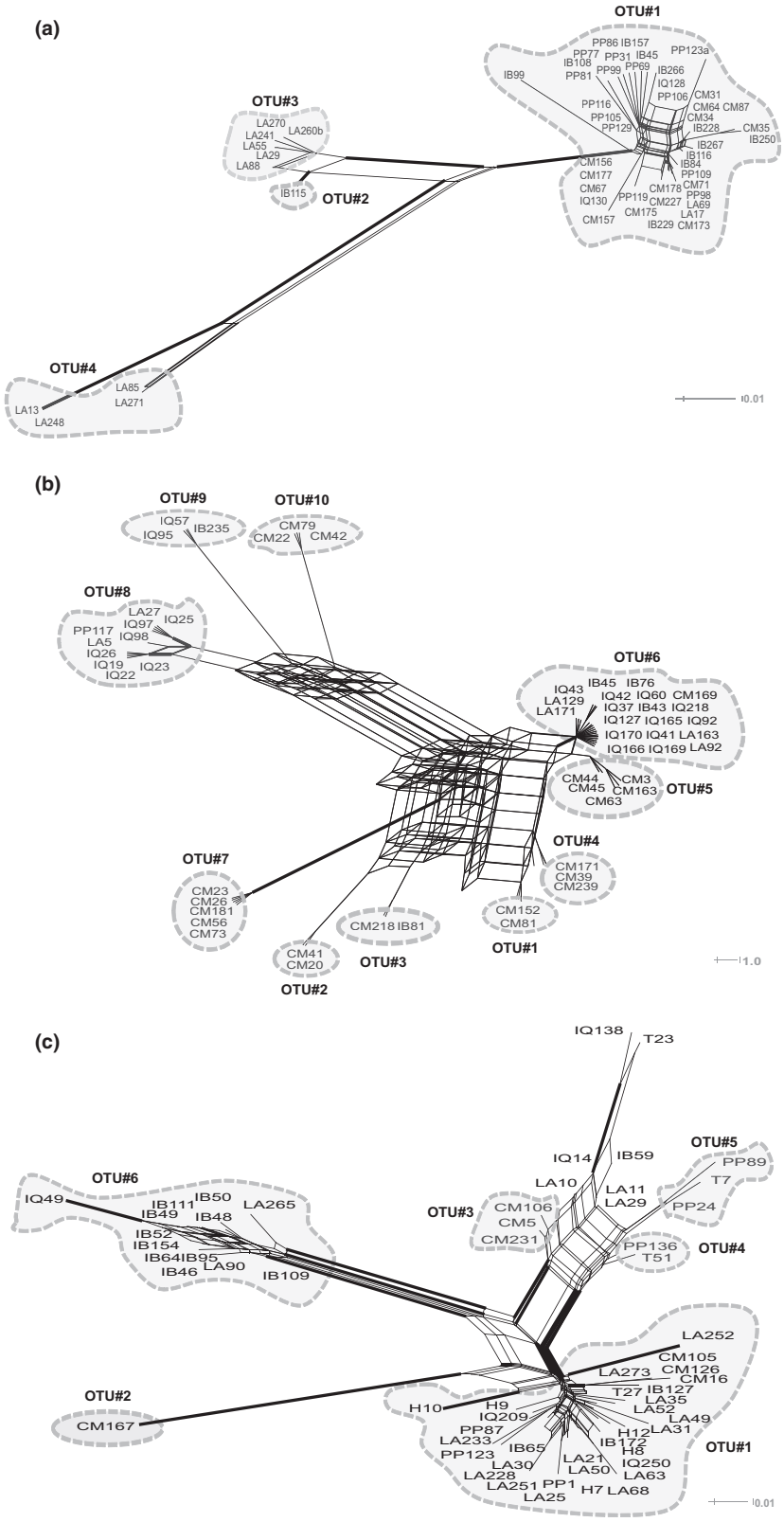

Fig. 3 Recombination network determined by SplitsTree (general time-reversible character transformation and reticulate network split transformation) from the concatenated data (*gpd*, internal transcribed spacer and *tef1*). (a) *Pestalotiopsis microspora* agg., (b) *Colletotrichum gloeosporioides* agg. and (c) *Trichoderma harzianum* agg. Split bootstrap support higher than 70% is represented by the bold lines. For each data set, putative phylogenetic species are enclosed within dotted lines.

collecting localities was not observed. In contrast, OTUs in *C.* cf. *gloeosporioides* did not display host specificity (Fig. 2b). Isolates from *H. guianensis* are intermixed with isolates from *H. brasiliensis*, and there are no OTUs that contain only endophytes from *H. guianensis*. Nevertheless, the combined analyses revealed a high geographic structure (Fig. 2b). For instance, OTUs 8 and 9 contained isolates from Peru and mainly from wild habitats. On the other hand, six lineages contain only Cameroonian isolates (OTUs 1, 2, 4, 5, 7 and 10). Host association was not apparent for any *T. harzianum* agg. OTU, although a correlation was found between geographic origin and management type (Fig. 2c). For instance, OTU3 is composed only of Cameroonian isolates, and OTUs 4 and 5 by isolates from wild trees distributed within the locality of Tambopata.

Effects of species delimitation on endophyte diversity measures

To assess the effect of ITS sequence similarity threshold on OTU diversity estimates, data for 106 endophytic Ascomycota collected from sapwood and leaves of wild *H. guianensis* were used. These data were selected because *Colletotrichum*, *Pestalotiopsis* and *Trichoderma* were all present in relatively high abundance at this site (Fig. S3, Supporting information). As DOTUR estimates are reported to be sensitive to the alignment (U'ren *et al*. 2009), the MAFFT alignment was not refined manually but subjected to analysis in DOTUR directly. The alignment for the analysis consisted of 730 bp including indels. For the phylogenetic analysis, two isolates belonging to the Basidiomycota, also collected from wild *H. guianensis*, were used as outgroup (LA216 and LA240, Table S2, Supporting information). The number of OTUs inferred by the analysis depended on the ITS similarity percentage threshold value. The resulting numbers of OTUs were as follows: 90%: 25; 91%: 25; 92%: 26; 93%: 26; 94%: 29; 95%: 31; 96%: 31; 97%: 33; 98%: 35; 99%: 38; and 100%: 50 (Table S5, Supporting information).

The following groups were affected by the increase in the ITS similarity threshold and thus an increase in their OTU numbers: *Beauveria* (95%: 2; 99%: 4), *Bionectria* (95%: 2; 99%: 4), *Fusarium* (95%: 2; 99%: 3), *Phomopsis* (95%: 1; 99%: 2) and *Xylaria* (95%: 1; 99%: 2) (Fig. S3, Supporting information). The groups in which OTU numbers were most affected are those with known low ITS interspecific variability (O'Donnell *et al*. 1998; Rehner & Buckley 2005; Tejesvi *et al*. 2009; Menezes *et al*. 2010). However, the majority of clades were unaffected and OTU estimates remained constant over different ITS thresholds.

The clades containing *Colletotrichum* and *Trichoderma* isolates were not affected by the increase in similarity threshold because of their extremely low interspecific ITS variability. The isolates belonging to each of the mentioned clades remained as one OTU even after increasing the cut-off to 99% similarity. However, as shown in Fig. 2b,c, *Colletotrichum gloeosporioides* agg. and *Trichoderma harzianum* agg. isolates from Los Amigos belong in 2 and 2–4 OTUs, respectively. In the case of *Pestalotiopsis microspora* agg., the use of 95% sequence similarity concurred with the combined analysis revealing two OTUs for the Los Amigos data set.

Discussion

The ITS region has been adopted by many fungal ecologists as the genetic marker of choice for species delimitation, even though its limitations for this purpose are widely acknowledged. ITS is useful for several reasons, among them the size of the available database in GenBank and the ease in which it is amplified along distant fungal lineages. However, several studies have demonstrated that for some groups (e.g. Ascomycota), the ITS region is insufficiently variable to resolve terminal species-level clades clearly resolved by other commonly used nuclear markers (O'Donnell *et al*. 1998; Inderbitzin *et al*. 2009; Pavlic *et al*. 2009; Druzhinina *et al*. 2010). In addition, field researchers face the challenge of having to examine large numbers of isolates; thus, multilocus analyses are usually not an affordable or efficient option for conducting pilot surveys of these organisms. Consequently, investigators need to compromise between the numbers of markers used to delimit species and the number of isolates included in the study. The choice of molecular marker and its use (e.g. distance vs. character based criterion, similarity threshold or phylogeny) is fundamental for accurately characterizing diversity. Segregation of species into units that reflect their genetic affinities can reveal patterns of biogeography and niche partitioning that may be overlooked otherwise. The failure to notice these patterns is generally a consequence of lumping potential species (or independent units) into large groups and assuming their homogeneity. In this study, all *Colletotrichum* sequences 'blasted' with *C. gloeosporioides* with 100% similarity in the GenBank database. Only in the combined gene analyses was it apparent that there are several endophytic species lineages and none are closely related to *C. gloeosporioides sensu stricto* (data not shown, but see Rojas *et al*. 2010 for an example). Assuming that the investigated endophytic strains were indeed *C. gloeosporioides* has further implications, because this species is an important and common pathogen of tropical crops. And, unless a multiapproach analysis was conducted, one could have assumed that this species is also a common endophyte of rubber trees.

Species delimitation in the selected species complexes

Experimental results reported here confirm that all three fungal genera represent species complexes. The combined analyses show that the isolates belonging to *P. microspora* agg. form four distinct clades (Fig. 2a). ITS distance analyses diagnosed just three of those putative species (OTU1, OTU2/OTU3 and OTU4) at a 95% similarity threshold. Only when the threshold was increased to 99%, did the number of putative species inferred from ITS sequences equal the number of species revealed by the combined analyses. Phylogenetic analyses based on the ITS locus did reveal a bifurcating pattern, although the reciprocal monophyly of OTU2 and OTU3 was only weakly supported (Fig. 1a). This might be a case of incomplete lineage sorting, in which some ancestral characters are being kept in both populations as a consequence of recent diversification events (Machado & Hey 2003; Koblmüller *et al*. 2010). The reticulate network analysis was mostly concordant with the multilocus phylogeny and grouped the isolates into the same OTUs as with the other species delimitation criteria. The only disagreement between these two methods was in the case of OTU2 and OTU3, where a nonsupported reticulation event was detected (Fig. 3a). This reticulation is reflected in the low MP and BI values estimated in the combined phylogeny. Unfortunately, with the available data, it is not possible to determine whether the latter represents recombination events between members of OTU2 and OTU3 or whether it is the case of incomplete lineage sorting (Morrison 2010). As the bootstrap value was low (<70%), the multilocus and *gsi* analysis was followed, dividing OTU2 and OTU3 into different entities.

The case of *C. gloeosporioides* agg. is more complex. The isolates formed ten well-supported lineages in the combined analysis (Fig. 2b). The only lineage that lacked significant support was OTU6 (Fig. 2b). This OTU was weakly supported by MP, ML and BI, but the *gsi* analysis revealed the monophyly of the group. For this group, the ITS marker showed extremely low interspecific variability (Table S7, Supporting information), and all the isolates were included in one OTU, even when using 99% of similarity. When sequences were grouped by unique haplotypes, eight groups were resolved. However, these ITS haplotype groups conflicted with the OTUs inferred by the phylogenetic multilocus approach (Fig. 2b). The reticulate network analysis showed similar results to those obtained by the combined phylogeny. The low branch support between OTU5 and OTU6 was concordant with a reticulation pattern between the mentioned OTUs (Fig. 3b). As mentioned before, with the available data, it is not possible to determine whether this pattern is because of recombination between populations or incomplete lineage sorting. However, in this case, incomplete lineage sorting seems to be the most plausible explanation. The latter because OTU5 is only composed by isolates from Cameroon and OTU6 by isolates from Peru, consequently because of the large geographic distance between these two populations, recent recombination events appear unlikely.

Trichoderma harzianum agg. presents an even more complicated scenario, which challenged the delineation of species within the complex. Even when using three independent loci and several approaches, it was not possible to assign some lineages to particular clades (or OTU). For some of these lineages, even though they had more than one representative, their monophyly was not supported by the analyses. Therefore, they were left as singletons until more isolates or additional markers are collected and relationships can be inferred more clearly. Nevertheless, the combined analysis resolved six well-supported clades. The ITS region in the *Trichoderma* data set displays low polymorphism (Table S8, Supporting information). Distance methods grouped all the isolates into one clade under 95% and 98% of similarity, irrespectively. Only when the threshold was increased to 99%, the isolates were grouped into three OTUs; however, this grouping did not correspond with the lineages inferred in the combined phylogenetic analysis. OTU6 encompassed the highest interspecific variability and was recognized even by the ITS phylogeny (Fig. 1c). Overall, phylogenetic analyses provided more resolution than distance methods (Table S4, Supporting information). As in the case of the *Colletotrichum* data set, all the *Trichoderma* sequences blasted to *T. harzianum* using the GenBank database. The latter illustrates the importance of using phylogenetic methods when identifying isolates. Reticulate network analysis was consistent with the multilocus phylogeny; however, it showed low bootstrap values for the delimitation of OTU3–5 (Fig. 3c). This might be a reflection of the lack of branch support for the internal nodes encompassing these three groups (Fig. 2c). All the 'singleton lineages' found in the combined analysis were also detected as singletons in the reticulate network but all of them were placed, with support, more closely to OTU3–5. OTU6 was clearly defined by this method, giving additional support to its segregation.

Does phylogeny and species delimitation reflect the ecology of endophytes?

Given the close relationship between ecological divergence and reproductive isolation and consequently the presumed role of ecological adaptation in the speciation process, an evaluation of the studied organism's

ecology should play a significant role in species delimitation (Chaverri *et al.* 2003; Nosil & Crespi 2006). The issue is whether genetic divergence in the absence of ecological change or adaptive divergence of some other type is sufficient for species delimitation (Davis *et al.* 2003; Bond & Stockman 2008; Frenkel *et al.* 2010). In the three species groups studied here, some OTUs of each group inferred in the combined analyses are characterized by unique ecological/biogeographic characteristics. For instance, OTU3 and OTU4 from the *P. microspora* agg. only include isolates collected from wild *H. guianensis*. This case demonstrates an agreement between genetic and ecological divergence, and the combined analysis revealed a host association between the species of *Hevea* and the endophytic group. Nevertheless, the majority of the *Pestalotiopsis* isolates were grouped under OTU1, composed of strains collected from different localities and management types. The combined analysis of the *C. gloeosporioides* agg. isolates also revealed a correlation between their genetic divergence and their ecological niche. For instance, OTUs 1–5, 7 and 10 include only isolates collected from trees in plantations mainly in Cameroon. On the other hand, OTU6 and 8 contain mainly isolates from wild trees (*H. guianensis* and *H. brasiliensis*). Even though OTU diagnosis in *T. harzianum* agg. was more challenging, the combined analyses also uncovered ecological affinities. This is the case of OTU6, composed only of isolates collected in Peru and mainly from plantations located in Iberia. All mentioned examples demonstrate how the use of phylogenetic approaches based on multiple loci assists in revealing ecological patterns of diversification. The latter is true especially for groups like *Colletotrichum* and *Trichoderma* that harbour a great genetic diversity and for which genetic sorting appears to be an active process.

Effects of species delimitation on endophyte diversity measures

There is no doubt that the ITS region will continue to be the genetic marker of choice for many future endophyte surveys. Therefore, this study investigated how an increase in the similarity threshold would impact the diversity estimation. Commonly, endophyte surveys use a conservative 95% similarity threshold for species delimitation. The effect of increasing the similarity percentages (90–100%) on the resulting species richness estimation was tested. It was found that most of the groups were not affected by the increase, because the majority of clades were phylogenetically distant, bearing many differences in nucleotides, and therefore, an increase in similarity did not affect their grouping (e.g. LA72a, LA89, LA227 and LA264a in Fig. S3, Supporting information). The group most affected was the Hypocreales (i.e. *Beauveria, Bionectria*, and *Fusarium*). These results coincide with others that also reported low ITS interspecific variability for the mentioned groups (O'Donnell *et al.* 1998; Davis *et al.* 2003; Rehner & Buckley 2005; Tejesvi *et al.* 2009). It is concluded that an increase in similarity threshold (based on ITS) has an important effect in the community's diversity estimate. For this example data set, the increase from 95% to 100% showed an overall increase of 19 OTUs, from 31 to 50 OTUs, respectively.

Is molecular data enough for species delimitation?

There is an ongoing controversy regarding how much (e.g. how many genes, how many morphological characters) and which type of data (e.g. molecular, morphological and ecological) should be used to circumscribe a species. For some species, morphologically distinctive characters were found only after those lineages were segregated and established by molecular data. A clear example can be drawn from this study. *Trichoderma harzianum* agg. is considered a species complex in which its members cannot be distinguished by using conventional morphological characters (e.g. conidiophore branching, conidia size/shape, phialide size/shape) (Chaverri *et al.* 2003; Druzhinina *et al.* 2010). Nevertheless, after the combined analysis confirmed the distinctiveness of one of the internal clades (OTU6 in Fig. 2c), a more intensive examination was performed to search for cryptic morphological differences that could support the molecular data. Few morphological differences between members of the clade OTU6 and the rest of the members of the *T. harzianum* complex were revealed. For instance, besides its endophytic habitat and its apparently specific host association (only found in *Hevea* species), members of OTU6 produce a diffusing brown pigment on artificial media and clustered chlamydospores. Recently, OTU6 has been described as the new species *Trichoderma amazonicum* (Chaverri *et al.* 2011). Even though for additional clades treated in this study monophyly was highly supported, the proposal of new species is beyond the scope of the present work.

Results from this and other studies advocate the use of more than one gene to delineate species. However, when many isolates are involved, sequencing several genes becomes a challenge. Therefore, for studies that aim to evaluate the diversity of a group of endophytes, the use of the ITS region is recommended for a first screening in which isolates can be sorted into different clades. Because many lineages will be singletons or will form definite clades, a more rigorous examination would only be necessary for speciose groups. On the other hand, if the objective of the study is to answer

ecological (e.g. comparison between sites, specificity evaluation, species range estimation) or evolutionary questions, the use of more than one gene is essential and several species limits approaches should be applied to the data. If the sequencing of additional markers is not feasible for a project, the use of a higher ITS similarity threshold (99% to almost 100% or unique haplotypes) is recommended for those speciose clades. The application of a higher ITS similarity threshold will most likely lead to a more accurate estimate of diversity.

Acknowledgements

We are grateful to Ikenna Okafor and Jonathan Agro for their assistance in laboratory work. We also thank Samuel Cunio, Yanet Barrios, Maribel Espinoza, and Aurelie Nyaka for their help during fieldwork. Funding was provided by a National Science Foundation grant to P. Chaverri 'Biodiversity of fungal endophytes in rubber trees: Towards understanding their role as plant protection agents' (DEB-0925672); a USDA-ARS Cooperative Agreement #5812758407 to P. Chaverri; and an Amazon Conservation Association graduate research grant to R. Gazis.

References

Agapow PM, Bininda-Emonds ORP, Crandall KA *et al.* (2004) The impact of species concepts on biodiversity studies. *The Quarterly Review of Biology*, **79**, 162–179.

Arnold AE (2008) Endophytic fungi: hidden components of tropical community ecology. In: *Tropical Forest Community Ecology* (eds Schnitzer R, Carson W), pp. 254–271. Blackwell Scientific, Hoboken, NJ.

Arnold AE, Lutzoni F (2007) Diversity and host range of foliar fungal endophytes: are tropical trees biodiversity hot spots? *Ecology*, **88**, 541–549.

Arnold AE, Maynard Z, Gilbert GS, Coley PD, Kursar TA (2000) Are tropical endophytes hyperdiverse? *Ecology Letters*, **3**, 267–274.

Avise JC, Ball RM (1990) Principles of genealogical concordance in species concepts and biological taxonomy. In: *Evolutionary Biology* (eds Futuyama D, Antonovics J), pp. 45–67. Oxford University Press, New York.

Baum D, Shaw KL (1995) Genealogical perspectives on the species problem. In: *Experimental and Molecular Approaches to Plant Biosystematics* (eds Hoch PC, Stephenson AG), pp. 289–303. Missouri Botanical Garden, St. Louis, MO.

Berbee ML, Pirseyedi M, Hubbard S (1999) *Cochliobolus* phylogenetics and the origin of known, highly virulent pathogens, inferred from ITS and glyceraldehyde-3-phosphate dehydrogenase gene sequences. *Mycologia*, **9**, 964–977.

Bond JE, Stockman AK (2008) An integrative method for delimiting cohesion species: finding the population–species interface in a group of Californian trapdoor spiders with extreme genetic divergence and geographic structuring. *Systematic Biology*, **57**, 628–646.

Burnett J (2003) *Fungal Populations and Species*. Oxford University Press, Oxford.

Carbone I, Kohn L (1999) A method for designing primer sets for speciation studies in filamentous Ascomycetes. *Mycologia*, **91**, 553–556.

Chaverri P, Castlebury LA, Samuels GJ, Geiser DM (2003) Multilocus phylogenetic structure of *Trichoderma harzianum/Hypocrea lixii* complex. *Molecular Phylogenetics and Evolution*, **27**, 302–313.

Chaverri P, Gazis R, Samuels GJ (2011) *Trichoderma amazonicum*, a new endophytic species on *Hevea brasiliensis* and *H. guianensis* from the Amazon basin. *Mycologia*, **103**, 139–151.

Cracraft J (1983) Species concepts and speciation analysis. *Current Ornithology*, **1**, 159–187.

Cracraft J (1989) Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In: *Speciation and its Consequences* (eds Otte D, Endler J), pp. 28–59. Sinauer Associates, Sunderland, MA.

Cummings MP, Neel MC, Shaw KL, Otto S (2008) A genealogical approach to quantifying lineage divergence. *Evolution*, **62**, 2411–2422.

Davis EC, Shaw AJ (2008) Biogeographic and phylogenetic patterns in diversity of liverwort-associated endophytes. *American Journal of Botany*, **95**, 914–924.

Davis EC, Franklin JB, Shaw AJ, Vilgalys R (2003) Endophytic *Xylaria* (Xylariaceae) among liverworts and angiosperms: phylogenetics, distribution, and symbiosis. *American Journal of Botany*, **90**, 1661–1667.

Dettman J, Anderson J, Kohn L (2008) Divergent adaptation promotes reproductive isolation among experimental populations of the filamentous fungus *Neurospora*. *BMC Evolutionary Biology*, **8**, 35.

Druzhinina IS, Kubicek CP, Komoń-Zelazowska M, Mulaw TB, Bissett J (2010) The *Trichoderma harzianum* demon: complex speciation history resulting in coexistence of hypothetical biological species, recent agamospecies and numerous relict lineages. *BMC Evolutionary Biology*, **10**, 94.

Evans HC, Holmes KA, Thomas SE (2003) Endophytes and mycoparasites associated with an indigenous forest tree, *Theobroma gileri*, in Ecuador and a preliminary assessment of their potential as biocontrol agents of cocoa diseases. *Mycological Progress*, **2**, 149–160.

Felsenstein J (1989, 2008) PHYLIP (phylogeny inference package). Version 3.68 PC. Available from: http://evolution.genetics.washington.edu/phylip.html.

Fisher MC, Koenig GL, White TJ, Taylor JW (2002) Molecular and phenotypic description of *Coccidioides posadasii* sp. nov., previously recognized as the non-California population of *Coccidioides immitis*. *Mycologia*, **94**, 73–84.

Frenkel O, Peever TL, Chilvers MI *et al.* (2010) Ecological genetic divergence of the fungal pathogen *Didymella rabiei* on sympatric wild and domesticated *Cicer* spp. (Chickpea). *Applied and Environmental Microbiology*, **76**, 30–39.

Frohlich J, Hyde KD (1999) Biodiversity of Palm fungi in the tropics: are global fungal diversity estimates realistic? *Biodiversity and Conservation*, **8**, 977–1004.

Gazis R, Chaverri P (2010) Diversity of fungal endophytes in leaves and stems of rubber trees (*Hevea brasiliensis*) in Tambopata, Peru. *Fungal Ecology*, **3**, 240–254.

Gotelli NJ, Colwell RK (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379–391.

Guo LD, Huang GR, Wang Y, He WH, Zheng WH, Hyde KD (2003) Molecular identification of white morphotype strains of endophytic fungi from *Pinus tabulaeformis*. *Mycological Research*, **107**, 680–688.

Hawksworth DL (2001) The magnitude of fungal diversity: the 1.5 million species estimated revisited. *Mycological Research*, **105**, 1422–1432.

Heath TA, Hedtke SM, Hillis DM (2008) Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematic and Evolutionary Microbiology*, **46**, 239–257.

Higgins KL, Arnold AE, Miadlikowska J, Sarvate SD, Lutzoni F (2007) Phylogenetic relationships, host affinity, and geographic structure of boreal and arctic endophytes from three major plant lineages. *Molecular Phylogenetics and Evolution*, **42**, 543–555.

Hoffman MT, Arnold AE (2008) Geographic locality and host identity shape fungal endophyte communities in cupressaceous trees. *Mycological Research*, **112**, 331–344.

Holmes EC, Worobey M, Rambant A (1999) Phylogenetic evidence for recombination in dengue virus. *Molecular Biology and Evolution*, **16**, 405–409.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

Huson DH, Bryant D (2010) User manual for Splits Tree4 V4.11.3 Available at : www.splitstree.org.

Inderbitzin P, Mehta YR, Berbee ML (2009) *Pleospora* species with *Stemphylium* anamorphs: a four locus phylogeny resolves new lineages yet does not distinguish among species in the *Pleospora herbarum* clade. *Mycologia*, **101**, 329–339.

Jacobs K, Bergdahl D, Wingfield M *et al.* (2004) *Leptographium wingfieldii* introduced into North America and found associated with exotic *Tomicus piniperda* and native bark beetles. *Mycological Research*, **108**, 411–418.

Jeewon R, Liew ECY, Hyde K (2004) Phylogenetic evaluation of species nomenclature of *Pestalotiopsis* in relation to host association. *Fungal Diversity*, **17**, 39–55.

Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*, **33**, 511–518.

Knowles LL, Carstens BC (2007) Delimiting species without monophyletic gene trees. *Systematic Biology*, **56**, 400–411.

Koblmüller S, Egger B, Sturmbauer C, Sefc KM (2010) Rapid radiation, ancient incomplete lineage sorting and ancient hybridization in the endemic Lake Tanganyika cichlid tribe Tropheini. *Molecular Phylogenetics and Evolution*, **55**, 318–334.

Lacap DC, Hyde KD, Liew ECY (2003) An evaluation of the fungal 'morphotype' concept based on ribosomal DNA sequences. *Fungal Diversity*, **12**, 53–66.

Lieckfeldt E, Seifert KA (2000) An evaluation of the use of ITS sequences in the taxonomy of the Hypocreales. *Studies in Mycology*, **45**, 35–44.

Lu G, Cannon PF, Reid A, Simmons CM (2004) Diversity and molecular relationships of endophytic *Colletotrichum* isolates from the Iwokrama Forest Reserve, Guyana. *Mycological Research*, **108**, 53–63.

Lutzoni F, Kauff F, Cox CJ *et al.* (2004) Assembling the fungal tree of life: progress, classification and evolution of subcellular traits. *American Journal of Botany*, **91**, 1446–1480.

Machado CA, Hey J (2003) The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 1193–1202.

Maddison WP, Maddison DR (2009) Mesquite: a modular system for evolutionary analysis. Version 2.6. Available from: http://mesquiteproject.org.

Magurran A (2008) *Measuring Biological Diversity*. Blackwell Publishing, Malden, MA.

Mason-Gamer RJ, Kellogg EA (1996) Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). *Systematic Biology*, **45**, 524–545.

Mayr E (1942) *Systematics and Origin of Species*. Columbia University Press, New York.

Mayr E (1963) *Animal Species and Evolution*. Harvard University Press, Cambridge, MA.

Menezes JP, Lupatini M, Antoniolli ZI, Blume E, Junges E, Manzoni CG (2010) Genetic variability in rDNA ITS region of *Trichoderma* spp. (biocontrol agent) and *Fusarium oxysporum* f. sp. *chrysanthemi* isolates. *Ciencia e agrotecnologia*, **34**, 132–139.

Miller AN, Huhndorf SM (2004) Using phylogenetic species recognition to delimit species boundaries and species relationships within *Lasiosphaeria*. *Mycologia*, **96**, 1106–1127.

Morrison DA (2010) Using data-display networks for exploratory data analysis in phylogenetic studies. *Molecular Biology and Evolution*, **27**, 1044–1057.

Murali TS, Suryanarayanan TS, Geeta R (2006) Endophytic *Phomopsis* species: host range and implications for diversity estimates. *Canadian Journal of Microbiology*, **52**, 673–680.

Nilsson H, Kristiansson E, Ryberg M, Hallenbergand N, Larsson KH (2008) Intraspecific ITS variability in the kingdom Fungi as expressed in the international sequence databases and it's implications for molecular species identification. *Evolutionary Bioinformatics*, **4**, 193–201.

Nosil P, Crespi BJ (2006) Ecological divergence promotes the evolution of cryptic reproductive isolation. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 991–997.

O'Donnell K, Cigelnik E, Nirenberg HI (1998) Molecular systematics and phylogeography of the *Gibberella fujikuroi* species complex. *Mycologia*, **90**, 465–493.

O'Meara BC (2010) New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology*, **59**, 59–73.

Pavlic D, Slippers B, Coutinho TA, Wingfield MJ (2009) Multiple gene genealogies and phenotypic data reveal cryptic species of the Botryosphaeriaceae: a case study on the *Neofusicoccum parvum/N. ribis* complex. *Molecular Phylogenetics and Evolution*, **51**, 259–268.

Polihronakis M (2009) Hierarchical comparative analysis of genetic and genitalic geographic structure: testing patterns of male and female genitalic evolution in the scarab beetle *Phyllophaga hirticula* (Coleoptera: Scarabaeidae). *Biological Journal of the Linnean Society*, **96**, 135–149.

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.

Promputtha I, Lumyong P, Dhanasekaran V, McKenzie EHC, Hyde KD, Jeewon R (2007) A phylogenetic evaluation of whether endophytes become saprotrophs at host senescence. *Microbial Ecology*, **53**, 579–590.

Rambaud A, Drummond AJ (2007) Tracer, version 1.4 [Internet]. Available from: http://beast.bio.ed.ac.uk/Tracer.

Rehner SA, Buckley EP (2005) Cryptic diversification in *Beauveria bassiana* inferred from nuclear ITS and EF1-alpha phylogenies. *Mycologia*, **97**, 84–98.

Reynolds DR (1993) The fungal holomorph: an overview. In: *The Fungal Holomorph: A Consideration of Mitotic and Pleomorphic Speciation* (eds Reynolds DR, Taylor JW), pp. 15–26. CAB International, Wallingford, UK.

Rojas EI, Rehner SA, Samuels GJ *et al.* (2010) *Colletotrichum gloeosporioides* s.l. associated with *Theobroma cacao* and other plants in Panamá: multilocus phylogenies distinguish host-associated pathogens from asymptomatic endophytes. *Mycologia*, **102**, 1318–1338.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Rosenberg NA (2002) The probability of topological concordance of gene trees and species trees. *Theoretical Population Biology*, **61**, 225–247.

Rubini MR, Silva-Ribeiro RT, Pomella AWV *et al.* (2005) Diversity of endophytic fungal community of cacao (*Theobroma cacao* L.) and biological control of *Crinipellis perniciosa*, causal agent of Witches' Broom Disease. *International Journal of Biological Sciences*, **1**, 24–33.

Schloss PD, Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Applied and Environmental Microbiology*, **71**, 1501–1506.

Schoch CL, Shoemaker RA, Seifert KA, Hambleton S, Spatafora JW, Crous PW (2006) A multigene phylogeny of the Dothideomycetes using four nuclear loci. *Mycologia*, **98**, 1041–1052.

Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML web-servers. *Systematic Biology*, **57**, 758–771.

Swofford DL (2002) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4 (Computer Program)*. Sinauer Associates, Sunderland, MA.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.

Taylor JW, Jacobson D, Fisher MC (1999) The evolution of asexual fungi: reproduction, speciation and classification. *Annual Review of Phytopathology*, **37**, 197–246.

Taylor JW, Jacobson DJ, Kroken S *et al.* (2000) Phylogenetic species recognition and species concepts in fungi. *Fungal Genetics and Biology*, **31**, 21–32.

Tejesvi MV, Tamhankar SA, Kini KR, Rao VS, Prakash HS (2009) Phylogenetic analysis of endophytic *Pestalotiopsis* species from ethnopharmaceutically important medicinal trees. *Fungal Diversity*, **38**, 167–183.

Thomas SE, Crozier J, Aime MC, Evans HC, Holmes KA (2008) Molecular characterization of fungal endophytic morpho-species associated with indigenous forest tree, *Theobroma gileri*, in Ecuador. *Mycological Research*, **112**, 852–860.

U'ren JM, Dalling JW, Gallery RE *et al.* (2009) Diversity and evolutionary origins of fungi associated with seeds of a Neotropical pioneer tree: a case study for analyzing fungal environmental samples. *Mycological Research*, **113**, 432–449.

Vilgalys R (2003) Taxonomic misidentification in public DNA databases. *New Phytologist*, **160**, 4–5.

Wang Y, Guo LD (2007) A comparative study of endophytic fungi in needles, bark, and xylem of *Pinus tabulaeformis*. *Canadian Jorunal of Botany*, **85**, 911–917.

White TJ, Bruns T, Lee S, Taylor JW (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In:*PCR Protocols: A Guide to Methods and Applications* (eds Innis MA, Gelfand DH, Sninsky JJ, White TJ), pp. 315–322. Academic Press, New York.

Zervakis GI, Moncalvo JM, Vilgalys R (2004) Molecular phylogeny, biogeography and speciation of the mushroom species *Pleurotus cystidiosus* and allied taxa. *Microbiology*, **150**, 715–726.

R.G. is a mycologist, with a special interest in ecology and diversity of tropical endophytes; S.R. is a mycologist, with interests on phylogenetics, population genetics and genomics of entomopathogenic and plant pathogenic fungi; P.C. research focuses on the systematics (taxonomy), evolution, and biodiversity of ascomycetous fungi.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Phylogenies resulting from maximum parsimony analyses of the *tef1* nrDNA (a) *Pestalotiopsis microspora* agg., (b) *Colletotrichum* cf. *gloesporioides*, and (c) *Trichoderma harzianum* agg.

Fig. S2 Phylogenies resulting from maximum parsimony analyses of the *gdp* nrDNA (a) *Pestalotiopsis microspora* agg., (b) *Colletotrichum* cf. *gloesporioides*, and (c) *Trichoderma harzianum* agg.

Fig. S3 A maximum-likelihood ITS (ITS1 + 5.8S + ITS2) phylogeny of 106 endophytic ascomycetes collected from leaves and sapwood of wild *Hevea guianensis*.

Table S1 All strains used in the study (including outgroups), their origin, and their Genbank accession number for the three loci.

Table S2 All strains used in the study case (LA), their origin, and their abundance.

Table S3 Characteristics of the data and results of phylogenetic analyses.

Table S4 Number of OUTs inferred by the different phylogenetic approaches based on single and multiple loci.

Table S5 Comparison of the number of operational taxonomic units (OTUs) based on complete ITS locus for each data set included in the study.

Table S6 ITS p-distances *Pestalotiopsis microspora* agg.

Table S7 ITS p-distances *Colletotrichum gloeosporioides* agg.

Table S8 ITS p-distances *Trichoderma harzianum* agg.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.