American Journal of Botany 101(8): 1362–1374, 2014.

GENETIC VARIATION IN HORIZONTALLY TRANSMITTED FUNGAL ENDOPHYTES OF PINE NEEDLES REVEALS POPULATION STRUCTURE IN CRYPTIC SPECIES[1]

RYOKO OONO[2,3,6,7], FRANÇOIS LUTZONI[3], A. ELIZABETH ARNOLD[4], LAUREL KAYE[3], JANA M. U'REN[4], GEORGIANA MAY[5], AND IGNAZIO CARBONE[6]

[2]University of California–Santa Barbara, Santa Barbara, California 93106 USA; [3]Duke University, Durham, North Carolina 27708 USA; [4]University of Arizona, Tucson, Arizona 85721 USA; [5]University of Minnesota, St. Paul, Minnesota 55108 USA; and [6]North Carolina State University, Raleigh, North Carolina 27695 USA

- *Premise of the study:* Fungal endophytes comprise one of the most ubiquitous groups of plant symbionts, inhabiting healthy leaves and stems of all major lineages of plants. Together, they comprise immense species richness, but little is known about the fundamental processes that generate their diversity. Exploration of their population structure is needed, especially with regard to geographic distributions and host affiliations.
- *Methods:* We take a multilocus approach to examine genetic variation within and among populations of *Lophodermium australe*, an endophytic fungus commonly associated with healthy foliage of pines in the southeastern United States. Sampling focused on two pine species ranging from montane to coastal regions of North Carolina and Virginia.
- *Key results:* Our sampling revealed two genetically distinct groups within *Lophodermium australe*. Our analysis detected less than one migrant per generation between them, indicating that they are distinct species. The species comprising the majority of isolates (*major* species) demonstrated a panmictic structure, whereas the species comprising the minority of isolates (*cryptic* species) demonstrated isolation by distance. Distantly related pine species hosted the same *Lophodermium* species, and host species did not influence genetic structure.
- *Conclusions:* We present the first evidence for isolation by distance in a foliar fungal endophyte that is horizontally transmitted. Cryptic species may be common among microbial symbionts and are important to delimit when exploring their genetic structure and microevolutionary processes. The hyperdiversity of endophytic fungi may be explained in part by cryptic species without apparent ecological and morphological differences as well as genetic diversification within rare fungal species across large spatial scales.

Key words: diversity; foliar fungal endophyte; *Lophodermium*; *Pinus*; Rhytismataceae; speciation.

Every major plant lineage examined to date hosts a rich community of fungal species that live asymptomatically within their tissues as endophytes (Arnold and Lutzoni, 2007; Rodriguez et al., 2009; U'Ren et al., 2012). Fungal endophytes in healthy foliage of woody plants are often horizontally transmitted (HT) and especially diverse, spanning at least eight classes of Fungi (Arnold et al., 2009; Gazis et al., 2012; class 3 endophytes, sensu Rodriguez et al., 2009; hereafter, endophytes). The community structure of HT endophytes often reflects environmental variables and host genotypes (Seghers et al., 2004; Hoffman and Arnold, 2008; Saunders and Kohn, 2009; Pancher et al., 2012; U'Ren et al., 2012; Zimmerman and Vitousek, 2012), despite the potential for species to interact with multiple host species via wind and air-dispersed propagules over various distances (see Rodriguez et al., 2009). It is unclear, however, whether single species of HT endophytes also harbor genetic structure that can be attributed to environmental variables, geographic distance, or host genotypes. Investigating population structure and genetic diversity within a widespread endophyte species as a function of host and geography is key to understanding important microevolutionary processes underlying the hyperdiversity of fungal endophytes at local and global scales.

Whereas vertically transmitted symbionts are often highly structured at the population and species level as a function of host specificity and limited dispersal (Sullivan and Faeth, 2004), population genetic structure in HT symbionts varies depending on life history processes and adaptation to host populations. Some populations may be panmictic, reflecting sexual reproduction and long-distance dispersal, or generalist associations with multiple host species (Hoffmann et al., 2010; Vincenot et al., 2012; Dunham et al., 2013). Alternatively, HT populations can be highly structured by mating strategies (Cubeta and Vilgalys, 1997), specificity to host taxa or genotypes (Ahlholm et al., 2002; Kaci-Chaouch et al., 2008; LaJeunesse et al., 2010), or adaptation to local environments (Kellermann et al., 2009; Keshavmurthy et al., 2012). Such structure can also reflect dispersal limitation and geographic structure (i.e., isolation by distance; Grubisha et al., 2007; Dunham et al., 2013) or several of these factors operating over time (Munkacsi et al., 2008). Whereas community-level studies of HT endophytes have become relatively prevalent in the literature (see references above) and population-level studies of HT pathogens

[1]Manuscript received 27 March 2014; revision accepted 16 July 2014.

The authors thank T. Truong, T. Gleason, E. M. Medina Tovar, J. Miadlikowska, K. T. Picard, M. T. Buntaine, S. Jiang, A. Simha, A. Gartin, and P. Manos for their cooperation or assistance in laboratory- and fieldwork. This research was supported by the Molecular Mycological Pathogenesis Training Program (MMPTP, Duke University) to RO, and NSF grants DEB-1046065 to FL, DEB-1046167 to IC, and DEB-1045766 to AEA.

[7]Author for correspondence (e-mail: ryoko.oono@lifesci.ucsb.edu)

doi:10.3732/ajb.1400141

have a rich history (reviewed in Milgroom, 1996; McDonald and Linde, 2002; Giraud et al., 2010), little is known about the processes that might structure populations of HT endophytes, and indeed, whether such structure can provide insights into the evolution of these hyperdiverse symbionts.

Here, we examine the genetic diversity and population structure of a widespread species of HT endophyte in coniferous hosts to test the dual hypotheses of structure as a function of host species and geography. We focus on *Lophodermium* (Rhytismataceae), a common genus of endophytes, saprotrophs, and pathogens in diverse woody plants (Tehon, 1935; Sieber-Canavesi et al., 1991; Stone et al., 2000; Müller et al., 2007; Lantz et al., 2011; Salas-Lizana et al., 2012). Endophytic species of *Lophodermium* that inhabit conifer foliage appear to be phylogenetically distinct from congeneric pathogens and endophytes of angiosperms (e.g., Poaceae, Rosaceae, and Ericaceae; Carroll, 1988; Ortiz-Garcia et al., 2003; Ganley et al., 2004; Lantz et al., 2011). Many occur in dead foliage of conifers and have been classified ecologically as saprotrophs, but in general these appear to represent strains that initially colonize host tissues as endophytes (Müller et al., 2001): they are not known to colonize and form sexual reproductive structures on pine needles that are already dead on forest floors (Osono and Hirose, 2011). We propose the endophytic *Lophodermium* species as a model for understanding the evolution and ecological specialization of endophytes with limited or absent stages as (1) latent pathogens or as (2) saprotrophs on hosts in which they are not endophytic. Moreover, in some cases they have an antagonistic effect on congeneric pathogens (see Carroll, 1988), suggesting a potentially important symbiosis in ecologically and economically important conifers.

Lophodermium species, which are heterothallic, frequently have high within-species genetic diversity even among isolates from the same leaf (Wilson et al., 1994; Deckert et al., 2002; Müller et al., 2007). In general, this high genetic diversity within the species may be explained by large population sizes, large distributions, or population-level specialization on diverse host taxa or physiographic regimes. The microevolutionary processes affecting this genetic diversity are still unknown, raising the question of whether several cryptic species are commonly analyzed as one (see Taylor et al., 2000, 2006; Müller et al., 2001). Past studies that have focused on population structure in endophytic *Lophodermium* (Müller et al., 2007; Salas-Lizana et al., 2012) either have used only fast-evolving markers (e.g., random amplified polymorphic DNA, RAPD) or have not explored population-species boundaries explicitly. Salas-Lizana et al. (2012) reported two groups of *L. nitens* that were genetically isolated by the Chihuahuan Desert, but it is unclear if this divergence is within or between species. Using the nuclear internal transcribed spacer (ITS) and two RAPD markers, Müller et al. (2007) found high genetic diversity but no evidence of cryptic species in *L. pinastri* from *Pinus sylvestris* in southern Sweden. In this case, it is unclear if high genetic diversity reflects a large population size or undiscovered population structure due to homoplasy in RAPD markers.

Detecting population structure and testing hypotheses regarding the relationship of endophyte population structure to host species or geography require careful delimitation of endophyte species. Many recent studies of endophyte diversity have used operational taxonomic units (OTUs) defined by sequence similarity (e.g., Zimmerman and Vitousek, 2012) or highly supported phylogenetic clades to estimate species boundaries (e.g., Gazis et al., 2011), but the structure within those OTUs is rarely explored.

In this study, we used sequence data from multiple loci to delimit species boundaries and explore the population genetic structure of a focal *Lophodermium* species associated with conifers in the southeastern United States. We concentrated on the common and widespread species *L. australe* Dearn. (Dearness, 1926), which was isolated from healthy foliage of *Pinus taeda* (loblolly pine) and *P. virginiana* (Virginia pine) (subgenus *Pinus*; section *Trifoliae*; subsections *Australes* and *Contortae*, respectively; Gernandt et al., 2005) in three major physiographic regions (mountain, Piedmont, and coast) in North Carolina and Virginia. To enhance our inferential power, we used a combination of molecular markers with high and low nucleotide substitution rates. Loci with high substitution rates will display greater homoplasy among groups with older divergence times, which can lead to overestimates of genetic diversity within species (Jarne and Lagoda, 1996; Neu et al., 1999) but will allow for the discovery of genetic structure among recently diverged groups. Conversely, loci with relatively low substitution rates may not reveal the fine-scale population structure necessary to understand diversity at the population level, but are necessary for phylogenetic inferences of closely related species. Thus, our work provides a perspective on appropriate molecular markers for resolving cryptic species or population structure as a function of host genotype or geography and also explores the scale of genetic diversity within and between species of a representative HT fungal endophyte.

MATERIALS AND METHODS

***Collection of* Lophodermium *endophytes*—**We collected healthy, second-year needles still attached to adult trees (>20 yr old) of *P. taeda* and *P. virginiana*. *Pinus echinata* (shortleaf pine) was sampled in regions where *P. taeda* was rare because they are closely related species and frequently hybridize (Fig. 1; Appendix S1, see Supplemental Data with the online version of this article; Gernandt et al., 2005). We collected across three physiographic regions of the southeastern United States: the Appalachian Mountains, the Piedmont Plateau, and the Atlantic Coastal Plain (http://pubs.usgs.gov/ha/ha730/ch_l/gif/L003.GIF). Two to three sites were sampled from each region for a total of eight sites (Fig. 1B, Table 1). In each site, at least three trees from three different locations (at least 5 km apart) were sampled when possible. The sampling breadth was designed to discover genetic structuring that was not evident in our preliminary ITS and microsatellite sequencing survey of *Lophodermium* isolates collected by Arnold et al. (2007; data not shown).

Needles were surface-sterilized following Arnold et al. (2007), cut into 2 mm sections, and plated on 2% malt extract agar (MEA). The number of needle sections plated per tree was approximately the same: because *P. virginiana* needles are approximately half the length of *P. taeda* or *P. echinata* needles, we sampled twice as many *P. virginiana* needles (10) as *P. taeda* or *P. echinata* needles (5) per tree. Fungal mycelia were allowed to grow for 2 to 4 wk until *Lophodermium* cultures could be distinguished morphologically from other common fungi.

***DNA extraction, PCR amplification, and sequencing*—**Small pieces of mycelium (ca. 1 × 1 mm) were placed in 0.6-mL tubes with 50 μL of glass beads (400 μm VWR silica beads) and 50 μL of Tris-EDTA (TE) buffer. DNA was extracted by vortexing the tubes at maximum speed (3200 rpm on a vortex mixer) for 2 min. Products were diluted with 200 μL of TE buffer for direct use in PCR.

We sequenced the internal transcribed spacer (ITS) region for isolates that were morphologically consistent with *Lophodermium* based on samples from Arnold et al. (2007) and confirmed the identity of 154 isolates with BLASTN matches (query cover ≥ 99% and max identity ≥ 97%) to known *Lophodermium* species. From these, we selected 81 isolates that had 96–100% ITS similarity to a previously identified specimen of *L. australe* (Ortiz-Garcia et al., 2003), which was keyed to species using ascoma descriptions (Minter, 1981). These isolates represent an even distribution of ITS haplotypes, host tree species, host individuals, collection sites, and physiographic regions (Table 1; online

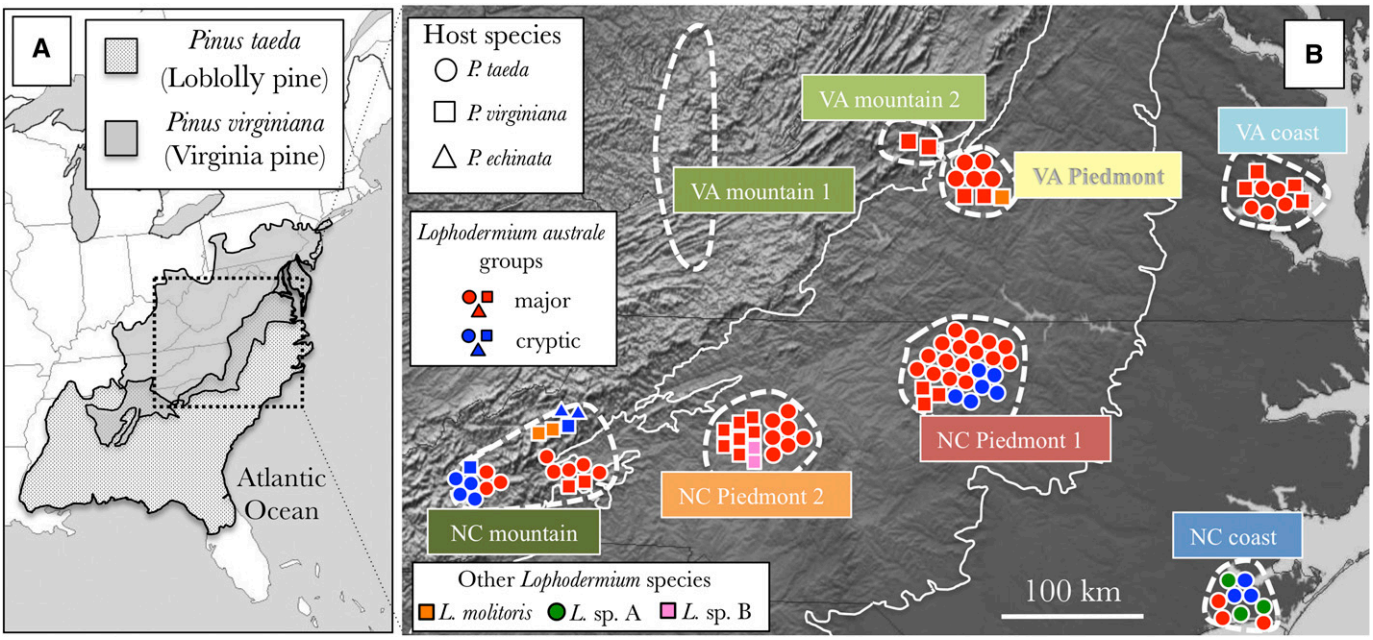

Fig. 1. Map of southeastern United States showing (A) geographic distributions of *P. taeda* and *P. virginiana* (efloras.org) and (B) sampling sites of 81 focal *Lophodermium* isolates that had high ITS similarity to *L. australe* (Ortiz-Garcia et al., 2003) and eight *Lophodermium* isolates distantly related to this focal group. *Lophodermium* endophytes were collected from eight sites (outlined by white dashes and labeled) that represent three physiographic regions (outlined by white borders): the Appalachian Mountains, the Piedmont Plateau, and the Atlantic Coastal Plain. No *Lophodermium* strains were isolated from the first Virginia mountain site (VA mountain 1; Table 1). Strains were placed in *major* (red) or *cryptic* (blue) groups, according to >50% membership coefficient by STRUCTURE (v2.3.1). Precise site coordinates are in online Appendix S1.

Appendix S1). We assessed genetic variation using five additional loci: the nuclear ribosomal DNA intergenic spacer (IGS) locus, three nuclear protein-coding loci (actin, calmodulin, and chitin synthase I), and one microsatellite locus. PCR conditions are listed in Appendix S2 (see online Supplemental Data). *Lophodermium* is haploid when hyphal, and haplotypes can be determined directly through sequencing.

PCR products were cleaned with ExoSAP-IT (Affymetrix, Santa Clara, California, USA) to remove unincorporated primers and dNTP and were sequenced with the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, California, USA) followed by analysis with an Applied Biosystems 3730xl DNA Analyzer. Sequences have been deposited in GenBank (KM106317-KM106821; online Appendix S1) and alignments have been submitted to the online database TreeBASE (http://treebase.org, S16106).

***Phylogenetic and clustering analyses*—**Sequences for each locus were aligned manually with the program MacClade 4.08a. Haplotype networks for each locus were inferred with the program TCS v1.21 (Clement et al., 2000) to visualize the genealogical relationships and frequencies among unique haplotypes at the population level. Gaps were considered as a fifth character state. TCS uses a Bayesian estimator to evaluate the number of mutational differences between haplotypes with greater than 95% parsimony probability. This number is then used to construct a network with a 95% probable connection limit following an algorithm outlined in Templeton et al. (1992). The resulting network (or set of networks) includes both parsimonious and nonparsimonious connections within the 0.95 probability limit and thus can accommodate low levels of homoplasy due to recombination. Haplotype networks allow assumptions that are invalid for typical phylogenetic methods, such as reticulate evolution, hybridization, and incomplete lineage sorting (Clement et al., 2000), which are possible in sexually reproducing species like *L. australe*. Geographic distributions and host affiliations were visualized with haplotype networks by color-coding respective collection sites and host species.

Sequences for the two rDNA loci and three protein-coding loci were converted into a binary character alignment for principal component analysis (PCA) and analysis in the program STRUCTURE v2.3.1 (Pritchard et al., 2000). The PCA was conducted for 70 of the 81 isolates for which all five loci were available to reveal genetic structure in a strictly nonparametric analysis. The number of significant axes of variation was determined using the Tracy–Widom statistic (Tracy and Widom, 1994). We used the Gap statistic, which compares

TABLE 1. Site and host origins of 81 *Lophodermium* isolates used for phylogenetic and population genetic analyses.

Sites	Area sampled (km²)	Meters above sea level	Trees from which isolates were obtained / Total trees sampled			No. of isolates		
			P. taeda	*P. virginiana*	*P. echinata*	*Major*	*Cryptic*	Total
NC mountain	1037	400–943	6/6	6/9	2/2	10	8	18
NC Piedmont 1	100	104–180	10/33	2/3	0/4	19	6	25
NC Piedmont 2	0.04	271	3/3	3/3	0	14	0	14
NC coast	930	4.3–8.6	4/9	0	0	3	3	6
VA mountain 1	90	324–598.5	0	0/9	0	0	0	0
VA mountain 2	1529	530–810	0	2/6	0	2	0	2
VA Piedmont	172	207–225	4/6	2/6	0	7	0	7
VA coast	594	16–24	5/9	2/2	0	9	0	9
			32/66	17/38	2/6	64	17	81

Notes: Isolates were grouped into *major* or *cryptic* based on STRUCTURE v2.3.1. NC = North Carolina, VA = Virginia.

within cluster variance to that expected under a reference null distribution based on PCA (Tibshirani et al., 2001), to infer the optimal number of clusters. To confirm the optimal cluster number and to assign isolates to these clusters, 10 independent STRUCTURE analyses of 1 million generations (following a burn-in of 100 000 generations) were run for each K value from 1 through 6, and results were analyzed using the Evanno method (Evanno et al., 2005) in STRUCTURE Harvester (Earl and von Holdt, 2012). Principal component, Tracy–Widom, and Gap statistics analyses, and inferences of population structure with STRUCTURE v2.3.1 were conducted in the Mobyle SNAP Workbench, a web-based analysis portal deployed at North Carolina State University (Price and Carbone, 2005, Monacell and Carbone, 2014).

After we observed distinct clusters in *L. australe* isolates using two independent approaches (PCA and STRUCTURE), we tested whether these clusters could be classified as different species following the genealogical concordance phylogenetic species recognition approach (GCPSR), which delimits species at the transitioning point from concordance to incongruity among branches of multiple-gene genealogies (Taylor et al., 2000). A maximum likelihood phylogeny was inferred using RAxML (Stamatakis, 2006) for all loci individually and together excluding the microsatellite locus. A general time-reversible model with a gamma distribution for rate heterogeneity was used with 1000 bootstrap replicates to assess support. The program PartitionFinder v1.0.1 (Lanfear et al., 2012) was used to divide the sequences into seven partitions and to identify optimum models of nucleotide substitution for phylogeny reconstruction using Bayesian methods (Appendix S3; see online Supplemental Data). Isolates that conflicted among individual gene trees using the reciprocal ≥70% bootstrap support criterion (Mason-Gamer and Kellogg, 1996) were excluded. Four separate Bayesian analyses were conducted in the program MrBayes v3.1.2 (Ronquist et al., 2012), each with four chains, 3 million generations, and sampling every 100 generations.

***Population genetic analyses of inferred* Lophodermium *groups*—**For each locus, we calculated the number of segregating sites (S), haplotype diversity (H_d), mean pairwise nucleotide differences (π), and nucleotide diversity (Nei and Jin, 1989) to evaluate diversity measures within the groups delimited by STRUCTURE and GCPSR. We calculated Wright's F_{ST} for each locus to evaluate divergence between groups. These statistics were inferred under the infinite-sites model of DNA sequence evolution in Arlequin 3.0 (Excoffier et al., 2005), implemented as a workflow in the Mobyle SNAP Workbench portal.

Poppr (Kamvar et al., 2013) was used to calculate the index of association (I_A) and the standardized index of association (r_D; Agapow and Burt, 2001), which estimates the degree of association of alleles at different loci (linkage disequilibrium) within and among groups compared to that observed in a permutated data set. P values were calculated by taking the fraction of permutated I_A values (of 99 999 permutations) greater than or equal to the observed I_A.

Gene flow between inferred groups was tested with the program IMa2 (Hey, 2010), implemented in Mobyle SNAP Workbench. The complete multilocus data set, including the flanking regions of the microsatellite and its repeat region for the 81 focal *Lophodermium* isolates, was fitted into an isolation-with-migration (IM) model using Markov chain Monte Carlo simulations (Nielsen and Wakeley, 2001). IMa2 generates posterior probability densities of gene genealogies for each locus and estimates population parameters such as migration rates, population sizes, and divergence times. The IM model is most appropriate for analysis of populations that have recently separated with possibilities of gene flow and incomplete lineage sorting (ancestral polymorphism) subsequent to divergence (Nielsen and Wakeley, 2001). IMa2 assumes the data set is selectively neutral; this was tested with Tajima's D (Tajima, 1989) and Fu's F_S (Fu, 1995) statistics, which infer deviations of the frequency distribution of polymorphic sites from neutral equilibrium expectations. IMa2 also assumes no recombination within loci; hence, sites that violated within-locus recombination were inferred and excluded by (1) collapsing each locus into unique haplotypes in SNAP Map (Aylor et al., 2006) with indels removed from alignments and infinite-sites violations excluded, (2) generating a site compatibility matrix in SNAP Clade (Aylor et al., 2006), and (3) excluding incompatible sites with SNAP CladeEx (Bowden et al., 2008). Appendix S4 (see online Supplemental Data) lists sites that were excluded from each locus. Simulations have shown that the IM model is generally robust to small violations of model assumptions, including population expansion, gene flow, and recombination (Strasburg and Rieseberg, 2010).

Short preliminary runs of IMa2 were used to assess priors for subsequent long runs. A uniform distribution with a maximum of three was used for the migration prior value. The maximum for population size parameter was set to 100, and the maximum time of splitting was set to five. Runs included a 5000 step burn-in followed by at least 100 000 sampled genealogies per locus. Each run used 40 chains with a geometric heating model (heating parameters: 0.999 and 0.888). Convergence to the stationary distribution of parameter values was confirmed with three separate runs with different random seeds. Effective sampling size for all parameters exceeded 100. To convert estimated parameters to demographic values, IMa2 uses the geometric average of mutation rates per locus. Rates were based on substitution rates estimated by Kasuga et al. (2002) and Tautz (1993) that were multiplied by the length of each locus.

***Genetic distance vs. geographic distance or host species*—**Correlation between the genetic and geographic distances of isolates was evaluated within each inferred *Lophodermium* group. The R package *ape* (Paradis et al., 2004) was used to calculate genetic distances between isolates using all nonmicrosatellite loci under the TN93 genetic model (Tamura and Nei, 1993), and *ade4* (Dray and Dufour, 2007) was used to perform Mantel tests with 9999 permutations. We calculated pairwise F_{ST} values (Weir and Cockerham, 1984) between sampling sites for each inferred group using the program Arlequin 3.0 (Excoffier et al., 2005) and tested for isolation by distance by regressing F_{ST} values [$F_{ST}/(1 - F_{ST})$] to the logarithmized geographic distances between sampling sites (Slatkin, 1993). We also tested overall F_{ST} for physiographic groups by dividing the isolates into coastal, Piedmont, and montane regions as well as for host species within each inferred group. Sites or groups with small sample sizes (<4) were not examined.

***Phenotypic characterization of genetically distinct* Lophodermium *strains*—**Discovery of distinctive phenotypic characters may be facilitated by delimitation of groups by molecular phylogenetic methods (Gazis et al., 2011; Hawksworth, 2012). We characterized mycelial growth patterns for a subsample of each *Lophodermium* group delimited by our analyses. In addition, in vitro pairwise interactions among a subsample of isolates from each group and two distantly related *Lophodermium* isolates (see online Appendix S1 for isolate references) were evaluated to gain insight into potential mechanisms by which genetically distinct groups may have arisen. Agar plugs containing actively growing mycelium were placed 3 cm apart on 2% MEA plates. Antagonistic interactions were scored on a scale of 1 to 5 after 4 wk at ca. 22°C (5: >3 mm clear zone between mycelia; 4: <3 mm clear zone but no contact between mycelia; 3: contact between mycelia but no overlap; 2: overlap of mycelia; 1: complete merging of mycelia) and averaged across three replicate trials. We tested the correlation between antagonistic scores and genetic distance using a linear function with a Mantel test. If mycelial antagonism is a reflection of genetic relationship, we expect greater antagonistic interactions among than within closely related groups.

Many *Lophodermium* species are saprotrophs with variable abilities to decompose lignin from dead pine needles (Osono and Hirose, 2011). They also tend to have high levels of extracellular cellulase production (Sieber-Canavesi et al., 1991). We compared cellulolytic and ligninolytic activity in vitro between subsamples of each delimited *Lophodermium* group. Agar plugs with actively growing mycelium were placed on 2% MEA plates amended with carboxymethylcellulose (cellulose substitute; Sigma-Aldrich, St. Louis, Missouri, USA) and water agar plates with 0.05% w/v indulin (lignin substitute; Mead-Westvaco, Richmond, Virginia, USA) (modified from Gazis et al., 2012; three replicates/isolate/medium type). The proportion of isolates that grew on each medium as well as the proportion that demonstrated diagnostic clearing of each medium were compared between groups using two-tailed Fisher's exact tests. Average growth rates (mm per day, obtained from measurements of colony diameter) were compared between groups on each medium using two-tailed t-tests. Where present, clearing at the periphery of mycelial growth was measured after 14 d, scaled by growth rate, and compared between groups using t tests. As needed, data were log-transformed prior to analysis.

RESULTS

In total, 154 isolates of *Lophodermium* were collected from healthy pine foliage in the southeastern USA (online Appendix S1). Of these, 146 isolates had 96–100% ITS similarity to the *L. australe* isolate (AY100647) of Ortiz-Garcia et al. (2003). We identified two genetically distinct groups within the 146 isolates according to ITS sequence variation (eight substitutions and two indel site polymorphisms). We sequenced five additional loci (see online Appendix S2 for sequence lengths) from 81 of these isolates representing 49 ITS haplotypes, including strains from

all three physiographic regions, seven sites, three host species, and 51 host individuals (32 *P. taeda*, 17 *P. virginiana*, and two *P. echinata*; Fig. 1, Table 1). One site in the Appalachian Mountains, far north of the natural range of *P. taeda*, revealed no *Lophodermium* isolates from *P. virginiana* (VA Mountain 1 in Fig. 1B) despite similar sampling efforts as other sites.

The other eight *Lophodermium* isolates had <87% ITS similarity to *L. australe* (above) and were used for outgroup taxa in phylogenetic analyses. Three of these isolates had 99% ITS sequence similarity to *L. molitoris* (AY100659, CBS597.84). The other five isolates, which represented two separate groups (*Lophodermium* sp. A and B) and had no BLASTN matches with >95% maximum identity, could not be identified to species. Strains belonging to *Lophodermium* sp. B and *L. molitoris* were isolated only from *P. virginiana* in more western sites (Fig. 1B; online Appendix S1).

***Phylogenetic and clustering analyses*—**The ITS haplotype network inferred by TCS suggested two genetic clusters (Fig. 2A), which were not as distinct in the networks of the other loci (Fig. 2B–D). The reconstructed haplotype networks did not suggest structure by geography or by host species for any loci (Fig. 2). The most common haplotype of each locus was found at all sites and in both *P. taeda* and *P. virginiana*. In the haplotype networks of ITS, IGS, chitin synthase I, and actin, many rare haplotypes are closely related to common haplotypes and form a star-like structure, suggesting population expansion (Slatkin and Hudson, 1991). Polymorphism for the calmodulin locus and the microsatellite flanking region demonstrated significant homoplasy arising from either high rates of recombination at the population scale or historical mutation at the species scale, preventing network reconstruction at the 95% limits of statistical parsimony.

The multilocus PCA revealed two genetic clusters (Fig. 3) corresponding to the distinct groups in the ITS haplotype network. The Tracy–Widom distribution estimated one significant axis (explaining 12.5% of the total variance) and the Gap statistic also estimated two clusters. STRUCTURE analyses also indicated

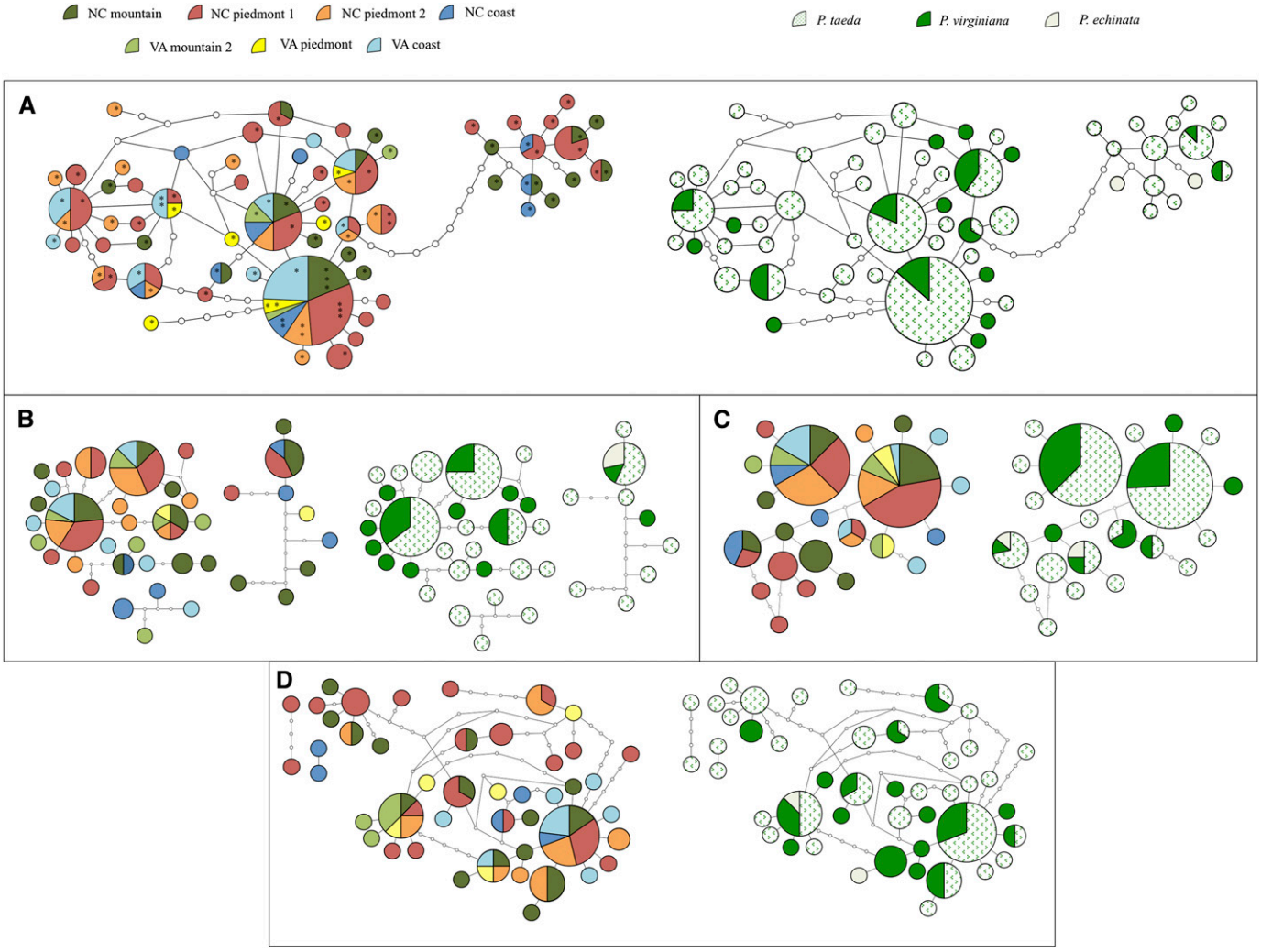

Fig. 2. Haplotype network for (A) ITS, (B) actin, (C) chitin synthase I, and (D) IGS. The ITS network includes 146 isolates; asterisks indicate the 81 isolates used in the analyses with other loci. Each circle represents a unique haplotype. The size of the circle is proportional to the number of isolates of the haplotype. The calmodulin locus and microsatellite flanking region had significant homoplasy among isolates, preventing network reconstruction at the 95% limits of statistical parsimony. Colors indicate geographic sites (left network) or host species (right network).

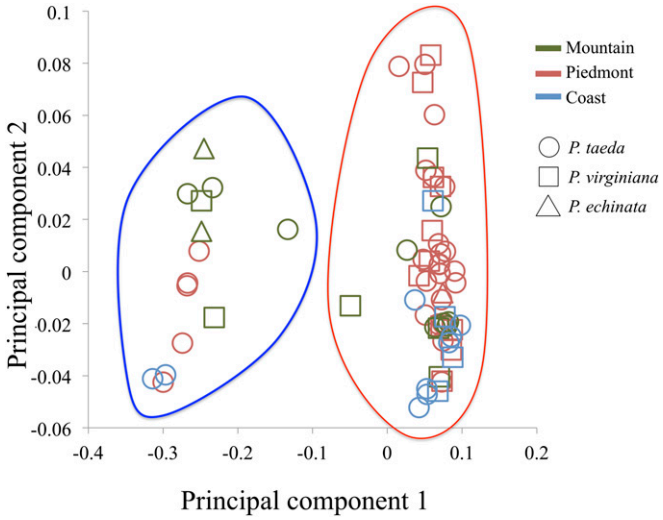

Fig. 3. Principal component analysis of 70/81 isolates for which five loci (two rDNA and three protein-coding loci) were sequenced. The first component explained 12.5% of the total variance and divided the isolates into two clusters: the *cryptic* cluster (left) and *major* cluster (right) according to >50% membership coefficient by STRUCTURE (v2.3.1).

that the ln [*P*(Data|*K*)] has the greatest change from $K = 1$ at −14 902 to $K = 2$ at −11 256 and then levels off with increasing *K* values. The peak for ΔK indicates two groups, and therefore, the isolates were delineated into two groups based on membership coefficients >50%. We named the more common group *major* (64 isolates) and the less common group *cryptic* (17 isolates). Four *major* isolates and one *cryptic* isolate, however, had relatively low membership coefficients (<90%) to either group and were designated as potential *hybrids*: DP_V1_9a (89.6% *major*), A_SL3_3a (87.8% *major*), 1VA3_10a (69.6% *major*), MS1_2_8 (59.3% *major*), and A_SL3_5a (50.5% *cryptic*). Two of these *hybrid* isolates were included in the PCA and appeared intermediate between the *major* and *cryptic* clusters. No structure was apparent for all isolates together on the basis of host species or geography, or within the *major* group, whereas members of the *cryptic* group appeared to be clustered by their geographic origins in the PCA (Fig. 3).

The percentage ITS sequence identity within each group ranged from 97.9–100% for the *major* group and 98.8–100% for the *cryptic* group. Between-group identity was 96.1–98.3%. *Cryptic* isolates represented 0–40% of the *L. australe* strains in each site, with five of the eight sites yielding no *cryptic* isolates (Table 1).

Isolates whose placement conflicted among individual gene trees (online Appendix S5) using the reciprocal ≥70% bootstrap support test (Mason-Gamer and Kellogg, 1996) had their respective sequences excluded from the concatenated phylogenetic analysis. Excluded sequences included IGS and chitin synthase I for some outgroup isolates (online Appendix S5b, c). Three *cryptic* isolates were highly supported (≥70% bootstrap) within the *major* clade in the IGS phylogeny; however, their IGS data were kept in the concatenated analysis because they were highly supported outside of this *major* clade in only one (calmodulin) or no other single-locus phylogenies. Although microsatellite loci have been used to diagnose phylogenetic species (Fisher et al., 2000), the microsatellite locus (flanking regions) that we amplified showed conflicting bootstrap support (online Appendix S5f) and was excluded from further analyses. Most single-locus trees separated the *cryptic* and *major* isolates from each other, but their monophyly was not highly supported (online Appendix S5).

A concatenated phylogeny across five loci revealed that the *cryptic* and the *major* clades are sister to each other, although neither group was strongly supported as monophyletic (Fig. 4A). The monophyletic group consisting of the *major* clade and all putatively *hybrid* isolates had moderate support with a posterior probability of 0.90 and bootstrap support of 75%. There was very little phylogenetic structure within the *major* clade. Within the *cryptic* clade, a branch consisting of isolates from the montane region was well supported (Fig. 4B), but the *cryptic* clade itself was not strongly supported by Bayesian or maximum likelihood inferences. *Hybrid* isolates were placed between the *major* and *cryptic* clades, indicative of their intermediate genetic composition between the two groups (Fig. 4A).

Population genetic analyses of major and cryptic groups—The ITS and actin loci were most diverged between the two groups ($F_{ST} > 0.700$). The other three loci had F_{ST} values ranging from 0.348 to 0.481. The flanking region of the microsatellite locus had the lowest F_{ST} value of 0.198 between groups. All F_{ST} values were significantly different from zero ($P < 0.0001$) with 1023 permutations. The haplotype and nucleotide diversities within each group appeared similar across loci (Table 2).

Tajima's *D* statistic did not suggest that any of the loci were under strong selection (online Appendix S6). The ITS and chitin synthase I loci in the *major* species had negative values, suggesting either population expansion or positive selection for low frequency mutations (Tajima, 1989). However, Fu's F_S test indicated departure from neutrality for most loci in the *major* species (online Appendix S6), suggesting that demographic effects, such as population expansion or genetic hitchhiking that affect all loci, are the cause for departure from neutrality rather than selection. The haplotype networks' star-like topology further suggests population expansion (Slatkin and Hudson, 1991) for ITS and chitin synthase I (Fig. 2). Furthermore, most segregating sites for protein-coding loci were either positioned within introns or were synonymous substitutions (Table 2).

IMa2 estimated the effective population size of the *major* group to be approximately three times that of the *cryptic* group ($\theta_{major} = 31.15$, $\theta_{cryptic} = 11.25$; Table 3; Fig. 5A, D). Based on our estimate of 5.14×10^{-6} mutations per locus per year (Tautz, 1993; Kasuga et al., 2002), this corresponds to approximately 3.0 million individuals in the *major* group and 1.1 million individuals in the *cryptic* group. The effective population size of the ancestor was estimated to be considerably smaller ($\theta_{ancestor} = 4.15$; approximately 400 000 individuals). Estimated rates of population migration ($2N_em$) were 0.914 (HPD 95%: 0.263, 1.978) migrants per generation from *major* to *cryptic* groups and 0.396 (HPD 95%: 0.033, 1.243) migrants per generation from *cryptic* to *major* groups (forward in time; Fig. 5B, D). Time of divergence was estimated at 1.038 in units of effective population size. According to our best estimate of the mutation rate, the divergence time is approximately 202 000 yr ago (Fig. 5C, D).

Genetic distance vs. geographic distance or host species—Mantel tests showed a significant correlation between genetic and geographic distances for the *cryptic* group ($r = 0.558$, $P = 0.001$) but not for the *major* group ($r = -0.009$, $P = 0.557$; Fig. 6).

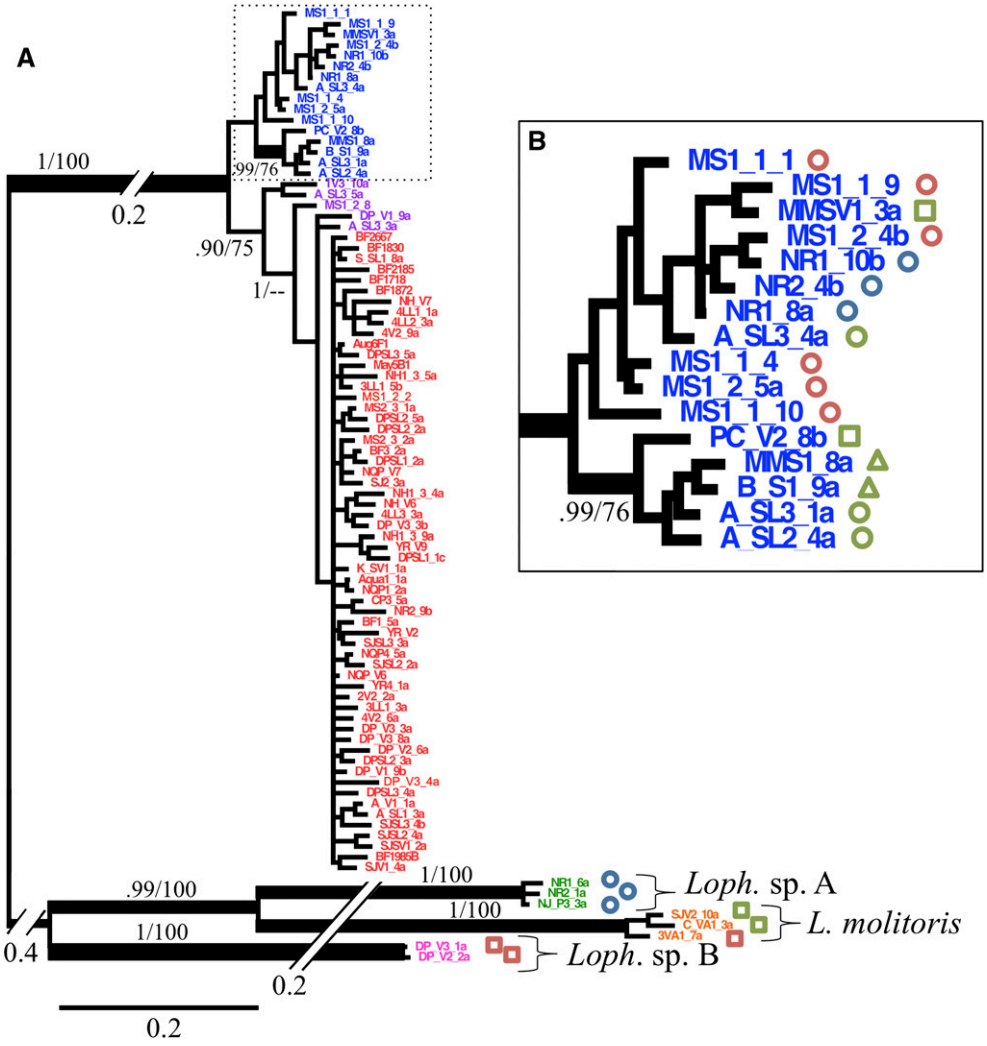

Fig. 4. (A) Concatenated multilocus Bayesian phylogeny with Bayesian posterior probabilities/maximum likelihood bootstrap scores at branches of interest. All 81 *Lophodermium* isolates were used, including ones with missing data. The scale bar corresponds to 0.2 substitutions per site. Blue taxa are *cryptic*, red taxa are *major*, and purple taxa are putative *hybrid* isolates as inferred by STRUCTURE v2.3.1. (B) Geographic and host origins of the *cryptic* isolates are indicated as in Fig. 3. Well-supported branch only contains isolates from the montane region.

The *major* group showed very low structure across sampling sites (F_{ST} = 0.04, $P < 0.0001$) and insignificant correlation between pairwise F_{ST} and geographic distance between sampling sites (online Appendix S7; r = 0.277, P = 0.212). Physiographic groups (coastal, Piedmont, montane) within the *major* species also showed a low level of genetic structure (F_{ST} = 0.049, $P < 0.05$). We could not reliably estimate F_{ST} among sites or physiographic regions for the rare *cryptic* group because of low sampling.

There was no indication of genetic structure between isolates from *P. taeda* and *P. virginiana* within the *major* group (F_{ST} = –0.01, $P > 0.05$). We could not reliably estimate F_{ST} for the rare *cryptic* group because of the low sampling within *P. virginiana* and *P. echinata*. Because *P. virginiana* and *P. echinata* appeared to be hosts to the *cryptic* isolates only in the mountains, genetic divergence as a function of geography could be artificially inflated. However, reanalysis without isolates from *P. echinata* and *P. virginiana* still indicated significant correlation between genetic and geographic distances (r = 0.599, P = 0.021).

The *major* and *cryptic* groups had high genetic divergence (F_{ST} = 0.50, $P < 0.001$) and were in linkage disequilibrium (I_A = 0.101, r_D = 0.033, P = 0.012), supporting our inference of limited gene flow between them. There was no deviation from linkage equilibrium to suggest restricted recombination within groups (*major* group, I_A = 0.002, r_D = 0.001, P = 0.476 and *cryptic*, I_A = –0.192, r_D = –0.051, P = 0.958).

***Phenotypic characterization of genetically distinct* Lophodermium *groups*—**We observed no substantive differences in hyphal and mycelial characteristics (e.g., growth rate, mean hyphal width, color, branching patterns) between members of the *major* and *cryptic* groups (online Appendix S8). For the in vitro mycelial interaction, no correlation was detected between antagonism scores and genetic distances (r = –0.129, P = 0.799). The average antagonism score (±SD) for self–self mycelia was 1.0 (± 0.3), whereas all other group interactions were scored at an average of 3.0 to 4.0, with no significant difference between interactions (e.g., *major* vs. *cryptic*, *major* vs. *outgroup*; online Appendix S9).

TABLE 2. Genetic variation and diversity estimates observed over six loci within *major* and *cryptic* groups of *Lophodermium australe*.

Locus	Segregating sites		Average pairwise differences (π)		Haplotype diversity		Nucleotide diversity		F_{ST}
	Major	*Cryptic*	*Major*	*Cryptic*	*Major*	*Cryptic*	*Major*	*Cryptic*	*M* vs. *C*
ITS	27	13	3.272	2.433	0.926	0.975	0.008	0.006	0.735 (0.713)
IGS	27	33	5.148	9.781	0.927	0.962	0.013	0.024	0.388 (0.313)
Actin	(0) 29	(1) 20	4.413	4.681	0.873	0.769	0.018	0.019	0.708 (0.682)
Calmodulin	(1) 114	(0) 39	18.57	12.41	0.999	0.981	0.045	0.03	0.348 (0.304
Chitin synthase I	(1) 11	(1) 5	1.036	1.175	0.705	0.775	0.005	0.005	0.481 (0.494)
Microsatellite flanking region	58	37	9.29	8.967	0.986	0.975	0.088	0.085	0.198 (0.177)

Notes: Potential *hybrid* isolates with <90% membership coefficient to either groups, according to STRUCTURE v2.3.1, were omitted from analysis for estimating segregating sites, average pairwise differences, haplotype diversity, and nucleotide diversity. Numbers in parentheses in front of protein-coding loci indicate the number of segregating sites leading to a nonsynonymous substitution. F_{ST} values were calculated between *major* and *cryptic* groups with *hybrid* isolates (indicated in parentheses) or without. All F_{ST} values differed significantly from zero with 1023 permutations ($P < 0.0001$).

The capacity to grow on carboxymethyl cellulose medium did not differ significantly between the genetically distinct groups: 16 of 16 *major* isolates grew successfully, as did 6 of 7 *cryptic* isolates (Fisher's $P = 0.30$). Among isolates that grew, average colony growth rates (±SD) did not differ significantly between the two groups: 2.37 mm/day (±1.61) vs. 3.58 mm/day (±1.14), respectively ($t = 1.964$, df = 20, $P = 0.110$). Of the isolates that grew and showed cellulase activity (7 of 16 *major* and 2 of 6 *cryptic* isolates), the average clearing score was 1.24 (±3.23) vs. 0.19 (±0.42) mm per cm of colony diameter, respectively, which did not differ significantly ($t = 1.666$, df = 7, $P = 0.240$).

The capacity to grow on lignin assay plates differed significantly between groups: 15 of 16 *major* isolates grew successfully vs. 3 of 7 *cryptic* isolates (Fisher's $P = 0.017$). However, isolates that did grow did not differ in growth rates (0.78 mm/day (±0.71) vs. 0.23 mm/day (±0.37), respectively) ($t = 0.499$, df = 16, $P = 0.624$). None of the isolates showed ligninase activity based on clearing of indulin from the media.

DISCUSSION

Horizontally transmitted fungal endophytes represent one of the most prevalent symbioses on earth, yet little is known about the population-level factors that might underlie their immense species richness. Our analyses revealed that the majority of strains identified as *Lophodermium australe* from healthy foliage of *Pinus* species in three physiographic regions of the southeastern United States belonged to one genetic cluster (*major*), and that a smaller cluster of closely related *Lophodermium* strains (*cryptic*) could be distinguished. We tested whether these clusters were different species using the criteria for GCPSR and estimated gene flow between the clusters with IMa2.

Phylogenetic analyses revealed genealogical concordance among two rDNA and three protein-coding loci, although there was low support for the monophyly of the *cryptic* group in the multilocus phylogeny. The presence of putative *hybrids* likely reduced the monophyletic support, but they were rare enough to suggest limited gene flow. The flanking region of the microsatellite used here demonstrated too much homoplasy to distinguish species. Perhaps greater sampling effort would allow this marker to be used to observe historical variance within species. Further studies are necessary to understand at what evolutionary scale or sample size microsatellites can be used to reveal genetic structure in *Lophodermium*.

The IMa2 analysis revealed asymmetric gene flow with less than one migrant per generation ($Nm < 1$) between the *major* and *cryptic* groups, suggesting that gene flow does not overcome divergence resulting from genetic drift (Wright, 1931). The boundary between species and populations can be subjective, as they can both be defined by limited gene flow or the criteria for GCPSR (Taylor et al., 2000). However, these two groups are geographically sympatric and occupy the same host species, suggesting that the limited gene flow and low rates of recombination between groups are due at least in part to a reproductive barrier. Hence, we consider them as two different species within the *L. australe* complex rather than highly structured populations of a single species.

We inferred the phylogenetic relationships of all *Lophodermium* isolates within the sampled geographic region. The eight *Lophodermium* isolates collected as outgroup taxa were too genetically distant to significantly influence the migration estimations in IMa2. However, we are aware that other *Lophodermium* species have as close as 95% ITS similarity to the *major* and *cryptic* groups (e.g., AY100644 from Mexico and KC283117 from Korea). Their absence in our sampling suggests that they are unlikely to have sufficient gene exchange with the *major* and *cryptic* species to influence the IMa2 analysis. However,

TABLE 3. Population parameters estimated in IMa2.

Estimates	t_0	θ_{major}	$\theta_{cryptic}$	$\theta_{ancestor}$	$m_{major \rightarrow cryptic}$	$m_{cryptic \rightarrow major}$
HiPoint	1.038	31.15	11.25	4.15	0.0495	0.2985
Lower bound of HPD 95%	0.7725	21.95	6.75	2.05	0.0045	0.0885
Upper bound of HPD 95%	1.458	44.75	19.25	7.05	0.1545	0.6735

Notes: t_0, estimated time since divergence of *cryptic* and *major* groups in units of effective populations ($N_e \cdot m$); θ_{major}, $\theta_{cryptic}$ and $\theta_{ancestor}$, estimates of $2N_e\mu$ where μ is the number of mutations per locus per generation for each group and the hypothetical ancestral species; $m_{(major \rightarrow cryptic)}$ and $m_{(cryptic \rightarrow major)}$, migration rate estimates per mutation where (major → cryptic) indicates gene flow from *major* to *cryptic* groups backward in time (in the coalescent). HPD = highest posterior density.

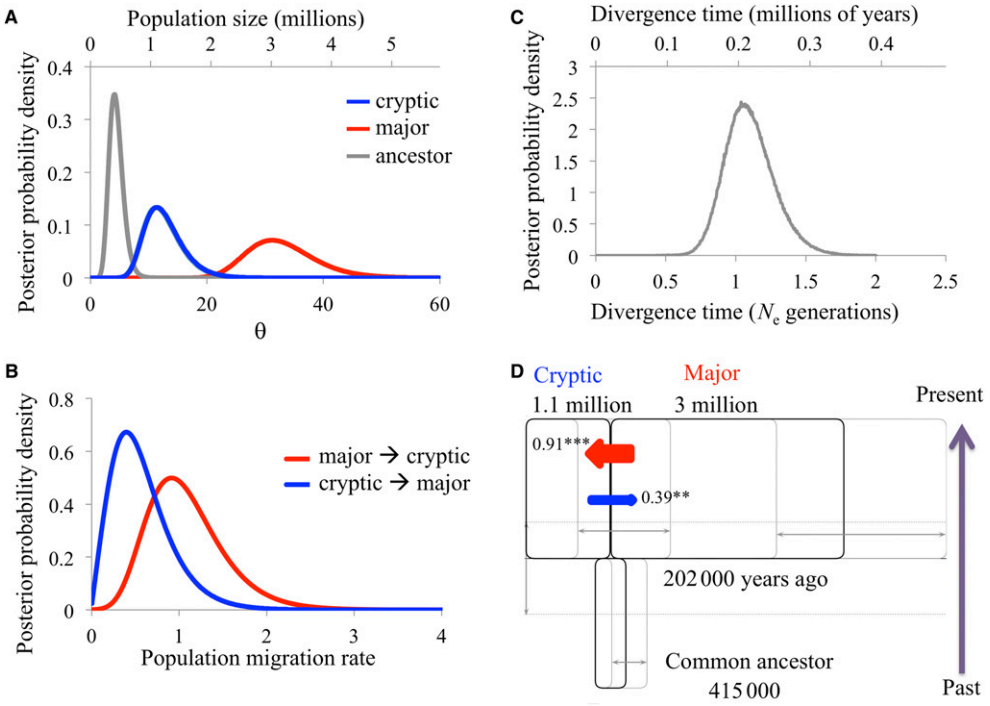

Fig. 5. Results of the IMa2 analysis of 81 *Lophodermium* isolates subdivided into *cryptic* and *major* groups according to STRUCTURE v2.3.1. (A) $\theta = 2N_e\mu$ because *Lophodermium* is haploid, N_e is the effective population size, and μ is the geometric mean of the mutation rates of all loci per year. In this analysis, $\mu = 5.14 \times 10^{-6}$ mutations per locus per year. Top scale is N_e inferred with the mutation rate. (B) Population migration rate, the effective rate at which migrants come into a population per generation, is calculated as $\theta \cdot m$ where m is the migration rate estimate per mutation and $m = m/\mu$, where m is the migration rate per gene per year (or generation in our case) from one population to another. IMa2 estimates population migration rates for diploid populations ($2Nm$). Hence, these values were divided by 2 and reported in the histogram. "Cryptic → major" indicates migration rate from *cryptic* to *major* groups in forward time. One generation was set as one year. (C) Divergence time is measured in years by dividing the estimated splitting time t_0 by μ, $t_0 = t \cdot \mu$. Mutation rates were modeled from Kasuga et al. (2002) and Tautz (1993); ITS (6.2×10^{-7}), IGS (4.26×10^{-7}), actin (2.5×10^{-7}), calmodulin (4.26×10^{-7}), chitin synthase I (1.5×10^{-6}), microsatellite flanking region (1×10^{-4}), and microsatellite repeat region (2.25×10^{-3}). Mutation rates were included in the analysis as mutations per year for each locus (not per base pair). Hence, the substitution rates estimated per site were multiplied by the length of each locus. Mutation rates are only used to calculate divergence time in years and to convert θ into effective population sizes. (D) Summary of the IMa2 results. Blocks represent sampled and ancestral populations. Arrows indicate migration rates between populations. Block height represents time of existence for the population and block width represents the effective population size. The confidence interval for a population size is given by a double-headed arrow extending from the right margin of that population's box and by boxes with thinner lines representing the lower and higher 95% highest posterior density (HPD) intervals. Splitting time 95% HPD intervals are given by dashed lines and double-headed arrows on the left. Time is represented as depth on the vertical axis.

the absence of more closely related outgroup taxa in the multilocus phylogenetic analysis may have contributed to the weak monophyletic support for the two groups. Future studies should aim to include *Lophodermium* populations from outside focal host and geographic boundaries for better confidence in gene flow estimations and thorough phylogenetic sampling.

***Major vs. cryptic groups*—**Members of the *cryptic* species appear to be less common and more unevenly distributed within their hosts' ranges than members of the *major* species (Fig. 1B; online Appendix S1). Although the *cryptic* species represented 21% of our sample in this study, we suspect its frequency in nature to be much lower. The lack of the *cryptic* genotype in the culture-based and culture-free surveys by Arnold et al. (2007) of >60 *Lophodermium* ITS genotypes from *P. taeda* further suggests that the *cryptic* species has a patchy distribution despite its extensive geographic range. Our results suggest that sampling more sites, rather than more intensive sampling within individual trees or sites, will increase the potential to recover greater species diversity in these hosts (see also Zimmerman and Vitousek, 2012; U'Ren et al., 2012) and uncover cryptic species.

We found that *cryptic* populations were structured with significant patterns of isolation by distance. However, the more common *major* species lacked population structure at the same geographic scale of this study. We expected the *major* species to have greater genetic diversity than the *cryptic* species due to the larger population sizes of the *major* species; however, both species showed high haplotype diversity (0.70–0.99), consistent with the study of *L. nitens* by Salas-Lizana et al. (2012). We hypothesize that the smaller population sizes of the *cryptic* species decrease gene flow among sites and increase genetic drift within sites, leading to greater population divergence than would be seen in the *major* species.

The higher rate of gene flow from the *major* species to the *cryptic* species may reflect the historical distribution pattern of the *cryptic* species. The majority of the *cryptic* isolates occur within the context of large numbers of *major* isolates, thereby increasing gene flow from the *major* species to populations of the *cryptic* species. In contrast, most members of the *major* species are more often closer to each other than they are to members of the *cryptic* species, thereby limiting gene flow from the *cryptic* species.

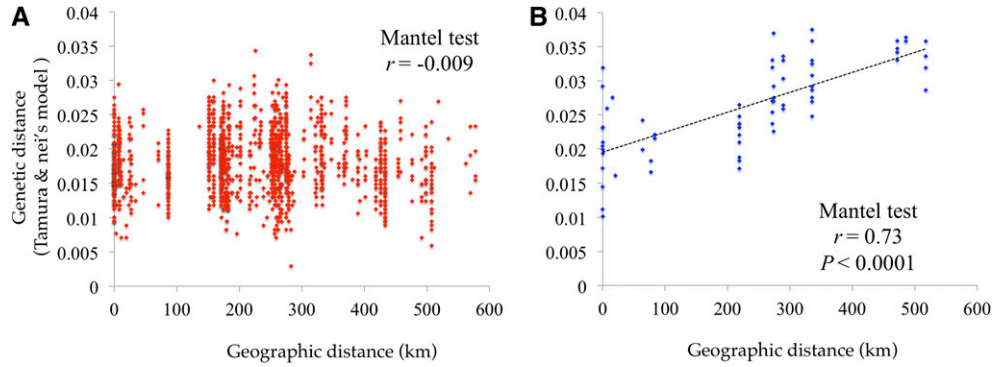

Fig. 6. Correlation between genetic and geographic distances was not evident for (A) the *major* group, but was evident for (B) the *cryptic* group.

We did not detect major differences between the *major* and *cryptic* species with regard to mycelial morphology, in vitro interactions, or cellulose degradation. However, a higher proportion of *major* than *cryptic* isolates grew on media with lignin as the primary carbon source, suggesting a difference in substrate use that may influence ecological breadth, relative abundance under natural conditions, and maintenance of species diversity in sympatry. In turn, variation among isolates of the *cryptic* species in their capacity to grow on the lignin medium may reflect variation in substrate use that leads to underestimates of its prevalence in culture-based studies. In the future, culture-free methods could confirm the population sizes of the *cryptic* species in nature. A detailed phenological study comparing developmental differences may also reveal mechanisms of reproductive isolation between species.

Lophodermium *species structure among host species*— Fungal endophytes that occur throughout a host's range and are common in host tissues, such as *Lophodermium* species in pines, may provide benefits to their hosts such as defense against pathogens, herbivores, and abiotic stress (Diamandis, 1981; Carroll, 1988; Arnold et al., 2003; Bae et al., 2009). Exploring the population structure of such recurring endophytic species may reveal ecological adaptations, genetic barriers, and host–symbiont coevolution at large and small spatial scales that are key to better understanding these widespread symbionts.

Although host genotypes have been shown to influence community structure of endophytes (e.g., Balint et al., 2013; Lamit et al., 2014), we did not detect any population structure within the *major* species with respect to host species. However, it is interesting to note the absence of *L. australe* from the VA Mountain 1 site, and the greater chances of isolating other *Lophodermium* species from *P. virginiana* in western sites where *P. taeda* is relatively rare (Fig. 1). Differences in host affinity measured by presence or relative abundances of endophytic species may be confounded when hosts' ranges overlap. Furthermore, endophyte communities and endophyte species' ranges are influenced by the range of their preferred host species as well as by the local plant community, including nonhost species (Shipunov et al., 2008).

To make inferences about host preferences and evolution of host affinity, taxonomic boundaries and species delimitation must be better defined for *Lophodermium*. *Lophodermium* taxonomy has changed markedly since the genus was first described (Chevallier, 1826, pp. 435–437), with recent surveys of endophytes (e.g., Ortiz-Garcia et al., 2003; Ganley et al., 2004; Arnold et al., 2007) discovering many taxa with similar or distinct ribosomal DNA sequence data relative to known taxa. When we summarized the host range of a particular endophyte species by using a 99% ITS similarity criterion to identify the same species across different studies, we discovered that *Lophodermium* species typically have finer preferences within conifer clades. For example, *L. australe* primarily associates with pines in the subgenus *Pinus* (e.g., *P. taeda* [Arnold et al., 2007], *P. elliottii* [U'Ren et al., 2012], and *P. pseudostrobus* and *P. radiata* [Ortiz-Garcia et al., 2003]) whereas *L. nitens* Darker (1932) associates with subgenus *Strobus* (e.g., *P. ayacahuite* and *P. strobiformis* [Salas-Lizana et al., 2012], and *P. chiapensis* and *P. monticola* [Larkin et al., 2012]). However, strict cospeciation between *Pinus* and *Lophodermium* species is not observed (Ortiz-Garcia et al., 2003). We suspect this is because all pine species can potentially host most *Lophodermium* species. This does not, however, exclude the possibility that host affinities vary for different *Lophodermium* species in different environments. Relative abundance, rather than presence or absence, of *Lophodermium* species on different host species may better reveal host affinities in such cases (see Arnold et al., 2003; Huang et al., 2008; Sun et al., 2012). Future studies that investigate host affinity should include abundance of each species across different host species once the interfaces of population-species boundaries are extensively explored.

***Implications for hyperdiversity of endophytes*—** This study revealed *Lophodermium australe* and a cryptic species have limited gene flow, may be sister species, and co-occur in the same pine hosts and geographic ranges. ITS delimitation served as an important starting point in identifying cryptic species. The additional use of multiple loci with a range of mutation rates allowed further investigation of population structure that identified isolation by distance in one species and panmixia in the other. This study illustrates the importance of multilocus analyses to delimit cryptic species in order to disclose the true diversity of foliar fungal endophytes as well as to study underlying microevolutionary processes that are fundamentally key to their hyperdiversity.

We propose that HT fungal endophytes have diverse microhabitats in which initial barriers for reproductive isolation can develop within genetically diverse populations, leading to sympatric speciation. In addition, many endophytic species may be so rare that their populations become isolated across large spatial scales over time and diverge genetically, leading to allopatric speciation. Both modes of speciation will give rise to high numbers of cryptic species and large species complexes. Additional population studies are necessary to understand whether the high

genetic diversity within the *Lophodermium* species we explored here is a common trait of endophytes in other lineages.

LITERATURE CITED

AGAPOW, P., AND A. BURT. 2001. Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes* 1: 101–102.

AHLHOLM, J. U., M. HELANDER, J. HENRIKSSON, M. METZLER, AND K. SAIKKONEN. 2002. Environmental conditions and host genotype direct genetic diversity of *Venturia ditricha*, a fungal endophyte of birth trees. *Evolution* 56: 1566–1573.

ARNOLD, A. E., D. A. HENK, R. L. EELLS, F. LUTZONI, AND R. VILGALYS. 2007. Diversity and phylogenetic affinities of foliar fungal endophytes in loblolly pine inferred by culturing and environmental PCR. *Mycologia* 99: 185–206.

ARNOLD, A. E., AND F. LUTZONI. 2007. Diversity and host range of foliar fungal endophytes: Are tropical leaves biodiversity hotspots? *Ecology* 88: 541–549.

ARNOLD, A. E., L. C. MEJÍA, D. KYLLO, E. I. ROJAS, Z. MAYNARD, N. ROBBINS, AND E. A. HERRE. 2003. Fungal endophytes limit pathogen damage in a tropical tree. *Proceedings of the National Academy of Sciences* 100: 15649-15654.

ARNOLD, A. E., J. MIADLIKOWSKA, K. L. HIGGINS, S. D. SARVATE, P. GUGGER, A. WAY, V. HOFSTETTER, ET AL. 2009. A phylogenetic estimation of trophic transition networks for Ascomycetous fungi: Are lichens cradles of symbiotrophic fungal diversification? *Systematic Biology* 58: 283–297.

AYLOR, D. L., E. W. PRICE, AND I. CARBONE. 2006. SNAP: Combine and Map modules for multilocus population genetic analysis. *Bioinformatics* 22: 1399–1401.

BAE, H., R. C. SICHER, M. S. KIM, S.-H. KIM, M. D. STREM, R. L. MELNICK, AND B. A. BAILEY. 2009. The beneficial endophyte *Trichoderma hamatum* isolate DIS 219b promotes growth and delays the onset of the drought response in *Theobroma cacao*. *Journal of Experimental Botany* 60: 3279–3295.

BALINT, M., P. TIFFIN, B. HALLSTROEM, R. B. O'HARA, M. S. OLSON, J. D. FANKHAUSER, M. PIEPENBRING, AND I. SCHMITT. 2013. Host genotype shapes the foliar fungal microbiome of Balsam poplar (*Populus balsamifera*). *PLoS ONE* 8: e53987.

BOWDEN, L. C., E. W. PRICE, AND I. CARBONE. 2008. SNAP Clade and Matrix, version 2. Available at http://snap.hpc.ncsu.edu, Department of Plant Pathology, North Carolina State University.

CARROLL, G. 1988. Fungal endophytes in stems and leaves: From latent pathogens to mutualistic symbiont. *Ecology* 69: 2–9.

CHEVALLIER, F. F. 1826. *Flore générale des environs de Paris* 1: 1–674.

CLEMENT, M., D. POSADA, AND K. CRANDALL. 2000. TCS: A computer program to estimate gene genealogies. *Molecular Ecology* 9: 1657–1659.

CUBETA, M. A., AND R. VILGALYS. 1997. Population biology of *Rhizoctonia solani* complex. *Phytopathology* 87: 480–484.

DARKER, G. D. 1932. They Hypodermataceae of conifers. *Contributions from the Arnold Arboretum of Harvard University* 1: 74.

DEARNESS, J. 1926. New and noteworthy fungi IV. Discomycetes including Phacidiaceae and Hypodermataceae. *Mycologia* 18: 236–255.

DECKERT, R., T. HSIANG, AND R. PETERSON. 2002. Genetic relationships of endophytic *Lophodermium nitens* isolates from needles of *Pinus strobus*. *Mycological Research* 106: 305–313.

DIAMANDIS, S. 1981. A serious attack on *Pinus brutia* L. in Greece. *In* C. S. Millar [ed.], Current research on conifer needle diseases. Proceedings of the IUFRO Working Party on Needle Diseases, 9–12, Sarajevo, 1980. Aberdeen University Press, Aberdeen, Scotland.

DRAY, S., AND A. B. DUFOUR. 2007. The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software* 22: 1–20.

DUNHAM, S. M., A. B. MUJIC, J. W. SPATAFORA, AND A. M. KRETZER. 2013. Spatial analysis of *Rhizopogon* genotypes within-population genetic structure differs between two sympatric sister-species of ectomycorrhizal fungi, *Rhizopogon vinicolor* and *R. vesiculosus*. *Mycologia* 105: 814–826.

EARL, D. A., AND B. M. VON HOLDT. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361.

EVANNO, G., S. REGNAUT, AND J. GOUDET. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* 14: 2611–2620.

EXCOFFIER, L., G. LAVAL, AND S. SCHNEIDER. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* 1: 47–50.

FISHER, M., G. KOENIG, T. WHITE, AND J. TAYLOR. 2000. A test for concordance between the multilocus genealogies of genes and microsatellites in the pathogenic fungus *Coccidioides immitis*. *Molecular Biology and Evolution* 17: 1164–1174.

FU, Y.-X. 1995. Statistical properties of segregating sites. *Theoretical Population Biology* 48: 172–197.

GANLEY, R., S. BRUNSFELD, AND G. NEWCOMBE. 2004. A community of unknown, endophytic fungi in western white pine. *Proceedings of the National Academy of Sciences, USA* 101: 10107–10112.

GAZIS, R., J. MIADLIKOWSKA, F. LUTZONI, A. E. ARNOLD, AND P. CHAVERRI. 2012. Culture-based study of endophytes associated with rubber trees in Peru reveals a new class of Pezizomycotina: Xylonomycetes. *Molecular Phylogenetics and Evolution* 65: 294–304.

GAZIS, R., S. REHNER, AND P. CHAVERRI. 2011. Species delimitation in fungal endophyte diversity studies and its implications in ecological and biogeographic inferences. *Molecular Ecology* 20: 3001–3013.

GERNANDT, D., G. LOPEZ, S. GARCIA, AND A. LISTON. 2005. Phylogeny and classification of *Pinus*. *Taxon* 54: 29–42.

GIRAUD, T., P. GLADIEUX, AND S. GAVRILETS. 2010. Linking the emergence of fungal plant diseases with ecological speciation. *Trends in Ecology & Evolution* 25: 387–395.

GRUBISHA, L. C., S. E. BERGEMANN, AND T. D. BRUNS. 2007. Host islands within the California Northern Channel Islands create fine-scale genetic structure in two sympatric species of the symbiotic ectomycorrhizal fungus *Rhizopogon*. *Molecular Ecology* 16: 1811–1822.

HAWKSWORTH, D. L. 2012. Global species numbers of fungi: Are tropical studies and molecular approaches contributing to a more robust estimate? *Biodiversity and Conservation* 21: 2425–2433.

HEY, J. 2010. Isolation with migration models for more than two populations. *Molecular Biology and Evolution* 27: 905–920.

HOFFMANN, M., M. FISCHER, A. OTTESEN, P. J. MCCARTHY, J. V. LOPEZ, E. W. BROWN, AND S. R. MONDAY. 2010. Population dynamics of *Vibrio* spp. associated with marine sponge microcosms. *ISME Journal* 4: 1608–1612.

HOFFMAN, M. T., AND A. E. ARNOLD. 2008. Geography and host identity interact to shape communities of endophytic fungi in cupressaceous trees. *Mycological Research* 112: 331–344.

HUANG, W. Y., Y. Z. CAI, K. D. HYDE, H. CORKE, AND M. SUN. 2008. Biodiversity of endophytic fungi associated with 29 traditional Chinese medicinal plants. *Fungal Diversity* 33: 61–75.

JARNE, P., AND P. LAGODA. 1996. Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution* 11: 424–429.

KACI-CHAOUCH, T., O. VERNEAU, AND Y. DESDEVISES. 2008. Host specificity is linked to intraspecific variability in the genus *Lamellodiscus* (Monogenea). *Parasitology* 135: 607–616.

KAMVAR, Z. N., J. F. TABIMA, AND N. J. GRÜNWALD. 2013. Poppr: An R package for genetic analysis of populations with mixed (clonal/sexual) reproduction. R package version 1.0.0. Available at http://cran.r-project.org/package=poppr.

KASUGA, T., T. WHITE, AND J. TAYLOR. 2002. Estimation of nucleotide substitution rates in eurotiomycete fungi. *Molecular Biology and Evolution* 19: 2318–2324.

KELLERMANN, V., B. VAN HEERWAARDEN, C. M. SGRO, AND A. A. HOFFMANN. 2009. Fundamental evolutionary limits in ecological traits drive drosophila species distributions. *Science* 325: 1244–1246.

KESHAVMURTHY, S., C.-M. HSU, C.-Y. KUO, P.-J. MENG, J.-T. WANG, AND C. A. CHEN. 2012. Symbiont communities and host genetic structure of the brain coral *Platygyra verweyi*, at the outlet of a nuclear power plant and adjacent areas. *Molecular Ecology* 21: 4393–4407.

LAJEUNESSE, T. C., D. T. PETTAY, E. M. SAMPAYO, N. PHONGSUWAN, B. BROWN, D. O. OBURA, O. HOEGH-GULDBERG, AND W. K. FITT. 2010. Long-standing environmental conditions, geographic isolation and host–symbiont specificity influence the relative ecological dominance and genetic diversification of coral endosymbionts in the genus *Symbiodinium. Journal of Biogeography* 37: 785–800.

LAMIT, L. J., M. K. LAU, C. M. STHULTZ, S. C. WOOLEY, T. G. WHITHAM, AND C. A. GEHRING. 2014. Tree genotype and genetically based growth traits structure twig endophyte communities. *American Journal of Botany* 101: 467–478.

LANFEAR, R., B. CALCOTT, S. Y. W. HO, AND S. GUINDON. 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* 29: 1695–1701.

LANTZ, H., P. R. JOHNSTON, D. PARK, AND D. W. MINTER. 2011. Molecular phylogeny reveals a core clade of Rhytismatales. *Mycologia* 103: 57–74.

LARKIN, B. G., L. S. HUNT, AND P. W. RAMSEY. 2012. Foliar nutrients shape fungal endophyte communities in Western white pine (*Pinus monticola*) with implications for white-tailed deer herbivory. *Fungal Ecology* 5: 252–260.

MASON-GAMER, R., AND E. KELLOGG. 1996. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). *Systematic Biology* 45: 524–545.

MCDONALD, B. A., AND C. LINDE. 2002. Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology* 40: 349–379.

MILGROOM, M. G. 1996. Recombination and the multilocus structure of fungal populations. *Annual Review of Phytopathology* 34: 457–477.

MINTER, D. W. 1981. *Lophodermium* species on pines. *In* C. S. Millar [ed.], Current research on conifer needle diseases. Proceedings of the IUFRO Working Party on Needle Diseases, 49–58, Sarajevo, 1980. Aberdeen University Press, Aberdeen, Scotland.

MONACELL, J. T., AND I. CARBONE. 2014. Mobyle SNAP Workbench: A web-based analysis portal for population genetics and evolutionary genomics. *Bioinformatics* 10.1093/bioinformatics/btu055.

MÜLLER, M., R. VALJAKKA, A. SUOKKO, AND J. HANTULA. 2001. Diversity of endophytic fungi of single Norway spruce needles and their role as pioneer decomposers. *Molecular Ecology* 10: 1801–1810.

MÜLLER, M. M., R. VALJAKKA, AND J. HANTULA. 2007. Genetic diversity of *Lophodermium piceae* in South Finland. *Forest Pathology* 37: 329–337.

MUNKACSI, A. B., S. STOXEN, AND G. MAY. 2008. *Ustilago maydis* populations tracked maize through domestication and cultivation in the Americas. *Proceedings. Biological Sciences* 275: 1037–1046.

NEI, M., AND L. JIN. 1989. Variances of the average numbers of nucleotide substitutions within and between populations. *Molecular Biology and Evolution* 6: 290–300.

NEU, C., D. KAEMMER, G. KAHL, D. FISCHER, AND K. WEISING. 1999. Polymorphic microsatellite markers for the banana pathogen *Mycosphaerella fijiensis. Molecular Ecology* 8: 523–525.

NIELSEN, R., AND J. WAKELEY. 2001. Distinguishing migration from isolation: A Markov chain Monte Carlo approach. *Genetics* 158: 885–896.

ORTIZ-GARCIA, S., D. GERNANDT, J. STONE, P. JOHNSTON, I. CHAPELA, R. SALAS-LIZANA, AND E. ALVAREZ-BUYLLA. 2003. Phylogenetics of *Lophodermium* from pine. *Mycologia* 95: 846–859.

OSONO, T., AND D. HIROSE. 2011. Colonization and lignin decomposition of pine needle litter by *Lophodermium pinastri. Forest Pathology* 41: 156–162.

PANCHER, M., M. CEOL, P. E. CORNEO, C. M. O. LONGA, S. YOUSAF, I. PERTOT, AND A. CAMPISANO. 2012. Fungal endophytic communities in grapevines (*Vitis vinifera* L.) respond to crop management. *Applied and Environmental Microbiology* 78: 4308–4317.

PARADIS, E., J. CLAUDE, AND K. STRIMMER. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20: 289–290.

PRICE, E., AND I. CARBONE. 2005. SNAP: Workbench management tool for evolutionary population genetic analysis. *Bioinformatics* 21: 402–404.

PRITCHARD, J., M. STEPHENS, AND P. DONNELLY. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

RODRIGUEZ, R. J., J. F. WHITE JR., A. E. ARNOLD, AND R. S. REDMAN. 2009. Fungal endophytes: diversity and functional roles. *New Phytologist* 182: 314–330.

RONQUIST, F., M. TESLENKO, P. VAN DER MARK, D. L. AYRES, A. DARLING, S. HOHNA, B. LARGET, L. LIU, M. A. SUCHARD, AND J. P. HUELSENBECK. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61: 539–542.

SALAS-LIZANA, R., N. S. SANTINI, A. MIRANDA-PEREZ, AND D. I. PINERO. 2012. The Pleistocene glacial cycles shaped the historical demography and phylogeography of a pine fungal endophyte. *Mycological Progress* 11: 569–581.

SAUNDERS, M., AND L. M. KOHN. 2009. Evidence for alteration of fungal endophyte community assembly by host defense compounds. *New Phytologist* 182: 229–238.

SEGHERS, D., L. WITTEBOLLE, E. M. TOP, W. VERSTRAETE, AND S. D. SICILIANO. 2004. Impact of agricultural practices on the *Zea mays* L. endophytic community. *Applied and Environmental Microbiology* 70: 1475–1482.

SIEBER-CANAVESI, F., O. PETRINI, AND T. N. SIEBER. 1991. Endophytic *Leptostroma* species on *Picea abies, Abies alba*, and *Abies balsamea*: A cultural, biochemical, and numerical study. *Mycologia* 83: 89–96.

SHIPUNOV, A., G. NEWCOMBE, A. K. H. RAGHAVENDRA, AND C. L. ANDERSON. 2008. Hidden diversity of endophytic fungi in an invasive plant. *American Journal of Botany* 95: 1096–1108.

SLATKIN, M. 1993. Isolation by distance in equilibrium and nonequilibrium populations. *Evolution* 47: 264–279.

SLATKIN, M., AND R. HUDSON. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129: 555–562.

STAMATAKIS, A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.

STONE, J. K., C. W. BACON, AND J. F. WHITE. 2000. An overview of endophytic microbes: endophytism defined. *In* C. W. Bacon and J. F. J. White [eds.], Microbial endophytes, 3–29. Marcel Dekker, New York, New York, USA.

STRASBURG, J. L., AND L. H. RIESEBERG. 2010. How robust are "Isolation with Migration" analyses to violations of the IM model? A simulation study. *Molecular Biology and Evolution* 27: 297–310.

SULLIVAN, T. J., AND S. H. FAETH. 2004. Gene flow in the endophyte *Neotyphodium* and implications for coevolution with *Festuca arizonica. Molecular Ecology* 13: 649–656.

SUN, X., Q. DING, K. D. HYDE, AND L. D. GUO. 2012. Community structure and preference of endophytic fungi of three woody plants in a mixed forest. *Fungal Ecology* 5: 624–632.

TAJIMA, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.

TAMURA, K., AND M. NEI. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10: 512–526.

TAUTZ, D. 1993. Notes on the definition and nomenclature of tandemly repetitive DNA sequences. *In* S. D. J. Pena, R. Chakraborty, J. T. Epplen, and A. J. Jeffreys [eds.], DNA fingerprinting: State of science, 21–28. Birkhauser, Basel, Switzerland.

TAYLOR, J., D. JACOBSON, S. KROKEN, T. KASUGA, D. GEISER, D. HIBBETT, AND M. FISHER. 2000. Phylogenetic species recognition and species concepts in fungi. *Fungal Genetics and Biology* 31: 21–32.

TAYLOR, J., E. TURNER, J. P. TOWNSEND, J. R. DETTMAN, AND D. JACOBSON. 2006. Eukaryotic microbes, species recognition and the geographic limits of species: Examples from the kingdom Fungi. *Philosophical Transactions of the Royal Society, B, Biological Sciences* 361: 1947–1963.

TEHON, L. R. 1935. A monographic rearrangement of *Lophodermium.* Illinois Biological Monographs 13(4). University of Illinois, Urbana, Illinois, USA.

TEMPLETON, A., K. CRANDALL, AND C. SINGH. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction

endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132: 619–633.

TIBSHIRANI, R., G. WALTHER, AND T. HASTIE. 2001. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society, B, Statistical Methodology* 63: 411–423.

TRACY, C., AND H. WIDOM. 1994. Level-spacing distributions and the airy kernel. *Communications in Mathematical Physics* 159: 151–174.

U'REN, J. M., F. LUTZONI, J. MIADLIKOWSKA, A. D. LAETSCH, AND A. E. ARNOLD. 2012. Host and geographic structure of endophytic and endolichenic fungi at a continental scale. *American Journal of Botany* 99: 898–914.

VINCENOT, L., K. NARA, C. STHULTZ, J. LABBÉ, M.-P. DUBOIS, L. TEDERSOO, F. MARTIN, AND M-A. SELOSSE. 2012. Extensive gene flow over Europe and possible speciation over Eurasia in the ectomycorrhizal basidiomycete *Laccaria amethystina* complex. *Molecular Ecology* 21: 281–299.

WEIR, B. S., AND C. C. COCKERHAM. 1984. Estimating *F*-statistics for the analysis of population structure. *Evolution* 38: 1358–1370.

WILSON, R., R. WHEATCROFT, J. D. MILLER, AND N. J. WHITNEY. 1994. Genetic diversity among natural populations of endophytic *Lophodermium pinastri* from *Pinus resinosa. Mycological Research* 98: 740–744.

WRIGHT, S. 1931. Evolution in Mendelian populations. *Genetics* 16: 97–0159.

ZIMMERMAN, N. B., AND P. M. VITOUSEK. 2012. Fungal endophyte communities reflect environmental structuring across a Hawaiian landscape. *Proceedings of the National Academy of Sciences, USA* 109: 13022–13027.