



Contents lists available at ScienceDirect

## Molecular Phylogenetics and Evolution

journal homepage: [www.elsevier.com/locate/ympev](http://www.elsevier.com/locate/ympev)

## Conserved genomic collinearity as a source of broadly applicable, fast evolving, markers to resolve species complexes: A case study using the lichen-forming genus *Peltigera* section *Polydactylon*

Nicolas Magain<sup>a,\*</sup>, Jolanta Miadlikowska<sup>a</sup>, Olaf Mueller<sup>b</sup>, Michael Gajdeczka<sup>c</sup>, Camille Truong<sup>a</sup>, Asaf A. Salamov<sup>d</sup>, Inna Dubchak<sup>d</sup>, Igor V. Grigoriev<sup>d</sup>, Bernard Goffinet<sup>e</sup>, Emmanuël Sérusiaux<sup>f</sup>, François Lutzoni<sup>a</sup>

<sup>a</sup> Department of Biology, Duke University, Durham, NC 27708, United States

<sup>b</sup> Duke Center for the Genomics of Microbial Systems, Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC 27708, United States

<sup>c</sup> Lambert Spawn, Coatesville, PA 19320, United States

<sup>d</sup> US Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, United States

<sup>e</sup> Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269, United States

<sup>f</sup> Evolution and Conservation Biology, University of Liège, Sart Tilman B22, B-4000 Liège, Belgium

## ARTICLE INFO

## Keywords:

Synteny  
Intergenic spacers  
Species delimitation  
Species discovery  
Species validation  
Lichen-forming fungi

## ABSTRACT

Synteny can be maintained for certain genomic regions across broad phylogenetic groups. In these homologous genomic regions, sites that are under relaxed purifying selection, such as intergenic regions, could be used broadly as markers for population genetic and phylogenetic studies on species complexes. To explore the potential of this approach, we found 125 Collinear Orthologous Regions (COR) ranging from 1 to > 10 kb across nine genomes representing the Lecanoromycetes and Eurotiomycetes (Pezizomycotina, Ascomycota). Twenty-six of these COR were found in all 24 eurotiomycete genomes surveyed for this study. Given the high abundance and availability of fungal genomes we believe this approach could be adopted for other large groups of fungi outside the Pezizomycotina. As a proof of concept, we selected three Collinear Orthologous Regions (COR1b, COR3, and COR16), based on synteny analyses of several genomes representing three classes of Ascomycota: Eurotiomycetes, Lecanoromycetes, and Lichinomycetes. COR16, for example, was found across these three classes of fungi. Here we compare the resolving power of these three new markers with five loci commonly used in phylogenetic studies of fungi, using section *Polydactylon* of the cyanolichen-forming genus *Peltigera* (Lecanoromycetes) – a clade with several challenging species complexes. Sequence data were subjected to three species discovery and two validating methods. COR markers substantially increased phylogenetic resolution and confidence, and highly contributed to species delimitation. The level of phylogenetic signal provided by each of the COR markers was higher than the commonly used fungal barcode ITS. High cryptic diversity was revealed by all methods. As redefined here, most species represent lineages that have relatively narrower, and more homogeneous biogeographical ranges than previously understood. The scabrosoid clade consists of ten species, seven of which are new. For the dolichorhizoid clade, twenty-two new species were discovered for a total of twenty-nine species in this clade.

## 1. Introduction

While next generation sequencing methods have proven to be very useful in accelerating data acquisition for fungi (e.g., Rokas et al., 2003; Fitzpatrick et al., 2006; Jeffroy et al., 2006; Kuramae et al., 2006; Robertse et al., 2006; Liu et al., 2009; Ebersberger et al., 2012; McDonald et al., 2013; Leavitt et al., 2016), selecting comparable sets of

molecular markers for phylogenetic studies remains problematic. Microsatellites have been often used to resolve species complexes, however, microsatellites selected for one species complex are usually not suitable for other species, even within the same genus. Although genome, transcriptome, and partial genome-wide (ddRadSeq, Peterson et al., 2012) sequencing can generate large numbers of highly variable sites to resolve species complexes, these techniques are not broadly

\* Corresponding author.

E-mail address: [nicolas.magain@duke.edu](mailto:nicolas.magain@duke.edu) (N. Magain).

<http://dx.doi.org/10.1016/j.ympev.2017.08.013>

Received 9 December 2016; Received in revised form 10 August 2017; Accepted 25 August 2017

1055-7903/© 2017 Elsevier Inc. All rights reserved.

available to systematists conducting large-scale revisionary studies that require the examination of hundreds to thousands of specimens. Molecular phylogenetic studies on fungi highlighted the urgent need for novel genetic markers to resolve phylogenetic relationships with high confidence (e.g., Crespo et al., 2010). Homologous regions that are both highly variable and can be amplified across genera, families, orders and classes, would provide an ideal set of markers for the study of populations and species complexes, as exemplified by the ITS region, which has been used as the primary fungal barcode marker for more than 20 years (Schoch et al., 2012). One advantage of using such markers is that the results, e.g., rates of nucleotide substitution and transversion:transition ratios, can be directly compared among studies. Fast evolving intergenic spacers in syntenic regions of genomes where collinearity, including gene orientation, is conserved across a broad array of fungal genera, families and orders (Hane et al., 2011), are likely to provide the homologous fast-evolving phylogenetic markers that are needed to discover, delimit and describe fungal biodiversity more efficiently. The sequencing of these regions can be done with next generation sequencing (e.g., low coverage genomics) for large sets of markers, or using PCR-based Sanger sequencing for a more restricted set.

Estimates of the global fungal biodiversity range from 1.5 to 5.1 million species (Hawksworth, 1991, 2001; O'Brien et al., 2005; Blackwell, 2011), yet only ca. 100,000 species of fungi have been described so far (Kirk et al., 2008). Bridging morphology-based species recognition with molecular approaches, and finding appropriate criteria to delimit species, defined as “separately evolving metalineages” (De Queiroz, 1998), are among the biggest challenges in modern fungal and lichen systematics (Taylor et al., 2000; Lumbsch and Leavitt, 2011). For example, recognizing biological species as “groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups” (Mayr, 1940), by testing mating compatibility is generally unachievable with lichens, because most lichenized fungi cannot be isolated in pure culture, or grow extremely slowly, and do not develop sexual reproductive structures in vitro (but see Zoller et al., 1999). Asexuality or homothallism, in some species, may also complicate the recognition of biological species in lichen-forming fungi (Taylor et al., 2000; but see Scherrer et al., 2005).

Species delimitation methods relying on molecular data are among the most widely used tools in modern systematics (see Carstens et al., 2013). Unlike theoretical species concepts, delimitation methods aim at testing operational criteria based on empirical data, i.e. properties acquired by the lineages during the course of divergence (De Queiroz, 2007). Therefore, all methods rely on a certain model and a limited set of parameters, which are not biased free (see for example Sukumaran and Knowles, 2017), but provide valuable information to test and characterize evolutionary meaningful lineages, and remove part of the subjectivity that is inherent to classical taxonomical work. Among the constantly increasing list of available methods, Structurama (Huelsenbeck et al., 2011) relies on multilocus genotype data to infer population structure and assign individuals to populations (Pritchard et al., 2000), where each population (or species) is modeled by a characteristic set of allele frequencies. bGMYC (Reid and Carstens, 2012) is a Bayesian implementation of the GMYC (General Mixed Yule Coalescent) model, which aims at finding a threshold between speciation and coalescent (inter- and intraspecific) branching patterns on ultrametric phylogenetic trees. bPTP (bayesian Poisson Tree Process, Zhang et al., 2013) implements a similar approach, but based on the number of substitutions (branch lengths) instead of time, considering that each substitution has a small probability to generate a speciation event. spedeSTEM (Kubatko et al., 2009; Ence and Carstens, 2011) calculates the maximum likelihood species tree from a set of gene trees with an a-priori assignment of species, and compares the likelihoods of the species tree using the proposed units as distinct species, with species trees for which the proposed units are merged following certain hierarchical permutations, to propose an optimal species delimitation. BPP

(Yang and Rannala, 2010) is a Bayesian approach that generates the posterior probabilities of species assignments, to test if lineages in the species tree should be considered as distinct lineages or merged into more inclusive species, based on multi-locus datasets, taking uncertainties due to unknown gene trees and the ancestral coalescent process into account.

The monophyletic section *Polydactylon*, one of eight sections currently recognized within the genus *Peltigera* (Miadlikowska and Lutzoni, 2000), is composed of three main, highly supported clades, the dolichorhizoid, polydactyloid and scabrosoid clades (Magain et al., 2017). *Peltigera* species (mycobionts) from this section are known to associate with one type of photosynthetic partner, the cyanobacterium *Nostoc* (cyanobiont). Species from section *Polydactylon* reproduce sexually through apothecia, dispersing the fungus without its cyanobiont. Some species also develop vegetative propagules, mostly marginal phyllidia, containing both the fungal and cyanobacterial partners. However, all these *Peltigera* species have the potential to reproduce asexually by thallus fragmentation and, consequently, transmit their cyanobiont vertically to the next generation.

Although section *Polydactylon* is almost cosmopolitan in distribution, the geographic ranges of individual species vary greatly, from extremely narrow (e.g., *P. melanorrhiza*, endemic to the Azores) to subcosmopolitan (e.g., *P. dolichorhiza* and *P. polydactylon* have been reported from five and six continents, respectively; Martínez et al., 2003; Magain et al., 2016). Prior to 2016, this section comprised fifteen species (Holtan-Hartwig, 1993; Vitikainen, 1994; Miadlikowska and Lutzoni, 2000; Sérusiaux et al., 2009), but the recognition of geographically structured morphotypes and chemotypes within some species, such as *P. neopolydactyla*, *P. scabrosa* (Holtan-Hartwig, 1993; Vitikainen, 1994), and *P. dolichorhiza* (Sérusiaux et al., 2009; Magain et al., 2017) strongly suggested the presence of multiple undescribed species.

A worldwide molecular phylogenetic study of section *Polydactylon* using five nuclear loci revealed the presence of at least 38 monophyletic putative, and predominantly cryptic, species (Magain et al., 2017). This preliminary species delimitation scheme was proposed based on two species discovery methods, bGMYC (Reid and Carstens, 2012) and Structurama (Huelsenbeck et al., 2011). Phylogenetic uncertainty, mainly in the dolichorhizoid clade, due to low variation in the three protein- and two ribosomal RNA-coding genes sequenced for that project, and incomplete sampling within the polydactyloid clade, especially in Asia and Australasia, limited our ability to delimit several putative species. New highly variable markers, applicable across the entire genus, were needed to confirm the delimitation of these new species.

The first aim of this study was to explore the potential of syntenic collinear regions conserved across orders and classes of fungi, for the selection of new loci for phylogenetic studies of species complexes. This was achieved by (1) conducting a synteny analysis on genomes of two lichen-forming fungi from two different orders within the Lecanoromycetes: we screened for collinear regions, where gene content, order, and orientation were maintained across genomes, and selected Collinear Orthologous Regions (COR) that included promising intergenic spacers, i.e., high variability, 600–1200 bp long, and highly conserved flanking protein-coding gene sequences, that are conducive to the development of universal primers for large fungal clades; (2) comparing these COR markers to widely-used loci in their contribution to phylogenetic studies and species delimitation methods; and (3) searching for additional COR regions that might be conserved across Lecanoromycetes and Eurotiomycetes, to gain a better idea of the global number of these intergenic markers that might be available across classes of fungi. Our second aim was to re-evaluate species delimitations within the scabrosoid and dolichorhizoid clades of section *Polydactylon* as defined by Magain et al. (2017), using the following approach: (1) develop *Peltigera*-specific primers for three COR markers (COR1b, COR3, and COR16) and add these three markers to the five

loci sequenced by Magain et al. (2017); (2) infer phylogenetic relationships for the entire section and implement three species discovery methods and two species validation methods to verify species delimitations presented in Magain et al. (2017); and (3) reassess the utility of species distributions and patterns of association with *Nostoc* phylogroups (sensu Magain et al., 2017) as a source of supplementary information supporting the newly circumscribed *Peltigera* species.

## 2. Materials and methods

### 2.1. Selection of Collinear Orthologous Regions (COR)

We explored the 30 largest scaffolds (macrosyntenic regions) of the *Cladonia grayi* genome assembly (Clagr3; <http://genome.jgi.doe.gov/Cladonia>), aligned to corresponding scaffolds of the *Xanthoria parietina* assembly (Xanpa1; <http://genome.jgi.doe.gov/Xanthoria>) in the JGI Vista Browser Synteny tool (Frazer et al., 2004; Grigoriev et al., 2011). We ranked nearly two hundred collinear syntenic blocks, including gene orientation, according to the following criteria: (1) greatest sequence variability in non-coding regions; (2) high to moderate sequence conservation in potential primer sites; (3) highest proportion of non-protein coding sequence in an amplicon of 600–1200 bp; (4) lack of obvious linkage to other markers; and (5) potential for development of internal primers for nested amplification. We used a custom implementation of the BLASTP 2.3X tool (Deng et al., 2007) to rank the 40 best regions according to shared homology (in conserved regions) with existing assemblies of metagenomic and metatranscriptomic data from low-coverage sequencing of three *Peltigera* lichen species: *P. dolichorhiza* (Magain et al., 2010), *P. membranacea* (Xavier et al., 2012), and *P. praetextata* (Hodkinson et al., 2014). To develop PCR-based markers, we selected 20 regions that shared the highest similarity in potential primer annealing sites, which were typically in exons regions flanking the intergenic spacer for at least one of the primers. To avoid non-specific annealing and amplification, one of the primers can be designed to target an untranslated region (UTR) or another variable region that is conserved for the targeted group but different from non-targeted fungal groups. Based on *Peltigera* data, when possible, we designed two to ten primers, resulting in a total of two to twelve primer combinations per targeted locus. We then tested these primers using PCR amplification and Sanger sequencing.

Using our comparative genomic synteny approach, three markers (COR1, COR3 and COR16) were found to be most promising for this project. COR1 is located on the 43th scaffold of the *Cladonia grayi* (JGI Mycoscosm Clagr3) genome assembly, between base pairs 224,478 and 226,844, and on scaffold 19 of the *Xanthoria parietina* (JGI Mycoscosm Xanpa2) assembly, between base pairs 632,451 and 634,199. COR3 is located within the eighth scaffold of the Clagr3 genome assembly, between base pairs 287,206 and 292,304, and in the seventh scaffold of the Xanpa1 assembly (between base pairs 1,104,171 and 1,108,382). COR16 is located within scaffold 9 of the Clagr3 assembly, between base pairs 225,330 and 230,241, and on scaffold 1 of Xanpa2, between base pairs 2,640,063 and 2,644,705. We looked for these COR regions across available genomes in the class Eurotiomycetes, Lecanoromycetes, and Lichinomycetes (Fig. 1, Supplementary Table S1). We identified syntenic blocks across these genomic regions with LASTZ (Harris, 2007) using *Cladonia grayi* as a reference genome. Primers were tested on *Peltigera* samples from most sections of the genus. Some primer pairs were successful in amplifying the targeted loci COR3 and COR16. COR1, as detected in Clagr3, which is composed of an intergenic spacer flanked by genes coding for the splicing factor *Ini1* and the nuclear protein *Sdk3* (sphingosine-dependent kinase 3), is not present in *Peltigera*. However, the primers designed for COR1 amplified a different intergenic locus that seems to be *Peltigera*-specific. We refer to this locus as COR1b, which includes an intergenic region flanked by genes coding for *Acl2* (ATP-citrate subunit 2 gene) and *Sdk3*. The sequences of the newly designed primers used for the amplification of the three new COR markers are shown in

Table 1. Primers specific to other genera within the Lecanoromycetes and closely related classes, i.e., Eurotiomycetes for COR3, Eurotiomycetes and Lichinomycetes for COR16 respectively, can be developed in the highly conserved flanking regions where the *Peltigera* primers are binding, by using the alignments provided in Supplementary Fig. S1.

### 2.2. Synteny analysis for discovering potential Collinear Orthologous Regions across Lecanoromycetes and Eurotiomycetes

Additional syntenic regions were identified based on conserved collinear gene orthologs found in nine genomes representing Lecanoromycetes and Eurotiomycetes (Fig. 2 and Supplementary Table S1). Initially for each pair of genomes, orthologous genes were selected as pairs of genes forming best bidirectional Blast hits (BBH) between each other. Subsequently, based on pairwise orthologs, we identified 3134 ortholog clusters using complete linkage analysis, containing one gene from each of the nine genomes. Based on the ortholog clusters, we searched for genomic regions that included two or more ortholog genes adjacent within each genome and collinear among all nine genomes. In total, we found 125 such syntenic regions, with a large majority (119) consisting of 2 collinear genes, 5 regions including 3 genes, and one four-gene region. Using genes from these 125 conserved regions in *Aspergillus nidulans*, we searched for similar regions in 24 other published Eurotiomycetes genomes (Fig. 2B, Supplementary Table S1), requiring adjacency and collinearity of corresponding best Blast hit genes.

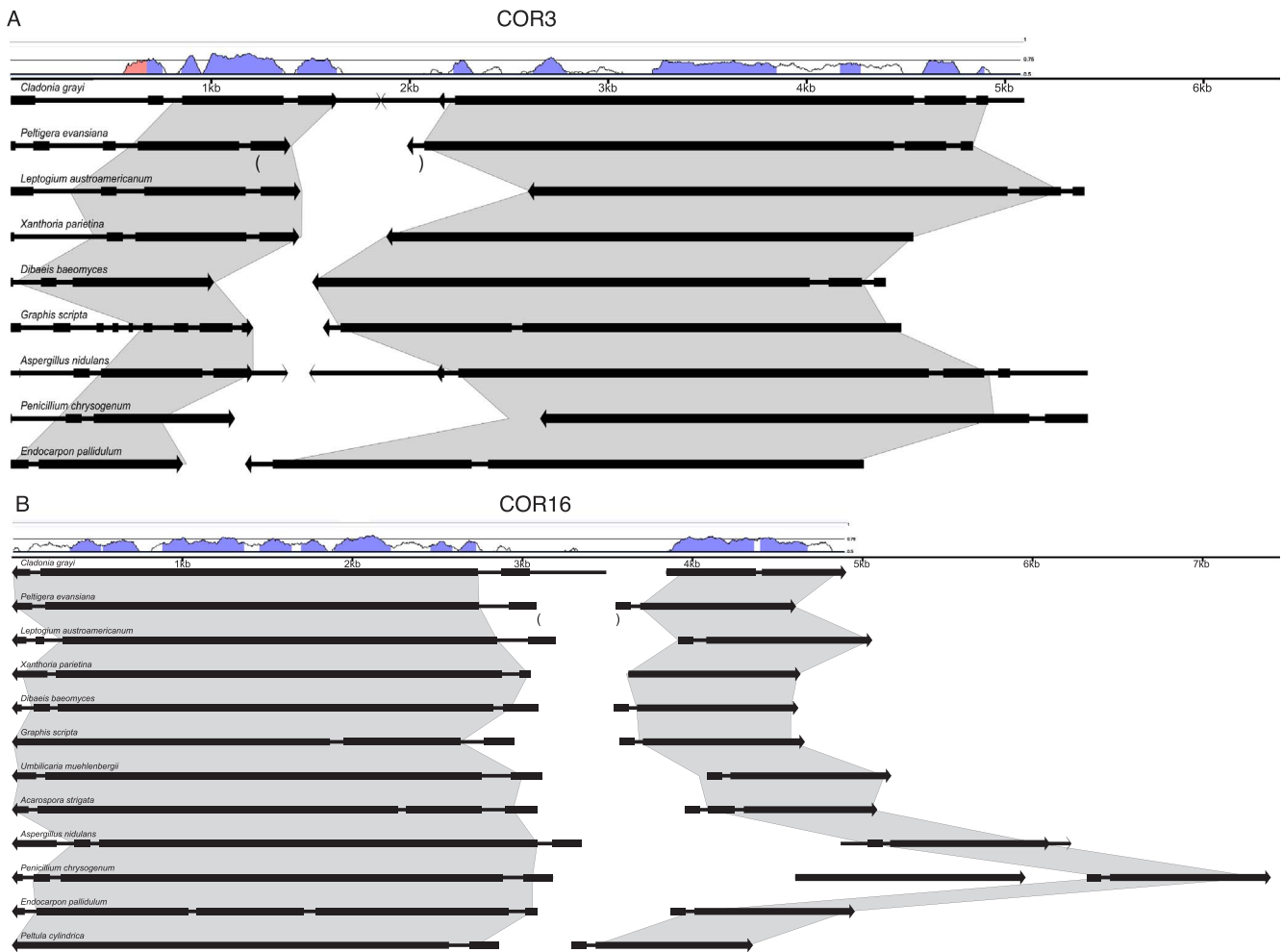
### 2.3. Taxon sampling and molecular data acquisition

The initial selection of specimens for DNA extraction and sequencing of the mycobiont (*Peltigera*) and cyanobiont (*Nostoc*) was based on morphological examination of over 2000 specimens obtained as loans from several herbaria world-wide (AMNH, B, BG, CGMS, CONN, DUKE, H, LG, MAF, MEXU, NSPM, NY, O, OSU, PTZ, QFA, UBC, UDBC, UGDA, UMEM, UPS), various private collections, and from our own field work (Reunion Island in 2009; China: Yunnan in 2010; Norway; Canada: Québec; USA: North Carolina and Alaska in 2011; Russia; Peru and Brazil in 2012). Here, we added 48 individuals to the 116 specimens included in Magain et al. (2017; Supplementary Table S2) for a total of 164 members from section *Polydactylon*. Three old herbarium specimens included in Magain et al. (2017) were not used in the current dataset due to missing, and/or low quality sequences: *P. neopolydactyla* 2 (P1760) from the dolichorhizoid clade; *P. oceanica* (P882) and *P. sumatrana* (P884) from the polydactyloid clade. Consequently, the latter two species were not represented in this study.

For all new samples and selected individuals with incomplete sequence data in Magain et al. (2017), we targeted the nuclear ribosomal loci nrLSU and ITS, as well as three nuclear protein-coding genes: *RPB1*, *EFT2.1* and  $\beta$ -tubulin. COR1b, COR3, and COR16 were targeted for all 164 specimens. A total of 489 new sequences (Supplementary Table S2) were obtained for the mycobionts: 27 nrITS, 39 nrLSU, 38  $\beta$ -tubulin, 47 *RPB1*, 36 *EFT2.1*, 75 COR1b, 116 COR3, and 111 COR16. We also sequenced the *rbcLX* region for the cyanobacterial partners of 21 of the 48 newly selected *Peltigera* specimens. All extraction, amplification and sequencing protocols follow Magain et al. (2017). For COR1b, COR3 and COR16, PCR conditions were: 94 °C for 30 s, 55 °C for 30 s (–0.4°/cycle), 72 °C for 1 min (+2 s/cycle) for 24 cycles; 94 °C for 30 s, 45 °C for 30 s, 72 °C for 2 min (+3 s/cycle) for 12 cycles; 72 °C for 10 min, followed by storage at 4 °C.

### 2.4. Datasets and phylogenetic analyses

Sequences were edited using Sequencher version 4.9 (Gene Codes Corporation, Ann Arbor, Michigan) and aligned manually with MacClade v. 4.08 (Maddison and Maddison, 2005). To reduce the number of ambiguously aligned sites excluded from phylogenetic analyses, single- and multi-locus datasets were assembled for the dolichorhizoid and scabrosoid clades separately.



**Fig. 1.** Syntenic conservation of COR3 (A), and COR16 (B) across genomes of two and three classes of fungi, respectively. *Cladonia grayi* to *Acarospora strigata* are classified within Lecanoromycetes. The genera *Aspergillus*, *Penicillium* and *Endocarpon* are members of the Eurotiomycetes, whereas *Peltula* represents the Lichinomycetes. Highly similar sequences are indicated in light grey. A scale in kb is shown at the top. Genes are marked as black lines, coding regions (CDS) as black blocks, and gene orientations are indicated by arrowheads. Similarity between the genomes of *Cladonia grayi* and *Xanthoria parietina* (50 bp sliding window, as determined by the JGI browser) are shown on the top row. Blue zones represent exons, with Conserved Non-Coding Sequences (CNS) indicated in red. Parentheses indicate the location of primers for this study, using *Peltigera evansiana* as an example.

**Table 1**

Primers used for the amplification and sequencing of the newly developed molecular markers COR1b, COR3, and COR16 for *Peltigera* section *Polydactylon*.

Primer name	Primer sequence (5'–3')
COR-1bF	GCTGTCGGCGAAGAGCTGAA
COR-1bR-B	CCATTTCTCCGCGTTCTGGTA
COR-3F-A	GGAGACGTTGCTAATGCATT
COR-3R-B	CCGAAGTCCGCTCTGAAGACA
COR-16Fout	GCGGAKGCGCAGATGATTG
COR-16Rmid1	TGTGGCACGGTGAACACTAG

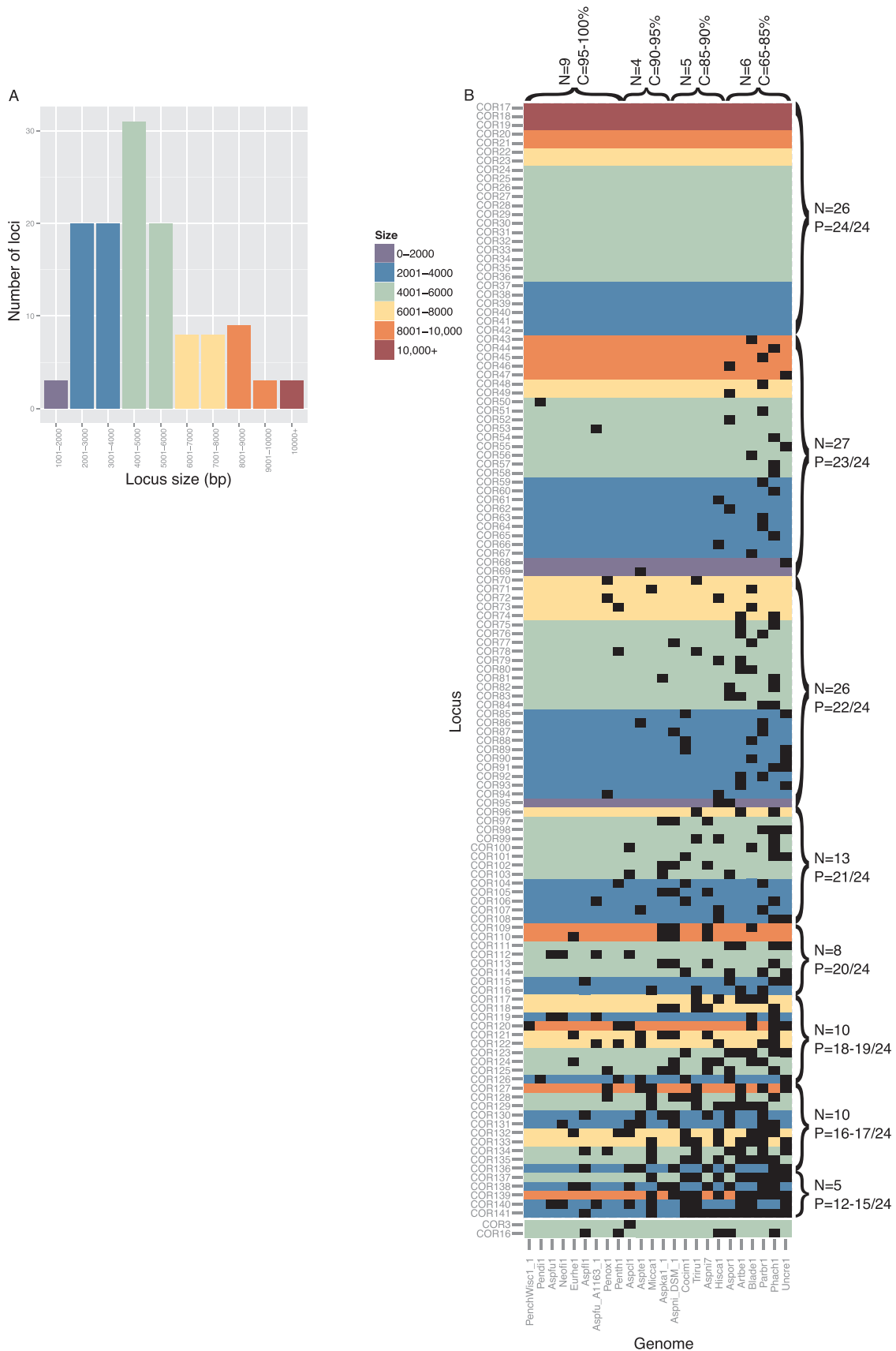
For each single-locus dataset the best model for nucleotide substitution was determined by MrModelTest v. 2.3 (Nylander, 2004). For all concatenated datasets, we used PartitionFinder (Lanfear et al., 2012; greedy algorithm and BIC criterion) to determine the best partitioning scheme and the optimal models to use in subsequent multi-locus phylogenetic analyses. The following 18 data subsets were pre-delimited: nrLSU, ITS1 + ITS2, 5.8S, COR1b (missing in the scabrosoid dataset), COR3, COR16, and, for each protein-coding gene (*RPB1*, *EFT2.1* and  $\beta$ -tubulin) four subsets corresponding to each codon position plus non-coding regions. Data partitions and optimal nucleotide substitution models used in various analyses are provided in Supplementary Table S3.

We generated single-locus phylogenetic trees for the entire section

and for the dolichorhizoid and scabrosoid clades using RAxML v. 7.4.2 and alternatively v. 8.0.9 (Stamatakis, 2006; Stamatakis et al., 2008) as implemented on the CIPRES portal (Miller et al., 2010). All optimal trees and bootstrap searches were conducted with the rapid hill-climbing algorithm for 1000 replicates with GTR substitution model (Rodriguez et al., 1990) and gamma distribution parameter approximated with four categories. We also run an analysis on *rbclX* with the same parameters, on a dataset containing sequences from Magain et al. (2017) and newly generated sequences, to determine the *Nostoc* phylogroups of the newly generated sequences (data not shown). To detect topological incongruence among single-locus datasets, a reciprocal 70% ML bootstrap support criterion was implemented (Mason-Gamer and Kellogg, 1996; Reeb et al., 2004). Maximum likelihood tree searches and bootstrap analyses were performed on concatenated datasets for the section and each clade separately, using the same settings as in single-locus analyses. Phylogenetic relationships that received bootstrap values of 70% and above were considered well supported. We evaluated the phylogenetic contribution of each marker by comparing the numbers of well-supported nodes that were in agreement among each single-locus phylogeny as well as the topology derived from the concatenated dataset, using the Mesquite module Hypha (Oliver et al., 2013).

We performed Bayesian analyses using MrBayes v. 3.2.2 (Huelsenbeck and Ronquist, 2001) as implemented on the CIPRES portal on the





**Fig. 2.** Syntenic collinear regions found across Lecanoromycetes and Eurotiomycetes and their presence/absence in 24 Eurotiomycetes genomes. (A) Size distribution of 125 Collinear Orthologous Regions (COR) present across nine fungal genomes spanning two classes of fungi (Supplementary Table S1). (B) Matrix showing the presence/absence of these 125 COR across 24 selected Eurotiomycetes genomes. COR loci (rows) are sorted from ubiquitous (top) to the least frequent (bottom) and genomes are sorted (left to right) from most inclusive (containing the highest number of the 125 COR) to the least inclusive. Black cells show COR loci absences. The presence of COR loci is color-coded according to their size as shown in (A). Within each COR frequency group (e.g., the 26 COR loci that were present in all 24 genomes), COR loci are ordered from the longest to the shortest size. The last two rows show the presence/absence of COR 3 and 16 across the 24 selected Eurotiomycetes genomes.

combined 8-locus dataset for the whole section (25 million generations, Estimated Sample Sizes [ESS] of the tree  $\ln L = 6369$ ) and on each clade separately (15 million generations, dolichorhizoid clade: ESS  $\ln L = 6789$ ; scabrosoid clade: ESS tree  $\ln L = 12022$ ) by completing 2 runs of 4 chains and sampling every 1000th generation. The first 25% of trees were discarded as burn-in, whereas the remaining trees were used to generate 50% consensus phylograms. We generated chronograms for the dolichorhizoid and scabrosoid clades with BEAST v. 1.8 (Drummond and Rambaut, 2007) as implemented on the CIPRES portal. For the concatenated analyses, we completed 50 million generations, sampling every 1000th generation, and discarded 20% of the trees as burn-in. For the single-locus analyses, we completed 15 million generations, sampling every 1000th generation, and discarded 20% of the trees as burn-in. We performed Likelihood Ratio Tests (LRT) on tree likelihoods generated from clocklike and non-clocklike models applied to the best ML trees, using PAML (Yang, 1997), to determine the adequate clock models for all analyses. A lognormal relaxed clock was selected for all analyses, except for the single-locus analyses of *RPB1* in the dolichorhizoid and scabrosoid clades, and the analysis of  $\beta$ -tubulin in the scabrosoid clade, which were run with a strict molecular clock model. Convergence of Bayesian results from all runs was explored using Tracer v. 1.5 (Rambaut and Drummond, 2007) and AWTY (Nylander et al., 2008) as implemented on the portal <http://king2.scs.fsu.edu/CEBProjects/awty/>. For the LSU analysis in the dolichorhizoid clade, ESS for the posterior probability and tree likelihood of the BEAST analysis were 347 and 228, respectively. The ESS for the tree likelihood and posterior probability were  $> 500$  in all other analyses. For the scabrosoid and the dolichorhizoid clades, uncorrected pairwise-distances (p-distances) among ITS sequences were computed using PAUP v. 4.0a134 (Swofford, 2003) and were used to calculate interspecific and intraspecific genetic distances depicted in heatmaps using R (2002) package ggplot2 (Wickham, 2009).

## 2.5. Species delimitation methods

For delimiting and validating species in the dolichorhizoid and scabrosoid clades, the following five methods were implemented: Structurama (Huelsenbeck et al., 2011), bGMYC (Reid and Carstens, 2012), bPTP (Zhang et al., 2013), spedeSTEM (Ence and Carstens, 2011), and BPP (Yang and Rannala, 2010). For the final delimitation of species, we used an integrative approach combining the results from each method with *Nostoc* phylogroup patterns of association, geographic and phenotypic data.

Before performing Structurama analyses, we coded the sequences of each individual (seven and eight loci in the scabrosoid and dolichorhizoid clades, respectively) into alleles with Sequencher using 100% similarity as the criterion for collapsing identical haplotype sequences into a single sequence representing each allele. We ran Structurama for one million generations, sampling every 1000th generation, and tested several gamma hyperpriors on the expected number of populations (a constant gamma scale of 1 and gamma shape values of 3, 5, 8, 10, 12, 15, 18, 22, 24, 27 and 30). We also tested the impact of each locus on species delimitation by removing each locus from the complete concatenated dataset and performing analyses on the reduced 6- or 7-locus datasets for the scabrosoid or the dolichorhizoid clade, respectively. We ran four analyses on each of these datasets, completing one million generations, sampling every 1000th generation, and implementing gamma shapes of 3, 8, 15 and 30 and gamma scale of 1 for the hyperprior on the expected numbers of populations.

For the scabrosoid clade, we performed bGMYC analyses on chronograms derived from BEAST analyses on each of the seven loci (nrITS, nrLSU,  $\beta$ -tubulin, *RPB1*, *EFT2.1*, COR3, COR16). For the dolichorhizoid clade, we used only the five most variable loci: nrITS,  $\beta$ -tubulin, COR1b, COR3, and COR16. From each single-locus BEAST analysis, we selected 200 chronograms from the posterior tree distribution. Each file initially contained 15,000 trees, of which we

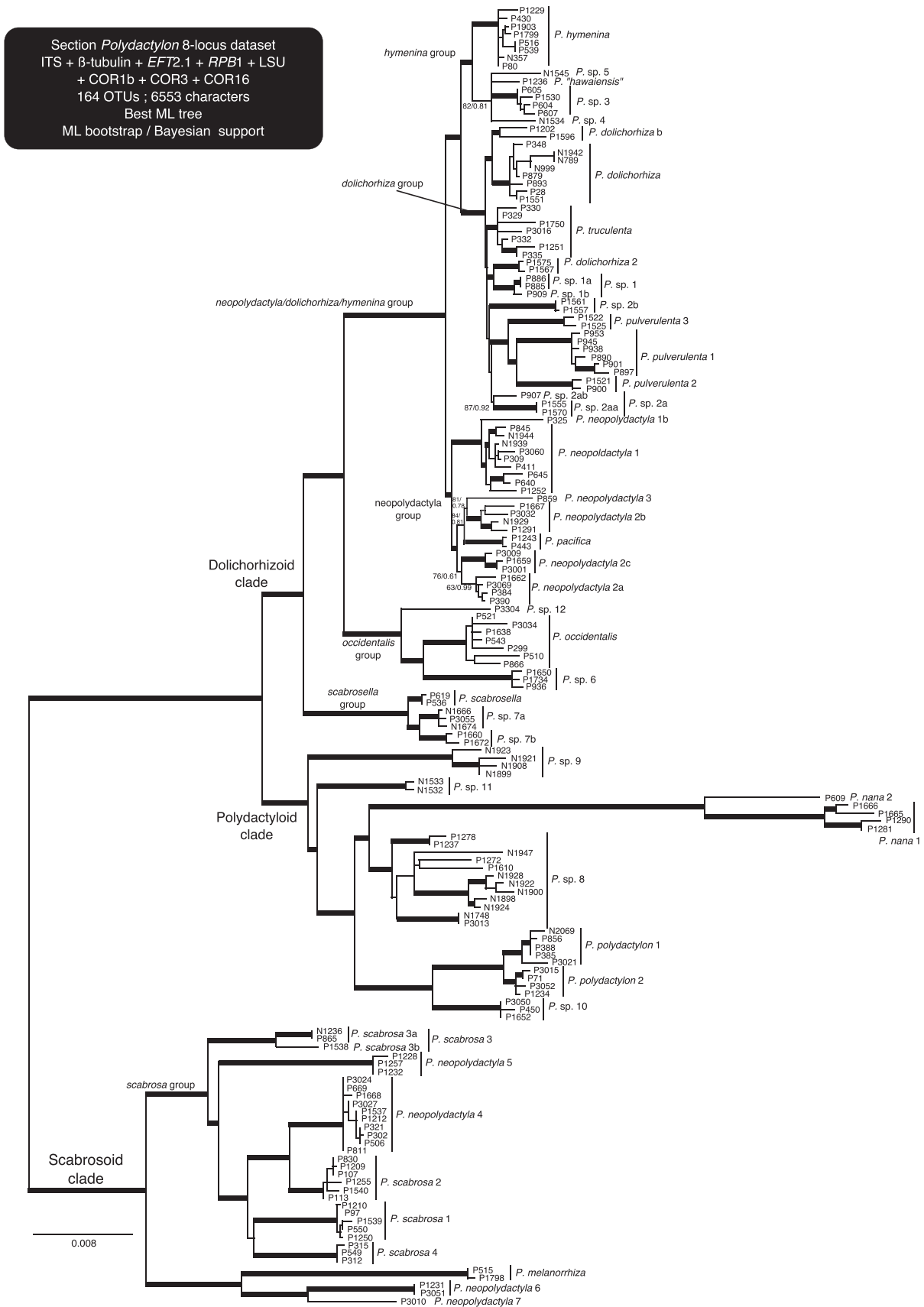
discarded the first 3000 trees as burn-in. From the remaining trees we kept every 60th sampled tree using R and the package APE (Paradis et al., 2004) to generate a final file of 200 trees. We ran bGMYC on each set of 200 trees for 50,000 generations per chronogram. We discarded 40,000 generations as burn-in with a thinning value of 100 and threshold values (corresponding to the interval of possible number of species) from 1 to 15 for the scabrosoid clade and from 2 to 40 for the dolichorhizoid clade. We considered a species to be well delimited by bGMYC when the probability of grouping haplotypes together was higher than the probability of any alternative grouping that included at least one haplotype from this putative species.

We performed bPTP as implemented on the website <http://species.h-its.org/> using the best ML trees that resulted from the RAxML analyses on the concatenated datasets for the scabrosoid and dolichorhizoid clades separately. We completed 500,000 generations with a thinning value of 1000 and discarded the first 25% as burn-in.

We performed spedeSTEM as implemented on the website <http://spedestem.osu.edu>, or alternatively with spedeSTEM v. 2.0 locally, using the discovery algorithm. Because spedeSTEM only tests broader species delimitation than specified a priori (does not split predefined species), our initial species delimitation corresponded to all individual lineages delimited in Fig. 3. For the scabrosoid clade, we used the seven single-locus chronograms generated with BEAST (used also for bGMYC analyses). We also completed the analysis without the *RPB1* chronogram because the level of genetic variation for this locus was very low across the scabrosoid clade. We estimated the relative rate of evolution for each locus based on substitution rates derived from maximum likelihood phylogenetic analyses of single loci. Because no estimate for our group of fungi was available, we tested a wide range (0.0001, 0.001, 0.005, 0.01, 0.05, and 0.1) of theta values, a parameter proportional to the population size and the substitution rate. We tested for the adequacy of the input parameters by completing an analysis using a trait file where *Peltigera neopolydactyla* 4 was randomly split into two species, and expecting them to be merged into one unit.

In the dolichorhizoid clade, single-locus chronograms were less resolved than the best ML trees. Therefore, we transformed the best ML single-locus phylogenies resulting from RAxML analyses into ultrametric trees by using non-parametric rate smoothing with TreeEdit (Rambaut and Charleston, 2008). In the test run, we assigned samples of *P. hymenina* to two distinct species as a sensitivity analysis for the adequacy of the parameters used. We tested the following theta values: 0.00001, 0.0001, 0.001, 0.005, 0.01, 0.05, 0.1, and 0.5.

We ran BPP v. 2.2 on the scabrosoid and dolichorhizoid clades using seven and eight single-locus alignments, respectively. For each clade, we used the same species assignment as for spedeSTEM (i.e., lineages defined in Fig. 3) and a guide-tree corresponding to the topology obtained from running MrBayes on the complete concatenated dataset (tree not shown). We used the species delimitation algorithm and kept all sites containing missing data. We estimated the relative rates of evolution for each locus based on substitution rates from the ML analyses. For the tau prior of the scabrosoid clade, we set the gamma shape to 4 and the gamma scale to 100, estimating the height of the root based on the branch lengths in the RAxML single-locus trees. For the theta prior, we set the gamma shape to 2, and tested scale values of 200,000, 20,000, 2000, and 200, so that the mean of the theta prior was 0.00001, 0.0001, 0.001, 0.001, and 0.01, respectively. We also tested the adequacy of the parameters by randomly splitting *P. neopolydactyla* 4 and *P. scabrosa* 2 under two theta priors: 0.0001 and 0.01. For the tau prior of the dolichorhizoid clade, we set the gamma shape to 3 and the gamma scale to 100, estimating the height of the root based on the branch lengths in RAxML single-locus trees. For the theta prior, we set the gamma shape to 2, and tested the scale values of 2000, 200, 100 and 40 so that the mean of the theta prior was 0.001, 0.01, 0.02 and 0.05, respectively. We tested the adequacy of the parameters by splitting *P. hymenina* into two groups.



**Fig. 3.** Phylogenetic relationships within *Peligeria* section *Polydactylon* as revealed by maximum likelihood (ML) analyses of a concatenated 8-locus dataset (nrITS,  $\beta$ -tubulin, *EFT2.1*, *RPB1*, nrLSU, COR1b, COR3 and COR16) representing 164 OTUs. Thick internodes received high ML bootstrap support (ML-bs  $\geq$  70%) and Bayesian posterior probabilities (pp  $\geq$  0.95). Support values are provided when internodes were highly supported by only one of the two analyses. The names of the three main clades refer to Magain et al. (2017). Putative species were defined based on monophyly, morphology, geography (Magain et al., 2017) and results from the species delimitation methods developed in this study.

### 3. Results

#### 3.1. Synteny and collinearity of COR markers

The synteny of COR1b is not well conserved compared to COR3 and COR16 and, therefore, is not likely to be recovered with genomic or metagenomic sequencing outside of the genus *Peltigera*. Contrary to the other two COR markers (COR3 and COR16), COR1b could only be amplified from *Peltigera* specimens using the primers we developed (Table 1).

The synteny of the genes flanking the intergenic spacer of COR3 (CLAGR\_001918 and CLAGR\_001919) and COR16 (CLAGR\_002094, CLAGR\_002095) in *Cladonia grayi* is maintained in several fungal genomes spanning two classes of fungi: Lecanoromycetes and Eurotiomycetes (Fig. 1A). The synteny of COR16 was found in two additional lecanoromycete genomes (*Umbilicaria muehlenbergii* and *Acarospora strigata*) and in one additional class of fungi: Lichinomycetes (*Peltula cylindrica*) (Fig. 1B). The lengths of the intergenic spacer regions are quite variable among species, from a few hundreds to more than one thousand base pairs (Fig. 1). Furthermore, in the genome of *Penicillium chrysogenum*, the synteny of COR16 (compared to *Cladonia grayi*) is not maintained due to the presence of an additional gene between the two flanking genes of COR16 (Fig. 1B). Therefore, synteny spanning multiple classes of fungi does not guarantee that it will be maintained in all species of these classes.

Some parts of the flanking genes of the three COR regions are very conserved among the investigated genomes of Lecanoromycetes, Eurotiomycetes and Lichinomycetes, and primers can be developed to amplify these regions across these classes, whereas other parts are extremely variable and impossible to align (Fig. 1, Supplementary Fig. S1). COR3 and COR16 were found in 23 and 19 of 24 eurotiomycete genomes surveyed, respectively (Fig. 2B). Variable regions can be used to develop primers targeting specific taxonomic groups (lower ranks) of fungi if their DNA was not isolated from axenic cultures. The list of references for all genomes used in this study can be found in Supplementary Table S1.

Our syntenic analysis of nine genomes, spanning Lecanoromycetes and Eurotiomycetes (Supplementary Table S1), revealed 125 syntenic collinear regions shared by all these genomes. Most of these regions (119) consist of two neighboring genes. Five regions included three collinear genes and one region consisted of four collinear genes. These 125 syntenic collinear regions range in size from one to more than ten kilobases (Fig. 2A). When screening 24 genomes from the Eurotiomycetes (Supplementary Table S1) we detected 26 of these 125 COR in every single genome (Fig. 2B), suggesting that they are universal (or at least nearly universal) across the Lecanoromycetes and Eurotiomycetes. These 26 COR include the largest syntenic collinear blocks revealed by this study (Fig. 2). Another set of 27 COR was present in all except one genome, and 26 of the 125 were present in 22 of 24 eurotiomycete genomes screened in this study (Fig. 2B).

#### 3.2. Among loci and datasets comparisons

All specimens of the scabrosoid clade dataset were represented by at least three loci, and most of them (89%) by at least four loci (Supplementary Fig. S2A). We were unsuccessful in sequencing COR1b for members of the scabrosoid clade. Within the dolichorhizoid clade, all 94 specimens were represented by four or more loci, and for almost half of them (49%) we could obtain sequences for all eight targeted markers. Overall, the ITS was sequenced for every individual sampled in this study and the nrLSU and *RPB1* loci were sequenced for 90% of all specimens (Table 2, Supplementary Fig. S2B).

After excluding ambiguously aligned regions, nrLSU had the lowest number of variable characters, whereas the overall level of genetic variation detected in COR1b, COR3 and COR16 was comparable to ITS (Supplementary Fig. S2C, and Table 2). Among protein-coding genes ( $\beta$ -tubulin, *EFT2.1*, and *RPB1*), *RPB1* was the most conserved marker.

Both *RPB1* and nrLSU have high informativeness (sensu Townsend, 2007) for resolving phylogenetic relationships at higher taxonomic levels (e.g., Hofstetter et al., 2007; Crespo et al., 2007, 2010; Schoch et al., 2009; Wedin et al., 2009), but were too conserved across our datasets to contribute substantially to resolving species complexes in section *Polydactylon* (Supplementary Fig. S2C, and Table 2).

The three new intergenic markers (COR1b, 3 and 16) contributed the highest number of variable characters to the phylogenetic analyses at the section level, outperforming ITS, the DNA barcode marker for fungi (Schoch et al., 2012). The ITS region contained many indels in its two spacers and, therefore, large portions of the alignment were excluded (Table 2).

#### 3.3. Contribution of the new COR loci to phylogenetic inference and species delimitation, compared to conventional markers

The great majority of the topological bipartitions that resulted from the analyses of each new COR marker were highly congruent among each other and with the other loci, which have been commonly used in molecular systematic studies of lichen-forming fungi at the intragenetic and species levels (e.g., Crespo et al., 2007; Leavitt et al., 2011b; Schmult et al., 2011; Miadlikowska et al., 2014). The new COR markers provided higher or equal (similar to ITS) levels of phylogenetic resolution compared to traditionally used loci (Table 2), especially within the clades representing the most recent radiations, e.g., the *neopolydactyla/dolichorhiza/hymenina* group.

When comparing phylogenetic support (i.e., number of internodes receiving ML bootstrap values  $\geq 70\%$ ) for interspecific relationships, nrLSU provided the least number of resolved (35%) and supported (25%) nodes (Table 2). The contributions of *EFT2.1* and *RPB1* were slightly higher, with 36% and 28% of nodes receiving high support, respectively (Table 2). Both, the ITS and  $\beta$ -tubulin were the best widely-used markers, with 37% and 45% well-supported nodes, respectively. However, the three new markers, COR1b, COR3 and COR16, outperformed all other loci in terms of support, with 60%, 59% and 53% of well-supported nodes, respectively. The phylogeny based on *EFT2.1* was in total agreement with the combined tree, whereas the remaining genes showed between one (COR3) to five conflicting relationships (*RPB1*, COR1b and COR16). However, these conflicts often involved very short branches.

#### 3.4. Single-locus phylogenies

The single-locus topologies were congruent across section *Polydactylon*, except for a few conflicting relationships within each of the three main clades (phylogenies not shown). The scabrosoid clade was strongly supported as monophyletic in all single-locus phylogenies, whereas the monophyly of the polydactyloid clade received high bootstrap support based on the three new intergenic markers,  $\beta$ -tubulin and *RPB1*, and weak support from *EFT2.1*. The ribosomal loci (ITS and LSU) resolved this clade as paraphyletic, however, without strong support (BS < 70). The dolichorhizoid clade including the *scabrosella* group was highly supported as monophyletic by *RPB1*, LSU, COR1b and COR3. For  $\beta$ -tubulin and COR16, the phylogenetic placement of the *scabrosella* group within section *Polydactylon* was not resolved, whereas *EFT2.1* placed it in the polydactyloid clade, but without strong support.

Most putative species within the scabrosoid clade are highly supported. Two occurrences of conflicting relationships include a close relationship of *P. scabrosa* 1 with *P. neopolydactyla* 4 + *P. scabrosa* 2, which was recovered with high support by *RPB1*, whereas LSU and COR3 strongly supported the affiliation of *P. scabrosa* 1 with *P. scabrosa* 4 (explaining the poor support the latter relationship received when the loci were concatenated; see Figs. 3 and 4A). Another conflict involved *P. scabrosa* 3, which resulted from the first split in the *scabrosa* group based on the ITS and COR3 phylogenies, while *EFT2.1* resolved *P. neopolydactyla* 5 as the earliest divergence event in this clade (Figs. 3 and 4A).



**Table 2**

Comparison among loci at the section and clade levels including the number of taxa sequenced for each locus, total number of characters, number of unambiguously aligned characters, total number of variable characters, number of variable characters included in the phylogenetic analyses, and their contribution to phylogenetic resolution and support. Results for the polydactyloid clade are shown only for presenting a global picture for section *Polydactylon* and to allow a comparison with the two clades targeted for this study – the dolichorhizoid and scabrosoid clades.

Section/clade	Locus	No. of specimens	% of specimens	Total no. of characters	No. of characters included/proportion <sup>a</sup>	Total No. of variable characters/proportion <sup>b</sup>	No. of variable characters included/proportion <sup>c</sup>	Proportion of resolved/supported nodes <sup>d</sup>
Section <i>Polydactylon</i>	$\beta$ -tubulin	139	85	760	578/0.76	230/0.30	153/0.26/0.67	0.62/0.45
	<i>EFT2.1</i>	120	73	820	800/0.98	149/0.18	141/0.16/0.95	0.54/0.36
	<i>RPB1</i>	153	93	761	678/0.89	114/0.15	95/0.14/0.83	0.48/0.28
	LSU	154	94	1426	1255/0.88	210/0.15	89/0.07/0.42	0.35/0.25
	ITS	164	100	750	521/0.69	318/0.42	258/0.50/0.81	0.57/0.37
	COR1b	75	46	1065	848/0.80	452/0.42	260/0.31/0.58	0.75/0.60
	COR3	116	71	850	815/0.96	351/0.41	316/0.39/0.90	0.76/0.59
	COR16	111	68	1130	1058/0.94	426/0.38	352/0.33/0.83	0.61/0.53
	Combined	164	100	NA	6553/NA	NA	1575/0.24/NA	NA
Polydactyloid clade	$\beta$ -tubulin	24	69	760	605/0.80	54/0.07	53/0.09/0.98	0.89/0.56
	<i>EFT2.1</i>	19	54	820	818/0.99	47/0.06	47/0.06/1.00	0.75/0.33
	<i>RPB1</i>	32	91	761	678/0.89	39/0.05	39/0.06/1.00	0.71/0.29
	LSU	33	94	1426	1307/0.92	32/0.02	32/0.02/1.00	0.25/0.25
	ITS	35	00	750	640/0.85	109/0.15	91/0.14/0.83	0.62/0.46
	COR1b	8	23	1065	788/0.74	142/0.13	93/0.12/0.65	1.00/1.00
	COR3	18	51	850	714/0.84	112/0.13	101/0.14/0.90	1.00/0.89
	COR16	17	49	1130	885/0.78	116/0.10	106/0.12/0.91	0.57/0.57
	Combined	35	100	NA	6435/NA	NA	562/0.09/NA	NA
Dolichorhizoid clade	$\beta$ -tubulin	91	97	760	575/0.76	140/0.18	96/0.17/0.69	0.55/0.39
	<i>EFT2.1</i>	73	78	820	817/0.99	64/0.08	60/0.07/0.94	0.44/0.27
	<i>RPB1</i>	89	95	761	678/0.89	50/0.07	44/0.06/0.88	0.33/0.18
	LSU	88	94	1426	1309/0.92	45/0.03	43/0.03/0.96	0.23/0.15
	ITS	94	100	750	620/0.83	171/0.23	111/0.18/0.65	0.45/0.22
	COR1b	67	71	1065	793/0.74	234/0.22	175/0.22/0.75	0.70/0.53
	COR3	75	80	850	771/0.91	157/0.18	152/0.20/0.97	0.66/0.48
	COR16	73	78	1130	946/0.84	195/0.17	184/0.19/0.94	0.55/0.43
	Combined	94	100	NA	6509/NA	NA	865/0.13/NA	NA
Scabrosoid clade	$\beta$ -tubulin	24	69	760	575/0.76	36/0.05	35/0.06/0.97	0.73/0.64
	<i>EFT2.1</i>	28	80	820	817/0.99	46/0.06	46/0.06/1.00	0.67/0.60
	<i>RPB1</i>	32	91	761	678/0.89	28/0.04	28/0.04/1.00	0.77/0.54
	LSU	33	94	1426	1313/0.92	47/0.03	37/0.03/0.78	0.75/0.56
	ITS	35	100	750	653/0.87	162/0.22	90/0.14/0.56	0.88/0.69
	COR1b	0	0	NA	NA	NA	NA	NA
	COR3	23	66	850	800/0.94	99/0.12	93/0.12/0.94	0.92/0.75
	COR16	21	60	1130	995/0.88	100/0.09	95/0.10/0.95	0.90/0.90
	Combined	35	100	NA	5831	NA	424/0.07/NA	NA

NA: not applicable.

<sup>a</sup> Proportion of characters included versus the total number of characters (after the slash).

<sup>b</sup> Proportion of the total number of variable characters versus the total number of characters (after the slash).

<sup>c</sup> Proportion of variable characters included in the analyses versus the total number of characters included (after the first slash) versus the total number of variable characters (after the second slash).

<sup>d</sup> Number of nodes that are resolved and supported in accordance with the concatenated topology, based on the ML analyses on the whole section.

Phylogenetic relationships among putative species in the dolichorhizoid clade received generally low support in the single locus phylogenies. The most robust phylogenies for this clade were obtained with COR3, COR16 and  $\beta$ -tubulin. The topologies derived for the three COR markers were congruent with the topologies from the commonly used markers.

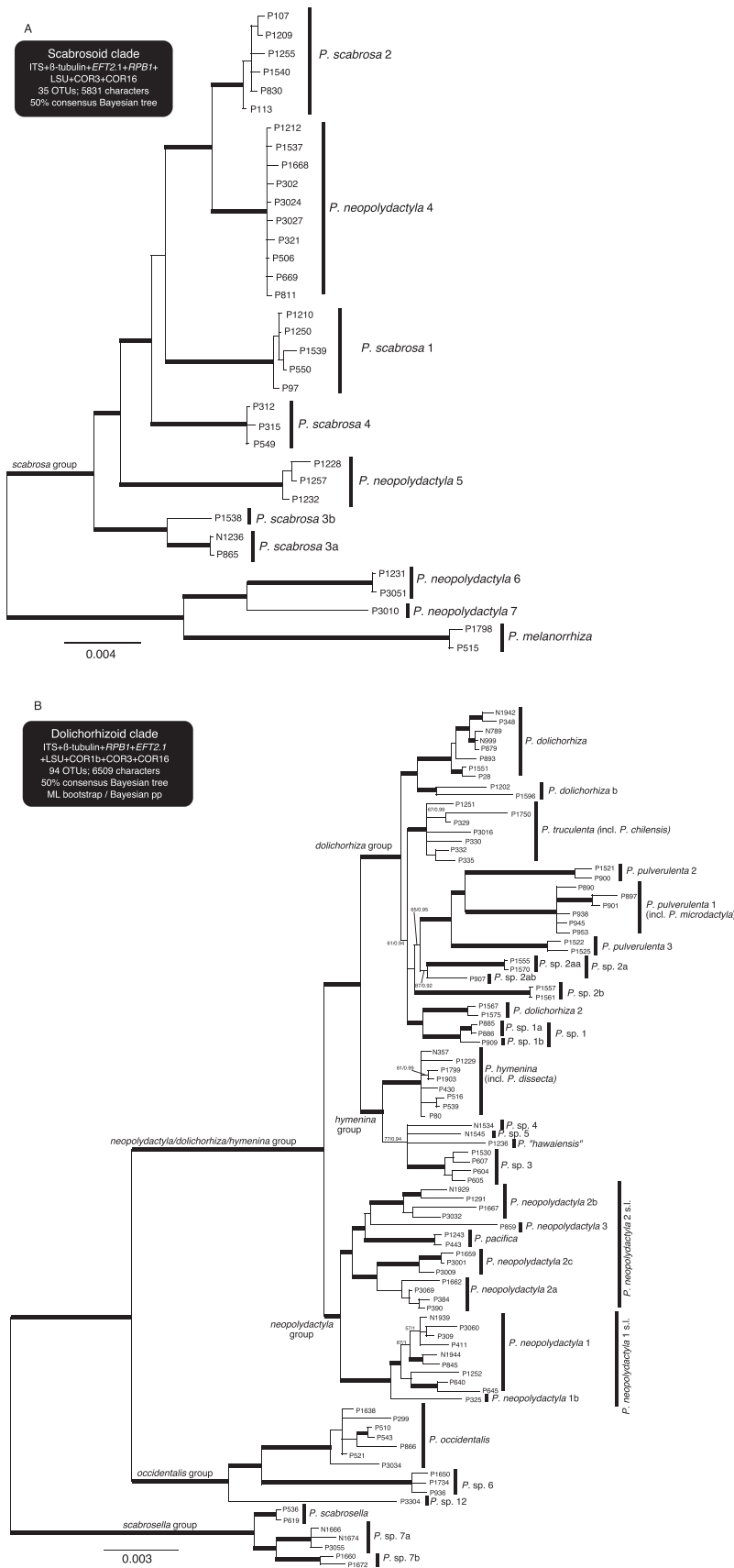
### 3.5. Multi-locus phylogenies for section *Polydactylon*

Compared to Magain et al. (2017), adding the three new COR loci and reducing the amount of missing sequences for the other loci, significantly improved the phylogenetic resolution and support, especially in the dolichorhizoid clade (Figs. 3 and 4). Most relationships among and within species complexes, including the monophyly of most putative species, were well-supported in multi-locus trees. The highest degree of phylogenetic uncertainty was restricted to the largest and most recent diversification, i.e., the *dolichorhiza* group, which is found mostly in Central and South America (Fig. 4B). We confirmed that *P. dissecta* is nested within the *P. hymenina* clade, *P. microdactyla* is intermixed with *P. pulverulenta* 1, and *P. chilensis* with *P. truculenta* (Fig. 4B). In the

scabrosoid clade, most putative species were resolved as monophyletic on long internodes supported by high bootstrap values. Three individuals that were not included in Magain et al. (2017) likely represent new species, namely *P. neopolydactyla* 7 and *P. scabrosa* 3b in the scabrosoid clade, and *P. sp.* 12 in the dolichorhizoid clade (Figs. 3 and 4).

### 3.6. Species delimitations with *Structurama*

The gamma shape hyperprior had a strong influence on the number of delimited species (see also Magain et al., 2017). In the scabrosoid clade, the number of species remained stable with gamma shape values  $\leq 5$  (Supplementary Fig. S3A). Higher gamma values (e.g., 8) led to a substantial increase in the number of delimited species, which is likely artefactual, including several singletons (i.e., species represented by only one individual). We chose the gamma shape value of 3 for the final analysis of the scabrosoid clade. In the dolichorhizoid clade, the number of putative species increased almost linearly with an increase of the gamma shape hyperprior. However, the number of singleton species rose drastically when the gamma shape hyperprior reached 10



**Fig. 4.** Bayesian consensus (50%) phylogenies based on concatenated datasets for (A) the scabrosoid clade (7 loci) and (B) the dolichorhizoid clade (8 loci). Thick internodes represent high ML bootstrap support values (ML-bs  $\geq 70\%$ ) and Bayesian posterior probabilities (pp  $\geq 0.95$ ). ML-bs values and pp are provided for internodes highly supported by only one of the two analyses. Both trees were rooted according to the phylogenetic tree for the entire section (Fig. 3). Vertical thick bars delimit putative species based on monophyly, phenotypic data, and geographic distribution, and correspond to a priori species assignment used in spedeSTEM and BPP analyses.

(Supplementary Fig. S3B). When using a gamma shape hyperprior of 3, some potentially distinct lineages (i.e., well defined phylogenetically, phenotypically and geographically) were grouped together within single putative species (e.g., *Peltigera* sp. 4 [N1534] with *P. pacifica*; Supplementary Fig. S4) even if these taxa do not share any allele. We selected the intermediary gamma shape value of 15 for the final analysis of the dolichorhizoid clade with Structurama.

No major discrepancies in the number of delimited putative species (8–10 species) were observed for the scabrosoid clade when removing one locus at a time from the Structurama analysis with a gamma shape of 3 (Supplementary Fig. S3C). Removal of each of the following loci: COR3, ITS, *EFT2.1* and LSU from the concatenated dataset resulted in the delimitation of the same nine putative species, whereas the exclusion of COR16 or  $\beta$ -tubulin decreased the number of putative species to 8 (*P. scabrosa* 4 and *P. neopolydactyla* 5 were merged) while the removal of *RPB1* resulted in one additional species.

In the dolichorhizoid clade, the exclusion of either COR1b or COR16 decreased the number of species, and the removal of COR3 had no effect, whereas the removal of *EFT2.1*, *RPB1*, or especially LSU strongly increased the number of putative species (Supplementary Fig. S3F). In this clade, many individuals were represented by a unique allele for all three COR intergenic spacers, but shared the same LSU, *RPB1* and *EFT2.1* alleles (Supplementary Fig. S3E). Overall, excluding a highly variable locus decreased the number of putative species, whereas the removal of a more conserved locus led to more putative species.

With a gamma shape of 3, all species within the scabrosoid clade corresponded to monophyletic lineages as revealed by our multi-locus phylogenetic analysis (Fig. 4A), except *P. neopolydactyla* 4 and *P. scabrosa* 1 clustered into one putative species, perhaps due to identical *RPB1* allele shared among the populations of these two taxa. When analyzing the dataset without *RPB1*, *P. scabrosa* 1 and *P. neopolydactyla* 4 were delimited as two putative distinct species with no effect on the delimitation of the remaining species. For the final Structurama analyses of the scabrosoid clade, the *RPB1* locus was not included in the combined dataset. With these settings, Structurama delimited 10 species in the scabrosoid clade (Fig. 5).

In the dolichorhizoid clade, many putative species were consistently defined as monophyletic regardless of the gamma shape of the hyperprior (Supplementary Fig. S4). Because Structurama is not a tree-based method, specimens from different clades, but sharing an allele, are grouped together, leading to broadly delimited polyphyletic species (shown in grey in Supplementary Fig. S4). This assembly of lineages representing one putative species was not corroborated by phylogenetic, phenotypic or geographical data. Some of the singleton species retrieved with Structurama were in agreement with phylogenetic, phenotypic and geographical evidence (e.g., *P. neopolydactyla* 1b [P325], *P. neopolydactyla* 3 [P859], *P. "hawaiensis"* [P1236], *P. sp. 4* [N1534], *P. sp. 5* [N1545], and *P. sp. 12* [P3304]; Fig. 4B, Supplementary Table S2), whereas other singletons or subgroups (e.g., P1291, N1929, P3032, P1662 [part of *P. neopolydactyla* 2 s.l.; Fig. 4B], P640 and P645 [part of *P. neopolydactyla* 1 s.l.; Fig. 4B]) likely represent intraspecific allelic variation among populations of the same species. With the final settings that we selected, Structurama delimited 26 species in the dolichorhizoid clade (Fig. 6).

### 3.7. Species delimitations with bGMYC and bPTP

Overall, in the scabrosoid clade, all bGMYC single-locus results were in agreement with the species delimitation revealed by the final Structurama analysis, with the exception of *P. scabrosa* 3a and *P. scabrosa* 3b, which were assigned to a single putative species based on ITS (Supplementary Fig. S5). Similar results were obtained from the bPTP analysis on the ML tree (RAxML phylogeny based on the concatenated 7-locus dataset; Fig. 4A) except for *P. scabrosa* 3a and *P. neopolydactyla* 5, which were both split into two putative species with low posterior probabilities (pp of 0.48–0.58).

To assess species delimitation in the dolichorhizoid clade, we completed bGMYC analyses on the five most variable loci (ITS,  $\beta$ -tubulin, COR1b, COR3 and COR16 [Supplementary Fig. S6A–E]), and bPTP analysis on the concatenated 8-locus dataset (Supplementary Fig. S6F). Many putative species delimited by Structurama were recovered by each analyzed locus, although the support varied depending on the dataset and analyses. For example, for *P. neopolydactyla* 3, posterior probabilities ranged from 0.34 ( $\beta$ -tubulin; Supplementary Fig. S6B) to 0.85 (ITS; Supplementary Fig. S6A) with bGMYC, whereas the posterior probability was equal to 1 with bPTP (Supplementary Fig. S6F). Inconsistencies between bGMYC and bPTP involved for example *P. occidentalis*, which was delimited as a species by all bGMYC analyses (pp from 0.30 to 0.51; Supplementary Fig. S6A–E) but not by bPTP (Supplementary Fig. S6F). Several other putative species were recovered by all but one locus (e.g., *P. pulverulenta* 1, *P. pulverulenta* 3, *P. pacifica*, *P. dolichorhiza*, and *P. neopolydactyla* 3 were not delimited by one of the COR loci), or their delimitation varied depending on the dataset (e.g., *P. scabrosella*, *P. sp. 7a* and *P. sp. 7b* were merged as one species by four loci, but  $\beta$ -tubulin and bPTP retrieved them as three distinct species, as did Structurama; Supplementary Figs. S4 and S6).

High level of uncertainty in species delimitation occurred within two putative species complexes, *P. neopolydactyla* 1 s.l. and *P. neopolydactyla* 2 s.l. (Supplementary Fig. S6). The delimitation of *P. truculenta* was also incongruent: it appeared as a putative species based on the ITS and COR1b (Supplementary Fig. S6A, C), but belonged to an unresolved lineage according to  $\beta$ -tubulin and COR3 (Supplementary Fig. S6B, D), and was divided into multiple singletons based on COR16 (Supplementary Fig. S6E) and multilocus-based bPTP analysis (Supplementary Fig. S6F).

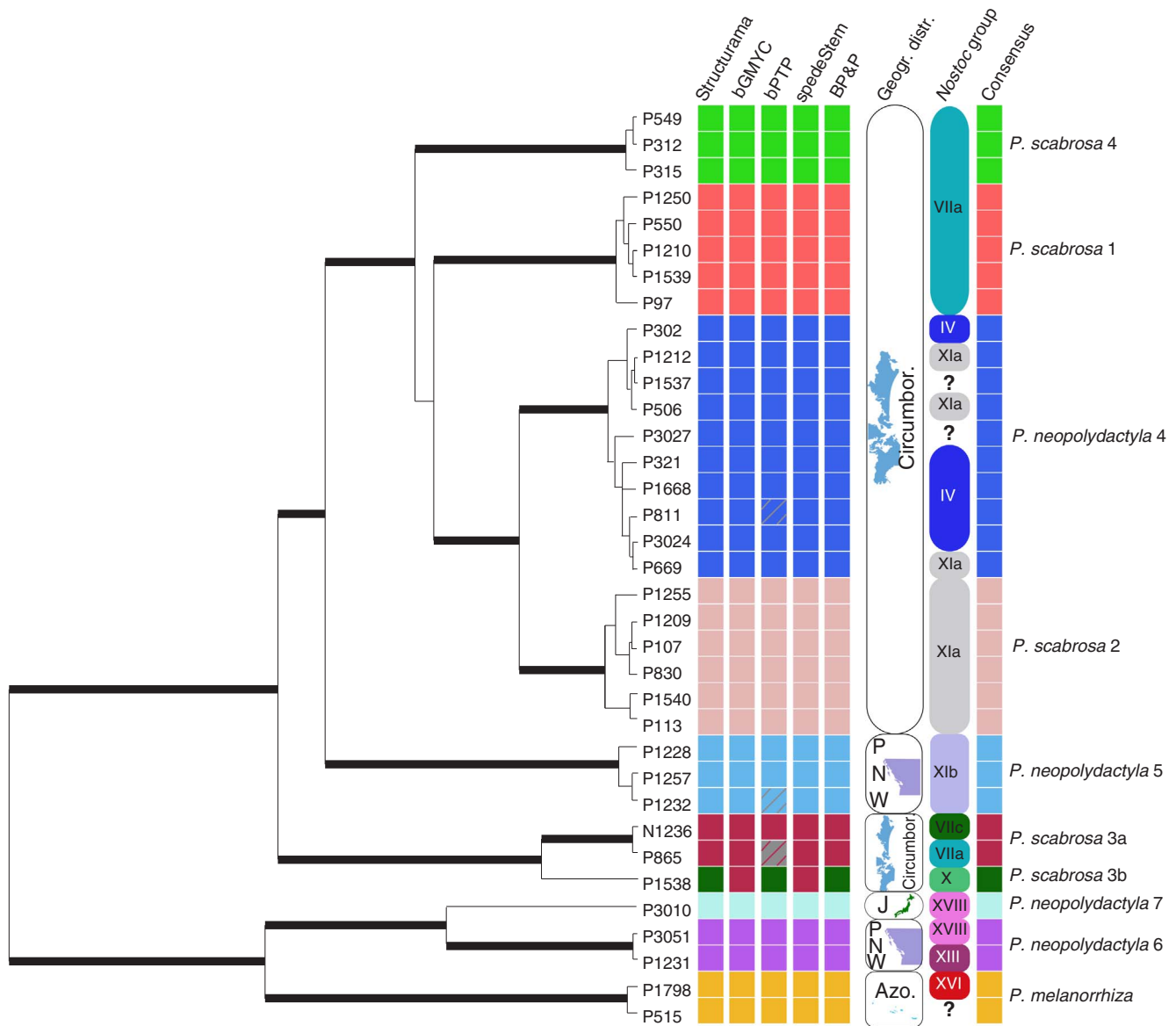
### 3.8. Species validations with spedeSTEM

When comparing species assignments in the scabrosoid clade by dividing *P. neopolydactyla* 4 randomly into two sets of five samples each (see Section 2), the two random sets were merged in one expected putative species for theta values  $\geq 0.020$  when *RPB1* was excluded, and  $\geq 0.035$  based on the complete multi-locus dataset. All remaining species were supported as distinct lineages in accordance to species assignment shown in Fig. 3, with only one uncertainty whether *P. scabrosa* 3a and 3b represent one or two putative species. For the final analysis with spedeSTEM, we selected the results from the analysis without *RPB1* using a theta value of 0.02, which resulted in the delimitation of 9 species (Fig. 5).

In the dolichorhizoid clade, the number of species supported by spedeSTEM varied considerably depending on the theta value used; from four species with theta values of 0.00001–0.000001, 12 species with theta values of 0.5, and 17–26 species for theta values of 0.0001–0.1. Theta values of 0.001, 0.005, 0.01, 0.02 provided congruent species delimitations, resulting in 21 species in the dolichorhizoid clade. The two parts of *P. hymenina* were delimited as a single species, for all tested theta values above 0.001 (see Section 2). We chose a theta value of 0.02 for our final species delimitation, which resulted in 18 species (Fig. 6). The only results that disagreed with most other methods were the merge of *P. dolichorhiza* with *P. truculenta* and *P. sp. 3–5*, respectively.

### 3.9. Species validation with BPP

When performing BPP analyses on the scabrosoid clade with an *a priori* species assignment where *P. neopolydactyla* 4 and *P. scabrosa* 2 were randomly split into two groups (to test the priors, see Section 2), the two species were not split into two in all tests. In all analyses, regardless of the mean of the theta prior, all lineages as delimited on Fig. 4A were significantly supported as putative species (Fig. 5), including *P. scabrosa* 3a and *P. scabrosa* 3b (pp = 0.97 for these two species, all other lineages pp = 1). In the dolichorhizoid clade, the two



**Fig. 5.** Comparison among five species assignment methods with geographical ranges, cyanobiont phylogroups and consensus species delimitations for the scabrosoid clade. This chronogram was part of the chronogram resulting from the BEAST analysis on the concatenated datasets (164 taxa, 8 loci, 6553 characters) of section *Polydactylon*. Thick branches indicate significant support ( $pp \geq 0.95$ ). Panels to the right of the chronogram (from left to right) are: species discovery by Structurama (with the gamma shape of 3); species discovery by bGMYC based on the ITS locus only; species discovery by bPTP on the concatenated dataset; species validation by spedeSTEM; species validation by BPP; geographic distribution of sampled specimens; associated *Nostoc* phylogroups as defined by Magain et al. (2017) with missing data represented by question marks; and the consensus species delimitation corresponding to the final species delimitation resulting from this study. Striped boxes indicate equally valid alternative species delimitations. Abbreviations for geographic zones are: Azo. = Azores, Circumbor. = Circumboreal, J = Japan, and PNW = Pacific Northwest.

subsets of *P. hymenina* were detected as distinct species (intermixed in the phylogeny; Figs. 3 and 4B, see Section 2) when the mean of the theta prior was  $\leq 0.001$ . Higher theta prior (e.g., 0.01–0.02) resulted in the placement of *P. hymenina* within one putative species. The remaining species were well-supported and their delimitations corroborated their monophyletic circumscription in the multi-locus phylogenies (Figs. 3 and 4). When the mean of the theta prior was 0.05, *P. scabrosella*, *P. sp. 7a* and *P. sp. 7b* were merged, while the two parts of *P. hymenina* were considered as distinct species. For our final BPP analysis we used the mean of the theta prior of 0.02, which resulted in the delimitation of 29 species (Fig. 6).

### 3.10. *rbclX* phylogeny

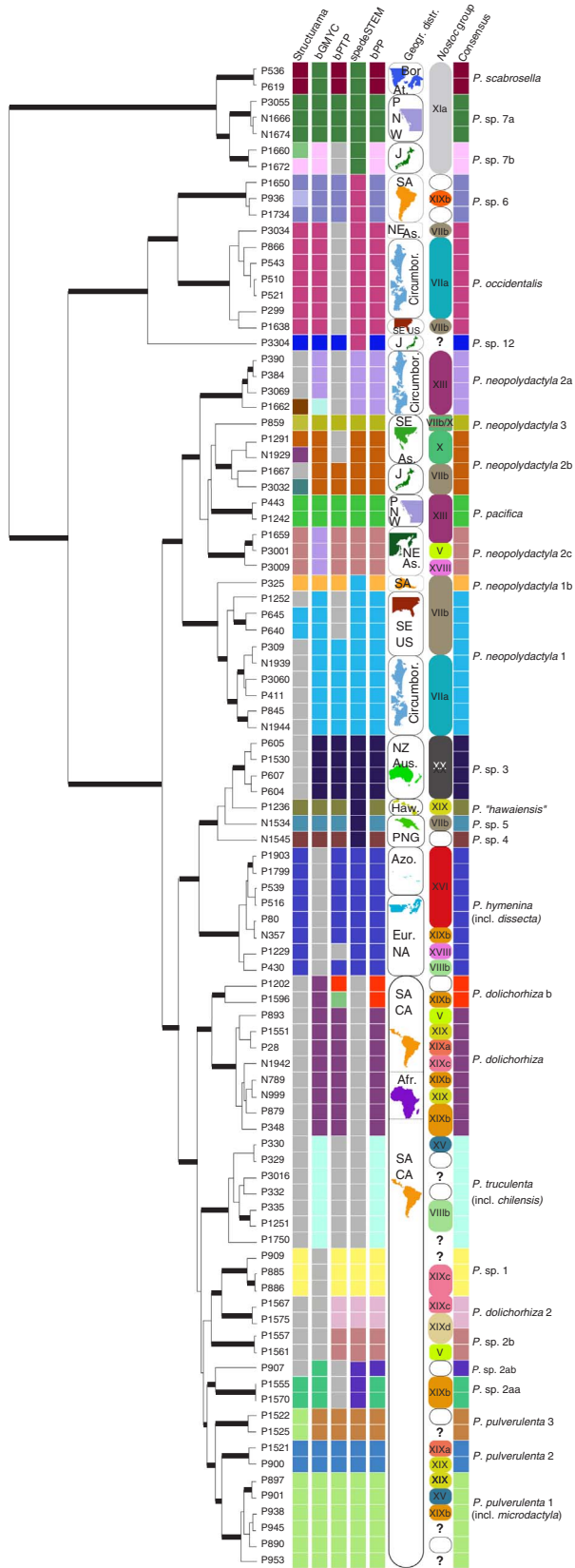
Based on an ML phylogenetic analysis of *rbclX* sequences (result not shown), the 20 *rbclX* sequences generated for this study represented

*Nostoc* phylogroups already recognized by Magain et al. (2017), except for P1202 which represented a unique haplotype. *Nostoc* phylogroup affiliations of all available *rbclX* sequences are indicated in Figs. 5 and 6.

### 3.11. Gap in the distribution of sequence similarities

We examined genetic variation in the ITS region, the official DNA barcoding marker for fungi (Schoch et al., 2012), within and among putative species, in order to better understand species boundaries within section *Polydactylon* (Fig. 7). For species in the scabrosoid clade, the intraspecific variation was noticeably lower (below 0.004) than the interspecific variation (above 0.006) (Figs. 7A and 8A), indicating that a DNA barcoding gap approach (i.e., greater genetic distance among than within species) using ITS could be implemented in this clade. No differences in the ITS locus were detected between *P. scabrosa* 3a and *P.*





**Fig. 6.** Comparison among five species assignment methods with geographical ranges, cyanobiont phylogroups and consensus species delimitations for the dolichorhizoid clade. This chronogram was part of the chronogram resulting from the BEAST analysis on the concatenated dataset (164 taxa, 8 loci, 6553 characters) of section *Polydactylon*. Thick branches indicate significant support ( $pp \geq 0.95$ ). Panels to the right of the chronogram (from left to right) are: species discovery by Structurama (with the gamma shape of 15); species discovery by bGMYC based on the ITS locus; species discovery by bPTP on the concatenated dataset; species validation by spedeSTEM; species validation by BPP; geographic distribution of sampled specimens; associated *Nostoc* phylogroups as defined in Magain et al. (2017) with missing data represented by question marks and white circles representing unique haplotypes outside of the color-coded *Nostoc* phylogroups; and the consensus species delimitation corresponding to the final species delimitation resulting from this study. Grey boxes indicate where species delimitation methods performed poorly, e.g., large polyphyletic species in Structurama, multiple putative species involving only one specimen with bGMYC and bPTP, and specimens failing to be assigned to a species following our criterion (see caption for Supplementary Fig. S6) by bGMYC. Abbreviations for the geographic zones: Afr. = Africa, As. = Asia, Aus. = Australia, Azo. = Azores, Bor At. = Boreal Atlantic, CA = Central America, Circumbor. = Circumboreal, Eur. = Europe, Haw. = Hawaii, J = Japan, NA = North America, NE = North-East, NZ = New Zealand, PNG = Papua New Guinea, PNW = Pacific Northwest, SA = South America, SE = South-East, and US = United States.

*scabrosa* 3b, both of which were supported as two distinct species by most species delimitation methods (Fig. 5).

A DNA barcoding gap approach was not suitable across the dolichorhizoid clade (Fig. 7B and 8B). Although most intraspecific and interspecific differences are below 0.002 and above 0.005 respectively, they overlap largely in the range between 0.002 and 0.005 (Fig. 8B). Some species clearly show a DNA barcoding gap (as defined above), e.g., *P. occidentalis*, *P. pulverulenta* 1, *P. sp. 6*, *P. sp. 12*, and *P. hymerina*, while several other species were not well-defined by the similarity of their ITS sequences, especially members of the *neopolydactyla*/*dolichorhiza*/*hymerina* group (e.g., *P. triculenta*, *P. neopolydactyla* 1; Fig. 7B and 8F). We detected a single case where two different putative species *P. neopolydactyla* 2a and *P. neopolydactyla* 2c were represented by identical ITS, but methods disagree as to whether they represent one or two species (Fig. 6).

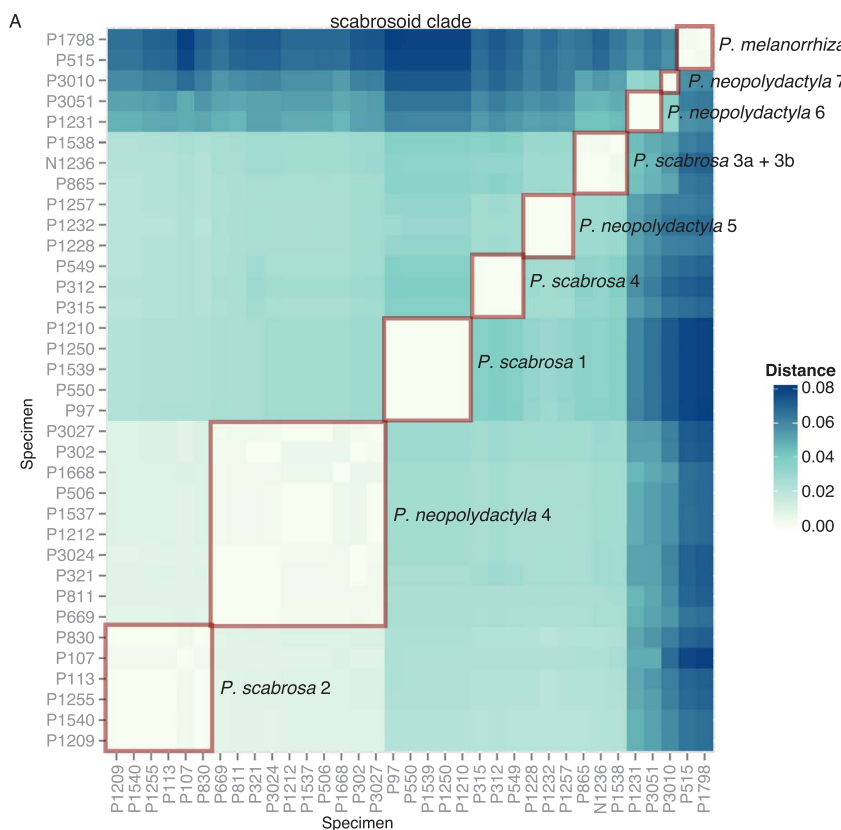
We further compared the intraspecific and interspecific distances in the dolichorhizoid clade for the highly variable COR1b, COR3 and COR16 markers in search for a gap in their similarity distributions. The patterns detected for these three markers are highly similar to ITS (Fig. 8C-E), showing an overlap for a small fraction of the interspecific and intraspecific genetic distances.

## 4. Discussion

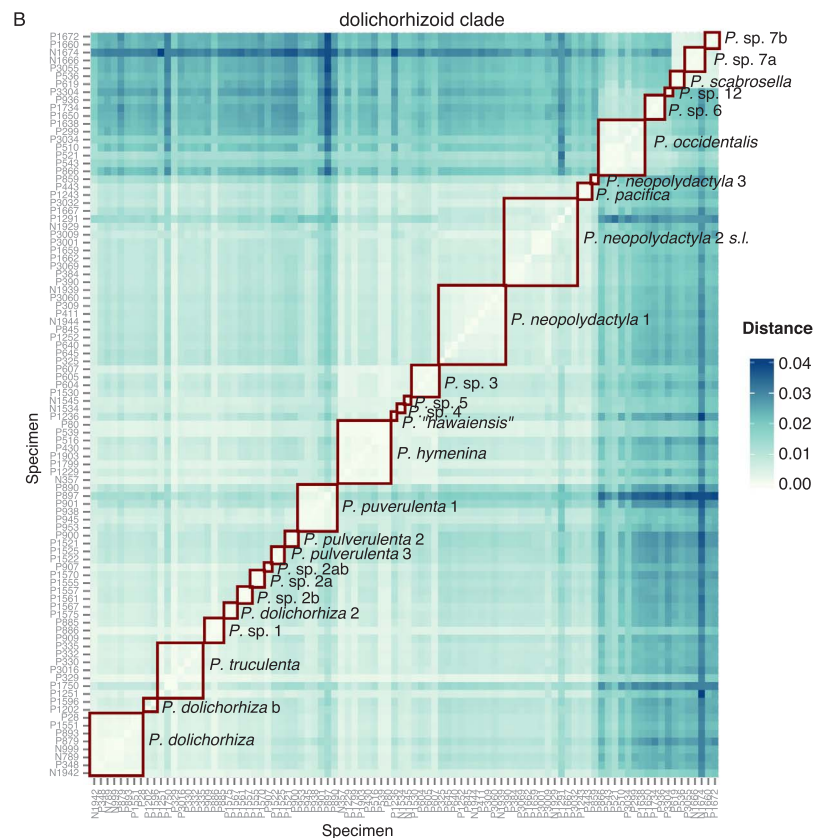
### 4.1. Contribution of COR markers to phylogenetic and population genetic studies at the interface of species and populations

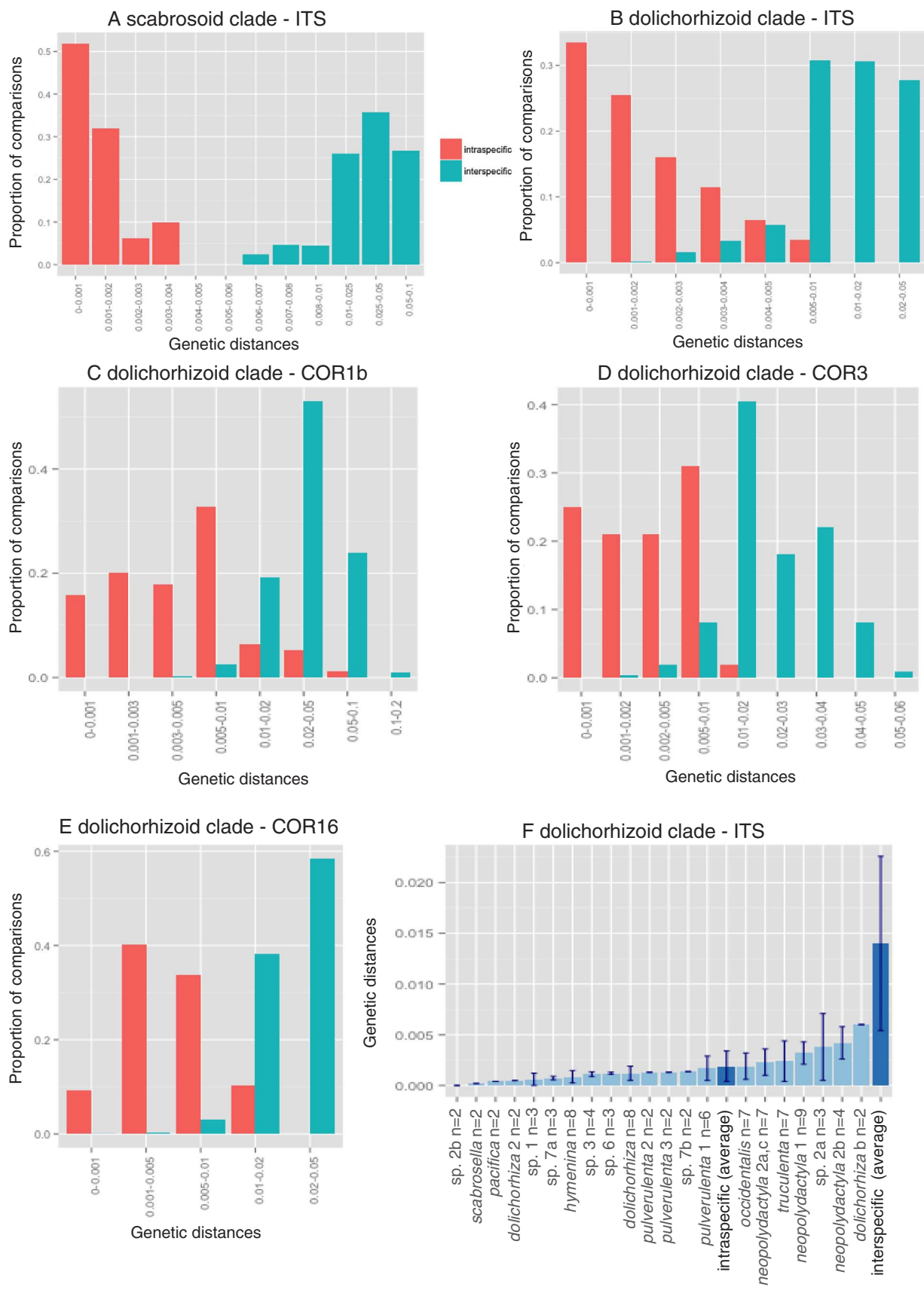
The three intergenic COR regions proved to be very useful to delimit *Peltigera* species and resolve phylogenetic relationships within section *Polydactylon*. The conservation of these markers across a wide range of fungal groups make them suitable for future studies across entire genera, families, or even orders. COR3 and COR16 could potentially delimit species and resolve evolutionary relationships in most species complexes of the classes Lecanoromycetes, Lichinomycetes and Eurotiomycetes. The 125 COR regions that we detected across Lecanoromycetes and Eurotiomycetes genomes, including 26 that are potentially universal for these two classes (Fig. 2), demonstrate the potential of our approach to identify new markers for the study of fungi.

By using the same set of genetic markers among the Lecanoromycetes and Eurotiomycetes, this COR-based approach could highly contribute to resolving species complexes and conducting powerful population genetic analyses, but also detecting global patterns of evolution, speciation and biogeography in these two classes of fungi. The sequences provided in the FASTA file (supplementary\_data\_S1.fas) can be used to screen for these 125 syntenic collinear regions, when using low coverage genome sequencing, or



**Fig. 7.** Heatmaps showing pairwise genetic distances among ITS sequences in the scabrosoid (A) and dolichorhizoid (B) clades. Red squares delimit ranges of genetic variation within putative species.





**Fig. 8.** Comparison of intraspecific (in red) and interspecific (in blue) pairwise genetic distances in the scabrosoid clade for ITS (A) and in the dolichorhizoid clade for ITS (B), COR1b (C), COR3 (D), COR16 (E). Comparison of ITS genetic distances for each species (light blue) and the average pairwise intraspecific and interspecific distances (dark blue) (F).

ddRadSeq, in members of the Lecanoromycetes and Eurotiomycetes. This would enhance parameter estimations and comparative studies among various genera across these two classes of fungi. Therefore, we recommend to use the names provided for each COR as shown in Fig. 2B and included in the [supplementary\\_data\\_S1.fas](#) of the Supplementary Material online. This approach could also be extended to other classes of fungi by conducting large-scale multi-taxon syntenic analyses to find sets of syntenic collinear regions that are specific to other major taxonomic groups of fungi.

#### 4.2. Comparison of species discovery and validation methods

Overall, all five methods were highly congruent for the scabrosoid clade (Fig. 5) where every delimited species was monophyletic and well resolved based on ITS alone (Figs. 7 and 8). The only incongruences involved the misplacement of a few specimens with bPTP and the recognition of the *P. scabrosa* 3 clade as one species with bGMYC and spedeSTEM, versus two species, i.e., *P. scabrosa* 3a and *P. scabrosa* 3b, with other methods (Fig. 5). Low taxon sampling for *P. scabrosa* 3 may have contributed to these ambiguities.

In the dolichorhizoid clade, the methods used disagreed for many species assignments (Fig. 6). Structurama gave noticeably different results from other methods, potentially due to insufficient number of representatives sampled from some of the putative species. Loci with lower numbers of alleles (less genetic variation; e.g., *EFT2.1*, *RPB1* and *LSU*) tended to diminish the number of putative species, while loci with a greater number of alleles (high genetic variation; e.g., COR markers and ITS) tended to increase the number of species. Highly variable loci (such as COR markers) could therefore overestimate the number of species. Yet, their high intra- and inter-specific variability is valuable to infer phylogenetic relationships within section *Polydactylon* (Table 2, Figs. 3 and 4). Consequently, when Structurama and tree-based methods are congruent, these putative species are more likely to represent meaningful evolutionary entities. For example, Leavitt et al. (2011a) used Structure as a species discovery tool, and BPP to validate species assignments. They concluded that the lichen-forming genus *Xanthoparmelia* included up to 21 species in North America, many of which were previously unrecognized, and their circumscriptions were not corroborated by the widely applied phenotypic characters. In the study of the genus *Letharia* by Altermann et al. (2014), Structure retrieved six monophyletic species in concordance with their phylogenetic study.

bGMYC provided congruent results for most species delimited, but noticeable differences were observed depending on the locus used. These discrepancies were partly due to low levels of variation of some loci, resulting in low and often erroneous phylogenetic resolution when using single loci, especially for recent diversification events, i.e. of the *dolichorhiza* group (Figs. 3 and 4B; Magain et al., 2017). The exclusion of the most variable sites that cannot be aligned reliably may also have contributed to this lack of resolution. For example, most of the characters segregating *P. hymenina*, *P. sp. 1* and *P. dolichorhiza 2* were excluded from ITS sequences due to alignment ambiguity. Because of this, bGMYC grouped these three putative species together (Supplementary Fig. S6A), whereas they were kept distinct in the bGMYC analyses based on other loci (Supplementary Fig. S6B–E). Such discrepancies may also reflect ongoing speciation events in this morphologically diverse group where certain taxa like *P. truculenta* are represented by a relatively high number of haplotypes. Delimiting species based on the GMYC model and a single locus is potentially misleading, considering that none of the analyzed loci provided the same number of putative species in this study (Supplementary Fig. S6). However, both methods have the potential to perform well if implemented on a robust phylogeny (i.e., inferred using phylogenetically informative markers) that is congruent with the species tree. Alternatively, applying bGMYC to several single-locus datasets and establishing a consensus across loci seems to be a reliable approach, especially when considering only the well-supported portions of single-locus phylogenies.

Parnmen et al. (2012) applied a GMYC approach based on five loci to delimit species in the *Cladia aggregata* group and found twelve putative species compared to the initial number of one (to eight) previously recognized species. These species were partially corroborated by geographical and phenotypic data (Parnmen et al., 2012). Similarly, Lücking et al. (2014) delimited 126 species based on the GMYC model and ITS sequences, for specimens from 16 previously described species within the two basidiolichen genera *Cora* and *Corella*.

By using multilocus phylograms, bPTP take advantage of the usually higher phylogenetic resolution associated with concatenated datasets. However, in our study, this method tended to over-split clades (Figs. 5 and 6, Supplementary Fig. S6F) and often considered that individuals on slightly longer branches inside a larger monophyletic putative species represented distinct species.

SpedeSTEM relies on supplied single-locus trees, and consequently, is influenced by the uncertainty associated with single-locus topologies. BPP requires alignments, and therefore it is less biased by the poor signal resulting from single-locus phylogenies, however, a guide-tree with predetermined relationships among individuals must be provided, and only the merging of predefined clades (in the guide tree) are being tested. In this study, spedeSTEM tended to delimit broader putative species compared to BPP (Fig. 6). These two methods require estimates of theta, which can be problematic when neither the estimation of the population sizes nor accurate estimates of nucleotide mutation rates are available, as it is the case in lichen-forming fungi. Sensitivity analyses across a broad range of theta values allowed us to evaluate the performance of both methods. Overall, BPP delimited putative species that corroborated better with species circumscriptions based on phenotypic and geographic distributions. BPP also seems to be less sensitive to the initial set of parameters. One advantage of spedeSTEM and BPP over other methods is that they take into consideration discrepancies in phylogenetic relationships among different loci, whereas bPTP and bGMYC delimit species based on a single tree at a time, even if bGMYC calculates probabilities by repeating the delimitation on multiple trees. Most studies on species delimitations inferred with discovery or validation methods in lichen-forming fungi (e.g., Leavitt et al., 2011a; Parnmen et al., 2012; Lücking et al., 2014; Schneider et al., 2016) have revealed high level of cryptic biodiversity. However, the majority of these studies made inferences using only a single method and/or a single locus.

All methods rely on different models and assumptions, and only test one or few properties of the lineages. Moreover, different loci can have different evolutionary histories. Therefore, different methods will often give different results, and we strongly recommend the use of several methods, as well as several loci, as previously suggested by Carstens et al. (2013), in order to reach a consensus, rather than building new taxonomies relying on a single result. We argue that species delimitation should not rely on DNA- and statistics-based methods alone, but that different lines of evidence, such as species morphology, distribution and patterns of interactions with other organisms, are integrated to characterize and consolidate species delimitations. We demonstrated such an approach in the present study, as illustrated in Figs. 5 and 6.

#### 4.3. Pattern of cyanobiont (*Nostoc*) associations in support of *Peltigera* species delimitations

Magain et al. (2017) reported that many *Peltigera* species from section *Polydactylon* show strong specificity towards their *Nostoc* symbiotic partners (cyanobionts). Most species associated with only one or two *Nostoc* phylogroups across their geographical distribution. Therefore, *Nostoc* identity can be a useful supplementary character for circumscribing species, even if traditional “lichen species” delimitations were based exclusively on traits of the fungal partner. For example, in this study, most methods performed poorly in delimiting species within *P. neopolydactyla* 2 s.l. (Fig. 6). However, according to Magain et al. (2017), *P. neopolydactyla* 2a was always found associated with *Nostoc*



phylogroup XIII, whereas *P. neopolydactyla* 2b associates with *Nostoc* phylogroup X or VIIb (Fig. 6). Differences in cyanobiont partnership between these two sister clades may suggest that mycobionts of *P. neopolydactyla* 2 s.l. belong to at least two distinct putative species. The identity of the *Nostoc* cyanobiont can therefore contribute to identify the species. For example, two pairs of morphologically similar and sympatric species, *P. neopolydactyla* 1/*P. neopolydactyla* 2a (Fig. 6) and *P. scabrosa* 1/*P. scabrosa* 2 (Fig. 5) were never found with the same *Nostoc* phylogroup, even when sampled in close spatial proximity. We noticed that all thalli of specimens associating with the *Nostoc* phylogroup VIIa, typically have an emerald green color when wet in nature, and some species (*P. neopolydactyla* 1, *P. occidentalis*, *P. scabrosa* 1, *P. scabrosa* 4) always associate with this phylogroup in the circumboreal zone, while other morphologically similar species (*P. neopolydactyla* 2a, *P. neopolydactyla* 4, *P. scabrosa* 2) never associate with it. The identity of the *Nostoc* can therefore be used to distinguish species within these two species pairs, especially in the field or when no molecular data are available. Further studies on *Nostoc* phylogroups, in concert with detailed phenotypic and geographical analyses, are needed to confirm the reliability of the color of cyanobionts and thalli as additional characters for identifying other *Peltigera* species.

#### 4.4. Species circumscription within the scabrosoid clade

Species delimitations in the scabrosoid clade corresponded to the circumscriptions by Magain et al. (2017) except for two additional lineages (*P. scabrosa* 3b and *P. neopolydactyla* 7), which were not included in that previous study. The scabrosoid clade, which traditionally includes three described species (*P. scabrosa*, *P. neopolydactyla*, and *P. melanorrhiza*) contained seven additional species supported by nearly all implemented methods (Fig. 5). Many of these newly delimited species were well-defined genetically (based on seven loci), geographically, and harbored specific *Nostoc* symbionts. On the other hand, they did not show obvious phenotypic morphological traits to distinguish them. Two species *P. neopolydactyla* 5 and *P. neopolydactyla* 6 are both endemic to the Pacific Northwest region of North America; however, *P. neopolydactyla* 6 showed a strong preference to two *Nostoc* phylogroups (XIII and XVIII), whereas *P. neopolydactyla* 5 was always found with *Nostoc* phylogroup XIb (Fig. 5). *Peltigera neopolydactyla* 7 is a rare species known only from one specimen collected in Japan. *Peltigera neopolydactyla* 4 is a widespread panboreal species, and most specimens identified by Holtan-Hartwig (1993) as *P. neopolydactyla* morphotype C correspond to this species.

Four newly circumscribed species (*P. scabrosa* 1, *P. scabrosa* 2, *P. scabrosa* 3 and *P. scabrosa* 4) were embedded in the former *P. scabrosa* s.l. (circumscribed by distinctly scabrous upper cortex of the thallus). Distinguishing them solely based on phenotypic traits is a difficult task. *Peltigera scabrosa* 4 is a rare species currently known from Norway and Québec (Canada), which might have a circumboreal distribution. Whether *P. scabrosa* 3 consists of one or two species (3a and 3b; Figs. 3, 4a, and 5) remains unclear. *Peltigera scabrosa* 1, *P. scabrosa* 2 and *P. scabrosa* 4 seem to differ slightly in their venation patterns. *Peltigera scabrosa* 2 can be separated from the remaining two species by the color of its thallus (brownish/grayish versus emerald green) and different *Nostoc* phylogroups (phylogroup VIId or XIa versus VIIa). Unfortunately, this characteristic can be obscured when the upper cortex of the thallus has a pronounced scabrosity. *Peltigera scabrosa* 3 is very rare and more material is needed to better understand its phenotypic traits. The circumscription of *P. melanorrhiza* (Purvis and James, 1993; Vitikainen, 1994) as a single species endemic to the Azores, was confirmed here.

#### 4.5. Species circumscription within the dolichorhizoid clade

In the dolichorhizoid clade, a total of 29 species were recognized in this study, compared to eleven currently accepted species. Three of the

previously described species, *P. dissecta*, *P. chilensis* and *P. microdactyla*, were found to be conspecific with *P. hymenina*, *P. triculenta*, and *P. pulverulenta*, respectively. Despite the addition of three new COR loci, new specimen collections, and the implementation of various species discovery and validation methods, we did not find major differences in the species delimited here compared to Magain et al. (2017).

The *occidentalis* species complex (including *Peltigera* sp. 6) was split here into three phylogenetically and morphologically distinct species (*P. occidentalis* s.str., *P. sp. 6* and *P. sp. 12*; Figs. 3, 4B, and 6). The panboreal species *P. occidentalis* s.str., which is easily identified morphologically, was always found in association with *Nostoc* phylogroup VIIa in the boreal zone, but was associated with *Nostoc* phylogroup VIIb in the Appalachians. Most specimens assigned by Holtan-Hartwig (1993) to *P. neopolydactyla* morphotype A correspond to *P. occidentalis*. *Peltigera* sp. 6 is one of the few generalist species (i.e., associates with multiple *Nostoc* phylogroups) as it was often the case for *Peltigera* species from tropical Central and South America (Magain et al., 2017), whereas *P. sp. 12* is known from a single specimen collected in Japan.

In the *hymenina* species complex, five distinct species (*P. hymenina*, *P. "hawaiensis"*, *P. spp. 3*, 4 and 5) were supported by most methods (Figs. 3, 4B, 6). Despite a unique ITS haplotype, morphology, and distribution restricted to the Azores (Purvis and James, 1993; Vitikainen, 1994), *P. dissecta* was nested within *P. hymenina*. The latter species has a broad distribution in the Northern Hemisphere, in (sub-) oceanic regions of North America, Europe and Atlantic Islands. *Peltigera hymenina* was however never found in the Pacific Islands, where the remaining species of this group occur. This species is easily recognized morphologically despite a relatively high degree of variation in the lobe size and shape. *Peltigera hymenina* tends to be a generalist, however, it was often found with phylogroup XVI. *Peltigera* sp. 3 is widespread in Australia and New Zealand, where it was usually identified as *P. dolichorhiza*, *P. polydactylon* or *P. nana*. *Peltigera "hawaiensis"* (informal name; Vitikainen pers. comm.) is known only from Hawaii. *Peltigera* spp. 4 and 5, both represented in this study by a single specimen from Papua New Guinea, were considered as morphotypes of *P. dolichorhiza* in the study by S erusiaux et al. (2009).

In the *neopolydactyla* species complex, *P. pacifica* (endemic to the Pacific Northwest of North America) is the only species recognized consistently by almost all methods. It is easily identified morphologically, partly because of the presence of marginal phyllidia. It also shows a high specificity to *Nostoc* phylogroup XIII. We found that *Peltigera neopolydactyla* s.l. represented an assemblage of phylogenetically unrelated lineages, seven of which were recognized as putative species inside the dolichorhizoid clade, and an additional four were detected inside the scabrosoid clade. *Peltigera neopolydactyla* 1b, which seems to be distinct from *P. neopolydactyla* 1, is known from a single specimen collected in Peru, whereas *P. neopolydactyla* 1 is broadly distributed in the Northern Hemisphere, including the Appalachian Mountains, Arizona, and boreal areas in Norway, Qu ebec and Russia. In the boreal zone, *Nostoc* phylogroup VIIa is its sole cyanobiont, whereas phylogroups VIIb and VIIc (not shown here) were found in thalli collected in the Appalachians and Arizona, respectively. *Peltigera neopolydactyla* 3 is known from Yunnan (China) and Vietnam, and is considered as a species by most methods. *Peltigera neopolydactyla* 2 s.l. (considered as one species in Magain et al., 2017) is a complex containing three lineages, *P. neopolydactyla* 2a, 2b and 2c, which were either considered as distinct species, unresolved cases, or were grouped together depending on the method used (Fig. 6). *Peltigera neopolydactyla* 2b could represent a distinct species according to most methods. This species is known mainly from Japan and the Yunnan Province in China and is associated with *Nostoc* phylogroup X and VIIb. The remaining two species *P. neopolydactyla* 2a (widespread in the boreal and sub-boreal regions and French Pyrenees and Vosges) and *P. neopolydactyla* 2c (found in Northern Japan and Eastern Russia) show the same association towards *Nostoc* phylogroup XIII. Their geographical ranges overlap in Japan and Eastern Russia. Because they share the same *Nostoc*

phylogroup, overlapping geographic ranges, and cannot be distinguished by ITS, it is extremely difficult to distinguish *P. neopolydactyla* 2a and 2c. These two taxa mostly differ by their COR sequences. It is unclear if this high variation in the COR loci is intraspecific or interspecific. Most specimens assigned to *P. neopolydactyla* morphotype B by Holtan-Hartwig (1993) belong to *P. neopolydactyla* 2a.

Depending on the method used (Fig. 6), the *scabrosella* species complex was delimited as a single species (as in Magain et al., 2017) or an assemblage of three species: the formerly recognized, *P. scabrosella* s.str. and two putative new species, *P. sp. 7a* and *P. sp. 7b*. All specimens from this complex were associated with *Nostoc* phylogroup XIa. Thalli vary from small, scabrid and almost veinless in *P. scabrosella* s.str. to wide, glabrous and distinctly veined in *P. sp. 7a* and *7b*, which can be distinguished from each other by the size and shape of their lobes. Moreover, each lineage has a distinct geographical distribution: *P. scabrosella* occurs in boreal zones of Scandinavia and the Atlantic coast of North America, *P. sp. 7a* is endemic to the Pacific Northwest of North America, whereas *P. sp. 7b* has been found only in Hokkaido (Japan). Although all populations were genetically similar, the three putative species had distinct haplotypes for most loci. This allelic exclusivity, combined with morphological differences and distinct geographical ranges, support the recognition of three distinct species within *P. scabrosella* s.l. (Fig. 6).

Most methods agreed in distinguishing three morphologically cryptic and sympatric species within *P. pulverulenta* s.l.: *P. pulverulenta* 1, 2, and 3. *Peltigera pulverulenta* 1 seems to be common in Central and South America (Supplementary Table S2), whereas the two other putative species potentially have more restricted distribution, having so far been found only in Colombia. They may represent sibling species (sensu Steyskal, 1972) defined as closely related and morphologically undistinguishable entities resulting from divergence and genetic isolation that are not correlated with visible phenotypic changes. Most methods also recognize the widespread *P. dolichorhiza* (found in Central and South America, as well as Africa, including Madagascar and Reunion Island), and the rare *P. sp. 1* (Colombia and Bolivia), *P. sp. 2a* (Colombia), *P. sp. 2b* and *P. dolichorhiza* 2 (Brazil) as separate species (Fig. 6). All these potential species are morphologically distinct, even though they are part of a recent radiation (Magain et al., 2017). The status of the morphologically variable *P. triculenta* (including the conspecific *P. chilensis*), known from southern regions of Chile and Argentina and several sub-antarctic islands (e.g., Kerguelen Islands), needs further assessment with additional specimens. The possibility that *P. dolichorhiza* b and *P. sp. 2ab* represent distinct species remains uncertain based on this study. Contrary to northern hemisphere species, none of the South-American putative new species (members of the former *P. dolichorhiza* s.l., *P. pulverulenta* s.l., and *P. triculenta*) shows a high degree of specialization in their *Nostoc* selection.

The dolichorhizoid clade represents a very recent radiation (Magain et al., 2017), probably still undergoing speciation processes. Therefore, it is possible that species have not diverged enough to detect a gap between their intra- and inter-species genetic distances. Artificial gap in similarity distributions can also result from insufficient sampling across taxa (Wiemers and Fiedler, 2007). Some of the species with low ITS interspecific variation, such as *P. dolichorhiza* 2 or *P. sp. 2b*, were poorly sampled in our analysis (Fig. 8F). However, ITS variation below the intraspecific average observed in well-sampled species, in terms of number of specimens and geographic origin, such as *P. dolichorhiza* or *P. hymenina* suggests that these taxa represent well-defined species, whereas high intraspecific dissimilarity in *P. neopolydactyla* 1, *P. triculenta*, *P. sp. 2a*, *P. dolichorhiza* b and *P. neopolydactyla* 2b might indicate that these species contain even more undiscovered biodiversity than revealed by our analyses or are still undergoing speciation (Fig. 8F).

#### 4.6. Are there cosmopolitan species in the section *Polydactylon*?

*Peltigera scabrosa*, *P. dolichorhiza*, *P. polydactylon* and *P. neopolydactyla* were traditionally considered as cosmopolitan, or sub-cosmopolitan species (Martínez et al., 2003). Our study revealed that each of these species represents a species complex containing multiple taxonomic entities, most of which were confirmed by multiple methods and criteria at the species level (formal descriptions of species will be included in a follow-up publication). For example, *Peltigera scabrosa* s.l. is composed of at least four (possibly five) panboreal species. *Peltigera dolichorhiza* s.str., is restricted to the Neotropics and Afrotropics, whereas similar morphotypes from other parts of the world represent putative undescribed species. Therefore, *Peltigera* section *Polydactylon* does not seem to include any cosmopolitan species. Most species occur within well-delimited geographic ranges. Many species are panboreal or widespread within a single continent, e.g., in South America, or Asia. The names used to identify similar morphotypes across continents and climatic zones were often found to represent multiple distinct lineages corresponding to different species in need of formal descriptions.

As circumscribed by molecular data and species delimitation methods, species in *Peltigera* section *Polydactylon* showed narrower ranges and higher specificity than traditionally delimited morphospecies. Species with limited distributions and high specificity towards their cyanobiont seem to be the rule in this group, while cosmopolitan species may not exist in section *Polydactylon*. We successfully developed fast evolving intergenic spacers identified from syntenic regions of a wide range of genomes, and demonstrated their usefulness in delimiting closely related species. We combined these new phylogenetic data with morphological, distribution and patterns of interactions with *Nostoc* photobiont to test the circumscription of *Peltigera* species using various species validation methods. We argue that integrating these various lines of evidence is critical to delimit species in recently diversified groups with complex morphological and DNA sequence patterns, such as section *Polydactylon*.

#### Acknowledgments

We would like to thank all the people and herbaria that sent us material. We thank Orvo Vitikainen and Trevor Goward for their assistance and expertise. We thank Peter Crittenden and Paul Dyer for allowing us to use unpublished genomic data of *Xanthoria parietina*. We also thank John Logsdon, Elizabeth Savelkoul and Cindy Toll, as well as support from grant NSF DEB 1011101 (to J.L. and E.S.) for the unpublished genomic data of *Peltigera evansiana*. This study was supported by the National Science Foundation (NSF) REVSYS award on the genus *Peltigera* DEB-1025930 to J.M. and F.L., NSF Dimensions of Biodiversity award DEB-1046065 to F.L., and NSF SG DEB-1556995 on phylogenetic and network analyses using *Peltigera* as a model system to F.L. and J.M. N.M. was supported by a BAEF (Belgian American Education Foundation) fellowship. B.G. was supported by NSF grant DEB-1354631. This work was supported in part by the Swiss National Science Foundation prospective researcher grant PBGP3\_145339 to C.T. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, was supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

#### Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ympmv.2017.08.013>.

## References

- Altermann, S., Leavitt, S.D., Goward, T., Nelsen, M.P., Lumbsch, H.T., 2014. How do you solve a problem like Letharia? A new look at cryptic species in lichen-forming fungi using Bayesian clustering and SNPs from multilocus sequence data. *PLoS One* 9 (5), e97556.
- Blackwell, M., 2011. The Fungi: 1, 2, 3, ..., 5.1 million species? *Am. J. Bot.* 98, 426–438.
- Carstens, B.C., Pelletier, T.A., Reid, N.M., Satler, J.D., 2013. How to fail at species delimitation. *Mol. Ecol.* 22, 4369–4383.
- Crespo, A., Kauff, F., Divakar, P.K., del Prado, R., Pérez-Ortega, S., de Paz, G.A., Ferencova, Z., Blanco, O., Roca-Valiente, B., Núñez-Zapata, J., Cubas, P., Argüello, A., Elix, J.A., Esslinger, T.L., Hawksworth, D.L., Millanes, A.M., Molina, M.C., Wedin, M., Ahti, T., Aptroot, A., Barreno, E., Bungartz, F., Calvelo, S., Candan, M., Cole, M.J., Ertz, D., Goffinet, B., Lindblom, L., Lücking, R., Lutzoni, F., Mattsson, J.E., Messuti, M.I., Miadlikowska, J., Piercey-Normore, M.D., Rico, V.J., Sipman, H., Schmitt, I., Spribille, T., Thell, A., Thor, G., Upreti, D.K., Lumbsch, H.T., 2010. Phylogenetic classification of parmelioid lichens (Parmeliaceae, Ascomycota) based on molecular, morphological and chemical evidence. *Taxonomy* 59, 1735–1753.
- Crespo, A., Lumbsch, H.T., Mattsson, J.E., Blanco, O., Divakar, P.K., Articus, K., Wiklund, E., Bawingan, P.A., Wedin, M., 2007. Testing morphology-based hypotheses of phylogenetic relationships in Parmeliaceae (Ascomycota) using three ribosomal markers and the nuclear RPB1 gene. *Mol. Phylogenet. Evol.* 44, 812–824.
- De Queiroz, K., 2007. Species concepts and species delimitation. *Syst. Biol.* 56, 879–886.
- De Queiroz, K., 1998. The general lineage concept of species, species criteria, and the process of speciation: a conceptual unification and terminological recommendations. In: Howard, Daniel J., Berlocher, Stewart H. (Eds.), *Endless Forms: Species and Speciation*. Oxford University Press, pp. 57–75.
- Deng, W., Nickle, D.C., Learn, G.H., Maust, B., Mullins, J.I., 2007. ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. *Bioinformatics* 23, 2334–2336.
- Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214.
- Ebersberger, I., de Matos Simoes, R., Kupczok, A., Gube, M., Kothe, E., Voigt, K., Haeseler, von A., 2012. A consistent phylogenetic backbone for the fungi. *Mol. Biol. Evol.* 29, 1319–1334.
- Ence, D.D., Carstens, B.C., 2011. SpedeSTEM: a rapid and accurate method for species delimitation. *Mol. Ecol. Resour.* 11, 473–480.
- Fitzpatrick, D.A., Logue, M.E., Stajich, J.E., Butler, G., 2006. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evol. Biol.* 6, 99.
- Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., Dubchak, I., 2004. Vista: computational tools for comparative genomics. *Nucleic Acids Res.* 32, W273–W279.
- Grigoriev, I.V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., Kuo, A., Minovitsky, S., Nikitin, R., Ohm, R.A., Otililar, R., Poliakov, A., Ratnere, I., Riley, R., Smirnova, T., Rokhsar, D., Dubchak, I., 2011. The genome portal of the department of energy joint genome institute. *Nucleic Acids Res.* gkr947.
- Hane, J.K., Rouxel, T., Howlett, B.J., Kema, G.H.J., Goodwin, S.B., Oliver, R.P., 2011. A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. *Genome Biol.* 12, R45.
- Harris, R.S., 2007. Improved pairwise alignment of genomic DNA. Ph.D. Thesis. The Pennsylvania State University.
- Hawksworth, D.L., 1991. The fungal dimension of biodiversity: Magnitude, significance, and conservation. *Mycol. Res.* 95, 641–655.
- Hawksworth, D.L., 2001. The magnitude of fungal diversity: the 1.5 million species estimate revisited. *Mycol. Res.* 105, 1422–1432.
- Hodkinson, B.P., Allen, J.L., Forrest, L.L., Goffinet, B., Sérusiaux, E., Andresson, O.S., Miao, V., Bellenger, J.-P., Lutzoni, F., 2014. Lichen-symbiotic cyanobacteria associated with *Peltigera* have an alternative vanadium-dependent nitrogen fixation system. *Eur. J. Phycol.* 49, 11–19.
- Hofstetter, V., Miadlikowska, J., Kauff, F., Lutzoni, F., 2007. Phylogenetic comparison of protein-coding versus ribosomal RNA-coding sequence data: a case study of the Lecanoromycetes (Ascomycota). *Mol. Phylogenet. Evol.* 44 (1), 412–426.
- Holtan-Hartwig, J., 1993. The Lichen Genus *Peltigera*, Exclusive of the *P. canina* Group, in Norway. Botanical Garden and Museum-University of Oslo.
- Huelsenbeck, J.P., Andolfatto, P., Huelsenbeck, E.T., 2011. Structurama: Bayesian inference of population structure. *Evolut. Bioinf. Online* 7, 55.
- Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755.
- Jeffroy, O., Brinkmann, H., Delsuc, F., Philippe, H., 2006. Phylogenomics: the beginning of incongruence? *Trends Genet.* 22, 225–231.
- Kirk, P.M., Cannon, P.F., Minter, D.W., Stalpers, J.A., 2008. *Dictionary of the Fungi*. CABI, Wallingford, U.K.
- Kubatko, L.S., Carstens, B.C., Knowles, L.L., 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics* 25, 971–973.
- Kuramae, E.E., Robert, V., Snel, B., Weiß, M., Boekhout, T., 2006. Phylogenomics reveal a robust fungal tree of life. *FEMS Yeast Res.* 6, 1213–1220.
- Lanfear, R., Calcote, B., Ho, S.Y., Guindon, S., PartitionFinder, S., 2012. Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* 29, 1695–1701.
- Leavitt, S.D., Fankhauser, J.D., Leavitt, D.H., Porter, L.D., Johnson, L.A., Clair, L.L., 2011a. Complex patterns of speciation in cosmopolitan “rock posy” lichens—discovering and delimiting cryptic fungal species in the lichen-forming *Rhizoplaca melanophthalma* species-complex (Lecanoraceae, Ascomycota). *Mol. Phylogenet. Evol.* 59, 587–602.
- Leavitt, S.D., Johnson, L.A., Goward, T., Clair, L.L., 2011b. Species delimitation in taxonomically difficult lichen-forming fungi: An example from morphologically and chemically diverse *Xanthoparmelia* (Parmeliaceae) in North America. *Mol. Phylogenet. Evol.* 60, 317–332.
- Leavitt, S.D., Grewe, F., Widhelm, T., Muggia, L., Wray, B., Lumbsch, H.T., 2016. Resolving evolutionary relationships in lichen-forming fungi using diverse phylogenomic datasets and analytical approaches. *Sci. Rep.* 6.
- Liu, Y., Steenkamp, E.T., Brinkmann, H., Forget, L., Philippe, H., Lang, B.F., 2009. Phylogenomic analyses predict sistergroup relationship of nucleariids and fungi and paraphyly of zygomycetes with significant support. *BMC Evol. Biol.* 9, 1.
- Lücking, R., Dal-Forno, M., Sikaroodi, M., Gillevet, P.M., Bungartz, F., Moncada, B., Yáñez-Ayabaca, A., Chaves, J.L., Coca, L.F., Lawrey, J.D., 2014. A single macrolichen constitutes hundreds of unrecognized species. *Proc. Natl. Acad. Sci.* 111, 11091–11096.
- Lumbsch, H.T., Leavitt, S.D., 2011. Goodbye morphology? A paradigm shift in the delimitation of species in lichenized fungi. *Fungal Diversity* 50, 59–72.
- Maddison, D., Maddison, W., 2005. *MacClade v. 4.08*.
- Magain, N., Forrest, L.L., Sérusiaux, E., Goffinet, B., 2010. Microsatellite primers in the *Peltigera dolichorhiza* complex (lichenized ascomycete, Peltigerales). *Am. J. Bot.* 97, e102–e104.
- Magain, N., Sérusiaux, E., Zhurbenko, M.P., Lutzoni, F., Miadlikowska, J., 2016. Disentangling the *Peltigera polydactylon* species complex by recognizing two new taxa, *P. polydactylon* subsp. *udeghe* and *P. seneca*. *Herzogia* 29, 514–528.
- Magain, N., Miadlikowska, J., Goffinet, B., Sérusiaux, E., Lutzoni, F., 2017. Macroevolution of specificity in cyanolichens of the genus *Peltigera* section *Polydactylon* (Lecanoromycetes, Ascomycota). *Syst. Biol.* 66, 74–99.
- Martínez, I., Burgaz, A.R., Vitikainen, O., Escudero, A., 2003. Distribution patterns in the genus *Peltigera* Willd. *Lichenologist* 35, 301–323.
- Mason-Gamer, R.J., Kellogg, E.A., 1996. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). *Syst. Biol.* 45 (4), 524–545.
- Mayr, E., 1940. Speciation phenomena in birds. *Am. Nat.* 1940, 249–278.
- McDonald, T.R., Mueller, O., Dietrich, F.S., Lutzoni, F., 2013. High-throughput genome sequencing of lichenizing fungi to assess gene loss in the ammonium transporter/ammonia permease gene family. *BMC Genomics* 14, 1.
- Miadlikowska, J., Lutzoni, F., 2000. Phylogenetic revision of the genus *Peltigera* (lichen-forming ascomycota) based on morphological, chemical, and large subunit nuclear ribosomal DNA data. *Int. J. Plant Sci.* 161, 925–958.
- Miadlikowska, J., Richardson, D., Magain, N., Ball, B., Anderson, F., Cameron, R., Lendemer, J., Truong, C., Lutzoni, F., 2014. Phylogenetic placement, species delimitation, and cyanobiont identity of endangered aquatic *Peltigera* species (lichen-forming Ascomycota, Lecanoromycetes). *Am. J. Bot.* 101, 1141–1156.
- Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Gateway Computing Environments Workshop (GCE)*. IEEE, pp. 1–8.
- Nylander, J., 2004. MrModeltest v2. Program Distributed by the Author, vol. 2. Evolutionary Biology Centre, Uppsala University.
- Nylander, J.A., Wilgenbusch, J.C., Warren, D.L., Swofford, D.L., 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in bayesian phylogenetics. *Bioinformatics* 24, 581–583.
- O'Brien, B.L., Parrent, J.L., Jackson, J.A., Moncalvo, J.M., Vilgalys, R., 2005. Fungal community analysis by large-scale sequencing of environmental samples. *Appl. Environ. Microbiol.* 71, 5544–5550.
- Oliver, J.C., Miadlikowska, J., Arnold, A.E., Maddison, D.R., Lutzoni, F., 2013. Hypha: A Mesquite Package for Support Value Integration. Version 1.0.
- Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290.
- Parmmen, S., Rangsiroji, A., Mongkolsuk, P., Boonpragob, K., Nutakki, A., Lumbsch, H.T., 2012. Using phylogenetic and coalescent methods to understand the species diversity in the *Cladia aggregata* complex (Ascomycota, Lecanorales). *PLoS One* 7 (12), e52245.
- Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* 7, e37135.
- Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multi-locus genotype data. *Genetics* 155, 945–959.
- Purvis, O.W., James, P.W., 1993. Studies on the lichens of the Azores. Part 1-Caldeira do Faial. *ARQUIPÉLAGO. Ciências Biol. Mar. = Life Mar. Sci.* 11, 1–15.
- R Development Core Team, 2002. *R: A Language and Environment for Statistical Computing*.
- Rambaut, A., Charleston, M., 2008. *Treeddit: phylogenetic tree editor v. 1.0 alpha 10*. See < <http://evolve.zoo.ox.ac.uk/software/TreeEdit/main.html> > .
- Rambaut, A., Drummond, A., 2007. *Tracer v1.5*.
- Reeb, V., Lutzoni, F., Roux, C., 2004. Contribution of RPB2 to multilocus phylogenetic studies of the euascomycetes (Pezizomycotina, Fungi) with special emphasis on the lichen-forming Acarosporaceae and evolution of polyspory. *Mol. Phylogenet. Evol.* 32 (3), 1036–1060.
- Reid, N.M., Carstens, B.C., 2012. Phylogenetic estimation error can decrease the accuracy of species delimitation: a Bayesian implementation of the general mixed yule-coalescent model. *BMC Evol. Biol.* 12, 196.
- Robbertse, B., Reeves, J.B., Schoch, C.L., Spatafora, J.W., 2006. A phylogenomic analysis of the Ascomycota. *Fungal Genet. Biol.* 43, 715–725.
- Rodriguez, F., Oliver, J., Marin, A., Medina, J.R., 1990. The general stochastic model of nucleotide substitution. *J. Theor. Biol.* 142, 485–501.
- Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425, 798–804.
- Scherrer, S., Zippler, U., Honegger, R., 2005. Characterisation of the mating-type locus in the genus *Xanthoria* (lichen-forming ascomycetes, Lecanoromycetes). *Fungal Genet.*

- Biol. 42, 976–988.
- Schmull, M., Miadlikowska, J., Pelzer, M., Stocker-Wörgötter, E., Hofstetter, V., Fraker, E., Hodkinson, B.P., Reeb, V., Kuwka, M., Lumbsch, H.T., Kauff, F., Lutzoni, F., 2011. Phylogenetic affiliations of members of the heterogeneous lichen-forming fungi of the genus *Lecidea* sensu Zahlbruckner (Lecanoromycetes, Ascomycota). *Mycologia* 103, 983–1003.
- Schneider, K., Resl, P., Spribille, T., 2016. Escape from the cryptic species trap: lichen evolution on both sides of a cyanobacterial acquisition event. *Mol. Ecol.* <http://dx.doi.org/10.1111/mec.13636>.
- Schoch, C.L., Crous, P.W., Groenewald, J.Z., Boehm, E.W.A., Burgess, T.I., De Gruyter, J., De Hoog, G.S., Dixon, L.J., Grube, M., Gueidan, C., Harada, Y., 2009. A class-wide phylogenetic assessment of Dothideomycetes. *Stud. Mycol.* 64, 1–15.
- Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., Griffith, G.W., 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci.* 109, 6241–6246.
- Sérusiaux, E., Goffinet, B., Miadlikowska, J., Vitikainen, O., 2009. Taxonomy, phylogeny and biogeography of the lichen genus *Peltigera* in Papua New Guinea. *Fungal Diversity* 38, 185.
- Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690.
- Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* 57, 758–771.
- Steyskal, G.C., 1972. The meaning of the term 'sibling species'. *Syst. Biol.* 21, 446.
- Sukumaran, J., Knowles, L.L., 2017. Multispecies coalescent delimits structure, not species. *Proc. Natl. Acad. Sci.* 114, 1607–1612.
- Swofford, D.L., 2003. PAUP\*. Phylogenetic analysis using parsimony (\* and other methods). Version 4.
- Taylor, J.W., Jacobson, D.J., Kroken, S., Kasuga, T., Geiser, D.M., Hibbett, D.S., Fisher, M.C., 2000. Phylogenetic species recognition and species concepts in fungi. *Fungal Genet. Biol.* 31, 21–32.
- Townsend, J.P., 2007. Profiling phylogenetic informativeness. *Syst. Biol.* 56 (2), 222–231.
- Vitikainen, O., 1994. Taxonomic revision of *Peltigera* (lichenized Ascomycotina) in Europe. *Acta Bot. Fennica* 152, 1–96.
- Wedin, M., Wiklund, E., Jørgensen, P.M., Ekman, S., 2009. Slippery when wet: phylogeny and character evolution in the gelatinous cyanobacterial lichens (Peltigerales, Ascomycetes). *Mol. Phylogenet. Evol.* 53, 862–871.
- Wickham, H., 2009. Ggplot2: Elegant Graphics for Data Analysis. Springer.
- Wiemers, M., Fiedler, K., 2007. Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae). *Front. Zool.* 4, 8.
- Xavier, B.B., Miao, V.P., Jonsson, Z.O., Andresson, O.S., 2012. Mitochondrial genomes from the lichenized fungi *Peltigera membranacea* and *Peltigera malacea*: Features and phylogeny. *Fungal Biol.* 116, 802–814.
- Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics* 13, 555–556.
- Yang, Z., Rannala, B., 2010. Bayesian species delimitation using multilocus sequence data. *Proc. Natl. Acad. Sci.* 107, 9264–9269.
- Zhang, J., Kapli, P., Pavlidis, P., Stamatakis, A., 2013. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 29, 2869–2876.
- Zoller, S., Lutzoni, F., Scheidegger, C., 1999. Genetic variation within and among populations of the threatened lichen *Lobaria pulmonaria* in Switzerland and implications for its conservation. *Mol. Ecol.* 8, 2049–2059.