Species delimitation in the cosmopolitan *Peltigera* section *Polydactylon* group (Peltigerales, Lecanoromycetes): comparison of methods based on molecular data and information about geography, morphol- ogy and association with the photobiont

Nicolas Magain, François Lutzoni, Michael Gajdeczka, Bernard Goffinet, Emmanuël Sérusiaux & Jolanta Miadlikowska

**Abstract**

We reconstructed the phylogeny of the mycobiont of *Peltigera* section *Polydactylon* based on molecular data from eight loci, including three newly designed intergenic *Peltigera*-specific markers (IGS1, IGS3 and IGS16) and applied five species delimitation methods. We focused on two major clades of the section: the Scabrosoid clade, where the lineages representing putative species are well delimited and most phylogenetic relationships among them are highly supported by bootstrap values; and the Dolichorhizoid clade, where substantially lower levels or resolution and bootstrap support was obtained and where species delimitation was more challenging. All methods resulted in mostly congruent species delimitations within the Scabrosoid clade. A total number of 12 species including 9 previously unrecognized species was detected. In the Dolichorhizoid clade, methods relying on different models and assumptions provided different species delimitations. The species delimitations we propose were based on a consensus among these various methods. We concluded that the Dolichorhizoid clade comprises 29 species, for which only 7 have already been described and named. The consensus approach revealed that most “evolutionary significant” species have relatively ￼well-defined distribution ranges (usually panboreal or restricted to a single biogeographic region), and cosmopolitan species names usually referred to assemblages of distinct evolutionary lineages. The majority of newly delimited species showed a high specificity towards their cyanobionts. Information about geographic origin and patterns of photobiont association can be useful for species delimitation and identification.

Key Words: Structurama, bGMYC, bPTP, spedestem, bPP, coalescence, mycobiont, lichen, *Nostoc*, biogeography, ascomycota, IGS

**Introduction**

Species are key units to understand relationships among organisms, ecosystem dynamics, as well as to understand the dynamics of evolution. Defining biologically significant species units is thus very important for many aspects of the study of life.

For long, the delimitation and identification of lichen-forming fungi was based solely on morphology.

However, molecular data highlighted the difficulty to correctly define boundaries among species (defined as "separately evolving metapopulation lineages"; De Queiroz 1998) in order to bridge the morphological and other commonly used species concepts (e.g., the biological species concept).

Indeed, the morphological species concept can be very difficult to apply in fungi, including lichens, because the absence of diagnostic traits can lead to the recognition of fewer species than phylogenetically defined (e.g., Crespo & Perez-Ortega 2009; Miadlikowska et al, 2014). Intraspecific plasticity can be often higher than interspecific differences, leading to the circumscription of species representing different phenotypes within a single evolutionary lineage (e.g., Pino-Bodas et al. 2011). It can also be very difficult to detect morphological convergence when characters lack distinct developmental signature, and as a consequence unrelated lineages were sometimes embedded within the same species (e.g., Lumbsch et al. 2005, Otalora & Wedin 2013, Passo et al. 2008). Moreover, cryptic species that cannot be recognized based solely on the morphology have been frequently detected in lichen-forming fungi, including well-studied taxa from well-sampled areas (e.g., Lumbsch & Leavitt 2011).Biogeographical factors shaping the systematics of lichen-forming fungi were often neglected. For example, the same species name was often applied to morphologically similar individuals from different continents, when they might be drastically different genetically (e.g., Leavitt et al. 2011). As a result, species based on morphological concepts (morphospecies), might not always represent biologically or phylogenetically meaningful units. Moreover, recognition and circumscription of morphological traits are sometimes arbitrary and authors may diverge on boundaries among morphologically defined species. Chemotypic variation (differences in the set of secondary compounds) as an alternative tool for species delimitation was proved to be unreliable because the chemical traits often vary depending on the stage of lichen development, the part of the thallus or the ecological conditions (Lumbsch 1998).

Recognizing biological species sensu Mayr (1940; “groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups”), by testing the mating compatibility, is problematic in lichens because most lichenized fungi do not grow in artificial conditions. Moreover, lichens grow very slowly and their spores are tiny, difficult to observe and can be carried on very long distances, resulting in a great difficulty to monitor and test lichen reproduction (but see Zoller et al. 1999). Asexuality or homothallism in some species are also factors that complicate mating tests on lichens (Taylor et al., 2000; but see Scherrer et al. 2005).

However, the availability of molecular data has made possible to apply in lichen-forming fungi the phylogenetic species concept ("a diagnosable cluster of individuals within which there is a parental pattern of ancestry and descent, beyond which there is not, and which exhibits a pattern of phylogenetic ancestry and descent among units of like kind’’; Eldredge and Cracraft, 1980) and the genealogical species concept ("basal group of organisms whose members are all more closely related to each other than they are to any organisms outside the group; Hudson & Coyne 2002) concepts. In particular, Taylor et al. (2000) recommended the use of the Genealogical Concordance species concept or "exclusive concordance of alleles, where different gene topologies have to be congruent for interspecific relationships" (Avise and Ball 1990) for species delimitation in fungi. Once phylogenetic or genealogical species are defined based on molecular data, it is possible to select a posteriori a set of phenotypic and chemotypic traits that are species-specific (Lumbsch & Leavitt 2011).

***Peltigera* section *Polydactylon***

*Peltigera* section *Polydactylon* is one of the eight sections defined by Miadlikowska et al. (2000). It is a well-supported monophyletic group (Miadlikowska et al. 2000), and is composed of three main clades, named the Dolichorhizoid (containing *P. dolichorhiza*), Polydactyloid (containing *P. polydactylon*) and Scabrosoid (containing *P. scabrosa*) clades (Magain et al. 2015). It only comprises species associating solely with the cyanobacterial genus *Nostoc*.

Most species from this section reproduce sexually, but a few members also produce vegetative propagules (mostly phyllidia). This section as a whole has a broad, almost cosmopolitan, distribution. However, the distribution range of individual species varies greatly within the group ranging from endemism in small geographical areas such as the Azores to broad distributions covering North America, Europe and Asia (Martinez et al. 2003). About fifteen species have been described in this group (Holtan-Hartwig 1993, Vitikainen 1994, Miadlikowska 2000) but it has been suspected that section *Polydactylon* contains several species complexes encompassing cryptic as well as morphologically distinct but unrecognized species. Because most species in this group have relatively uniform morphology, the implementation of a morphological species concept is likely to lead to an underestimation of the number of species. However, the recognition of geographically structured morphotypes and chemotypes within broadly distributed taxa such as for *P. neopolydactyla*, *P. scabrosa* (Holtan-Hartwig 1993, Vitikainen 1994), and *P. dolichorhiza* (Serusiaux et al. 2009) strongly suggest the presence of multiple undescribed species.

Distinct morphological, chemical and geographical patterns detected within the section *Polydactylon* makes it a good model system for testing if cosmopolitan lichen species represent single “evolving metapopulation lineage" (De Quieroz 1998) or assemblages of morphologically similar but evolutionary distinct lineages.

**Objectives**

We selected eight loci (three protein-coding and two ribosomal genes and three newly developed *Peltigera*-specific molecular markers; intergenic spacers; IGS), to reconstruct the phylogeny of the mycobiont based on a worldwide sampling of most of the known species from the section *Polydactylon*. The multi-locus data was used to test and compare several species delimitation methods on the Scabrosoid clade, where lineages representing putative species are well defined; and on the Dolichorhizoid clade, where boundaries among species are not clear because of their putative recent origin. The tree topologies for these two groups, are very different (see Magain et al. 2015); the Scabrosoid clade is well-resolved, with many long, well-supported branches representing different lineages, whereas the Dolichorhizoid clade includes several polytomies, many short branches, and unsupported relationships. We also reconstructed a species tree, once species were defined, and compared its topology with phylogenies obtained from concatenated datasets.

We tested whether cosmopolitan species represent single evolutionary lineages or the assemblages of several distinct entities and if unrecognized, cryptic or morphologically meaningful species are present in this group. We assessed, the taxonomic status of phylogenetic lineages derived most likely from recent radiation events within the South American clade. I also tested if geographical data on species distributions and patterns of association with *Nostoc* phylogroups (sensu Magain et al. 2015) can be used, in addition to the morphology to support newly delimited species.

Because molecular data were more limited in the Polydactyloid clade (higher proportion of missing sequences, fewer representatives per species, lack of data for species from remote regions of the world with no recent material available) Ithis clade was not included in the comparison of the species delimitation methods.

**Species delimitation methods and approaches tested**

**Structurama** (Huelsenbeck et al. 2011) is a software using multilocus genotype data to infer population structure and assign individuals to populations (Pritchard et al., 2000). Each cluster (population) is modeled by a characteristic set of allele frequencies. Its main modeling assumptions are Hardy-Weinberg equilibrium within populations and complete linkage equilibrium between loci within populations. Under these assumptions each allele at each locus in each genotype is an independent draw from the appropriate frequency distribution. It attempts to find population groupings that are not in desequilibrium.

The difference between the popular software Structure and Structurama is that, while Structure can only assign specimens to a number of populations fixed by the user, Structurama can estimate the number of populations, based on the data and on priors determined by the users(Huelsenbeck et al., 2011).

Structurama has been widely used for species delimitation, assuming that the reconstructed populations in equilibrium are distinct, isolated species (see for instance Pinzon 2011, Salicini 2011).

The **GMYC** (General Mixed Yule Coalsecent) model considers branching between species as a Yule model (Yule, 1924), a stochastic birth-only model, which allows to calculate the likelihoods of the times before a new species appears in a phylogeny with assumed constant average speciation rates.

The GMYC model assumes neutral coalescence within each species; and a coalescent branching rate parameter for each species. It attempts to fit the location of the switches from speciation to coalescent nodes; which corresponds to the most recent common ancestral node defining each species.

In particular, it assumes that there is a threshold time before which all nodes reflect interspecific relationships and after which all nodes reflect intraspecific variation (Pons et al. 2006). This model can be tested on ultrametric phylogenetic trees.

**bGMYC** (Reid and Carstens, 2012) is a bayesian implementation of the GMYC approach that account for tree uncertainty.

bPTP (bayesian Poisson Tree Process) implements a similar approach, but models speciations using the number of substitutions (based on branch lengths) instead of the time. It considers that each substitution has a small probability of generating a speciation, and that it follow a Poisson distribution in continuous time. It assumes that a tree has been generated by two distinct Poisson process classes, one describing speciation, and the other describing within-species branching events and searches for the transition points where the branching pattern changes from an among-species to a within-species branching pattern. This model can be tested on non-ultrametric phylogenetic trees.

spedeSTEM (Ence et al. 2011) incorporates the program STEM (Kubatko et al. 2009) which calculates the maximum likelihood species tree from a set of gene trees, under the assumption that the incongruences between gene trees are due to coalescence only. It takes an a-priori assignment of species and single-locus gene trees as input, and compares the likelihoods of the species tree when the units tested are considered as distinct species, or merged following certain hierarchical permutations, and proposes an optimal species delimitation to maximize the likelihood of the species tree.

bPP (Yang and Rannala 2010) is a bayesian approach that generates the posterior probabilities of species assignments based on multi-locus datasets,

taking the uncertainties due to unknown gene trees and the ancestral coalescent process into account . It requires the user to provide a species guide trees, and tests if the lineages defined in the species tree should be considered as distinct lineages or merged into more inclusive species.

These two methods thus try to define the best species delimitation to maximize a species tree in a coalescent framework. One of the main difference of approach is that spedeSTEM takes fixed gene topologies into account, and don't take gene trees uncertainties into account, but computes different species trees, while on the other hand bPP takes the alignment as input, allowing to take gene topologies uncertainties into account, but requires a fixed species tree, so it doesn't consider species tree uncertainties.

These two methods try to rearrange predefined lineages among more inclusive species, but don't test the splitting of predefined lineages.

The concept of barcoding gap (Hebert et al. 2003) suggests that there should be a big difference between interspecific genetic variation and intraspecific variation so that there would be a gap, and therefore sequences with variation below a certain threshold could be assigned to a certain species.

# Materials and methods

## Development of three new markers: IGS1, IGS3, IGS16

Existing genetic markers do not provide a sufficient phylogenetic resolution and support for relationships among closely related individuals at the population and species levels in the genus *Peltigera*, as well as in other lichen genera in Lecanoromycetes.. With the aim of potentially discovering novel standing variation within genera, we used available genomic, metagenomic and metatranscriptomic data to develop and test three novel molecular markers (Gajdeczka et al., in prep). We sampled genomic sequence data for *Cladonia grayi* and *Xanthoria parietina* and three *Peltigera* species, largely from non-coding regions in order to choose potential regions that could be amplified. We targeted highly polymorphic, approximately neutrally evolving regions of the genomes.

We scanned the 30 largest scaffolds of the *Cladonia grayi* genome assembly (Clagr2; [http://genome.jgi.doe.gov/Clagr2/Clagr2.home.html](http://genome.jgi.doe.gov/Clagr2/Clagr2.home.html" \t "_blank)), as aligned to corresponding scaffolds of the *Xanthoria parietina* assembly (Xanpa1; <http://genome.jgi-psf.org/Xanpa1/Xanpa1.home.html>) in the DoE JGI Vista Browser Synteny tool (Grigoriev et al., 2011; Frazer et al., 2004). We ranked nearly two hundred conserved syntenic blocks according to the following criteria: 1) greatest sequence variability in non‑coding regions; 2) greatest sequence conservation in potential primer sites; 3) highest proportion of non-protein coding sequence; 4) lack of obvious linkage to other markers; and 5) potential for development of internal primers. We used a custom implementation of the BLAST‑P 2.3X tool (Deng et al. 2007) to rank the 40 most promissing regions according to shared homology (in conserved regions) with existing assemblies of meta‑genomic and meta‑transcriptomic data. These data were derived from low‑coverage sequencing of three *Peltigera*‑associated lichen species: *P. dolichorhiza* (Magain *et al.*, 2010), *P. membranacea* (Xavier et al., 2012)and *P. praetextata* (Hodkinson et al., 2014). For developing PCR-based markers, we selected 20 regions that shared the most homology in potential primer annealing sites (typically in flanking exons). Based on *Peltigera* data, if possible, we designed one to ten primers (including degeneracies) per Watson‑Crick strand, resulting in a total of two to twelve primer combinations per region.

We designed three markers, IGS1, IGS3 and IGS16 using our conserved synteny comparative genomic approach. Each of these markers comprises an intergenic region and two flanking gene parts were the primers were placed. IGS1 is located within a two-gene microsyntenic region on the first scaffold of the Clagr2 genome assembly (between base pairs 283,015 and 285,166) and on scaffold 19 of the Xanpa1 assembly (between base pairs 632,451 and 634,199). IGS3, is located within a microsyntenic region on the fourth scaffold of the Clagr2 genome assembly (between base pairs 508,539 and 513,282), and in the seventh scaffold of the Xanpa1 assembly (between base pairs 1,104,171 and 1,108,382). IGS16 is located within a microsyntenic region on scaffold 27 of the Clagr2 assembly (between base pairs 335,996 and 340,907) and on scaffold 1 of Xanpa1 (between base pairs 2,640,063-2,644,705). The sequences of the newly generated primers used for the amplification of the three new markers are shown in Table 1. These primers were tested on Peltigera samples from most sections of the genus and showed great success in amplifying the targeted loci.

|  |  |
| --- | --- |
| Primer name | Primer sequence (5’-3”) |
| IGS-1F | GCTGTCGGCGAAGAGCTGAA |
| IGS-1R\_B | CCATTTCTCCGCCGTTCTGGTA |
| IGS-3F\_A | GGAGACGTTGCTAATGCATT |
| IGS-3R\_B | CCGAAGTCCGCTCTGAAGACA |
| IGS-16Fout | GCGGAKGCGCAGATGATTTG |
| IGS-16Rmid1 | TGTGGCACGGTGAACACTAG |

## Taxon Sampling

Over 2000 specimens of *Peltigera* section *Polydactylon* (identified based on morphology) obtainedas loans from several herbaria world-wide (AMNH, B, BG, CGMS, CONN, DUKE, H, LG, MAF, MEXU, NSPM, NY, O, PTZ, QFA, UBC, UDBC, UGDA, UMEX, UPS) and various private collections, as well as collected during numerous field trips part of this study (Reunion Island in 2009; Norway, Canada: Québec, USA: North Carolina and Alaska in 2011; Russia, Peru and Brazil in 2012) were examined to select a set of representative specimens for DNA extraction and sequencing of the mycobiont and cyanobiont.

## DNA extraction and Sequencing

We extracted DNA from approximately 950 well-preserved lichen specimens lacking any visible symptoms of fungal infection following two extraction protocols: Cubero et al. (1999) or modified Zolan and Pukkila (1986) using a 2% sodium dodecyl sulphate (SDS) as the extraction buffer. We amplified the internal transcribed spacer (ITS) of the nuclear ribosomal tandem repeat of the mycobiont from about 950 lichen thalli representing a broad geographic and morphological diversity of the group, using the ITS1F (Gardes and Bruns 1993) and ITS4 (White et al. 1990) primers. The PCR conditions were as follows: 95°C for 5 min, followed by 25 cycles of 95°C for 45 sec, 52°C for 45 sec, and 72°C for 90sec, with a final extension of 72°C for 10 min. Based on these preliminary results, we further amplified 7 other loci on 164 specimens (94 for the Dolichorhizoid clade, 35 for the Scabrosoid clade, and 34 for the Polydactyloid clade) representing the diversity in ITS sequences. We also selected specimens with identical ITS haplotypes if they came from very distinct geographic regions (e.g. different continents) or exhibited very different morphology.

We amplified the nuclear ribosomal large subunit (LSU) using primers LR0R and LR7 (Vilgalys and Hester 1990), three protein-coding genes: RNA II polymerase largest subunit (*RPB1)* using primers RPB1-AF (Stiller and Hall 1997) and RPB1-CR (Matheny et al. 2002), elongation factor 2 region 1 (*EFT2*.1) using primers EFT2.1\_1F (Miadlikowska et al. 2014) and EFT2.1\_3R (Magain et al. 2015), and ß-tubulin using the reverse primer BT2B (Glass and Donaldson 1995) and the forward primer T1 (O’Donnell and Cigelnik 1997) or alternatively bt\_34F (O’Brien et al. 2009), and the three newly developed intergenic spacers, IGS1, IGS3 and IGS16 (primers are provided in Table 1). Amplification of RPB1 and LSU follows O'Brien et al. (2009) whereas the amplification of ß-tubulin, EFT2.1, IGS1, IGS3 and IGS16 were as follows: 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. All PCR amplicons were cleaned with ExoSAP (Affymetrix, Santa Clara, California, USA) following the manufacturer’s protocol.

Sequencing was carried out in 10 μL reactions using: 1 μL primer (10 μmol/L), 1 μL purified PCR product, 0.75 μL Big Dye (Big Dye Terminator Cycle sequencing kit, ABI PRISM version 3.1; Perkin-Elmer, Applied Biosystems, Foster City, California, USA), 3.25 μL Big Dye buffer, and 4 μL double-distilled water. Automated reaction clean-up and visualization was performed at the Duke Genome Sequencing and Analysis Core Facility of the Institute for Genome Sciences and Policies (for details see Gaya et al., 2012)

The list of specimens used for this study can be found in Supplementary Table S1.

## Alignment, model selection and partitioning

Sequences were edited using Sequencher version 4.9 (Gene Codes Corporation, Ann Arbor, Michigan) and aligned using MacClade v. 4.08 (Maddison and Maddison, 2005). In order to reduce the number of ambiguously aligned characters that must be excluded from phylogenetic analyses, single-locus and multi-locus separate datasets were assembled for the whole section, as well as for each of the three clades.

For each single-locus dataset the best model for nucleotide substitution was determined by MrModelTest v. 2.3 (Nylander, 2004) using the greedy algorithm and the BIC criterion on all models available. For all the concatenated datasets, we used PartitionFinder (Lanfear et al., 2012) to determine the best partition to use in subsequent phylogenetic analyses. The following 18 data subsets were pre-delimited: LSU, ITS1+ITS2, 5.8S, IGS1 (not included in dataset for Scabrosoid clade), IGS3, IGS16, and four subsets for each protein-coding locus, (each of the three codon positions, and the non-coding parts.

## Phylogenetic analyses

We generated single-locus phylogenetic trees for the section and for each of the three clades using RAxML v. 7.4.2 (Stamatakis et al. 2006) or alternatively RAxML v. 8.0.9 (Stamatakis, 2006; Stamatakis et al., 2008) as implemented on the CIPRES portal (Miller et al., 2010). Optimal tree and bootstrap searches were conducted with the rapid hill-climbing algorithm for 1000 replicates with GTR substitution model (Rodríguez et al., 1990) and gamma distribution parameter approximated with four categories in all analyses. Phylogenetic relationships that received bootstrap values of 70% and above were considered highly supported.. We also generated phylogenetic trees on the concatenated datasets, using the same settings as above, and the best schemes determined by PartitionFinder to partition the data.

We generated 50% consensus phylograms with MrBayes v. 3.2.2 (Huelsenbeck and Ronquist, 2001) on the CIPRES portal using the best BIC scheme determined by PartitionFinder to partition the data and determine the substitution models; completing 15 million generations for each clade, and 25 million generations for the whole Section; with 2 runs of 4 chains (3 cold chains and a heated one) each; sampling every 1000th generation; and discarding the 25% first trees as burn-in.

We generated chronograms for the Dolichorhizoid and Scabrosoid clades with BEAST v. 1.8 (Drummond and Rambaut, 2007) as implemented on the CIPRES portal by completing 50 million generations and discarding 20% of the trees as burn-in. For the concatenated datasets, we used the best BIC scheme determined by PartitionFinder to partition the data and determine the substitution models. For single-loci analyses on the Dolichorhizoid and the Scabrosoid clade, we applied the substitution models determined by MrModelTest. The concatenated and single-locus BEAST analyses on the Scabrosoid clade were performed using a strict molecular clock. For the Dolichorhizoid clade, the concatenated analysis, as well as the single-locus analysis on ß-tubulin were performed with a lognormal relaxed clock, while the remaining analyses were performed with a strict clock. Convergence of Bayesian results was explored using Tracer v. 1.5 (Rambaut and Drummond, 2003) and AWTY (Nylander et al., 2008) as implemented on the portal http://king2.scs.fsu.edu/CEBProjects/awty/.

### Pairwise distances

For the Scabrosoid and the Dolichorhizoid clades, pairwise-distances between ITS sequences were computed using PAUP v. 4.0a134 (Swofford, 2003).These distances were used to generate a heatmap using R (2012) package ggplot2 (Wickham, 2009).

### Species delimitation methods

### Structurama

Using Sequencher, for each individual we coded alleles represented in each locus sequenced (eight loci for Dolichorhizoid clade and seven loci for Scabrosoid clade) using 100% similarity as the criterion to collapse samples in a single allele. We ran Structurama (Huelsenbeck et al. 2011) 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, successively). We also tested the impact of each locus on species delimitation, by assembling datasets with one locus removed, resulting on eight 7-locus datasets for the Dolichorhizoid clade, and seven 6-locus datasets for the Scabrosoid clade. We ran four analyses on each of these datasets, completing one million generations, sampling every 1000th generation, with gamma shapes of 3, 8, 15 and 30 respectively, and gamma scale of 1 for the hyperprior on the expected numbers of populations.

### bGMYC

For the Scabrosoid clade, we ran a bGMYC (Reid and Carstens, 2012) analysis on chronograms derived from the seven loci (ITS, ß-tubulin, LSU, RPB1, EFT2.1, IGS3, IGS16). For the Dolichorhizoid clade, we ran a bGMYC analysis on the five most variable loci only: ITS, ß-tubulin, IGS1, IGS3, and IGS16. For each analysis, we selected 200 chronograms from the tree distribution resulting from the single-locus BEAST analyses. Each of the files contained 50.000 trees, of which we discarded the first 5000 trees as burn-in. We then selected one tree out of each 225th sample using R and the package APE (Paradis et al., 2004) to obtain a 200-trees file. We ran bGMYC on each set of 200 trees for 50,000 generations on each tree, 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.

### bPTP

For the the Scabrosoid and Dolichorhizoid clades, we ran bPTP (Zhang et al., 2013) as implemented on the website http://species.h-its.org/ on the best ML tree resulted from the RAxML analyses on the concatenated datasets. We completed 500,000 generations with a thinning value of 1000 and discarding the first 25% generations as burn-in.

### spedeSTEM

We ran spedeSTEM (Ence and Carstens, 2011) as implemented on the website http://spedestem.osu.edu, using the discovery method. Because spedeSTEM test the merging but don't split predefined species, we tested species corresponding to all the individual lineages delimited in Figure 3.

The program requires single-locus ultrametric trees as input. For the Scabrosoid clade, we used the seven single-locus chronograms generated with BEAST (see above). We also ran the analysis without the *RPB1* chronogram, because this locus is not variable enough in the Scabrosoid clade. We estimated the relative rates of each locus based on the substitution rates derived from phylogenetic analyses. We tested a wide range (0.0001, 0.001, 0.005, 0.01, 0.05, and 0.1) of theta values (parameter depending on the population size and the substitution rate) because good estimates for our group are not available. We also ran an analysis with specimens attributed to *P. neopolydactyla* 4 randomly split in two species in the traits file, to check if the method merges them into one unit, as a test for the adequacy of the input parameters.

Because single-locus chronograms were poorly resolved in the Dolichorhizoid clade, we generated ultrametric trees by transforming the best ML single-locus trees resulting from the RAxML analyses using non-parametric rate smoothing with TreeEdit (Rambaut and Charleston, 2002). We used the same approach as for the Scabrosoid clade for assigning a species trait to our samples. We also assigned samples of *P. hymenina* and *P. dissecta* (which have been shown to be conspecific in phylogenetic reconstructions and other species delimitation methods) to distinct species as a test for the adequacy of the parameters inputted. We tested the following theta values: 0.00001, 0.0001, 0.001, 0.005, 0.01, 0.05, 0.1, and 0.5.

### bPP

We ran bPP v. 2.2 (Yang & Rannala 2010) on the Scabrosoid and Dolichorhizoid clades using seven and eight single-locus alignments, respectively. We used the same species assignment as for spedeSTEM and a guide-tree reflecting the topology found in the MrBayes analyses on the concatenated datasets. We used the species delimitation algorithm, keeping all sites containing missing data. We estimated the relative rates of single loci based on substitution rates from the ML analyses.

For the Scabrosoid clade, for the tau prior, we set the gamma shape to 4 and the gamma scale to 100.

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 is 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 Dolichorhizoid clade, for the tau prior we set the gamma shape to 3 and the gamma scale to 100. 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 is 0.001, 0.01, 0.02 and 0.05 respectively.

We tested the adequacy of the parameters by splitting *P. hymenina* in two (*hymenina* and *dissecta* morphotypes).

### Final species delimitation and species tree

We used the consensus species delimitation resulted from all species delimitation methods as the species assignment for our specimens and generated species tree from eight (for the Dolichorhizoid clade) and seven (for Scabrosoid clade) loci using \*BEAST (Heled and Drummond, 2010). We ran the program for 50 million generations, sampling every 1000th generation and used lognormal relaxed clocks. For each locus, we attributed the nucleotide substitution model according to MrModelTest results (see Supplementary Table 2).

# Results and discussion

## Sequencing, alignment and concatenation

Within the Polydactyloid clade, for a total of 35 taxa, all of them are represented by at least three loci (100%), 30 taxa by four loci (85.7%), 24 taxa by five loci (68.6%), 15 taxa by six loci (42.9%), 9 taxa by seven loci (25.7%), and 3 taxa by eight loci (8.6%). Within the Scabrosoid clade, for a total of 35 taxa, all of them are represented by at least three loci (100%), 31 taxa by four loci (88.6%), 25 taxa by five loci (71.4%), 21 taxa by six loci (60%), and 14 taxa by seven loci (40%). We could not obtain IGS1 sequences for members of this clade. Within the Dolichorhizoid clade, for a total of 94 taxa, all of them are represented by at least four loci (100%), 87 taxa by five loci (92.6%), 75 taxa by six loci (79.8%), 67 taxa by seven loci (71.3%), and 46 taxa by eight loci (48.9%). The highest proportion of missing sequences occurs in the Polydactyloid clade whereas the Dolichorhizoid clade is represented by the most complete sequence data (Fig. 1.1).

Overall each specimen is represented by the ITS sequence and 90% of individuals have LSU and *RPB1* loci in all datasets. Polydactyloid clade contains the highest proportion of missing sequences for *EFT2*.1, IGS3 and IGS16. Dolichorhizoid clade has the best coverage of IGS1 locus, which is the least represented marker across the clades (in less than 50% of taxa) while *EFT2*.1, IGS3 and IGS16 were successfully sequenced for 65-75% of targeted taxa (Fig. 1.2).

### Comparison of length and variability of loci, including the IGS markers

For each targeted locus, we compared: (a) the total number of characters (the length of the longest sequence, excluding introns in the IGS markers); (b) the total number of characters included in the phylogenetic analyses (after the exclusion of ambiguous regions); (c) the total number of variable characters before the exclusion of ambiguous regions; (d) the total number of variable characters after the exclusion of ambiguous regions for the whole section and (e) for the Dolichorhizoid clade only. This information can be found in Table 2 and figure 1.3.

The largest amplicons among our targeted loci and after the removal of ambiguously aligned regions of the alignments were the LSU (1255bp), however this locus contained the least number of variable characters included in phylogenetic analyses (89 characters). Besides LSU, the three IGS markers were the longest (from 1058 bp to 815 bp), and overall the most variable ones, ranging from 452 variable characters in IGS1, 426 in IGS16 to 351 in IGS3, and therefore outperforming the level of genetic variation provided by the ITS (351 characters) even after the exclusion of ambiguously aligned regions (260 character in IGS1 versus 258 in the ITS). ß-tubulin and LSU were much less variable (230 and 210 variable characters, respectively), while EFT2.1 and RPB1 contained the lowest numbers of variable characters (149 and 114, respectively). The three protein-coding genes and the ITS were of similar total length (750-820bp) but the number of unambiguously aligned characters for the EFT2.1 and RPB1 (800 and 678bp respectively) was greater than for ITS and ß-tubulin (521 and 578 characters at the section level, respectively), which contain large regions that could not be unambiguously aligned, including introns. Following LSU, the locus with the least number of variable characters kept in phylogenetic analyses was RPB1 (95 characters) whereas EFT2.1 and ß-tubulin contributed equally (141 and 153 characters, respectively).

Among the three new markers, IGS16 has the highest number of variable characters included in the analyses at the section level (352 characters) followed by IGS1 (316 characters) and IGS3 (260 characters). At the level of the Dolichorhizoid clade only, IGS16 is still the one with the most variable characters kept in the analyses (184), followed this time by IGS1 (175) then IGS3 (152).

All other markers are significantly less variable: ITS, the most variable of the remaining loci, has 111 variable characters, followed by ß-tubulin (96 characters).

All three IGS markers are extremely useful at this taxonomic level, compared to the other loci. They are among the longest fragments amplified, with LSU, but have much more variable characters that can be kept for phylogenetic analyses. They are the three markers with the highest number of variable characters, at the section level, and more strikingly at the clade level, where ITS loses much of its variation. These markers are thus of great interest when working at this very low taxonomic level, at the specific or intraspecific level. ITS, ß-tubulin and to some extent EFT2.1 still have enough variable characters to be useful at this level, whereas RPB1, and especially LSU, have for long proven their use for studies at higher taxonomic levels, but are less helpful when working at the specific or intraspecific level, as in the present study.

## Phylogenetic reconstructions

The list of all the performed phylogenetic analyses along with the nucleotide substitution models determined by MrModelTest, as well as the partition and nucleotide substitution models determined by PartitionFinder for the concatenated datasets can be found in Supplementary Table 2. Single-locus topologies are available in the online supplementary material.

**1. Single-locus phylogenies**

Within the Polydactylon section, the Scabrosoid clade was revealed as a strongly supported monophyletic group in all single locus ML trees. The monophyly of the Polydactyloid clade received high support based on the analyses of the three IGS markers, ß-tubulin and *RPB1* and was weakly supported in EFT2.1 tree. In ITS and LSU trees, this clade is paraphyletic, however, without strong support. The Dolichorhizoid clade is monophyletic and includes the *scabrosella* group strongly supported in the RPB1, LSU, IGS1 and IGS3 trees but not in ITS phylogeny (without strong support). However, based on the ß-tubulin, EFT2.1 and IGS16 topologies, the placement of the *scabrosella* group is unresolved. No significant conflict was detected among the main clades within the section based on the single locus phylogenies.

Most putative species within the Scabrosoid clade represent well-defined and highly supported lineages. Two conflicting relationships include close relationship of *P. scabrosa* 1 with *P. neopolydactyla* 4 and *P. scabrosa* 2 in the *RPB1* phylogeny (highly supported), whereas LSU and IGS3 strongly support the affiliation of *P. scabrosa* 1 with *P. scabrosa* 4. Similarly, *P. scabrosa* 3 represents the first split in a clade containing *P. neopolydactyla* 5 and other species from the *scabrosa* group based on the ITS and IGS3 phylogenies, while EFT2.1 placed *P. neopolydactyla* 5 as the early diverged lineage in this clade.

A few cases of conflicting relationships among single locus topologies occurred in the Dolichorhizoid clade. For example, *P*. sp. 7a and *P.* sp. 7b in the *scabrosella* group are sister based on *RPB1* and IGS3 phylogenies whereas close relationship between *P. sp*. 7a and *P. scabrosella* was revealed by ß-tubulin. The IGS1 and IGS3 grouped the south-american species together with the *hymenina* group, and each of them are strongly supported clades in the ß-tubulin, IGS3 and IGS16 phylogenies. Overall phylogenetic relationships among the putative species in the Dolichorhizoid clade received a low support is the single locus phylogenies.

The most robust phylogenies for this clade were obtained based with the

IGS3, IGS16 and ß-tubulin loci.

The large amount of missing data in the Polydactyloid clade makes it difficult to test the discrepancies between the single loci.

The majority of conflicts detected among the single locus phylogenies may occur due to complex gene histories, but more likely because most of the loci are not very variable at this taxonomic level, and therefore include a small number of phylogenetically informative characters that are crucial for inferring relationships among individuals that diverged recently.

The topologies resulted from the new IGS markers were highly congruent with the remaining loci that have been commonly used in molecular systematic studies in lichen-forming fungi at the intrageneric and species levels. These three markers provided higher level of the resolution compared to the traditional loci, especially within the clades representing the most recent radiations, e.g., the *neopolydactyla/dolichorhiza/hymenina* group.

### 2. Multi-locus phylogenies

Figure 2 shows the 50% consensus trees derived from the MrBayes analyses for the Dolichorizoid and the Scabrosoid clades.

In the Scabrosoid clade, most putative species, resolved as monophyletic groups on long, well-supported branches, seem to perfectly fit the concept of "distinct evolutionary lineages" (sensu De Queiroz 1998).

The first split in the clade divides a group of 3 distinct lineages corresponding to *P. neopolydactyla* 6 and *P. neopolydactyla* 7 supported together, and *P. melanorrhiza*; from the rest of the group, referred to as the the *scabrosa* group.

In the *scabrosa* group, *P. scabrosa* 3 is the most basal lineage, and possibly composed of 2 distinct lineages (*P. scabrosa* 3a :P865-N1236 and *P. scabrosa* 3b: P1538). *P. neopolydactyla* 5 is basal to remaining species: *P. scabrosa* 4, *P. scabrosa* 1, *P. neopolydactyla* 4 and *P. scabrosa* 2, the two latter being grouped together with support. The only branch without support is the one holding *P. scabrosa* 1 with *P. neopolydactyla* 4 + *P. scabrosa* 2. In single gene phylogenies several loci place *P. scabrosa* 1 as sister to *P. scabrosa* 4, whereas others place it in the same position as in the concatenated analysis.

In the Polydactyloid clade, P*. polydactylon* from North America and Europe are in two well-supported distinct groups, possibly representing distinct evolutionary lineages. They appear as sister to P. sp. 10. P. nana 1 and P. nana 2, grouped together, are sister to the P. sp. 8 complex. P. sp. 11 and P. sp. 9 are basal. In the tree of the section, P. sp. 9 is the most basal species of the Polydactyloid clade, but without support (pp=0.86)

In the Dolichorhizoid clade, the *scabrosella* group is basal (pp=1) to the rest of the clade. Then the next split divides the *occidentalis* group (pp=1) from the *neopolydactyla/dolichorhiza/hymenina* group. In the *scabrosella* group, *P*. sp 7a and *P.* sp7b are more closely related (pp=1) than they are to *P. scabrosella* (there were conflicts about this relationships in the single gene topologies). In the *occidentalis* group*, P. occidentalis* and *P.* sp. 6 are more closely related than they are to *P.* sp. 12 .

In the *neopolydactyla/dolichorhiza/hymenina* group, the *hymenina* and South American groups are resolved together (pp=1), the *neopolydactyla* group is basal. In the *neopolydactyla* group, *P. neopolydactyla* 1 s.l. is basal (pp=1), *P neopolydactyla* 1b is sister to the rest of *P. neopolydactyla* 1 s. l. (pp=1). *P. neopolydactyla* 2b, *P. neopolydactyla* 3 and *P. pacifica* group together (pp=0.997), *P. neopolydactyla* 2a and *P. neopolydactyla* 2c also group together (pp=0.994). In the *hymenina* group, *P.* sp. 3, *P.* sp. 4, *P*. sp. 5 and *P. hawaiensis* group together with pp=0.94, this group is sister to *P. hymenina*. *P. dissecta* is nested inside *P. hymenina* with no structure to suggest that it might represent a distinct species, as already shown in Magain et al. (2015).

In the South American group, *P. dolichorhiza* is basal, sister to P1202 and P1596 ("*P. dolichorhiza* b"). The rest of the group is supported as monophyletic at pp=0.94, and forms a three-branches polytomy: *P. truculenta* (including *P. chilensis*, confirming the findings from Magain et al. 2015) on one branch, *P. dolichorhiza* 2 and *P.* sp. 1 supported together on the second branch, and the third branch composed of *P. pulverulenta* 1, *P. pulverulenta* 2, *P. pulverulenta* 3 (the three resolved together at pp=1); *P.* sp. 2a (grouped with them at pp=0.95); and *P.* sp. 2b resolved with them at pp=0.62.

For the Scabrosoid clade, the topology of the BEAST chronogram is in agreement with the MrBayes phylogram. In the chronogram of the Dolichorhizoid clade alone, the *occidentalis* group and the *scabrosella* group appear grouped together (without support, pp=0.52) as opposed to the *neopolydactyla/dolichorhiza/hymenina* group. This is due to a rooting issue, and is even the case with a lognormal relaxed clock. In the whole Section tree, the *scabrosella* group appears as basal to the *occidentalis* group and the *neopolydactya/dolichorhiza/hymenina* group, in agreement with the Mr Bayes phylograms. The rest of the topologies are congruent.

## Species delimitation: comparison of the methods`

### Structurama

**1) Impact of the gamma shape hyperprior and individual loci on the number of delimited species**

In the Scabrosoid clade, the number of species remains stable until the gamma shape value reaches 5. Higher values (e.g., 8) lead to splitting lineages and increase the number of delimited species, especially singletons (species represented by only one individual). We chose the gamma shape value of 3 for the final analysis. In the Dolichorhizoid clade, the number of species increases almost linearly with the increase of gamma shape hyperprior, but the number of singleton species goes up drastically when gamma shape reaches 8 and fluctuates above this value.

Analysis with a low gamma shape hyperprior (= 3) groups some potentially unrelated (well accepted and circumscribed morphologically and geographically) lineages, e.g., *P.* sp. 5 (N1534) with *P. pacifica*, and *P. neopolydactyla* 3 (P859) with a subset of *P. neopolydactyla* 1 and *P. hymenina* despite that these species do not share any allele. We selected the intermediary gamma shape value of 15 for the final analysis.

In the Scabrosoid clade, in general a single locus does not affect the number of delimited species in a meaningful way when gamma shape prior is low (= 3). Removal of each of the following loci: IGS3, ITS, EFT2.1 or LSU from the combined dataset resulted in the same nine species delimitation; the exclusion of IGS16 or ß-tubulin decreased the number of species by one (*P. scabrosa* 4 and *P. neopolydactyla* 5 collapsed); without *RPB1* the number of species increased to ten.

A similar pattern of fluctuation in the number of recognized species (from one less to two extra species) was obtained with the gamma shape value of 8. For higher gamma values, the removal of one locus increased the number of species regardless of the locus removed (for the gamma shape of 30, the number of species can increase by six ). In general, removing a single locus from the combined dataset leads to greater number of species in the Scabrosoid clade. It is very likely that having fewer loci decreases the chance that specimens share an allele at any locus and lowers the probability that they will cluster together.

In the Dolichorhizoid clade, the exclusion of one of the three IGS markers (or ITS/ ß-tubulin for low gamma shape values) usually slightly decreases the number of species, whereas removing *EFT2*.1, *RPB1* or especially LSU strongly increases the number of species retrieved. Overall, removing a variable locus results in less species, meaning that these loci tend to increase the number of species, whereas removing a less variable locus results in more species, meaning that these loci tend to decrease the number of species. It makes sense as in a less variable locus, the probability that samples share alleles is higher, so the probability that they will cluster together is higher too.

For the IGS regions, almost every individual is represented by a unique allele while for the LSU, *RPB1* and *EFT2*.1 many specimens share the same alleles (Fig. 3f). It is clear that the loci with the lower number of alleles within the sampled individuals (the left bar) are the ones who tend to reduce the number of species (Fig. 3e) whereas the ones with the highest number of alleles (the right bar) tend to increase the number of species (Fig. 3e).

Removing EFT2.1, RPB1 or LSU increases the number of species, but especially the number of singletons (Fig. 3g). For example, a broadly delimited paraphyletic species (corresponding to the grey zone in Figure 4), which members share one or several alleles for these three loci was reconstructed, regardless of the gamma shape value On the other hand IGS markers are too variable (almost every individual has a different allele), and therefore contribute to the overestimation of the number of species. However, their high variability makes them suitable loci for the phylogenetic reconstructions at the species level.

To improve Structurama performance on the *Polydactylon* section, more markers with an intermediary genetic variation (like ß-tubulin and ITS) are needed. In addition including more loci and more representatives from each putative species should increase the chance of specimens from an actual species to share alleles and cluster together.

2) **Species delimitation**

In the Scabrosoid clade, when including all loci, the only difference when rising the gamma shape from 3 to 8 is the split of *P. melanorrhiza* in two. Then from a gamma shape of 8 to 15, the split of *P. neopolydactyla* 5 in two.

At a gamma shape of 3, all lineages retrieved but one coincide with monophyletic groups and correspond to *P. scabrosa* 2, *P. scabrosa* 3a and *P. scabrosa* 3b (as two distinct species), *P. scabrosa* 4, *P. neopolydactyla* 5, *P. neopolydactyla* 6, *P. neopolydactyla* 7 and *P. melanorrhiza*. The only non-monophyletic group retrieved is the clustering of *P. neopolydactyla* 4 and *P. scabrosa* 1 together. These two species share the same allele for the locus *RPB1*, which exhibits a low variation in this group. When analyzing the dataset without *RPB1*, *P. scabrosa* 1 and *P. neopolydactyla* 4 appear as distinct species, the rest of the lineages delimited are the same. We therefore decided to implement the Structurama analysis without *RPB1* for our final consensus.

In our case, with few loci and haploid data, it seems that in most cases, when distinct lineages share an allele, they are clustered together.

Figure 4 shows the species delimitation by Structurama in the Dolichorhizoid clade, in function of four different gamma shapes.

Some of our putative species are always well-defined regardless on the gamma shape of the hyperprior, this is the case of *P. scabrosella*, *P.* sp. 7a, *P. occidentalis* (at gamma shape of 18, P3034 is however considered as a singleton), *P. pulverulenta* 2, *P. neopolydactyla* 1b, *P.* sp. 5 and *P. hawaiensis*.

On the one hand, some putative species are well defined with low gamma values, but are splitted in several species with higher gamma values, like *P.* sp. 6 (split in two with a gamma shape of 8), or *P.* sp. 7b (split in two with a gamma shape of 15).

On the other hand, some putative species are well-defined at high gamma values, but are clumped with other unrelated taxa at low gamma values (*P. hymenina*, *P*. sp. 1, *P. sp*. 2a, are well defined with a gamma shape of 8 and above, *P. neopolydactyla* 3, *P. pacifica* and *P*. sp. 5 at a gamma shape of 15 and above, *P. neopolydactyla* 2c at a gamma shape of 18 or above, but these species are part of non-monophyletic assemblages at lower gamma values).

If Structurama seems to perform well in some parts of our tree, the fact that it is not tree-based, and that specimens which share an allele will often be grouped together, result in a large paraphyletic species (in grey in figure 4) comprizing *P.* sp. 3, *P. neopolydactyla* 2a, a subset of *P. neopolydactyla* 1, *P. truculenta*, *P. dolichorhiza*, *P. dolichorhiza* b, *P. dolichorhiza* 2 and *P.* sp. 2b. This paraphyletic species has no geographical or morphological significance. Therefore, Structurama alone can't infer species boundaries in our whole group.

Some singleton species retrieved by Structurama are credible based on phylogeny, morphology and geography, as P325 (*P. neopolydactyla* 1b), P859 (*P. neopolydactyla* 3), P1236 (*P. hawaiensis*), N1534 (*P.* sp4), N1545 (*P*. sp5), P3304 (*P*. sp. 12). On the contrary, we consider that other singletons such as P1291, N1929, P3032, P1662 (parts of *P. neopolydactyla* 2 s.l.), the pair P640-645 (parts of *P. neopolydactyla* 1 s.l.) more likely represent samples from species with intraspecific allelic variation, that Structurama fails to cluster together.

At the low gamma shape of 3, in the Dolichorhizoid clade, some samples even cluster while why don't even share an allele (e.g., the very distantly related *P.* sp. 7a and *P.* sp. 2a cluster together). The value of the gamma shape of the hyperprior is thus very important, a value too low will result in paraphyletic assemblages of unrelated taxa, whereas a value too high will result in splitting some species in several singletons.

Removing IGS3, IGS16 don't affect the species delimitation, suggesting that their high variability is not informative in this analysis.

## bGMYC and bPTP

Posterior probabilities to support a species in bGMYC are usually relatively low, because it takes the tree uncertainties into account. Indeed, in the present case, a posterior probability of 0.5 means that half of the 200 trees tested support the delimitation of a certain species, which is considerable, especially considering all the other possible combinations.

We will consider that a species is delimited by bGMYC when the probability of grouping only these samples together is higher than all the probabilities of other groupings including at least one of these samples.

Missing samples differ from one locus to the other but if the bGMYC analysis on a locus assigns A, B and C to a species, and the analysis on another locus where C is missing assigns A and B to a species, we will consider that bGMYC is congruent on these two loci.

Supplementary Table S4 contains the posterior probabilities for each species delimited by bGMYC.

1) Scabrosoid clade

Figure 4 shows the species delimitation by bGMYC in the Scabrosoid clade.

bGMYC on each locus agrees with the final species delimitation from the Structurama analysis (performed without *RPB1*), at the exception of the analysis on ITS, where *P. scabrosa* 3a and *P. scabrosa* 3b are assigned together as a single species.

The delimitation from bPTP on the best ML tree from the RAxML analysis on the concatenated dataset gives a similar result, at the exception of *P. scabrosa* 3a, which is split in two with a probability of 0.5 and *P. neopolydactyla* 5, which is also split in two. This is a very surprising result, as the topology of the tree does not support such delimitation (Figure 2).

By comparison, when applying bPTP to the 50% bayesian consensus tree, the species delimitation is the same as in bGMYC and Structurama.

2) Dolichorhizoid clade

In the Dolichorhizoid clade, we tested bGMYC on the 5 most variable loci: ITS, ß-tubulin, IGS1, IGS3 and IGS16.

Figure 6 shows the results of the species delimitation by bGMYC on each locus, as well as the species delimitation by bPTP.

Some species are retrieved by the analyses on every locus: it is the case of the two singleton species P. neopolydactyla 3 (pp varying from 0.19 to 0.86 in bGMYC, pp=1 in bPTP) and P. sp. 12 (pp from 0.53 to 0.84 in bGMYC, pp=1 in bPTP). P. sp. 6 is also retrieved as a species according to all loci (pp ranging from 0.56 to 0.81, bPTP pp=0.86), as well as P. pulverulenta 2 is supported as a species by all loci (pp ranging from 0.25 to 0.55, bPTP pp=0.64).

P. occidentalis is also retrieved as a species by all bGMYC analyses (pp from 0.3 to 0.51) but not by bPTP.

P. sp. 2a (P1555 and P1570) is supported as a species by all loci (pp 0.28-0.59), but in ITS, P907 is included in the species (pp=0.38).

Several species are supported by all loci but 1: it is the case of P. pulverulenta 1 (pp from 0.28 to 0.56, bPTP pp=0.75 but not retrieved in IGS1); P. pulverulenta 3 (pp from 0.35 to 0.54 and bPTP pp=0.65, but retrieved in IGS16); P. pacifica (pp from 0.27-0.47, bPTP pp=0.88 but not retrieved in IGS3); P. dolichorhiza (pp from 0.18-0.46, bPTP pp=0.62, not retrieved in IGS16); P. neopolydactyla 3 (pp varying from 0.19 to 0.86, not retrieved in IGS3).

Among the disagreements between loci, P1202 and P1596 (P. dolichorhiza b) are part of a same species in ITS and ß-tubulin, but separate species in IGS3 and IGS16.

Similarly, P. scabrosella, P. sp. 7a and P. sp. 7b are merged in a single species by 4 loci (with pp from 0.52-0.8) but ß-tubulin and bPTP consider them as 3 distinct species, as did Structurama .

There are a lot of uncertainties in the species delimitation in the group of P. neopolydactyla 1 s.l. and P. neopolydactyla 2 s.l. (see figure 7).

For instance, *P. neopolydactyla* 2a and *P. neopolydactyla* 2c are grouped together in ITS, but not in the other loci. *P. neopolydactyla* 1 is delimited without *P. neopolydactyla* 1b in ITS and IGS1, but with *P. neopolydactyla* 1b in IGS3 and ß-tubulin.

*P. truculenta* is also problematic, it appears as a distinct species in ITS and IGS1 but is grouped with *P.* sp1 in ß-tubulin**,** part of a non-resolved group in IGS3, and is divided in many singletons in IGS16 as well as in the bPTP analysis. It could be due to the lack of resolution in single locus trees, especially because there is not much variation in the South American group. It could also reveal that speciation is under process in this little group that has a high degree of haplotype and morphological diversity.

*P.* sp. 1 is defined as a species in ITS, IGS16 and bPTP, but is grouped with *P. truculenta* in ß,-tubulin and with *P. dolichorhiza* 2 in IGS1, and part of a non-resolved group in IGS3. This is probably linked with the lack of resolution in single loci in the South American group due to the very recent radiation.

The situations where bGMYC performs poorly are in most cases due to uncertainties in the trees analyzed. For instance in ITS, most of the characters segregating *P. hymenina*, *P.* sp. 1 and *P. dolichorhiza* 2 are excluded, resulting in a lack of resolution in the tree, and bGMYC group them together, whereas these species are resolved as distinct in the bGMYC analyses on the other loci.

bGMYC gives mostly congruent results in the majority of cases, but shows differences depending on the loci, so delimiting species based on GMYC or bGMYC on a single locus can be inaccurate, as we haven't found two loci giving the exact same results.

Moreover, single gene topologies don't always match the topology of the "real" species tree, so GMYC or bGMYC must be used with caution, even if in most cases it gave us good and congruent results. Applying bGMYC to several loci and establishing a consensus might be a more effective approach.

Among the advantages of bGMYC, it can be used on a single locus, and therefore is more cost-effective than multi-loci methods. It is tree-based so it will always return monophyletic species.

However, it assumes that all the transitions from interspecific to intraspecific events took place at the same time, whereas it is not always the case (the example of a rapid radiation) and that all the species have similar patterns of intraspecific variation. A multi-threshold implementation of the GMYC model exists, where several transitions can be implemented, but this model did not improve the species delimitation results in our group.

Different results from a locus to another might be explained by different gene histories, but more likely in our case by the lack of resolution in the single-locus topologies, due to their lack of variation. The more variable the locus is, the best it is for this method (as long as there is no saturation and that it can be used to accurately reconstruct the phylogeny).

bPTP seems to perform well in most cases and is usually congruent with bGMYC. Its advantage is that it can be used on multi-loci phylograms, using the best resolution available from the tree provided. However, in several cases in our study, it split species that seemed to represent homogeneous, well-supported lineages in many singletons. It is the case for *P. occidentalis*, *P. neopolydactyla* 1, *P. neopolydactyla* 2 s.l. and *P. truculenta*, which were split in several singletons. It seems that even in what appears to be species, if one tip is a little longer, it has a high probability to be considered as a distinct species. This must be due to the model, with the probability of being part of a distinct species following a Yule process, depending on the number of substitutions, i.e. the branch lengths.

## spedeSTEM

1. Scabrosoid clade

1.1) Without RPB1

When testing the species assignment with *P. neopolydactyla* 4 split in two random sets of 5 samples each, the two parts of *P. neopolydactyla* 4 appears as a single species for theta values of 0.02 and above. For these values, *P. scabrosa* 3a and *P. scabrosa* 3b are merged as a single species. Other lineages are considered as distinct species, even for high values of theta. The species delimitation is the same with the species assignment following the lineages represented in figure 3.

1.2) With RPB1

Testing the species assignment with *P. neopolydactyla* 4 split in two random sets of 5 samples each, *P. neopolydactyla* 4 appears as a single species for values of theta of 0.035 and above. At these theta values, all the species are supported as distinct lineages.

With these values of theta and the species assignment following figure 3, all species tested are supported as distinct lineages.

There is thus a discrepancy when testing spedeSTEM with or without RPB1, on whether *P. scabrosa* 3a and *P. scabrosa* 3b represent one or two distinct species.

For the final consensus, we selected the results from the analysis without RPB1 and a theta value of 0.02

2. Dolichorhizoid clade

The number of species supported by spedeSTEM varies quite much depending on the theta value, from 12 species with theta=0.5; 4 species with theta = 0.00001 or 0.000001; to 17 to 26 species for theta values between 0.0001 and 0.1. Actually, theta values of 0.001, 0.005, 0.01, 0.02 return the same species delimitation, that we will keep for the final consensus.

We selected a theta value of 0.02 (the same value as the one we selected for the Scabrosoid clade). This species delimitation merges *P. hymenina* and *P. dissecta*, which was our test for the adequacy of the parameters as these two OTUs are conspecific based on phylogenetic reconstructions. It also merges *P. scabrosella*, *P.* sp. 7a and *P.* sp. 7b; and *P.* sp. 1 and *P.* sp. 1b (as most methods). *P. dolichorhiza* is merged with *P. dolichorhiza* b and also with *P. truculenta*. The merging of *P. dolichorhiza* with *P. truculenta* is the only result that is in disagreement with most methods tested as well as with the phylogenetic reconstructions.

### bPP

In the Scabrosoid clade, when running bPP with the species assignment where *P. neopolydactyla* 4 and *P. scabrosa* 2 are randomly split in two groups, the two halves of each species are merged, suggesting that the method performs well. In this analysis and in all other analyses regardless of the mean of the theta prior, every lineage is supported as a distinct species, including *P. scabrosa* 3a and *P. scabrosa* 3b, which are supported as distinct species with pp=0.97, all other species are supported as distinct with pp=1.

In the Dolichorhizoid clade, when the mean of the theta prior equals 0.001, *P. dissecta* and *P. hymenina* appear as distinct species, suggesting that the priors are not adequate.

When the mean of the theta prior equals 0.01 and 0.02, *P. hymenina* and *P. dissecta* are merged in a single species, whereas all the other species are supported as distinct lineages.

When the mean of the theta prior equals 0.05, *P. scabrosella*, *P*. sp. 7a and *P*. sp. 7b are merged, but *P. hymenina* and *P. dissecta* are considered as distinct species.

We decided to select the analysis with the mean of the theta prior equal to 0.02, as for the other analyses.

### Molecular distances: existence of barcoding gap?

Figure 7 shows the heatmaps summarizing the pairwise differences between ITS sequences within each clade.

In the Scabrosoid clade, no real difference can be seen between *P. scabrosa* 3a and *P. scabrosa* 3b (otherwise supported by most methods as two distinct species).

For the other species, the intraspecific variation is way smaller (light colors inside the squares) than the interspecific variation (darker colors outside the squares

In the Scabrosoid clade, it therefore seems that a barcoding gap approach could be implemented.

However, it seems difficult to apply the concept of barcoding gap in the Dolichorhizoid clade. If some species clearly appear to have lower intraspecific variation than interspecific variations with other species (e.g., *P. occidentalis*, and to some extent *P. hymenina*), in other groups such as in the South-American group or in *P. neopolydactyla* 2. s.l. we can't see a difference between intraspecific variation inside putative species and interspecific variation.

In the *neopolydactyla/dolichorhiza/hymenina* group, there is clearly no such thing as a barcoding gap between ITS intraspecific and interspecific variation, probably due to the fact that the different species diverged too recently.

**Comparison of the species delimiting methods**

The delimitation by each method and the final consensus are shown in Figure 8.

In our study, most methods performed well on the Scabrosoid clade, which shows that they all perform well when the species have split for large amounts of time, and are well-separated (based on the topology of the phylogenetic tree and the ITS distances between species).

In the Dolichorhizoid clade, Structurama doesn't seem to perform well. It is probably because we don't include enough taxa per putative species.

We would also probably need more loci, especially loci with an intermediary level of variation, to obtain better results. Further, this approach is not based on phylogenetic data and results in the delimitation of non-monophyletic species, unlike the other methods tested. However, in the case of the Scabrosoid clade, and in parts of the Dolichorhizoid clade, species are accurately delimited (based on the phylogenetic reconstructions and the other species delimitation methods). When Structurama and tree-based methods are congruent, it is a strong indication that our putative species are genuine species, because different assumptions led to the same delimitation.

bGMYC performed well in most cases, even if it is sometimes negatively impacted by the lack of resolution of single-loci trees, when there is not enough variation in the locus. However, bGMYC can be a very powerful tool if implemented on very variable markers that are congruent with the real species tree.

bPTP also performed well and has the advantage that it can be ran on multi-locus phylograms, allowing to use a tree with all the the resolution provided by the different loci. However, it apparently has a tendency to over-split some species, when branches inside species are short.

These two methods are not requiring big computational power, and just take a tree as input, which make them very inexpensive.

spedeSTEM takes gene trees as input, therefore uncertainty in the single-locus topologies cannot be taken into account. However, the species tree is inferred by the program, so relationships between species can vary.

It is the opposite approach in bPP: the alignments are provided, and therefore single-locus topologies are not fixed; but a guide-tree is provided, and the relationships between species have thus to be predetermined.

In our case, it seems that spedeSTEM has a higher tendency to merge species, whereas bPP keeps them separate.

These two methods require estimates of theta, which is problematic in many cases including ours, because neither the estimation of the population sizes nor the accurate estimates of the substitution rates exist for lichen-forming fungi. However, testing a variety of theta values allowed us to test these methods.

Overall, bPP gave estimates that better match our species concepts based on morphology and geographic distributions, and seem to be less sensitive to the parameters it takes as input.

The advantage of these two methods is that they take conflicts between different loci in consideration whereas bPTP and bGMYC work either on a single tree or on trees from the same locus, respectively.

However, more loci would probably be needed to provide more information, and so that the impact of a single locus be smaller. But the more loci used, the more expensive and computationally intensive these methods are.

## Comparison of the species tree and concatenated 8-locus tree

Figure 9 shows the species trees generated by \*BEAST .

In the Scabrosoid clade, the species tree has the same topology as the concatenated tree, except that *P. scabrosa* 1 is grouped with *P. scabrosa* 4 in the species tree (pp=0.49), whereas it is grouped with *P. scabrosa* 2 and *P. neopolydactyla* 4 in the concatenated analyses (for instance, pp=0.7 in the MrBayes analysis). However, none of these relationships are supported, and this emphasizes the uncertainty about this relationship.

In the Dolichorhizoid clade, the topology is globally the same. There is one difference in the *P. neopolydactyla* 2. s.l. group, where *P. neopolydactyla* 2b is grouped with *P. neopolydactyla* 2a and *P. neopolydactyla* 2c in the species tree but without support (pp=0.36). These three lineages are then grouped with *P. pacifica* (pp=0.54) and with *P. neopolydactyla* 3 (pp=0.91).

In the concatenated analysis, *P. neopolydactyla* 2b is grouped with *P. neopolydactyla* 3 (pp=0.84) then with *P. pacifica* (pp=0.99), whereas *P. neopolydactyla* 2a and *P. neopolydactyla* 2c are grouped together (pp=0.99).

This conflict actually mostly concerns the position of *P. neopolydactyla* 2b, whereas both methods agree to group these 5 species together.

This difference might also be caused by the ITS sequence of P859 (*P. neopolydactyla* 3) which places *P. neopolydactyla* 3 at the base of the *neopolydactyla/dolichorhiza/hymenina* group in the ITS phylogeny, because it is very different (it has many unique substitutions), unlike other loci which place *P. neopolydactyla* 3 as closely related to *P. neopolydactyla* 2. s.l. The influence of this different topology in the concatenated analysis might not influence the final topology, because many characters, especially in the IGS markers, support the alternative topology, whereas single-locus topologies have a bigger influence in the species tree.

The *hymenina* group is well supported in both analyses, but *P. hawaiensis* is grouped with *P. hymenina* in the species tree whereas it is grouped with *P. sp*. 3, *P.* sp. 4 and *P.* sp. 5 in the concatenated analysis. However none of these relationships are supported (pp=0.38 in the species tree, pp=0.94 in the Bayesian analysis).

Finally in the South-American group, the topologies are congruent, which is remarkable given the recent radiation and the lack of resolution and amount of short branches in the single locus trees. The only difference is the position of *P.* sp. 2a and *P.* sp. 2b, grouping with *P. pulverulenta* 1, 2 and 3 in the concatenated analysis, and with *P.* sp. 1 and *P. dolichorhiza* 2 in the species tree, but none of these relationships are supported (pp=0.33 in the species tree, pp=0.62 in the Bayesian concatenated tree).

In general, posterior probabilities are lower in the species tree, which makes sense because it takes the uncertainty of the conflicting single-locus topologies into account, whereas in concatenated analyses, the tendency is that many characters from one or few loci supporting one relationship will mask the impact of the conflicting characters in the other loci.

Examples include the grouping of *P*. sp. 7a and *P.* sp. 7b (pp=0.75 in the species tree, pp=1 in the MrBayes tree), *P. occidentalis* with *P.* sp. 6 (pp=0.9 in the species tree, pp=1 in the concatenated dataset), *P. pulverulenta* 1 with *P. pulverulenta* 2 (pp=0.78 in the species tree vs pp=1 in the concatenated analysis). For the two latter cases, conflicts came from only one locus (see the results about the single locus topologies), but still significantly affect the posterior probabilities.

## Consensus on species delimitation

Influence of the *Nostoc* on the phenotype and on species delimitation

We noticed that all specimens associating with the *Nostoc* phylogroup VIIa had a typical emerald green color when wet, especially when growing in their natural habitat. Because some species (*P. neopolydactyla* 1, *P. occidentalis*, *P. scabrosa* 1, *P. scabrosa* 4) always associate with this phylogroup in the panboreal zone, and other morphologically similar species (*P. neopolydactyla* 2, *P. neopolydactyla* 4, *P. scabrosa* 2) never associate with it, the identity of the *Nostoc* phylogroup can be very useful to identify the species, especially in the field or when no molecular data is available.

More detailed study of the influence of other *Nostoc* phylogroups on the phenotype of the thallus should be conducted, to determine if this character could be used for the identification of a wide variety of species.

As many species have shown a strong specificity towards the *Nostoc* phylogroup they associate with, this information might also be helpful as a criterion for species delimitation. For instance, methods performed poorly to delimit species in the group of *P. neopolydactyla* 2 s.l., however *P. neopolydactyla* 2a and *P. neopolydactyla* 2c were always found associating with *Nostoc* phylogroup XIII, whereas *P. neopolydactyla* 2b was always found with *Nostoc* phylogroup X. This different pattern of association might be a good indication that they actually belong to different species, even if it could also be specialization to different ecological conditions. The fact that the sympatric and morphologically similar pair *P. neopolydactyla* 1 and *P. neopolydactyla* 2a, or *P. scabrosa* 1 and *P. scabrosa* 2 actually belong to different species is consolidated by the fact that they never share the same *Nostoc* phylogroup even in the same localities.

**Species delimitation, geographic ranges, *Nostoc* specificity and morphological variation in the Scabrosoid clade**

The delimitation of species in the Scabrosoid clade was congruent in most methods, samples are assigned to species as follow:

- P1798 and P515 as *P. melanorrhiza*, one rare species endemics to the Azores. This species is easy to recognize morphologically.

- P1231 and P3051 as *P. neopolydactyla* 6, a species endemics to the Pacific Northwest region of North America. It is not a strict specialist, but shows affinity to two *Nostoc* phylogroups (XIII and XVII, for the numbering of phylogroups, see Magain et al. 2015). It corresponds to very specific morphotypes of *P. neopolydactyla* s.l. and is easy to recognize morphologically.

- P3010 as *P. neopolydactyla* 7, a rare species only known from one specimen from Japan.

- P1228, P1232 and P1257 as *P. neopolydactyla* 5, a species endemics to the Pacific Northwest region of North America. It is highly specialized, always found with *Nostoc* phylogroup XIb. It corresponds to very specific morphotypes of *P. neopolydactyla* s.l. and is easy to recognize morphologically.

- P312, P315 and P549, as *P. scabrosa* 4, a rare species known from Québec and Norway, restricted to boreal zones. It is highly specialized, always associating with *Nostoc* phylogroup VIIa.

- P97, P550, P1210, P1250 and P1539 as *P. scabrosa* 1, a widespread species from the panboreal zone. Specimens included in this study come from Québec, Alberta, Norway, British Columbia and Siberia, Russia. It is highly specialized, always associating with *Nostoc* phylogroup VIIa.

- P107, P113, P830, P1209, P1255 as *P. scabrosa* 2, a widespread species from the panboreal zone. Specimens included in this study also come from Québec, Norway, British Columbia and Siberia, Russia. It is not a strict specialist, but shows high selectivity towards *Nostoc* phylogroup XIa, even if it has also been found with phylogroup VIId.

- P302, P321, P506, P669, P811, P1212, P1537, P1668, P3024, P3027 as *P. neopolydactyla* 4, a widespread panboreal species. Specimens included in this study come from Québec, Norway, Michigan, Siberia, Eastern Russia and Japan. It is not a strict specialist, but shows very high specificity towards *Nostoc* phylogroup IV and XIa. It morphologically resembles *P. neopolydactyla* 1 s.l. and *P. neopolydactyla* 2 s.l.

The only open question concerns *P. scabrosa* 3a and *P. scabrosa* 3b: several methods considered these two entities as one species (bGMYC on ITS, spedeSTEM on 6 loci), others as two species (Structurama, , bGMYC on the other loci, bPTP, spedeSTEM on 7 loci, bPP). Most methods placed them in two distinct lineages, and it makes sense when looking at the concatenated phylogenetic analyses. However, these species are very rare and we only had 3 samples (2 of *P. scabrosa* 3a and one of *P. scabrosa* 3b, respectively) included in the present study. We suggest it might represent two distinct species and we included two lineages in the species tree but more material should be examined before the final conclusion is made.

**Identification of scabrosa morphotypes.**

Recognizing *P. scabrosa* 1, *P. scabrosa* 2, *P. scabrosa* 3 and *P. scabrosa* 4 is very difficult based on traditional morphological characters alone. *P. scabrosa* 3 is very rare and more material should be studied before drawing conclusions.

*P. scabrosa* 1, *P. scabrosa* 2 and *P. scabrosa* 4 seem to slightly differ in their vein patterns. More interestingly, the fact that *P. scabrosa* 1 and *P. scabrosa* 4 always associate with *Nostoc* phylogroup VIIa make them easy to segregate from *P. scabrosa* 2, as this phylogroup gives a specific emerald green color to the thallus whereas specimens of *P. scabrosa* 2 exhibit a more classical brownish/grayish color. However, specimens with strong scabrosity can be problematic as it gives a more brownish color to all the thalli.

**2. Species delimitation in the Dolichorhizoid clade**

In the Dolichorhizoid clade, species delimitation is easy in some groups:

In the *occidentalis* group, *P. occidentalis*, *P.* sp. 6 and *P.* sp. 12 are three distinct species. Even if some methods splitted *P. occidentalis* (bPTP) or *P.* sp. 6 (Structurama) in several singletons, we believe that this is more a problem with the datasets and the methods, as these species are well-defined, based on the phylogeny and the morphology, and don't appear to include cryptic variation, based on the sequences.

*P. occidentalis* has a stable morphology and is easy to identify. It always associates with the *Nostoc* phylogroup VIIa in the boreal zone, and has a panboreal distribution, and it can also very rarely be found in the Appalachians, where it associates with *Nostoc* phylogroup VIIb. In our study, specimens from Norway, Québec, Alaska, Japan and North Carolina were included.

P936, P1650 and P1734 belong to *P.* sp. 6, which is only found in tropical Central and South America, and doesn't seem to have specificity in its association with *Nostoc*. Our study included specimens from Peru, Colombia and Honduras

P3304 belong to *P.* sp. 12, which is a very rare species, only known from one specimen from Japan.

In the *hymenina* group, *P. hymenina*, *P. hawaiensis*, *P.* sp. 3, *P.* sp. 4 and *P.* sp. 5 are supported as distinct species by most methods. *P. dissecta* appears as conspecific to *P. hymenina*, despite his unique ITS haplotype and its very different morphology.

P604, P605, P607 and P1530 represent *P*. sp. 3, only known from, but widespread in Australia and New Zealand, where it was usually identified as *P. dolichorhiza*, *P. polydactylon* or *P. nana*. It has an important ITS haplotype diversity and morphological variation.

P1236, *P. "hawaiensis"* is only known from Hawaii. *P.* sp. 4 (N1534 in this study) and *P.* sp. 5 (N1545) are only known from Papua New Guinea, where they were described as variants of *P. dolichorhiza* in Sérusiaux et al. (2009).

P80, P430, P516, P539, P1229, P1799, P1903 belong to *P. hymenina*, which has a wide distribution, but is only present in Northern regions, in North America, Europe and Atlantic Islands; and was never found in Pacific Islands, where the other species of this group can be found. It is easy to recognize morphologically despite a quite high degree of variation in the lobes size and shape, but is not specialized to a single *Nostoc* phylogroup, even if it is often found with phylogroup XVI. Specimens from our study came from the Pacific Northwest of North America, Newfoundland (Canada), Iceland, Norway, Tenerife and the Azores.

In the *neopolydactyla* group, *P. pacifica*, composed of P443 and P1243, might be the only non-problematic species. It was delimited as a species by almost every method. It is easy to recognize morphologically, endemic to the Pacific Northwest of North America, and is specialized toward *Nostoc* phylogroup XIII.

*P. neopolydactyla* 1b (P325) seems to be a distinct species from *P. neopolydactyla* 1 (N1939, P309, P411, P640, P645, P845, P1252, .P3060).

*P. neopolydactyla* 1b is only known from one specimen from Peru whereas *P. neopolydactyla* 1 is only found in Northern regions, including the Appalachians, Arizona, and boreal zones in Norway, Québec and Russia. In the boreal zone, it always associates with *Nostoc* phylogroup VIIa, whereas it was found with phylotype VIIb in the Appalachians and VIIc in Arizona.

P859, *P. neopolydactyla* 3 is also considered as a distinct species by most methods. It is known from Yunnan (China) and Vietnam.

*P. neopolydactyla* 2 s.l. is a very complicated group. It is mainly composed of three lineages: *P. neopolydactyla* 2a, 2b and 2c, which are considered as distinct species in several methods, unresolved in others, rarely grouped together. Specimens such as P1662 were also considered as distinct singleton species in some methods.

It seems that *P. neopolydactyla* 2b (N1929, P1291, P1667, P3032), known mainly from Yunnan (China) and Japan and associating with *Nostoc* phylogroup X would be a distinct lineage according to most methods, but it is also split in several species by other methods.

*P. neopolydactyla* 2c (P1659, P3001, P3009), known mainly from Northern Japan and Russia, and *P. neopolydactyla* 2a (P384, P390, P3069, P1662), widespread in the boreal zone, including Norway, Canada, Russia, Japan, but also moutainous regions from France, for instance, both show specificity towards phylotype XIII. If they are distinct species, they would sympatric in Japan.

The *scabrosella* group (composed of *P. scabrosella*: P536 and P619, *P.* sp. 7a: N1666, N1674 and 3055; *P.* sp. 7b: P1660 and P1672) is also problematic. Some methods assigned them to a single species, whereas other assigned them to three distinct species. They all associate with *Nostoc* phylogroup XIa. However, given that *P. scabrosella* have a very different morphology (small scabrid lobes without veins, *P.* sp. 7a and 7b have wide glabrous lobes with veins) and thalli from *P.* sp. 7a and *P.* sp. 7b also differ considerably from each other. Moreover, they have distinct geographical distributions. *P. scabrosella* occurs in boreal zones of Scandinavia and Atlantic coast of North America, *P.* sp. 7a is endemic to the Pacific Northwest of America, whereas *P.* sp. 7b have only been found in Hokkaido. Moreover, if their haplotypes are always genetically very similar, they are never identical from one species to the other. Therefore, allelic exclusivity, morphological differences and distinct geographic ranges make us consider them as three distinct species.

In the South-American group, finally, we didn't find any species with specialization in their *Nostoc* selection.

Most methods agreed to distinguish *P. pulverulenta* 1 (P890, P897, P901, P938, P945, P953), *P. pulverulenta* 2 (P900 and P1521) and *P. pulverulenta* 3 (P1522 and P1525) as three distinct lineages, while they are apparently sympatric (*P. pulverulenta* 1 is widespread, found in Mexico, Guatemala, Colombia, Bolivia, Brazil ; whereas the two other lineages are rare, and were only found in Colombia. They cannot be recognized morphologically and they might represent sibling species (sensu Steyskal 1972), closely related species where divergence and genetic isolation was not correlated with significant change in morphology

Most methods also agreed to recognize the widespread *P. dolichorhiza* (N789, N999, N1942, P348, P28, P879, P893, P1551) found in Central and South America, from Mexico to Brazil, including Colombia, Bolivia, Galapagos, as well as in Africa, including Rwanda, South Africa, Madagascar and Reunion Island), and the rare *P.* sp. 1 (P885, P886, P909, from Colombia and Bolivia), *P.* sp. 2a (P1555, P1570 from Colombia), *P.* sp. 2b (P1557, P1561 from Brazil), and *P. dolichorhiza* 2 (P1567, P1575 from Brazil). These are usually very distinct morphologically. The problem lies in the resolution of the phylogenetic tree, as they originated from a very recent radiation, and further testing should be investigated, to determine if they are completely reproductively isolated.

The status of *P. truculenta* (including *P. chilensis*) known from the Neantarctic parts of Chile and Argentina as well as several remote islands (e.g., Kerguelen islands), and exhibit significant morphological plasticity, should be further assessed.

The status of the very rare P1202, P1596 ("*P. dolichorhiza* b") on whether they are one or two species, or if they represent genetically distinct variants of *P. dolichorhiza*; and P907 ("*P*. sp. 2ab" from Colombia) as whether it belongs with *P.* sp. 2a (P1555 and P1570 from Brazil) are open questions.

**Morphotypes in *P. neopolydactyla* s.l.**

*P. neopolydactyla* 5 and *P. neopolydactyla* 6 only occur in the Pacific Northwest and are easy to identify morphologically.

In the boreal zone, *P. neopolydactyla* 1, *P. neopolydactyla* 2a and *P. neopolydactyla* 4 are common and frequently occur in the same localities. They are morphologically similar, but *P. neopolydactyla* 1 can be easily identified as it associates with *Nostoc* phylogroup VIIa, which gives a specific emerald green color to the thallus. *P. neopolydactyla* 2a and *P. neopolydactyla* 4 are very difficult to distinguish, which is interesting because they belong to different clades, but they slightly differ in vein patterns, rhizine lenghts and thickness of the thallus. They can further be easily identified based on their TLC profiles and they never associate with the same *Nostoc* phylogroup so in the future we might be able to use that information to recognize them.

In temperate Asia, only *P. neopolydactyla* 2b and *P. neopolydactyla* 3 occur. *P. neopolydactyla* 3 has very wide lobes, resembling strongly to *P. neopolydactyla* 2a, whereas *P. neopolydactyla* 2b ususally has narrower lobes, resembling *P. polydactylon*. In this geographic region, it is thus possible to identify them.

In Northern Japan and Eastern Russia, *P. neopolydactyla* 2a and *P. neopolydactyla* 2c occur and they might be difficult to identify. More work on morphology will be needed to distinguish these two lineages.

*P. dolichorhiza* also resembles these morphotypes, but geographic information can easily allow us to identify it.

**Occurence of cosmopolitan species in the section Polydactylon?**

*P. scabrosa*, *P. neopolydactyla*, *P. dolichorhiza* and *P. polydactylon* were the taxa which were considered to have a cosmopolitan, or almost cosmopolitan distribution.

*P. scabrosa* s.l. is actually composed of 4 or 5 lineages. But they all have a panboreal distribution.

*P. dolichorhiza* is only present in the Neotropics and Afrotropics. Morphotypes identified as *P. dolichorhiza* in other parts of the world don't belong to the actual species.

*P. neopolydactyla* 1, 2a and 4 are panboreal, sometimes reaching mountain chains in temperate zones,

*P. neopolydactyla* 2b and *P. neopolydactyla* 3 seem restricted to temperate Asia, and *P. neopolydactyla* 2c to Hokkaido and Eastern Russia.

*P. neopolydactyla* 5 and *P. neopolydactyla* 6, as well as *P.* sp. 7a are endemics to the Pacific Northwest region of North America, *P.* sp. 12 and *P.* sp. 7b to Japan.

*P. polydactylon* is actually composed of three different geographically distinct lineages; in Europe (*P. polydactylon* 1), East Coast of North America (*P.* sp. 10), and West Coast of America (*P. polydactylon* 2). , Morphotypes identified as *P. polydactylon* in Asia or South America belong to other species.

It therefore seems that cosmopolitan lineages don't actually exist in *Peltigera* section *Polydactylon*. Widespread species having a panboreal distribution, or widespread in South America, or in Asia exist, but no species have been found with an actual or an almost cosmopolitan distribution. The names used to identify similar morphotypes across continents and climatic zones were assigned to distinct evolutionary species, which need formal description.

Now that it is clear that most species only occur within well-delimited geographic range, this information can also be used to facilitate species delimitation. Once that morphologically similar species that are known not to occur in a specific zone have been dismissed, it will be easier to identify *Peltigera* species.

Supplementary Material

Supplementary Material from this chapter can be downloaded online at https://github.com/NicolasMagain/ThesisOnlineSupplementary/.