

Species boundaries in the messy middle – testing the hypothesis of micro-endemism in a recently diverged lineage of coastal fog desert lichen fungi

Jesse Jorna¹, Jackson Linde¹, Peter Searle¹, Abigail Jackson¹, Mary-Elise Nielsen¹, Madeleine Nate¹, Natalie Saxton¹, Felix Grewe², María de los Angeles Herrera-Campos³, Richard Spjut⁴, Huini Wu², Brian Ho², Steven Leavitt¹, and Thorsten Lumbsch²

¹Brigham Young University

²Field Museum of Natural History

³Universidad Nacional Autónoma de México Instituto de Biología

⁴World Botanical Associates

April 5, 2021

Abstract

Species delimitation among closely related species is challenging because traditional phenotype-based approaches, e.g., morphology, ecological, or chemical characteristics, often produce conflicting results. With the advent of high-throughput sequencing, it has become increasingly cost-effective to acquire genome-scale data which can resolve previously ambiguous species boundaries. As the availability of genome-scale data has increased, numerous species delimitation analyses, such as BPP and SNAPP+Bayes factor delimitation (BFD*), have been developed to delimit species boundaries. However, even empirical molecular species delimitation approaches can be biased by confounding evolutionary factors, e.g., hybridization/introgression and incomplete lineage sorting, and computational limitations. Here we investigate species boundaries and the potential for micro-endemism in a lineage of lichen-forming fungi, *Niebla Rundel & Bowler* in the family Ramalinaceae. The species delimitation models tend to support more specious groupings, but were unable to infer robust, consistent species delimitations. The results of our study highlight the problem of delimiting species, particularly in groups such as *Niebla*, with complex, recent phylogeographic histories.

Species boundaries in the messy middle – testing the hypothesis of micro-endemism in a recently diverged lineage of coastal fog desert lichen fungi

Jesse Jorna^{1,*}, Jackson B. Linde^{1,*}, Peter C. Searle^{1,*}, Abigail C. Jackson¹, Mary-Elise Nielsen¹, Madeleine S. Nate¹, Natalie A. Saxton¹, Felix Grewe², María de los Angeles Herrera-Campos³, Richard W. Spjut⁴, Huini Wu², Brian Ho², H. Thorsten Lumbsch², Steven D. Leavitt^{1, 5, *}

¹Department of Biology, Brigham Young University, Provo, UT, USA

²Science & Education, The Grainger Bioinformatics Center, The Field Museum, 1400 S. Lake Shore Drive, Chicago, IL, USA

³Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, Apartado Postal 70-367, 04510, Ciudad De México, Mexico

⁴World Botanical Associates, PO Box 81145, Bakersfield, CA, USA

⁵Monte L. Bean Life Science Museum, Brigham Young University, Provo, UT, USA

*first three authors contributed equally

Abstract

Species delimitation among closely related species is challenging because traditional phenotype-based approaches, e.g., morphology, ecological, or chemical characteristics, often produce conflicting results. With the advent of high-throughput sequencing, it has become increasingly cost-effective to acquire genome-scale data which can resolve previously ambiguous species boundaries. As the availability of genome-scale data has increased, numerous species delimitation analyses, such as BPP and SNAPP+Bayes factor delimitation (BFD*), have been developed to delimit species boundaries. However, even empirical molecular species delimitation approaches can be biased by confounding evolutionary factors, e.g., hybridization/introgression and incomplete lineage sorting, and computational limitations. Here we investigate species boundaries and the potential for micro-endemism in a lineage of lichen-forming fungi, *Niebla* Rundel & Bowler in the family Ramalinaceae. The species delimitation models tend to support more specious groupings, but were unable to infer robust, consistent species delimitations. The results of our study highlight the problem of delimiting species, particularly in groups such as *Niebla*, with complex, recent phylogeographic histories.

Key words: Baja California, Bayes factor delimitation (BFD*), genealogical divergence index (*gdi*), RAD-seq, species delimitation

Introduction

Species delimitation has never been a simple or straightforward task, due, in part, to how species are conceptualized (Mayden, 1997; Wilkins, 2011). However, conceptualizing species as “separately evolving metapopulation lineages” – the unified species concept (USC) – shifts the debate away from species concepts to how species boundaries can be empirically delimited (de Queiroz, 2007). Under the USC, rather than weighing which property, such as morphology, monophyly, or reproductive isolation is indicative of a population representing an independent lineage, these properties simply become lines of evidence that can be used within an empirical framework for delimiting species (Sites and Marshall, 2004). Different properties may arise at different times or in different orders during the process of speciation (de Queiroz, 1998; de Queiroz, 2005; de Queiroz, 2007). Thus, early in the process of speciation, there is often a grey area where species boundaries are difficult to ascertain and potential for conflicting signals among independent lines of evidence (Cars- tens et al., 2013; Huang and Knowles, 2015; Mallet, 2008). Hence, one of the biggest challenges of species delimitation is in determining when differences in intra-specific population structure becomes species-level divergence.

With closely related or recently diverged species, delimitation of species boundaries using traditional phenotype-based approaches, e.g., morphology, ecological, or chemical characteristics, becomes challenging (Poelstra et al., 2020; Printzen, 2010; Wagner et al., 2013). The inclusion of molecular sequence data can be a critical tool for inferring robust species boundaries, especially within an integrative framework (Dayrat, 2005; Fujita et al., 2012; Will et al., 2005). Particularly, high-throughput sequencing technologies allow for genome-scale data to be generated using a wide range of time- and cost-effective strategies, generating hundreds to thousands of loci (Hale et al., 2020; McKain et al., 2018). However, even when using genome-scale data and empirical species delimitation analyses, the species boundary can remain blurred, especially in recently diverged species, hybrid zones, cases of introgression, and when species boundaries remain semipermeable (Harrison and Larson, 2014; Hey et al., 2003; Leaché et al., 2018). Furthermore, harnessing genome-scale data creates computational challenges and divergent approaches for selecting appropriate genomic loci (Faircloth et al., 2012; Kapli et al., 2020). While additional loci allows for the detection of subtle differences in population structure, potentially resolving recent species-level divergence (Huang, 2020), these data may also lead to over-splitting populations and inappropriately recognizing populations as species (Rannala, 2015; Sukumaran and Knowles, 2017).

Similar to other non-model organismal groups, interpreting species-level diversity in lichen-forming fungi has been revolutionized by incorporating DNA sequence data (Lumbsch and Leavitt, 2011; Printzen, 2010; Schneider et al., 2016). Interpreting anatomical, chemical, and morphological characters of lichens within

a phylogenetic framework has transformed our understanding of the role of phenotypic characters used for lichen-forming fungal taxonomy (Printzen, 2010). While portions of the ribosomal cistron, including the internal transcribed spacer (Schoch et al., 2012) – the standard barcoding marker for fungi, and regions of the mitochondrial genome, have been the standard genetic loci for lichen fungi systematics, genomic data have become increasingly useful when studying closely related species or recently diverged populations (Allen et al., 2018; Alonso-García et al., 2021; Grewe et al., 2017; Leavitt et al., 2016; Widhalm et al., 2021).

Here we investigate species boundaries in a lineage of fruticose lichen-forming fungi endemic to coastal fog deserts along the Pacific coasts of the New World – *Niebla* Rundel & Bowler in the family Ramalinaceae. Members of this genus form conspicuous, fruticose lichens, expressing an impressive range of morphological and chemical diversity (Fig. 1). This phenotypic variation has resulted in two widely different interpretations of species-level diversity, either a few, morphologically polymorphic species (Bowler and Marsh, 2004) or a much finer subdivision, two genera, *Niebla* with 42 species and one variety, and *Vermilacinia* Spjut & Hale with 28 species (Spjut, 1996). These genera were recently investigated within a multi-locus, phylogenetic framework (Spjut et al., 2020). *Niebla* diversified more recently, estimated at ca. 13 Mya, with speciation largely impacted by the climatic oscillations since the Miocene (Spjut et al., 2020). While evidence from Spjut et al. (2020) support high levels of diversity in *Niebla*, most phenotypically circumscribed *Niebla* species were not recovered as monophyletic using a multi-locus dataset. These results have been explained by model of micro-endemism of allopatric species in *Niebla*, with convergent forms occurring in separate populations, especially different geographic regions within the coastal fog deserts of California, Baja California, and Baja California Sur (Spjut et al., 2020 – e.g., *N. homalea*, *N. undulata*, Fig. 7). A – a Furthermore, despite generating a robust, multi-locus dataset and implementing multiple DNA-based empirical species delimitation analyses, species boundaries remain uncertain due to discrepancies among different analytical approaches (Carstens et al., 2013; Spjut et al., 2020).

Given the incongruence between what has been considered taxonomically diagnostic phenotypic characters and results from multi-locus phylogenetic inference and species delimitation analyses, here we aim to explore the utility of genome-scale data to resolve species boundaries in this fascinating lineage of lichen-forming fungi. In addition, we test the hypothesis of micro-endemism in this clade of symbiotic fungi. Specifically, we used restriction site-associated DNA sequencing (RADseq) to generate tens of thousands of short loci distributed across the genome to infer evolutionary relationships. Several species delimitation analyses, including BPP and SNAPP+Bayes factor delimitation (BFD*), were used to infer species boundaries. BPP is a Bayesian MCMC program for analyzing genomic sequence data under the multispecies coalescent model and has been shown to be able to delimit species boundaries even among closely related lineages (Flouri et al., 2018; Yang and Rannala, 2010, 2014; Zhang et al., 2011). Given that genomic data be can effective in delimiting fine-scale population structure, rather than species, we use the recently heuristic genealogical divergence index (gdi) to distinguish between population structure and species boundaries (Jackson et al., 2017; Leaché et al., 2018). With BPP's estimates of coalescent node heights (τ) and ancestral effective population size (θ), gdi scores can be calculated to estimate the divergence of two taxa (Leaché et al., 2018; Poelstra et al., 2020). Complementarily, we used SNAPP+BFD* to rank different species delimitation models. SNAPP attempts to decrease computation time for large amounts of phylogenetic datasets by bypassing the need for gene trees, and species tree estimates are directly inferred from biallelic markers, such as SNP data (Bryant et al., 2012). For each model, Bayes factor scores are calculated based on their marginal likelihood values using a path-sampling analysis and subsequently compared (Leaché et al., 2014). While this comparison is useful in identifying the best-fitting species delimitation models, it computationally limited to only smaller datasets with fewer than 10 species and under 100 individual samples (Leaché et al., 2014).

Our results provide crucial insights into the diversification history of *Niebla* and highlight the power and limitations of using genome-scale data to infer species boundaries in recently diverged lineages.

2. Material and Methods

2.1 Specimen sampling

To investigate evolutionary relationships and the potential for micro-endemism in *Niebla* species occurring in coastal fog deserts in North America, 182 *Niebla* specimens were collected from Baja California Norte, Mexico in December 2016. Secondary metabolites were identified from all specimens using thin layer chromatography (TLC) and following standard methods with solvent systems ‘C’ and ‘G’ (Culberson, 1972; Orange et al., 2001). Specimens were determined by R. Spjut according to his classification (Spjut, 1996), with exception to three additional specimens collected in the Channel Islands. Three additional *Niebla* specimens from the Southern Islands (San Clemente and Santa Barbara Islands), USA, were also included and determined following Spjut (1996). From the initial collections, 40 specimens representing 27 phenotypically circumscribed *Niebla* species were selected for restriction site-associated DNA sequencing (RADSeq) to infer evolutionary relationships and delimit species boundaries in this group (Appendix 1). When possible, multiple representatives of each species collected from different sampling localities were selected. Seven specimens representing *Vermilicinia*, recently shown to be sister to *Niebla*, were used as the outgroup (Spjut et al., 2020). We note that specimens were selected and sequenced before a broad molecular phylogeny was available for *Niebla* (Spjut et al., 2020).

We attempted to amplify the fungal internal transcribed spacer region (ITS), the standard DNA barcoding marker for fungi (Schoch et al., 2012), for the *Niebla* specimens selected for RADSeq. From these specimens, a small portion of the thallus free of visible contamination was excised, and total genomic DNA was extracted using the E.Z.N.A. Plant DNA DS Mini Kit (Omega Bio-tek), following the manufacturers’ recommendations. Extractions of specimens for RADseq were done using ZR Fungal/Bacterial DNA MiniPrep Kit (Zymo Research, Irvine, CA, USA) as previously described (Grewe et al., 2017). We amplified the ITS region (ITS1, 5.8S & ITS2) using primers ITS1 (Gardes and Bruns 1993) with ITS4 (White et al. 1990). Polymerase chain reaction (PCR) amplifications were performed using Ready-To-Go PCR Beads (GE Healthcare, Pittsburgh, PA, United States), with cycling parameters following a 66–56°C touchdown reaction (Lindblom and Ekman 2006). PCR products were visualized on 1% agarose gel and enzymatically cleaned using ExoSAP-IT Express (USB, Cleveland, OH, United States). Complementary strands were sequenced using the same primers used for amplifications, and sequencing reactions were performed using BigDye 3.1 (Applied Biosystems, Foster City, CA, United States). Products were run on an ABI 3730 automated sequencer (Applied Biosystems) at the DNA Sequencing Center at Brigham Young University, Provo, UT, United States.

2.2 Reference Sequencing and Assembly

We first deep-sequenced and assembled a reference draft genome from a specimen representing *N. homalea* (Ach.) Rundel & Bowler for identifying mycobiont loci by mapping during the processing of metagenomic RADseq data (*Niebla* culture #47: USA, California, Sonoma Co., Sonoma Coast State Park, Shell Beach, 38.4225, -123.114, on large rock, 25 July 2018, Coll: MH Huhndorf). DNA from this specimen was isolated using the ZR Fungal/Bacterial DNA MiniPrep Kit (Zymo Research, Irvine, CA, USA), converted into libraries with the KAPA Hyper Prep Kit (KAPA Biosciences, Wilmington, MA, USA) and sequenced at the University of Illinois at Chicago Research Resource Center on Illumina’s NextSeq platform as before for lichen-forming fungi (Grewe et al., 2018). High-molecular weight DNA isolation, long-read sequencing on a Nanopore GridIONx5 sequencer, and assembly of *N. homalea* was done as described before for the lichen fungal culture of *Physcia stellaris* (Wilken et al., 2020). The pipeline used canu v1.8 (Koren et al., 2017) for a long-read assembly with a genome size estimation of 26 megabases. The raw contigs were corrected twice with racon v1.3.2 (Vaser et al., 2017) and subsequently polished twice with the Illumina short reads of *N. homalea* using Pilon v1.23 (Walker et al., 2014). Finally, we created a Bowtie2 (Langmead and Salzberg, 2012) database from the selected scaffolds for the mapping approach to filter for fungal RAD loci.

2.2.1 RADseq Library Preparation and Sequencing

RADseq libraries were prepared from the isolated DNA as described following (Grewe et al., 2017). In summary, DNA isolations were pooled with sequence adapters (Rubin and Moreau, 2016), digested with the restriction enzyme ApeKI (New England Biolabs, Ipswich, MA, USA) and ligated using T4 ligase (New England Biolabs). All samples with compatible barcodes were pooled and selected for fragment sizes between 300 and 500 bp using the BluePippin DNA size selection system (Sage Science, Beverly, MA, USA). The

pooled libraries were amplified using the REDTaq ReadyMix (Sigma-Aldrich, St. Louis, MO, USA) prior to sequencing on an Illumina MiSeq using the MiSeq Reagent Kit v3 for 150 cycles (Illumina, San Diego, CA, USA) to produce single-end sequences with a length of 150 bp.

2.2.2 Assembly of RADseq datasets

The raw reads from the MiSeq sequencing were processed and assembled Stacks v2.3 (Rochette et al., 2019) as described earlier for metagenomic datasets of lichens (Alonso-García et al., 2021). In short, we demultiplexed reads of individuals from the pool of raw sequence reads based on their barcodes with the script ‘process-radtags’. The demultiplexed reads of each individual were aligned to the reference genome database using Bowtie2 (Langmead and Salzberg, 2012). The script ‘gstacks’ with default parameters was used to identify SNPs in the reads aligned to the reference genome. The SNP data was then analyzed and filtered with the script ‘populations’ considering each individual a single population. The final dataset did not allow heterozygosity at a locus (`-max-obs-het 0`) and was filtered for a minimum minor allele frequency of 5% (`-min-maf .05`). Only the first SNP per locus (`-write-single-snp`) was retained when present in at least 30% (`-R 0.3`) of the individuals.

2.3 Phylogenomic reconstructions

To explore species monophyly and patterns of micro-endemism using genome scale data (e.g., Wagner et al., 2013), evolutionary relationships among *Niebla* specimens were inferred from RADseq data using two strategies, (i) a supermatrix approach from the concatenated RADseq loci and (ii) a computationally efficient species tree inference approach accounting for incomplete lineage sorting (Chifman and Kubatko, 2014) using variable sites extracted from RADseq loci. Concatenation, or phylogenomic supermatrix approaches, have been shown to accurately infer relationships across a range of scenarios (Tonini et al., 2015). We inferred relationships from both the concatenated RADseq loci, and a separate dataset comprised of only variable sites (SNP dataset). Phylogenetic trees were reconstructed using maximum likelihood as implemented in IQ-TREE v 1.6.12 (Nguyen et al., 2015), with 1,000 ultra-fast bootstrap replicates (Hoang et al., 2018) to assess nodal support. For the SNP dataset comprised of 198K SNPs (see results), we used the GTR+ASC to avoid overestimated branch lengths due to the SNP alignment comprised only of variable sites. For the concatenated RADseq loci dataset, comprised of 25,086 loci, total alignment length 3.6 Mb (see results), we used the GTR+I+G substitution model. Because standard concatenation approaches may return incorrect trees with high support in the presence of incomplete lineage sorting (Edwards, 2009; Kubatko and Degnan, 2007), we used SVDQuartets, as implemented in PAUP*, a method that infers relationships among quartets of taxa under the multispecies coalescent model (Chifman and Kubatko, 2014). SVDQuartets+PAUP* is computationally efficient with large genome-scale datasets and able to accurately infer relationships under a range of scenarios (Chou et al., 2015). All possible quartets were evaluated under the multispecies coalescent tree model, and nodal support was performed using 100 bootstrap replicates.

2.4 Species Delimitation analyses

2.4.1 Single locus species delimitation using the standard fungal DNA barcode

To compare species boundaries inferred using the standard fungal barcoding marker with inferences from genome-scale data, we used Assemble Species by Automatic Partitioning (ASAP; Puillandre et al., 2021). ASAP is a recently developed method that circumscribes species partitions using an implementation of a hierarchal clustering algorithm based on pairwise genetic distances from single-locus sequence alignments (Puillandre et al., 2021). The pairwise genetic distances are used to build a list of partitions ranked by a score, which is computed using the probabilities of groups to define panmictic species. ASAP provides an objective approach to circumscribe relevant species hypotheses as a first step in the process of species delimitation. Therefore, we used ASAP to circumscribe candidate *Niebla* species from a multiple sequence alignment of the ITS barcoding region. ITS sequences generated for this study were combined with those from Spjut et al. (2020) and aligned using the program MAFFT v7 (Katoh et al. 2005, Katoh and Toh 2008). We implemented the G-INS-i alignment algorithm and ‘1PAM / K=2’ scoring matrix, with an offset value of 0.1, the ‘unalignlevel’ = 0.2, and the remaining parameters were set to default values. The multiple

sequence alignment was analyzed using the ASAP Web Server (<https://bioinfo.mnhn.fr/abi/public/asap/>), with the ‘asap-score’ considered to select the optimal number of species partitions (Puillandre et al.).

2.4.2 Genomic species delimitating under the multispecies coalescent model using BPP/GDI

Incorporating genome-scale data with the multispecies coalescent model has provided unprecedented insight into species boundaries (Flouri et al., 2018). The Bayesian program BPP includes a full-likelihood implementation of the multispecies coalescent model (Yang, 2015), providing powerful approaches for delimiting species and populations (Degnan and Rosenberg, 2009; Knowles and Carstens, 2007; Yang and Rannala, 2010). Here we use BPP to test the hypothesis of micro-endemism vs. phenotypically polymorphic species in *Niebla*. Due to computational limitations, we were unable to analyze the complete RADseq dataset – 25,086 RADseq loci – using BPP. Therefore, we chose to explore species delimitation results using various subsets of the RADseq data. We identified the number of informative sites per RADseq locus by running the complete, partitioned dataset in IQ-TREE v. 2.1.1 (Minh et al., 2020). Based on these results, we selected data subsets comprised of the 10 most variable loci, all loci with at least 20 informative sites (163 of the 25,086 original loci), and the 500 most variable loci. After excluding loci with no variation, we also generated data subsets comprised of 100, 500, 1000 and 5000 randomly sampled loci, in replicates of three. Finally, we included a subset of 5649 loci that included between 10 and 20 variable sites across the entire locus. To prepare this data for implementation in BPP, we used a “.fasta” file of the RAD-Seq data, a “.txt” file with the names of the partitions of interest, and the “.txt” partition file itself. We created a custom python script (Dryad file S51) which uses those three files to output a .phy file of the RAD-Seq data partitions of interest to make them compatible with BPP. Further, it creates a .txt file with the names of candidate species from the “.phy”. From this, a designation was made for the candidate species, e.g. A1, A2, B, C, etc., to create the imap file for BPP. Afterwards, these files were used to run “A00” and “A10” analyses in BPP (described below).

BPP analysis ‘A00’ estimates tau (τ) and theta (θ) parameters under the multispecies coalescent model when the species phylogeny is given (Yang, 2015). Priors for the *Niebla* species population metrics, θ and τ , were estimated from each subset of RADseq loci using the BPP software’s built-in function for parameter estimation to avoid any existing taxonomic bias. Following Flouri et al. (2018), an initial naïve analysis was run using the θ and τ estimates of the software-provided dataset of Asiatic brown frogs where $\theta = 0.004$ and $\tau = 0.002$. For each subsequent BPP analysis, the model for each data subset was specified to use the corresponding θ and τ priors ($\theta = 0.0001$ and $\tau = 0.00003$), as well as using BPP’s built-in function to re-estimate theta depending on the dataset chosen. Note, these priors differ widely from the parameters of the BPP analyses of Ramalinaceae reported in Spjut et al. (2020), although specific effects of different θ and τ priors on species delimitation models were not reported therein. Each analysis of the 10-, 163- and 500-most variable loci datasets was replicated six times total to account for variation in the GDI estimates, while we independently took a random sample three times and analyzed those in four replications, for a total sample of 12 per randomly sampled subset. Subsequently, we performed the ‘A10’ analysis in BPP for species delimitation using a fixed guide tree, which provides posterior probabilities for different estimated species delimitation models using a Bayesian modelling approach (Yang and Rannala, 2010). All subsets were run for a total of 500,000 MCMC simulations, with the exception of the 5000 and 5649 subsets which were run for 100,000 MCMC simulations to avoid computational constraints. MCMC simulations were guided by a ‘burn-in’ of an extra 10% of the total run length, e.g., 50,000 iterations for most runs. RADseq data can recover populations and recently diverged species as reciprocally monophyletic (Hou et al., 2015; Wagner et al., 2013). Recent studies have also shown that multispecies coalescent approaches to species delimitation can overestimate the number of species by delimiting population structure, rather than species-level divergence when using genome-scale data (Leaché et al., 2018; Sukumaran and Knowles, 2017). Therefore, 17 candidate species were initially circumscribed as reciprocally monophyletic clades consistent between the IQtree and SVDQuartets+PAUP* inferences and nested within the candidate species inferred from the single locus ASAP partitions, e.g., phylogenetic substructure within the ASAP partitions – see results. The IQtree inference was used as the initial guide tree, collapsing nodes comprising multiple specimens within each candidate species. To assess the effect of increased coverage of RADseq loci in BPP species delimitation

analyses, we used the different subsets of loci as defined above for our ‘A10’ BPP analyses. Parameters were consistent with the “frog.cti” found in BPP v. 4.3.0 tutorial (Flouri et al., 2018), with the exception of θ and τ priors, as mentioned above. The same 17 species phylogenetic tree prior was included in all analyses. For each dataset, MCMC chains were run for 100,000 generations, including 20,000 burn-in iterations (Dryad file S52), and the same sample size as used in the ‘A00’ analysis was used to control for variation between runs (six independent runs for the most informative loci datasets, and three independent subsets repeated four times each for the randomly selected loci subsets). The posterior probability of each species’ delimitation model was then pooled for an aggregate possibility between the six replicates of highly informative loci datasets, or 12 replicates of the randomly sampled loci datasets. Resulting log files were analyzed in Tracer v.1.7.1 (Rambaut and Drummond, 2005) to assess ESS values and convergence of runs.

Because genome-scale data sets may delimit population boundaries and not species limits, we coupled our BPP analyses with the recently proposed heuristic empirical genealogical divergence index (*gdi*; Jackson et al., 2017; Leaché et al., 2018). With the proposed pattern of micro-endemism and with putatively sympatric species common in our sampled candidate species, we used BPP’s estimates of τ and θ to calculate the *gdi* for pairwise species comparisons (Poelstra et al., 2020). We used the equation in Leaché et al. (2018), $gdi = 1 - e^{-2\tau/\theta}$, to calculate the probability that two sequences coalesce before reaching species divergence (τ) when the genealogy is traced backwards in time. We used a custom python script to extract τ and θ values from the BPP analyses (Dryad file S53), treating well-supported phylogenetic substructure in the ML topology as candidate species – see Results. The script takes in an ‘A00’ BPP output file, calculates the *gdi* for each candidate species comparison, then writes a matrix to a .csv file. The ‘A00’ analyses were executed for the same subsets of RADseq loci in six or 12 replicates, and *gdi* scores were calculated individually for each run and averaged across replicates. Resulting *gdi* scores for each comparison between either single species’ group or composite species’ groups were displayed in a matrix. To assess potential biases in *gdi* scores resulting from differences in the number and types of sampled RADseq loci, we visualized the distribution of *gdiscore* results from the different sample subsets using a normalization function in Excel. As a rule of thumb, *gdi* values <0.2 suggest a single species and *gdi* values >0.7 suggest two distinct species, while *gdi* values within the range indicate ambiguous delimitation (Jackson et al., 2017). The resulting BPP inferred *gdi* delimitations were compared to the initial ‘A10’ analysis performed as described above.

We also used an iterative, ‘tip-down’ approach using BPP+*gdiscores* following Jackson et al. (2017) and Leaché et al. (2018). After collapsing shallow clades comprised of morphologically identical species collected from the same location (*N. cornea* [sl16863BF & sl16857BF]; *N. dilatata* [sl16940 & sl16946]; *N. aff. isidiosa* [sl16775 & sl16853BF]; *N. juncosa* var. *juncosa* [sl16705 & sl16707]; and *N. juncosa* var. *spinulifera* [sl16706 & sl16719] – see Results), the topology inferred using IQtree was used as the initial guide tree, treating each terminal node as a candidate species – 35 candidate species total; and specimens representing phenotype-based species were split into multiple candidate species when not recovered as monophyletic. In BPP, the ‘A00’ analysis was performed, and we extracted the θ , τ and taxa labels as described above. Candidate species were collapsed into a single species group if the average of the pairwise comparisons of *gdi* scores between sister groups was ≥ 0.2 . The process was performed iteratively until the *gdi* scores no longer suggested the joining of any candidate species in the topology. A total of ten analyses per guide tree were completed to ensure consistent results by the comparing means and standard deviations of each *gdi* score. The analyses were run for 2,000,000 iterations, with ten independent runs, calculating the *gdi* score from the mean θ and τ values.

2.4.2 Bayes Factor species delimitation implemented in SNAPP

As an alternative to BPP+*gdi*, we use Bayes factor species delimitation (BFD*) analyses in SNAPP v.1.3 (Bryant et al. 2012). For the SNAPP analyses, 500 variable sites evenly distributed across the dataset were extracted using a custom script and uploaded to a SNAPP template in BEAUTi v.2.6.3 (Bouckaert et al., 2019). Alternative species models were subsequently assigned based on prior species delimitation inferences from ASAP, BPP, and well-supported clades inferred from in the ML analyses (see results) recovered within distinct ASAP partitions. Due to computational limitations in SNAPP, we divided the ML tree into three

major clades, with each clade run as a separate analysis in BEAST v2.6.3 (see Dryad file S54).

Substitution rates were calculated directly in BEAUTi using the entire 500 SNP dataset. Coalescence rate was set at 10.0 and specified to sample. Priors for both lambda and theta were set using a gamma distribution. Lambda was set to have a distribution of $G(2, 500)$. Theta was set to have a distribution of $G(1, 250)$. The chain length was set to 1,000,000, sampling every 1,000 generations. We utilized the BFD* methods outlined by Leaché et al. (2014). Marginal likelihood values were obtained using a path-sampling analysis in BEAST with 48 steps, chain length of 1,000,000 generations and 10% burn-in. Resulting log files were analyzed in Tracer v1.7.1 (Rambaut and Drummond, 2005) to assess ESS values and convergence of runs. Each model was run four times and the marginal likelihood value was averaged across each clade in order to further ensure convergence.

3. Results

The number of Illumina reads for each sample varied from 58,930 to 1,151,143, with an average of 326,580 sequences per sample, with 7.3 to 75.3% of reads mapped to the *Niebla* reference using Bowtie2 (Appendix 1). The number of *Niebla* RADseq loci (within sample clusters) that STACKS generated from these sequences ranged from 2,534 to 39,941 (mean=22,373 loci/sample). Our final RADseq dataset resulted in a total of 25,086 143-bp loci. The total length of the concatenated RADseq loci was 3.6 Mb, including 198K variable sites.

3.1. Phylogenetic reconstructions

The ML topologies inferred from both the concatenated RAD-Seq data and the SNP data recovered identical, well-supported topologies (Fig 2). Major clades recovered here coincided with the production of distinct secondary metabolites, although specimens producing depsides of either sekikaic acid or divaricatic acid were each recovered in two polyphyletic clades (Fig. 2) with a much larger divaricatic acid clade sister to the depsidone (nonterpenoid) clade, or a clade divided into sekikaic acid and depsidone subclades. Phenotypically circumscribed species represented by multiple samples were not recovered as monophyletic except those identified as *N. cornea*; they were collected from the Channel Islands, California, USA and from near San Quintín, Baja California, Mexico (Appendix 1; Fig. 2). Additionally, specimens collected in close proximity having the same chemotype—but differentiated morphologically as distinct species or varieties—occurred in different clades (Fig. 2); examples are divaricatic acid specimens that include saxicolous sl-16775 (*N. aff. isidiosa*) and sl-16778 (*N. laminaria*), sl-16705 (*N. juncosa* var. *juncosa*) and sl-16706 (*N. juncosavar. spinulifera*), sl-16707 (*N. juncosa* var. *juncosa*) and sl-16708 (*N. undulata*). Those specimens evidently growing in close proximity differentiated as species more by their secondary metabolites—obviously in different clades—include, for example, terricolous sl-161004 (*N. palmeri*, sekikaic acid), sl-161005 (*N. arenaria*, salazinic acid), and saxicolous sl-161006 (*N. rugosa*, divaricatic acid). As many as eight species have been recognized at a location (Spjut et al., 2020). SVDQuartets+PAUP* resulted in a generally well-supported topology, although in some cases backbone relationships differ strikingly from the ML reconstruction (Fig. 2; Dryad file S55).

3.2. Species delimitation

3.2.1. ASAP species delimitation using the standard fungal barcode

A maximum of 32 species partitions were inferred from the combined ITS dataset (ITS sequences generated for this study + ITS sequences reported in Spjut et al. 2020) using ASAP, 13 of which were represented by the samples selected for RADseq. The majority of ASAP partitions containing at least two samples were comprised of morphologically/chemically polymorphic specimens, e.g., specimens identified as different species (Fig. 2). The ASAP partitions inferred from the standard fungal barcode marker largely coincided with reciprocally monophyletic clades inferred for RADseq data in the SVDQuartets+PAUP* and ML topologies (Fig. 2). Specimen sl16849BF (*Niebla* “sp. nov.”) was inferred as a distinct species partition in ASAP but was combined with another partition comprising the remainder of the specimens in the same clade for all subsequent analyses. This specimen was morphologically similar to *N. versiforma* but contains salazinic

acid, rather than divaricatic acid (with triterpenes), and represents a putative undescribed species. Specimens sl16775 & sl16853BF, both representing isidiate *Niebla* specimens (Fig. 1E &K) and collected from the same locality, were inferred as separate ASAP partitions but were combined for all subsequent analyses tentatively identified as *N.aff. isidiosa*.

3.2.2 Species delimitation using RADseq loci under the multi-species coalescent using BPP

Estimates of θ values under the multispecies coalescent model – analysis ‘A00’ in BPP – were on the same order of magnitude regardless of different RADseq loci subsamples, although standard deviation was much greater in the 163- and 500-most variable loci data subsets. Greater variation in τ values was observed depending on the subsampled RADseq loci (Table 1). Results using the same subset of RADseq loci generally converged on similar *gdi* estimations. Species delimitation using a fixed guide tree – analysis ‘A10’ in BPP – and the candidate species shown in Fig. 3 inferred up to 17 species, the maximum subdivision specified in the guide tree. Aggregate probabilities across all runs resulted in the highest probabilities across the 13–17 species range, with a 15 species model being most likely across all analyses (Fig. 4). However, the results of the ‘A10’ species delimitation analyses in BPP varied across both subsets of data and independent runs, with the randomly sampled subsets showing less proclivity of an unrealistic one-species model. (Table 1; Dryad file S56).

While the candidate species collapsed under each analysis varied widely, the most frequent candidate species groups that were collapsed included: ‘O’, ‘P’ and ‘Q’ collapsed into 1 or 2 species; ‘J’, ‘K’, ‘L’, ‘M’ and ‘N’ were frequently collapsed into 1, 2 or 3 different species groups; and ‘A’, ‘B’, ‘C’ and ‘D’ which were frequently collapsed into 1 or 2 species groups, sometimes including ‘E’ (Fig. 3).

3.2.3 Species delimitation validation using *gdi* scores

The *gdi* scores inferred from parameters under a multispecies coalescent model – analysis ‘A00’ in BPP – generally did not provide justification for collapsing any nodes in the 17-species model – based on 17 well-supported clades identified in the ML topology, the maximum subdivision specified in the guide tree (Fig. 3; Dryad file S56), suggesting that our sampling includes at least 17 species-level lineages. However, in the analyses of the 10 and 163 most informative RADseq loci, a number of candidate species were collapsed based on *gdi*. In the 10-loci subset, candidate species ‘A’, ‘B’, ‘C’, and ‘D’ were combined (*gdi* scores <0.2), as were candidate species ‘F’ and ‘G’, candidate species ‘J’ and ‘K’, candidate species ‘M’ and ‘N’, and finally candidate species ‘P’ and ‘Q’ (Fig. 3). In the 163-loci subset analysis, candidate species ‘A’, ‘B’, and ‘C’ were collapsed together, as well as ‘F’ and ‘G’. The *gdi* scores calculated from all other analyses suggested maintaining the 17-species model. In the “tip-down”, iterative approach, *gdi* scores supported collapsing a number of tips resulting in a 29-species model from the original 35 tips after combining morphologically similar specimens from the same location (Fig. 3; Dryad file S57). Collapsed specimens included two of the three specimens identified as *N. cornea* from the Channel Islands with two specimens from Baja California (in clade ‘O’), specimens identified as *N. fimbriata*, *N. palmeri*, and *N. suffnessii* (in clade ‘M’), *N. flagelliforma* with *N. juncosa* var. *juncosa* (in clade ‘H’), and *N. flagelliforma* with *N. rugosa* (in clade ‘B’) (Fig. 3).

To further characterize the impact of data subsampling on *gdi* scores, we report average *gdi* scores across all nodes for each of the RADseq data subsets. The 10- and 163-locus datasets comprising the most variable sites had lower *gdi* scores when compared with the remaining subsets (Fig. 5). Average *gdi* values generally increased with an increased coverage of genetic markers, although using the most informative markers rather than a random subset seems to limit this trend. This trend is primarily caused because the τ estimate is much higher when estimated using only the most variable loci, with the exception of a very small dataset of 10 loci. Estimates for θ of each individual candidate species were also higher when using the most informative subsets, but so is the standard deviation of the estimated and θ values.

3.2.4 BFD* implemented in SNAPP

In all cases, BFD* delimited more putative species than the single-locus ASAP analysis (Fig. 2; Dryad file S54). BFD* analyses supported the most divisive species delimitation models in two of the three subclades –

clades ‘1’ and ‘3’, and the intermediate species delimitation model for ‘Clade 2’. Marginal likelihood values across each of the four independent runs were largely consistent but tended to be slightly more variable with an increasing number of species.

4. Discussion

Our study provides the first genome-scale perspective into diversification of the lichen forming fungal genus *Niebla* – endemic to the Pacific coast of Baja California, highlighting dynamic speciation in this clade of symbiotic fungi. Our data support the perspective of a species-rich genus, and that many species are indeed sympatric, occurring in close association, while single species populations have also been recognized, mostly near the range limits of the genus (Spjut 1996; Spjut et al., 2020). Although candidate species inferred from genome-scale do not support current phenotype-based species circumscriptions, morphologically similar specimens having the same chemotype collected from distinct localities along the coast of Baja California were consistently delimited as separated species-level lineages using RADseq data, further supporting a model of micro-endemism in *Niebla* (Spjut et al., 2020). On the other hand, morphological parameters applied to distinguish species could be revised to agree with molecular data, while chemical variation in lichen metabolites need further study such as triterpenes in *N. lobulata* (Spjut, 1996) that may correlate with species differences (Spjut et al., 2020), and isodivarcic acid found in type specimens of *N. eburnea* (Spjut-Rakatondraibe comm., Oct 2016) instead of divarcic acid as reported for *N. homalea* from the putative type locality (Zhang et al. 2020). Nevertheless, similar to an earlier study using multi-locus sequence data (Spjut et al., 2020), we found that robust, consistent species delimitations were difficult to infer, even with thousands of RADseq loci. Below we discuss the implications of our findings in relationship to the challenge of species delimitation and characterizing diversity in *Niebla*.

4.1 Species delimitation in the “messy middle”

Differences in species delimitation studies, ranging from the number and type of loci selected, how candidate species are initially identified, e.g., morphology-based vs. sequenced-based, to empirical analytical approaches, can result in discrepancies in the inferred species boundaries (Carstens et al., 2013; Luo et al., 2018). Furthermore, differences in diversification histories likely require different types of data and different analytical approaches in order to accurately delimit species boundaries (Camargo and Sites, 2013; Leavitt et al., 2015). With an increasing number of species inferred from phylogenomic datasets, the problem of inferring population structure versus actual species boundaries becomes acute (Leaché et al., 2018). Therefore, how researchers select the appropriate number of loci and analytical approaches becomes increasingly important as analyses continue to increase in scale. Recognizing that any species delimitation approach infers hypotheses of species boundaries, rather than providing the ultimate answer/solution, is imperative (Matute and Sepúlveda, 2019). Hence, integrating independent lines of evidence with phylogenomic datasets can provide increasingly robust hypotheses of species boundaries (Fujita et al., 2012; Yang et al., 2019).

The results of our study highlight the problem of delimiting species, particularly in groups such as *Niebla*, with complex, recent phylogeographic histories. Despite the informed perspective from genome-scale data and long-term investigations of phenotype-based circumscriptions, species boundaries in *Niebla* lichens are far from being resolved. While differences in the number of delimited candidate species in *Niebla*, depending on the analyzed loci and species delimitation analysis, resulted in no clear favored species delimitation model, our analyses supported high species-level diversity in *Niebla*. Both BPP species delimitation analyses and BFD* analyses implemented in SNAPP provided strong evidence for multiple species-level lineages in ASAP species partitions inferred from the standard fungal barcoding marker, the ITS region (Figs. 2 & 3; Dryad file S54). The ‘tip-down’ approach using BPP+*gdi* scores (Jackson et al., 2017; Leaché et al., 2018), supported the highest number of candidate species from our dataset, a 29-species model (Fig. 3). The BFD* analyses also generally favored the most divisive species delimitation models (Dryad file S54), although independent runs did not always converge with the more divisive species models. While the BFD* provides a direct, quantitative method for comparing species models (Leaché et al., 2014), with increasing numbers of sampled loci and candidate species, computational limitations may constrain the utility of this methodological approach (Leaché et al., 2014). In this study, we attempted to circumvent these limitations

by subdividing the topology in order to analyze computationally feasible data subsets. Despite those efforts, convergence among independent runs remained an issue in models with even a modest number of species (Dryad file S54).

While the heuristic *gdi* criterion for species delimitation provides a metric to guide the distinction of population structure from species boundaries (Jackson et al., 2017; Leaché et al., 2018), we show here that locus selection can have a substantial impact on *gdiscores*. In the empirical scenario of *Niebla*, selecting limited, but highly variable, RADseq loci resulted in the lowest *gdiscores*, specifically in the 10- and 163-marker datasets (Fig. 5). In contrast, randomly selected RADseq loci, resulted in largely consistent, higher *gdi* scores; and these *gdi* scores were more consistent with the 500-most variable RADseq loci dataset. These results suggest that when using short RADseq loci, it is important to maximize the number of analyzed loci if possible, but that lower number of randomly sampled loci can result in consistent inferences. The heuristic *gdi* criterion may also be biased in cases where a population was established by a few founder individuals and N_A and θ_A may be very small. Because the criterion depends on the population divergence time relative to population size ($2\tau/\theta_A$), the use of *gdi* may lead to claims of species status even if the populations are recently diverged (Leaché et al., 2018). We speculate that this may be the case for *Niebla* populations in Baja California. Here we inferred recent divergence among *Niebla* lineages, and it would not be unreasonable to suspect that independent populations were founded by few individuals. If this is the case, it may be necessary to consider both the absolute population divergence, as well as the population divergence relative to population size (Leaché et al., 2018; Yang and Rannala, 2010). Finally, given the large range of indecision using the *gdi*, it may not be possible to infer with confidence species boundaries given the nature of the speciation process in these lichen-forming fungi.

4.2 Concordance between the standard DNA barcoding marker and clades inferred from phylogenomic data

The standard DNA barcoding marker – the ITS – has been a powerful tool for understanding species diversity in lichen-forming fungi (Lücking et al., 2020; Schoch et al., 2012), although issues with using this marker in some clades have been demonstrated (Pino-Bodas et al., 2013). In taxonomically challenging groups, the ITS can provide an important first pass for grouping closely related species (Leavitt et al., 2013; Lücking et al., 2014; Moncada et al., 2020). Our results show that species partitions inferred from the ITS marker coincided with clades comprised of closely related species-level lineages in *Niebla* inferred from phylogenomic data (Fig. 2). These results suggest that in this speciose genus, the ITS can play a crucial role in assigning specimens to species complexes and provide guidance for future studies characterizing species-level diversity. However, our analyses of ITS alignments indicate that the ITS alone is not able to diagnose species-level lineage in *Niebla*. Extensive sampling of specimens representing ITS-delimited species, coupled with genome-scale data, will likely be essential to robustly characterize diversity in this genus. Similarly, focusing on specific habitats in Baja California supporting diverse *Niebla* communities, including Mesa Camacho, Mesa Santa Catarina, ridges between Punta Santa Rosalillita and Punta Negra, and Morro Santo Domingo, will be crucial to infer diversification processes in important genus of lichen-forming fungi occurring only in coastal fog deserts.

While the results of our analysis are somewhat ambiguous in terms of species delimitation, they provide a valuable perspective of the use of these empirical species delimitation methods in a non-model system. Furthermore, they add to the growing body of literature on the all-too-often “messy” species boundaries in organisms that further complicate our understanding of their complex evolutionary history. Future work should build upon the hypotheses presented here to add additional lines of evidence for specific species models in order to resolve the taxonomy in this group. Lichen-forming fungi are notoriously difficult to define in terms of species delimitation, and in some cases, traditional morphological and chemical characters used for taxonomy fail to reflect natural lineages (Lumbsch and Leavitt, 2011). Nevertheless, more thorough chemical investigations may clarify some species in *Niebla*. In addition, ecological and paleoclimatic feature should be considered in future work with speciation in *Niebla*.

Data Accessibility:

All supporting files are available on Dryad: https://datadryad.org/stash/share/XD_kUF-

pRIek8qAngIa3OIEclxC0JePCDmZAYCx1ClyI (temporary link)

- **Dryad file S1** . ‘Script 1’ – custom python script for concatenating FASTA sequences into BPP-style files using a “.fasta” file of the RAD-Seq data, a “.txt” file with the names of the partitions of interest, and the “.txt” partition file itself. Those three files are used to output a .phy file of the RAD-Seq data partitions of interest to make them compatible with BPP. Further, it creates a .txt file with the names of candidate species from the “.phy” file.
- **Dryad file S2** . ‘Script 2’ – control file for the A10-type analysis in BPP.
- **Dryad file S3** . ‘Script 3’ – custom python script to extract τ and θ values from the BPP analyses to calculate *gdi* scores directly from BPP ‘A00’-type analysis output.
- **Dryad file S4** . Results of the SNAPP+BFD* analyses. Page 1 – the five different species delimitation models for ‘clade 1’, ranging from 3 to 16 species assessed; the 16-species model was selected using Bayes factors (BF). Page 2 – the three different species delimitation models for ‘clade 2’ and ‘clade 3’; in ‘clade 2’, the 5-species model was selected over the 2- and 13-species models; in ‘clade 3’, the 7-species model was selected over the 2- and 3-species models. Page 3 – BF scores and rankings from the clade specific SNAPP+BFD* analyses
- **Dryad file S5** . Side-by-side comparison of the ML topology inferred using IQtree and the species tree inferred using PAUP+SVDquartets (summarized in Fig. 2). Values at nodes indicate bootstrap support when less than 100%.
- **Dryad file S6** . *gdi* scores & MCMC runs from the ‘A10’ species delimitation analyses using different subsets of RADseq loci.
- **Dryad file S7** . Results from the BPP + *gdi* “tip-down” species delimitation approach.

All RADseq reads are deposited in NCBI Short Read Archive under project # pending.

Acknowledgements

The field work portion of this study was funded by The Negaunee Foundation. We gratefully acknowledge support from Terra Peninsular for facilitating field collections in Reserva Natural Punta Mazo and Enrique Alfaro for his expertise and guidance. We also acknowledge support for the College of Life Sciences and Department of Biology at Brigham Young University and the Grainger Bioinformatics Center at the Field Museum. We thank Jacob Jensen, Zachary Muscavitch, and Emmanuel Serusiaux for fruitful discussions.

References

- Allen, J.L., McKenzie, S.K., Sleith, R.S., Alter, S.E., 2018. First genome-wide analysis of the endangered, endemic lichen *Cetradonia linearis* reveals isolation by distance and strong population structure. *American Journal of Botany* 105, 1556-1567.
- Alonso-García, M., Grewe, F., Payette, S., Villarreal A., J.C., 2021. Population genomics of a reindeer lichen species from North American lichen woodlands. *American Journal of Botany* 108, 159-171.
- Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F.K., Müller, N.F., Ogilvie, H.A., du Plessis, L., Poppinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M.A., Wu, C.-H., Xie, D., Zhang, C., Stadler, T., Drummond, A.J., 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology* 15, e1006650.
- Bowler, P.A., Marsh, J.E., 2004. *Niebla* . In: Nash, T.H., III, Ryan, B.D., Diederich, P.C., Gries, C., Bungartz, F. (Eds.), *Lichen Flora of the Greater Sonoran Desert Region*, Vol. 2. Lichens Unlimited, Arizona State University, Tempe, Arizona, pp. 36-80.
- Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A., 2012. Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis. *Molecular Biology and Evolution* 29, 1917-1932.

- Camargo, A., Sites, J.W., 2013. Species Delimitation: A Decade After the Renaissance. In: Pavlinov, I. (Ed.), The Species Problem - Ongoing Issues. InTech.
- Carstens, B.C., Pelletier, T.A., Reid, N.M., Satler, J.D., 2013. How to fail at species delimitation. *Molecular ecology*, n/a-n/a.
- Chifman, J., Kubatko, L., 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30, 3317-3324.
- Chou, J., Gupta, A., Yaduvanshi, S., Davidson, R., Nute, M., Mirarab, S., Warnow, T., 2015. A comparative study of SVDquartets and other coalescent-based species tree estimation methods. *BMC Genomics* 16.
- Culberson, C.F., 1972. Improved conditions and new data for identification of lichen products by standardized thin-layer chromatographic method. *Journal of Chromatography A* 72, 113-125.
- Dayrat, B., 2005. Towards integrative taxonomy. *Biological Journal of the Linnean Society* 85, 407-415.
- de Queiroz, K., 1998. The general lineage concept of species, species criteria, and the process of speciation: a conceptual unification and terminological recommendations. Oxford University Press, Oxford.
- de Queiroz, K., 2005. Ernst Mayr and the modern concept of species. *Proc Natl Acad Sci USA* 102, 6600 - 6607.
- de Queiroz, K., 2007. Species Concepts and Species Delimitation. *Systematic Biology* 56, 879-886.
- Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 24, 332-340.
- Edwards, S.V., 2009. Is a new and general theory of molecular systematics emerging? *Evolution* 63, 1-19.
- Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology* 61, 717-726.
- Flouri, T., Jiao, X., Rannala, B., Yang, Z., 2018. Species Tree Inference with BPP Using Genomic Sequences and the Multispecies Coalescent. *Molecular Biology and Evolution* 35, 2585-2593.
- Fujita, M.K., Leaché, A.D., Burbrink, F.T., McGuire, J.A., Moritz, C., 2012. Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution* 9, 480-488.
- Grewe, F., Huang, J.-P., Leavitt, S.D., Lumbsch, H.T., 2017. Reference-based RADseq resolves robust relationships among closely related species of lichen-forming fungi using metagenomic DNA. *Scientific Reports* 7, 9884.
- Grewe, F., Lagostina, E., Wu, H., Printzen, C., Lumbsch, H.T., 2018. Population genomic analyses of RAD sequences resolves the phylogenetic relationship of the lichen-forming fungal species *Usnea antarctica* and *Usnea aurantiacoatra*. *MycoKeys*, 91-113.
- Hale, H., Gardner, E.M., Viruel, J., Pokorny, L., Johnson, M.G., 2020. Strategies for reducing per-sample costs in target capture sequencing for phylogenomics and population genomics in plants. *Applications in Plant Sciences* 8, e11337.
- Harrison, R.G., Larson, E.L., 2014. Hybridization, Introgression, and the Nature of Species Boundaries. *Journal of Heredity* 105, 795-809.
- Hey, J., Waples, R.S., Arnold, M.L., Butlin, R.K., Harrison, R.G., 2003. Understanding and confronting species uncertainty in biology and conservation. *Trends Ecol. Evol.* 18, 597.
- Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* 35, 518-522.

- Hou, Y., Nowak, M.D., Mirré, V., Bjorå, C.S., Brochmann, C., Popp, M., 2015. Thousands of RAD-seq Loci Fully Resolve the Phylogeny of the Highly Disjunct Arctic-Alpine Genus *Diapensia* (Diapensiaceae). *PLOS ONE* 10, e0140175.
- Huang, J.-P., 2020. Is population subdivision different from speciation? From phylogeography to species delimitation. *Ecology and Evolution* 10, 6890-6896.
- Huang, J.-P., Knowles, L.L., 2015. The species versus subspecies conundrum: quantitative delimitation from integrating multiple data types within a single Bayesian approach in Hercules beetles. *Systematic Biology* in press, 10.1093/sysbio/syv1119.
- Jackson, N.D., Morales, A.E., Carstens, B.C., O'Meara, B.C., 2017. PHRAPL: Phylogeographic Inference Using Approximate Likelihoods. *Systematic Biology* 66, 1045-1053.
- Kapli, P., Yang, Z., Telford, M.J., 2020. Phylogenetic tree building in the genomic age. *Nature Reviews Genetics* 21, 428-444.
- Knowles, L.L., Carstens, B.C., 2007. Delimiting Species without Monophyletic Gene Trees. *Systematic Biology* 56, 887-895.
- Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome research* 27, 722-736.
- Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of Phylogenetic Estimates from Concatenated Data under Coalescence. *Systematic Biology* 56, 17-24.
- Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9, 357-359.
- Leaché, A.D., Fujita, M.K., Minin, V.N., Bouckaert, R.R., 2014. Species Delimitation using Genome-Wide SNP Data. *Systematic Biology* 10.1093/sysbio/syu018.
- Leaché, A.D., Zhu, T., Rannala, B., Yang, Z., 2018. The spectre of too many species. *Systematic Biology* 68, 168-181.
- Leavitt, S.D., Fernández-Mendoza, F., Pérez-Ortega, S., Sohrabi, M., Divakar, P.K., Lumbsch, H.T., St. Clair, L.L., 2013. DNA barcode identification of lichen-forming fungal species in the *Rhizoplaca melanophthalma* species-complex (Lecanorales, Lecanoraceae), including five new species *MycosKeys* 7, 1–22.
- Leavitt, S.D., Grewe, F., Widhalm, T., Muggia, L., Wray, B., Lumbsch, H.T., 2016. Resolving evolutionary relationships in lichen-forming fungi using diverse phylogenomic datasets and analytical approaches. *Scientific Reports* 6, 22262
- Leavitt, S.D., Moreau, C.S., Lumbsch, H.T., 2015. The Dynamic Discipline of Species Delimitation: Progress Toward Effectively Recognizing Species Boundaries in Natural Populations. In: Upreti, D.K., Divakar, P.K., Shukla, V., Bajpai, R. (Eds.), *Recent Advances in Lichenology*. Springer India, pp. 11-44.
- Lücking, R., Aime, M.C., Robbertse, B., Miller, A.N., Ariyawansa, H.A., Aoki, T., Cardinali, G., Crous, P.W., Druzhinina, I.S., Geiser, D.M., Hawksworth, D.L., Hyde, K.D., Irinyi, L., Jeewon, R., Johnston, P.R., Kirk, P.M., Malosso, E., May, T.W., Meyer, W., Öpik, M., Robert, V., Stadler, M., Thines, M., Vu, D., Yurkov, A.M., Zhang, N., Schoch, C.L., 2020. Unambiguous identification of fungi: where do we stand and how accurate and precise is fungal DNA barcoding? *IMA Fungus* 11, 14.
- 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. *Proceedings of the National Academy of Sciences* 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.

- Luo, A., Ling, C., Ho, S.Y.W., Zhu, C.-D., 2018. Comparison of Methods for Molecular Species Delimitation Across a Range of Speciation Scenarios. *Systematic Biology* 67, 830-846.
- Mallet, J., 2008. Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philos Trans R Soc Lond B* 363, 2971 - 2986.
- Matute, D.R., Sepúlveda, V.E., 2019. Fungal species boundaries in the genomics era. *Fungal Genetics and Biology* 131, 103249.
- Mayden, R.L., 1997. A hierarchy of species concepts: The denouement in the saga of the species problem. In: Claridge, M.F., Dawah, H.A., Wilson, M.R. (Eds.), *Species: The Units of Biodiversity*. Chapman & Hall, London, pp. 381-424.
- McKain, M.R., Johnson, M.G., Uribe-Convers, S., Eaton, D., Yang, Y., 2018. Practical considerations for plant phylogenomics. *Applications in Plant Sciences* 6, e1038.
- Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* 37, 1530-1534.
- Moncada, B., Sipman, H., Lücking, R., 2020. Testing DNA barcoding in *Usnea* (*Parmeliaceae*) in Colombia using the internal transcribed spacer (ITS). *Plant and Fungal Systematics* 65, 358-385.
- Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32, 268-274.
- Orange, A., James, P.W., White, F.J., 2001. *Microchemical methods for the identification of lichens*. British Lichen Society, London.
- Pino-Bodas, R., Martín, M.P., Burgaz, A.R., Lumbsch, H.T., 2013. Species delimitation in *Cladonia* (Ascomycota): a challenge to the DNA barcoding philosophy. *Molecular Ecology Resources* 13, 1058-1068.
- Poelstra, J.W., Salmons, J., Tiley, G.P., Schüßler, D., Blanco, M.B., Andriambeloson, J.B., Bouchez, O., Campbell, C.R., Etter, P.D., Hohenlohe, P.A., Hunnicutt, K.E., Iribar, A., Johnson, E.A., Kappeler, P.M., Larsen, P.A., Manzi, S., Ralison, J.M., Randrianambinina, B., Rasoloarison, R.M., Rasolofson, D.W., Stahlke, A.R., Weisrock, D.W., Williams, R.C., Chikhi, L., Louis, E.E., Jr., Radespiel, U., Yoder, A.D., 2020. Cryptic Patterns of Speciation in Cryptic Primates: Microendemic Mouse Lemurs and the Multispecies Coalescent. *Systematic Biology*.
- Printzen, C., 2010. *Lichen Systematics: The Role of Morphological and Molecular Data to Reconstruct Phylogenetic Relationships*. pp. 233-275.
- Rambaut, A., Drummond, A.J., 2005. Tracer v1.3: MCMC Trace Analysis Tool.
- Rannala, B., 2015. The art and science of species delimitation. *Current Zoology* 61, 846-853.
- Rochette, N.C., Rivera-Colón, A.G., Catchen, J.M., 2019. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology* 28, 4737-4754.
- Rubin, B.E.R., Moreau, C.S., 2016. Comparative genomics reveals convergent rates of evolution in ant-plant mutualisms. *Nature Communications* 7, 1-11.
- Schneider, K., Resl, P., Spribille, T., 2016. Escape from the cryptic species trap: lichen evolution on both sides of a cyanobacterial acquisition event. *Molecular Ecology* 25, 3453-3468.
- Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., Consortium, F.B., 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences* 109, 6241-6246.

- Sites, J.W., Marshall, J.C., 2004. Operational criteria for delimiting species. *Annual Review of Ecology, Evolution, and Systematics* 35, 199-227.
- Spjut, R., Simon, A., Guissard, M., Magain, N., Sérusiaux, E., 2020. The fruticose genera in the Ramalinaceae (Ascomycota, Lecanoromycetes): their diversity and evolutionary history. *MycoKeys* 73, 1.
- Spjut, R.W., 1996. *Niebla* and *Vermilacinia* (Ramalinaceae) from California and Baja California. BRIT Press.
- Sukumaran, J., Knowles, L.L., 2017. Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences* 114, 1607.
- Tonini, J., Moore, A., Stern, D., Shcheglovitova, M., Orti, G., 2015. Concatenation and species tree methods exhibit statistically indistinguishable accuracy under a range of simulated conditions. *PLoS Curr* 7.
- Vaser, R., Sović, I., Nagarajan, N., Šikić, M., 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research* 27, 737-746.
- Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., Sivasundar, A., Seehausen, O., 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology* 22, 787-798.
- Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS One* 9, e112963.
- Widhelm, T.J., Grewe, F., Huang, J.-P., Ramanauskas, K., Mason-Gamer, R., Lumbsch, H.T., 2021. Using RADseq to understand the circum-Antarctic distribution of a lichenized fungus, *Pseudocyphellaria glabra*. *Journal of Biogeography* 48, 78-90.
- Wilken, P.M., Aylward, J., Chand, R., Grewe, F., Lane, F.A., Sinha, S., Ametrano, C., Distefano, I., Divakar, P.K., Duong, T.A., Huhndorf, S., Kharwar, R.N., Lumbsch, H.T., Navathe, S., Pérez, C.A., Ramírez-Berrutti, N., Sharma, R., Sun, Y., Wingfield, B.D., Wingfield, M.J., 2020. IMA Genome - F13. *IMA Fungus* 11, 19.
- Wilkins, J.S., 2011. Philosophically speaking, how many species concepts are there?
- Will, K., Mishler, B., Wheeler, Q., 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Syst Biol* 54, 844 - 851.
- Yang, L., Kong, H., Huang, J.-P., Kang, M., 2019. Different species or genetically divergent populations? Integrative species delimitation of the *Primulina hochiensis* complex from isolated karst habitats. *Molecular Phylogenetics and Evolution* 132, 219-231.
- Yang, Z., 2015. The BPP program for species tree estimation and species delimitation. *Current Zoology* 61, 854-865.
- Yang, Z., Rannala, B., 2010. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences* 107, 9264-9269.
- Yang, Z., Rannala, B., 2014. Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution*.
- Zhang, C., Zhang, D.-X., Zhu, T., Yang, Z., 2011. Evaluation of a Bayesian coalescent method of species delimitation. *Systematic Biology* 60, 747-761.

Figure 1 . Typic coastal fog desert community in Punta Mazo Nature Reserve, near San Quintín, Baja California, supporting diverse *Niebla* communities. A. *Niebla* communities on volcanic slopes on Volcán Sudoeste, Punta Mazo. B. soil dwelling *Niebla arenaria* near Bahía Falsa, San Quintín. C. *Niebla* communities on West-facing slopes of Monte Ceniza, Punta Mazo. D. *Niebla flagelliforma* (Leavitt 16856BF). E. *Niebla*

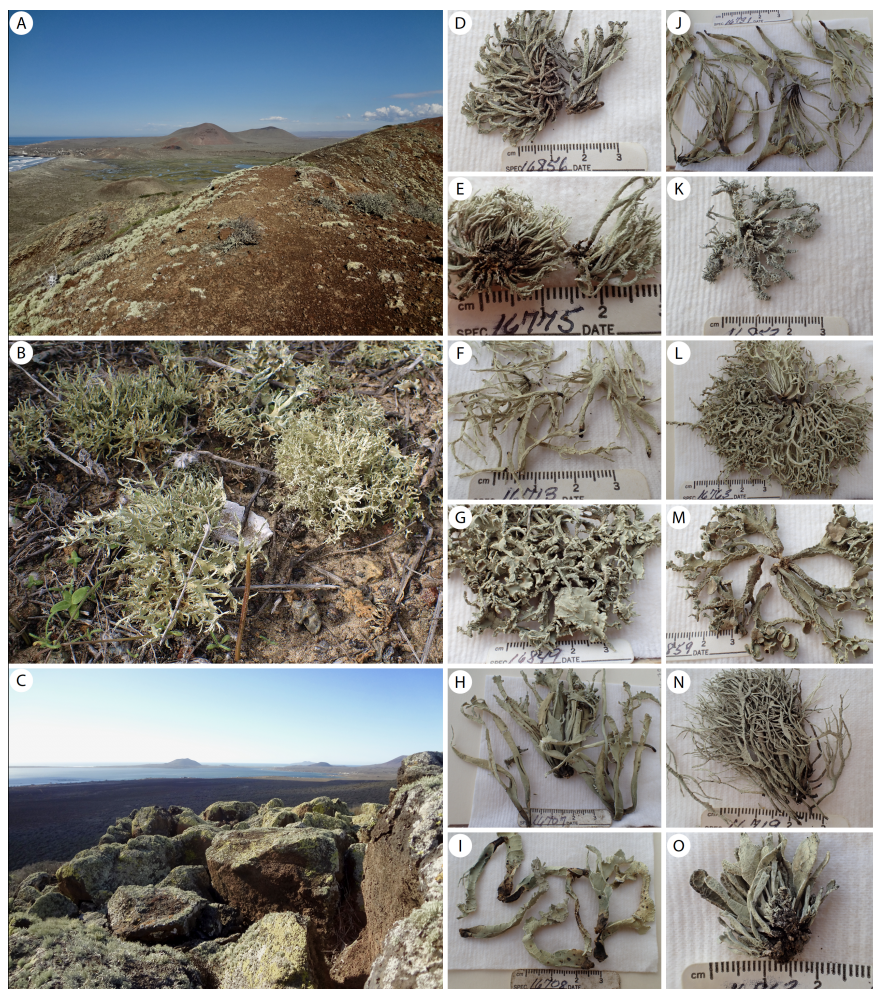
aff. isidiosa ; this specimen was identified as *N. aff. isidiaescens* and differs from *N. aff. isidiosa* in key morphological traits by the branches curved downwards near apex in contrast to the rigid terminal divaricate to reflex segments shown in panel ‘K’ (*Leavitt 16775*), which closely resembles the type from Isla Guadalupe. F. *Niebla flabellata* (*Leavitt 16713*). G. *Niebla* “sp. nov”. (*Leavitt 16849BF*). H. *Niebla juncosa* var. *juncosa* (*Leavitt 16707*). I. *Niebla undulata* (*Leavitt 16708*). J. *Niebla flagelliforma* (*Leavitt 16731*). K. *Niebla aff. isidiosa* ; (*Leavitt 16853*). L. *Niebla marinii* (*Leavitt 16763*). M. *Niebla lobulata* (*Leavitt 16859BF*). N. *Niebla juncosa* var. *spinulifera* (*Leavitt 16719*). O. *Niebla cornea* (*Leavitt 16863BF*). All specimens identified by R. Spjut.

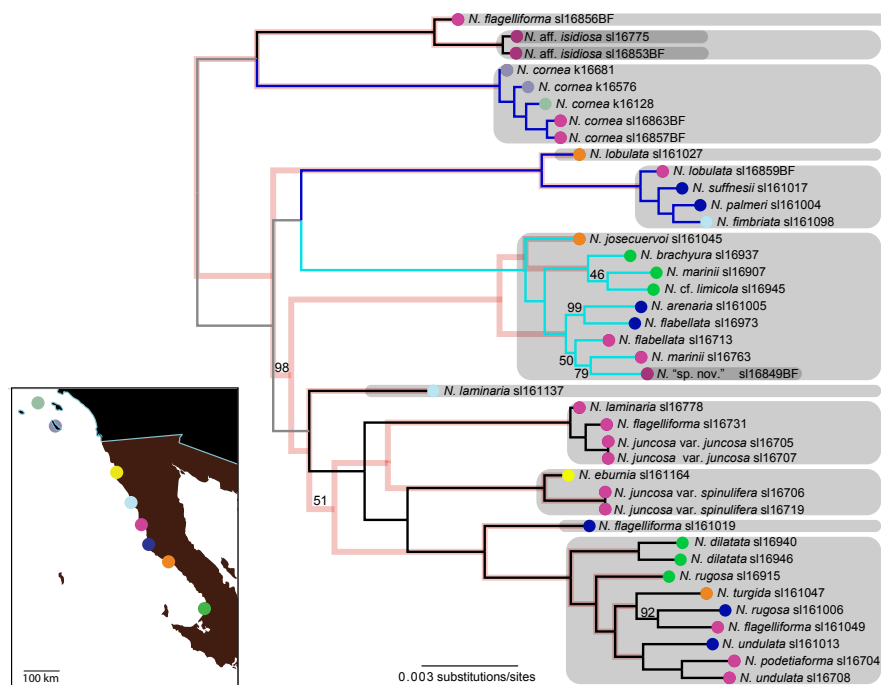
Figure 2 . ML topology inferred from 298,000 variable sites distributed across over 25,000 RADseq loci. Pink “shadow” tree is the species tree inferred under the multispecies coalescent model in PAUP+SVDquartets. Colors at tips correspond to sampling sites in map in bottom-left panel. Chemical clades are indicated with colored branches on the ML topology: black branches indicate lichens producing divaricatic acid (two clades), dark-blue branches indicate lichens producing sekikaic acid (two clades); and light blue branches indicate lichens producing salazinic acid. Grey boxes correspond to ASAP partitions from analyses of the standard DNA barcoding marker for fungi (ITS), with double shaded tips representing separate ASAP partitions that were combined for all subsequent analyses. Outgroup – *Vermalicinia* specimens – not shown.

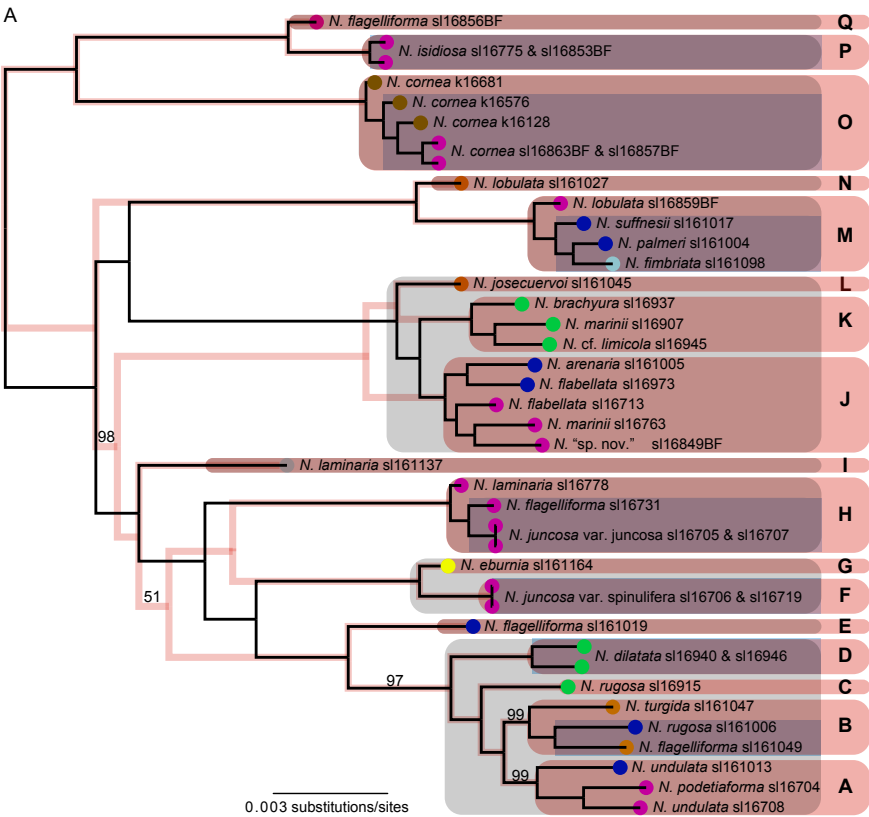
Figure 3 . Candidate species and species delimitation analyses using BPP+*gdi* . A. Clades labeled ‘A’ through ‘Q’, and highlighted in pink boxes, represent candidate species defined based on phylogenetic subdivision inferred from the ML and SVDquartets+PAUP tree inferences (Pink “shadow” tree is the species tree inferred under the multispecies coalescent model in SVDquartets+PAUP); tips collapsed in the “tip-down” BPP+*gdi* approach are highlighted with purple boxes. B. Heuristic *gdi* values inferred from different candidate species comparisons – see candidate species in panel ‘A’ – and using different subsampled RADseq datasets (5649 variable RADseq loci; three random subsets of 1000 variable loci; three random subsets of 500 variable loci; the 500 most variable loci; the 163 most variable loci; three random subsets of 100 variable loci; and the ten most variable loci); following Jackson et al. 2017, *gdi* scores [?] 0.2 indicate a single species (highlighted in blue), *gdi* scores [?] 0.7 indicate distinct species highlighted in yellow, and status for comparisons with *gdi* scores between 0.2 and 0.7 are ambiguous (highlighted in grey).

Figure 4 . Probability of different species delimitation models – species ‘A’ – ‘Q’, Fig. 3 – inferred using the ‘A10’ species delimitation model in the program BPP. The results are summarized from all analyses of RADseq data subsets, with the Y-axis shows the probability, in aggregate, for the inferred number of species.

Figure 5 . Distribution of *gdi* scores estimated from different subsets of RADseq loci (143 bp each).







B

	5649 loci	1000-a	1000-b	1000-c	500-a	500-b	500-c	500-most	163-most	100-a	100-b	100-c	10-most
Clade 1													
A-B	0.51	0.35	0.18	0.34	0.19	0.45	0.2	0.2	0.09	0.32	0.24	0.25	0.05
AB-C	0.55	0.51	0.6	0.73	0.39	0.58	0.53	0.31	0.17	0.54	0.43	0.56	0.1
ABC-D	0.54	0.72	0.56	0.73	0.72	0.73	0.73	0.33	0.22	0.65	0.53	0.61	0.16
ABCD-E	0.8	0.79	0.83	0.76	0.76	0.93	0.82	0.76	0.41	0.82	0.78	0.78	0.21
F-G	0.96	0.71	0.88	0.81	0.79	0.9	0.82	0.33	0.11	0.34	0.74	0.69	0.12
ABCD-EFG	0.8	0.88	0.82	0.82	0.75	0.94	0.88	0.77	0.6	0.88	0.84	0.87	0.26
ABCD-EFG-H	0.84	0.91	0.89	0.81	0.85	0.95	0.91	0.86	0.64	0.94	0.94	0.89	0.31
ABCD-EFGH-I	0.59	0.92	0.84	0.77	0.89	0.99	0.94	0.83	0.58	0.95	0.94	0.92	0.34
Clade 2													
J-K	0.42	0.34	0.41	0.42	0.57	0.41	0.57	0.36	0.41	0.57	0.51	0.25	0.1
JK-L	0.53	0.7	0.62	0.64	0.54	0.7	0.57	0.31	0.4	0.7	0.67	0.51	0.22
M-N	0.9	0.77	0.86	0.87	0.75	0.98	0.67	0.67	0.41	0.73	0.74	0.69	0.16
JKL-MN	0.92	0.91	0.87	0.91	0.95	0.96	0.91	0.93	0.57	0.95	0.95	0.93	0.32
Clade 2													
P-Q	0.95	0.66	0.68	0.76	0.7	0.82	0.17	0.63	0.3	0.39	0.75	0.29	0.14
PQ-O	0.98	0.99	0.96	0.99	0.99	0.99	1	0.98	0.49	0.99	0.95	0.935	0.29

