**Lineage and role in integrative taxonomy of a heterotrophic orchid complex**

Running title: **Lineage and role in integrative taxonomy**

Craig F. Barrett1, Mathilda V. Santee1, Nicole M. Fama1, John V. Freudenstein2, Sandra J. Simon3, Brandon T. Sinn4,5

1 Department of Biology, West Virginia University, 53 Campus Drive, Morgantown, West Virginia, USA 26506.

2 Department of Evolution, Ecology, and Organismal Biology, Ohio State University, 1315 Kinnear Rd., Columbus, Ohio, USA 43212.

3 Department of Biology, West Virginia University Institute of Technology, Beckley, WV, USA

4 Department of Biology and Earth Science, Otterbein University, Westerville, OH, USA.

5 Department of Botany and Ecology, University of Latvia, Jelgavas iela 1, Riga, LV-1004, Latvia

**ORCiD**:

Craig Barrett: 0000-0001-8870-3672

Mathilda V. Santee: 0000-0001-5349-724X

John V. Freudenstein: 0000-0001-9410-0624

Sandra J. Simon: 0000-0002-7928-3503

Brandon T. Sinn: 0000-0002-5596-6895

**Author for correspondence:** Craig F. Barrett, email: [craig.barrett@mail.wvu.edu](mailto:craig.barrett@mail.wvu.edu)

ABSTRACT

Lineage-based species definitions applying coalescent approaches to species delimitation have become increasingly popular. Yet, the application of these methods and the recognition of lineage-only definitions have recently been questioned. Species delimitation criteria that explicitly consider both lineages and evidence for ecological role shifts provide an opportunity to incorporate ecologically meaningful data from multiple sources in studies of species boundaries. Here, such criteria were applied to a problematic group of mycoheterotrophic orchids, the *Corallorhiza striata* complex, analyzing genomic, morphological, phenological, reproductive-mode, niche, and fungal host data. A recently developed method for generating genomic polymorphism data–ISSRseq–demonstrates evidence for four distinct lineages, including a previously unidentified lineage in the Coast Ranges and Cascades of California and Oregon, USA. There is divergence in morphology, phenology, reproductive mode, and fungal associates among the four lineages. Integrative analyses, conducted in population assignment and redundancy analysis frameworks, provide evidence of distinct genomic lineages and a similar pattern of divergence in the ‘extended’ data, albeit with weaker signal. However, none of the extended datasets fully satisfy the condition of a significant role shift, which requires evidence of fixed differences. The four lineages identified in the current study are recognized at the level of variety, short of comprising different species. This study represents the most comprehensive application of ‘lineage+role’ to date and illustrates the advantages of such an approach.

**Keywords**: Integrative species delimitation, ISSRseq, SNP data, species concept, species boundaries, Orchidaceae

**1. Introduction**

Species are a (or perhaps *the*) fundamental unit of biodiversity, conservation, evolution, and ecology; our understanding of biodiversity, and the scientific studies that deal with it, depend on what we empirically designate as species (e.g. Wright, 1940; Mayr, 1942; Dobzhansky, 1950; Hennig, 1966; Sneath and Sokal, 1973; Donoghue, 1985; Sterelny, 1999; Wilson, 1999; Sites and Marshall, 2003; Wiens, 2007; Camargo and Sites, 2013; Wilson, 2017; Pedraza-Marrón et al., 2019; Hillis et al., 2021). Focusing on *the species concept* (what species are; Mayden, 1997) most often points to the idea of an evolutionary lineage (metapopulation through time) as key, as described by Simpson (1951) in his *Evolutionary Species Concept* (ESC). Correspondingly, lineage-based empirical approaches to the practice of species delimitation (how we find species) have become popular (XXXX de Queiroz, 1998; 2005; 2007; Reeves and Richards, 2007) in part because genomic data make reconstructing them increasingly tractable, even under a variety of confounding processes [incomplete lineage sorting (ILS), gene flow, gene transfer (Jackson et al., 2017; Jiao and Yang, 2021)]; other species concepts or lines of evidence may also support the idea of an independently evolving lineage. Equating species with historical lineages as identified by genomic data forms the foundation of multispecies coalescent (MSC) delimitation models (Yang and Rannala, 2010; Rannala and Yang, 2020). Recent studies have questioned the relevance of ‘lineage’ as the sole criterion for recognizing species, however, and some methods designed to delimit species, e.g. the MSC (Freudenstein et al., 2017; Sukumaran and Knowles, 2017; Sukumaran et al., 2021; Wells et al., 2021), because they may not circumscribe the units we care about. Arguments against lineage-only approaches center around observations that such methods may tend to delimit population structure rather than species and are sensitive to biases in range-wide sampling (Sukumaran and Knowles, 2017; Chambers and Hillis, 2020; Mason et al., 2020; Hillis et al., 2021; Sukumaran et al., 2021). Ignoring aspects of phenotype described as inherent in the ESC by Simpson (1951) [and advocated more recently in de Queiroz (2007) and Freudenstein et al. (2017)] may leave us with units that are little more than historical constructs without meaning to biodiversity.

Over the last two decades, there has been an appreciation of the potential for integrative taxonomy to inform more multifaceted, robust estimates of biodiversity than those based on one-dimensional representations of variation (morphology or genetic sequences alone; Padial et al., 2005; Padial and de la Riva, 2011; Schlick-Steiner et al., 2010; Carstens et al., 2013; Krug et al., 2013; Edwards and Knowles, 2014; Leaché and Fujita, 2014; Solís‐Lemus et al., 2016; Cicero et al., 2021; Padial and de la Riva, 2020; Wells et al., 2021). Integrative taxonomy considers multiple data sources, ideally in an evolutionary context, and provides multiple lines of evidence in supporting or failing to support delimitations, including genomics, morphology/anatomy, ultrastructure, development/phenology, behavior, metabolites, and ecological interactions (Fujita et al., 2012; Carstens et al., 2013). A further distinction can be made between *iterative* and *integrative* taxonomy, which in the former, datasets are analyzed separately or in a specific order (Padial and de la Riva, 2010a,b), and in the latter they are analyzed simultaneously in a single framework (Yeates et al., 2011; Leaché and Fujita, 2012; Edwards and Knowles, 2014; Solís‐Lemus et al., 2016; Sukumaran et al., 2021).

Freudenstein et al. (2017) emphasized the co-equal importance of lineage and phenotype in their explication of Simpson’s ESC. Simpson’s use of the term *role* in his statement of the concept embodied the phenotypic aspect, which Freudenstein et al. (2017) viewed broadly as a distinct extended phenotype (Dawkins, 1982). Role here can be viewed as the interplay between extended phenotype and the evolutionary trajectory of a lineage, the basis of which is the hypothesis of a shift in character states to influence diversification by facilitating novel ecological interactions or ‘roles.’ Role has a long history, rooted in species concepts that incorporate evolutionary and ecological distinctness (Simpson, 1951; 1961; Van Valen, 1976; Mayr, 1982; 1998; Levin, 2000). ‘Extended phenotype,’ is an intentionally broad category, encompassing any fixed, observable differences among lineages in phenotype, ecological interactions (including symbionts), behavior, reproductive biology, etc., with the requirement that these differences are hypothesized to underlie the evolution of distinct ecological roles and therefore species. How can researchers diagnose role in the practice of species delimitation? Freudenstein et al. (2017) state: “...any fixed change in expressed organismal properties provides evidence for a hypothesis of role shift.” The emphasis on ‘fixed’ organismal properties provides a criterion for assessing patterns of variation among lineages, similar to the empirical concept of ‘diagnosability’(e.g., Nixon and Wheeler, 1990; Davis and Nixon, 1992), which is broader in that it accepts *any* fixed difference (such as a codon base change that is synonymous). This view of species has been applied or invoked in a handful of studies (Folk et al., 2017; Sinn, 2017; da Cruz et al., 2018; Zachos, 2018; Ackerfield et al., 2020; Frolov, 2021), but a comprehensive application is lacking.

Some organisms are inherently taxonomically challenging. Parasites, with reduced morphology and genomes, are especially problematic (Molvray et al., 2000; Cameron et al., 2004; Freudenstein et al., 2004; Evans et al., 2008; McNeal et al., 2013; Wicke et al., 2013; Haag et al., 2014; Li et al., 2019). Mycoheterotrophic plants are parasites specializing on and deriving nutrients from their fungal hosts (Bidartondo, 2005; Leake et al., 2005; Merckx, 2013). Mycoheterotrophs convergently display morphological reductions, lacking leaves or other features that contain character information for species delimitation (Merckx and Freudenstein, 2010; Tsukaya, 2018). They display gene losses (e.g. plastid and nuclear genomes), accelerated substitution rates, and horizontal gene transfer, which complicate phylogenomic analyses (Merckx et al., 2013; Barrett et al., 2014; Yang et al., 2016; Braukmann et al., 2017; Graham et al., 2017; Su et al., 2018; Yuan et al., 2018; Barrett et al., 2019). Parasites continue to be difficult to place, having been historically lumped into ‘trash bin’ morphologically-defined taxa, due to convergent losses (morphological and genomic), long branches in phylogenetic analyses, and missing data, requiring novel or specialized analytical approaches (e.g. Philippe et al., 2005; Pagel and Meade, 2008; Darriba et al., 2016; Givnish et al., 2018; Lam et al., 2018; Young and Gillung, 2019; Crotty et al., 2020). An integrative approach to taxonomy is essential for such groups, as it maximizes potential information for their placement, including species delimitation (Barrett et al., 2011; Broe, 2014; Freudenstein and Barrett, 2014).

One problematic group is the *Corallorhiza striata* complex, a wide-ranging, variable orchid of uncertain status (Fig. 1). This complex has a broad but patchy distribution, is locally rare, has reduced morphology (leaflessness and rootlessness), and displays clinal patterns of morphological variation (Freudenstein, 1997). Recent analyses have recognized three species: *Corallorhiza bentleyi*, a threatened species restricted to the eastern USA, known from ~12 populations; *C. involuta* (formerly *C. striata* var. *involuta*), a poorly known, disjunct relative of *C. bentleyi* endemic to Mexico; and *C. striata*, a wide-ranging, variable species distributed from Mexico to Canada (Freudenstein, 1997; Magrath and Freudenstein, 2002; Barrett & Freudenstein, 2011). *Corallorhiza bentleyi* and *C. involuta* share a close relationship yet are divergent from *C. striata* based on nuclear markers and plastid genomes (Barrett and Freudenstein, 2011; Barrett et al., 2018). Here we focus exclusively on *C. striata*, within which two varieties have been recognized: var. *vreelandii*, distributed among sky island forests of Mexico and the southwestern USA, and var. *striata*, distributed from the northern-US Rocky Mountains into southern Canada.

Taxonomic delimitation within *C. striata* is challenging. Freudenstein (1997) noted clinal patterns of morphological variation across North America, with *C. striata* var. *vreelandii* (Fig. 1A) having smaller, less-open flowers, suggesting a reproductive mode of self-pollination. *Corallorhiza striata* var. *striata* (Fig. 1B) has large, open flowers, pollinated by ichneumonid wasps (Freudenstein, 1997; *personal observation*). DNA sequences and morphology showed distinctness among them, but with some overlap in morphological features, making field- or herbarium-based identification difficult (Barrett and Freudenstein, 2009; 2011). Adding to the confusion, populations in California, USA (Fig. 1C, D), are morphologically intermediate between the two recognized varieties(Barrett and Freudenstein, 2009; 2011), having been proposed as introgressants by some researchers (e.g. Magrath and Freudenstein, 2002). Plastid genomes revealed that populations sampled from the Sierra Nevada (California) comprise a unique lineage, sister to the clade of vars. *striata* and *vreelandii*, within which these two varieties display a pattern of reciprocal monophyly (Barrett and Freudenstein, 2009; 2011). Further, analysis of fungal hosts across North America revealed specificity among the *C. striata* complex, and particularly among Sierran populations, with this plastid clade exclusively associating with a single clade of ectomycorrhizal fungi within *Tomentella fuscocinerea*, whereas the two currently recognized varieties associate more broadly with genotypes of the same species (Barrett et al., 2010). *Tomentella* is a diverse, globally distributed genus of ~80 species of ectomycorrhizal fungi within the family Thelephoraceae, and is often common in forest communities, especially in those dominated by gymnosperms (Larsen, 1974; Jakucs and Erős-Honti, 2008).

To date, few collections from the Coast Ranges of California or the Cascades of Oregon have been studied, calling into question their taxonomic affinities relative to Sierran populations and vars. *striata* and *vreelandii*. A recent analysis of plastid genomes included five accessions from the Coast Ranges (California) and Cascades (western Oregon), indicating that these accessions were nested among accessions of *C. striata* var. *striata* (Barrett et al., 2018). While plastid DNA is informative and displays accelerated evolutionary rates in *C. striata* (below the species level), it represents one, uniparentally inherited record of evolutionary history, and is subject to confounding effects of ILS and past/present gene flow (Willyard et al., 2009; Folk et al., 2017; Garcia et al., 2017; Gernandt et al., 2018; Doyle, 2021; Rose et al., 2021). Further, the Coast Range/Cascade populations have been observed to flower earlier than those in the rest of North America (February-April in the California Coast Ranges vs. May-July elsewhere; Coleman, 1995), possibly representing a temporal barrier to gene flow. Lastly, none of the Coast/Cascades populations have been sampled for fungal associates, calling into question whether they show a similar pattern of specificity as do those in the Sierra Nevada relative to vars. *vreelandii* and *striata*.

Here we analyze data from genome-wide single nucleotide polymorphisms, floral morphometry, phenology, reproductive mode, abiotic niche, and patterns of fungal host specificity to quantify and analyze variation in *Corallorhiza striata* within the framework of ‘lineage+role.’ The rationale for including each of these datasets is as follows. Genomic data can provide a powerful source of evidence for distinct lineages (de Queiroz, 2007; Leache et al., 2014), whereas phenotypic data can provide the basis for hypotheses of role shifts (*sensu* Freudenstein et al., 2017). Further, shifts in role may be attributed to changes in reproductive features, forming barriers to gene flow (phenology, selfing vs. outcrossing), or differences in features hypothesized to be driven by variation in abiotic niches. Lastly, mycoheterotrophic orchid lineages are known to have undergone significant host shifts from saprotrophic ‘rhizoctonia,’ the typical associates of photosynthetic orchids, to ectomycorrhizal fungi, and are often highly specific towards their host fungi (e.g. Taylor and Bruns, 1997). This potentially equates to novel and highly specific interactions that have been hypothesized to drive speciation (e.g. Taylor et al., 2004).

Specifically, we ask: How many species comprise this complex, aside from *C. bentleyi* and *C. involuta*? Can an integrative approach to taxonomy shed light on species-level delineation in this complex? How can role be applied in this complex as a defining factor in delimiting species, and what is role here? We conduct the first in-depth analysis and application of ‘ISSRseq’ (Sinn, Simon et al., 2021), an economical, straightforward approach to generating SNP data, and demonstrate that these data are highly informative within the *C. striata* complex. We hypothesized that morphology, phenology, reproductive mode, and fungal associates may each satisfy role in this complex. Here, we consider a sufficient definition of species boundaries to be one that requires separately evolving lineages [*sensu* de Queiroz (2007)], with the added requirement of displaying evidence of fixed differences in some aspect of the ‘extended phenotype.*’* To our knowledge, this study represents the most comprehensive application of ‘lineage+role’ to date in integrative species delimitation, the implications of which are discussed.

**2. Materials and Methods**

**2.1 Sample collection and DNA extraction**

Tissues were collected across the USA, Mexico, and western Canada between 2005 and 2017 (1-5 individuals each among 53 localities, 200 individuals total; Table S1). We aimed to maximize broad geographic sampling, focusing on the center of diversity in western North America, but also sampled in putative contact zones in northern Utah and California/Oregon (Fig. 1E). A CTAB protocol was performed to isolate shoot (floral) and rhizome (plant+fungal) DNA (Doyle and Doyle, 1987).

**2.2 ISSRseq and SNP analyses**

In order to determine evidence for distinct lineages within *C. striata*, we used a recently published protocol, ISSRseq, as an effective method of reduced-representation sequencing, outlined in Sinn, Simon et al. (2021). Briefly, this method involves PCR amplification and high-throughput sequencing of inter-simple sequence regions (ISSR), amplified with single primers that bind to microsatellites motifs (e.g. Zietkiewicz et al., 1994). Instead of scoring presence/absence of bands as dominant markers, as is traditionally done with ISSR, we conduct library preparations and Illumina sequencing of these regions, resulting in genomic-scale datasets of codominant SNP markers. ISSR regions were amplified for 87 *C. striata* individuals from 26 localities with eight primers in single-primer reactions, with a focus on western North America, specifically California. Amplification, library preparation, Illumina sequencing, bioinformatic processing, and the data themselves are described in Sinn, Simon et al. (2021). Biallelic SNPs were filtered using “phrynomics” in R to remove non-binary and non-informative SNPs (Leaché et al., 2015). SNPs were used to infer relationships among individuals with RAxML-NG using 1,000 bootstraps with the best-fit model: GTR+G4m+ASC (Kozlov et al. 2019). Bootstrap support was assessed with 1,000 standard bootstrap pseudoreplicates. Another tree was built using IQtree2 (Nguyen et al., 2015; Minh et al., 2020) with best fit model TVM+F+ASC+R3 as determined by ModelFinder (Kalyaanomoorthy et al., 2017). The ‘ASC’ option accounts for ascertainment bias in SNP-only data (Lewis, 2001; Leaché et al., 2015). Support was assessed with 1,000 ultrafast bootstrap pseudoreplicates. A coalescent-based “species tree” was estimated with SVDquartets in PAUP v.4 with 1,000 bootstraps (coalescent model, sampling all quartets, Erik+2 normalization; Swofford, 2003; Chifman and Kubatko, 2014; Chifman and Kubatko, 2015).

To test among alternative species delimitation scenarios using genomic data, Bayes Factor Delimitation (Leaché et al., 2014) was conducted via SNAPP (SNAPP-BFD\*) in BEAST2 (Bouckaert et al., 2019). Path sampling was conducted via stepping stone analysis (Kass and Raftery, 2005; Baele et al., 2012; Grummer et al., 2014) to obtain maximum likelihood estimates (MLE) under four different scenarios: 1) four species (*striata*, *vreelandii*, Coast/Cascades, Sierra), 2) three species (*striata*, *vreelandii*, Coast/Cascades+Sierra, 3) three species (*striata*+*vreelandii*, Coast/Cascades, Sierra), and 4) two species (*striata*+*vreelandii*, Coast/Cascades+Sierra). Due to the computational requirements of SNAPP, we reduced the 83-accession dataset to 26, randomly choosing one per locality. As above, sites were removed in ‘phrynomics’ to reduce the missing data to <5 % (Leaché et al., 2014). We ran 48 ‘steps’ for each of three independent runs for each scenario to check convergence among replicates (30 cores, 512 Gb RAM). Each chain was run for 250,000 generations, with 20% burn-in, after pre-burn-in of 50,000 generations. We verified stationarity and convergence across runs in Tracer v.1.7.1 (Rambaut et al., 2018), requiring effective sample sizes >200. Results were summarized by ranking each model/scenario based on the lowest MLE score, using the formula: BF=2×(MLE1–MLE0), in which model0 had the highest MLE. Bayes Factors < –10 indicate ‘decisive’ support for model0 over model1.

Population structure was characterized using Principal Components Analysis (PCA) in Adegenet v.2.1.5 (Jombart, 2008; Jombart and Ahmed, 2011), after linkage disequilibrium (LD) thinning to one SNP per locus in VCFTOOLS (Danecek et al., 2011); thinning was set to 3kb, greater than the length of the longest assembled contig. Global FST among lineages was assessed with hierfstat (Goudet, 2005). Analysis of admixture was conducted in DyStruct v1.1.0 (Joseph and Pe’er, 2019), with K=1-8, to determine the most likely number of ancestral clusters. Runs were replicated ten times for each K-value, specifying a hold-out of 0.05, starting each replicate from a random seed, and increasing the epochs to 100. MLEs across each value of K were used to calculate ‘Delta K’ to determine the most likely number of clusters (Evanno et al., 2005).

To compare resolution of SNPs and microsatellites from ISSRseq data, we used ‘SSRGenotyper’ (Lewis et al., 2020; <https://github.com/dlewis27/SSRgenotyper>), which identifies 2-, 3-, and 4-mer motifs with MISA (Beier et al., 2017) from .sam files and a reference sequence, requiring >100 bp of flanking sequence. Cleaned reads were mapped to the de novo reference sequence (see Sinn, Simon et al., 2021 for details) with BWA-MEM (Li and Durbin, 2009), and processed with SAMTOOLS (sorting, removing duplicates, filtering on mapping quality of -q 45). GENEPOP output was read into ‘adegenet’ v.2.1.3 (Jombart, 2008) in R, as was the LD-thinned SNP dataset. We used Discriminant Analysis of Principal Components (DAPC; Jombart et al., 2010), choosing the optimal number of PCs to retain via cross-validation and a-score optimization.

**2.3 Morphology**

To test for morphological distinctness among previously identified lineages, and whether they constitute evidence for role shifts, we measured fourteen floral characters for a single flower from each of 127 individual plants (as in Barrett and Freudenstein, 2011) using a stereo microscope (Fisher Scientific, Waltham, Massachusetts, USA). All flowers were field-collected and fixed in 10% formalin, 50% ethanol, and 5% glacial acetic acid. Measurements were calibrated to a 2 mm slide and analyzed with Image J v.1.53e (Abràmoff et al., 2004). PAST v.4.09 software (Hammer et al., 2001) was used to conduct PCA of log10-transformed data with a variance-covariance matrix and Multivariate Analysis of Variance (MANOVA). Plots were constructed using ‘ggpubr’ v.0.4.0 in R (Kassambara, 2018; Wickham et al., 2019).We further constructed a ‘phylomorphospace’ with the R package ‘phytools’ v.0.7.47 (Revell, 2012), using the RAxML-NG tree and morphological PC scores.

**2.4 Abiotic niche**

We used publicly available collections records to test whether var. *striata*, var. *vreelandii*, Coast/Cascades, and Sierran populations occupy different abiotic niche space, to explore whether differences in occupancy of niche space constitute evidence (as a proxy) for shifts in role. All records from the Global Biodiversity Information Facility (<https://www.gbif.org>; accessed 05 January, 2022) were downloaded specifying georeferenced records and preserved specimens, with the R package ‘rgbif’ v.3.6.0 (Chamberlain et al., 2022).’ Occurrences were then filtered for coordinate uncertainty with CoordinateCleaner v.2.0.20 in R (Zizka et al., 2019). Records from US States/Counties, Mexican States, and Canadian Provinces where each of the aforementioned entities are known to be distributed were manually curated and assigned to one of the four groups. When available, digitized images were inspected to verify records. Nineteen BIOCLIM variables (<https://worldclim.org>) were downloaded at 10 km resolution in R via the ‘raster’ v.3.5.2 package (Hijmans, 2021). The full BIOCLIM dataset was subjected to Principal Components Analysis in Past v.4 (Hammer at al., 2001), specifying a correlation matrix to account for differences in the scale of the variables.

The full BIOCLIM dataset was then subjected to pairwise Pearson correlation analysis using ENMTools v.1.0.5 (Warren et al., 2021), retaining only uncorrelated BIOCLIM variables at a 0.7 threshold. Data were spatially thinned to a minimum distance of 10 km to reduce the effects of autocorrelation with ‘SPthin’ v.0.2.0 in R (Aiello-Lammens et al., 2015). The ‘ecospat’ function (Di Cola et al., 2017) via ENMTools was used to conduct niche identity and niche background/similarity tests in R, using two metrics: Schoener's *D* (Schoener, 1968) and Warren’s *I* (Warren et al., 2008). Ecospat identifies available and occupied environmental niche space as an N-dimensional representation via kernel density (Broennimann et al., 2012; Di Cola et al., 2017). We performed pairwise identity tests among each of the four entities, choosing 1,000 background points at 10 km resolution with 999 permutations. The niche identity test generates a null distribution of overlaps and asks whether two entities are drawn from the same distribution. Further, we conducted background/similarity tests, which address expected similarity between groups based on environments available to them, and are more robust to situations of allopatry or parapatry (Warren et al., 2008).

**2.5 Phenology and reproductive mode**

To investigate reproductive evidence for shifts in role, records from the Global Biodiversity Information Facility (GBIF) were searched to visualize differences in flowering time among var. *vreelandii*, var. *striata*, Sierran, and Coast Range/Cascade populations. Only records post-1940 were retained, as those before this time had poorer annotation data and may not reflect current phenologies due to climate change. Flowering is not known to occur after mid-August in *C. striata* (Freudenstein, 1997), so records up to August 1 were kept. Collection dates were converted to days from January 1 within each year in R with ‘lubridate’ v.1.7.10 (Grolemund and Wickham, 2011). We searched individual records via the SEINet collections database (<https://swbiodiversity.org/seinet>) to compile information on pollination frequency. Specifically, we quantified the proportion of flowers per specimen that showed evidence of fruiting (large, swollen ovaries). Because populations from the California Coast Ranges typically flower much earlier (late February to mid-March; Coleman, 1995; Coleman et al., 2012), we included specimens collected after April 1 for these populations. One-way Analysis of Variance was used to test for significant differences in flowering time and the proportion of pollinated flowers among the groupings with Tukey comparisons in R.

**2.6 Fungal DNA analyses**

In order to quantify patterns of specificity and divergence in fungal associations among lineages, we extracted DNAs from rhizome tissues (mixed orchid and fungal DNAs) of 11 accessions form the Coast Ranges/Cascades were amplified and sequenced using fungal-specific primers ITS1F/ITS4 (White et al., 1990), following Barrett et al. (2010), and combined with sequence data sampled more broadly in the latter study. Electropherograms edited in Geneious v.10 (Drummond et al., 2011) and further aligned using the MAFFT v.7.490 plugin (‘auto’ algorithm; Katoh and Standley, 2013). In addition, we included >1,400 ITS reference sequences from the fungal family Thelephoraceae (NCBI GenBank), excluding ‘environmental’ sequences, for phylogenetic placement of fungal sequences derived from *C. striata*. *Thelephora terrestris* was chosen as an outgroup taxon (NCBI GenBank accession AF272923). *Corallorhiza*-derived fungal sequences were realigned with reference sequences in MAFFT. Ambiguously aligned regions with >95% gaps or missing data were removed in Geneious. Phylogenetic analysis was conducted with FastTree v.2.1.10 under a GTR model with 1,000 bootstrap pseudoreplicates (Price et al., 2009).

To test for phylogenetic signal in associations among the four groupings (i.e. nonrandom associations), we used the sub-clade from the fungal ITS tree containing all *C. striata* accessions, pruning reference *Tomentella* and associates of *C. bentleyi*+*involuta* with ‘ape’ v.5.5 in R (Paradis and Schliep, 2019). Associations were coded as a multistate, unordered character corresponding to each grouping. Phylogenetic signal was tested using a randomization test in R, 'phylo.signal.disc' (Bush et al., 2016; <https://github.com/juliema/publications/blob/master/BrueeliaMS/Maddison.Slatkin.R>), with 999 permutations to generate a null distribution of states on the tree. The script uses ‘ape,’ ‘geiger’ v.2.0.7, ‘phangorn’ v.2.5.5, and ‘phylobase’ v.0.8.10 (Pennell et al., 2014; Schliep et al., 2017; R Hackathon et al., 2020) to compare the null distribution of the number of character state transitions with the observed number on the tree.

**2.7 Integrative analyses**

Lastly, we conducted integrative analyses of multiple datasets to quantify their contribution to divergence among lineages, and to identify evidence for shifts in role. In order to analyze multiple datasets integratively, population assignment analyses were conducted with the R package ‘assignPOP’ v.1.2.4 (Chen et al., 2018). This software allows simultaneous integration of different data types, including SNP data, morphology, etc., by conducting principal components analysis on each dataset and then using linear discriminant analysis or various machine learning algorithms to estimate population assignments and ancestry coefficients. A proportion of the data is specified to train predictive models using Monte Carlo and K-fold cross-validation. We ran assignPOP for a reduced dataset of 47 accessions, including only accessions for which genomic (LD-thinned SNP data), morphological, abiotic niche, and fungal data were available for each accession, with each lineage representing a population cluster (i.e., K = 4). The fungal ITS subtree containing only *C. striata* accessions was transformed into Abouheif root-to-tip distances (Abouheif, 1999) in the R package ‘adephylo’ v.1.1.11 (Jombart and Dray, 2008). Abouheif distance relies on a phylogenetic proximity matrix, considering the inverse product of all direct descendants of a node, and thus captures the tree topology as opposed to more standard distance metrics. All data were log10-transformed.

Analyses were conducted using the ‘assign.MC’ (MCMC, Markov Chain Monte Carlo) and ‘assign.kfold’ cross-validation algorithms under LDA, support vector machine (SVM), and random forest (RF) algorithms. After testing different trial parameter sets and assignment accuracy with ‘accuracy.plot’, we specified 10% of SNP loci (minor allele frequency of < 0.05), the FST method, retaining the first ten principal components, setting K-fold=5, and running for 100 iterations (MCMC). We repeated each analysis three times with: (1) SNP data, (2) SNP + ‘extended’ data (morphological+abiotic niche+fungal data), and (3) only ‘extended’ data. The rationale was to test whether addition of the latter datasets markedly changed assignment probabilities or ancestry coefficients estimated from SNP data alone.

We also tested species delimitations integratively using Redundancy Analysis (RDA; Rao, 1964; Legendre and Legendre, 1999) in the R package ‘vegan’ v.2.5.7 (Oksanan et al., 2020), to address the hypothesis that ‘extended’ datasets provide evidence for role shifts with respect to the genomic lineages identified. We tested ‘four species’ vs. ‘three species’ delimitations, which in the latter case accessions from the Coast/Cascades and Sierra Nevada were considered a single species. Delimitations were coded as ‘dummy’ variables (0/1; Zapata and Jiménez, 2012; Papakostas et al., 2016; Firneno et al., 2021), and were used as potential explanatory variables in four RDA scenarios with different datasets as the response variables (representing role): (1) morphology only, (2) abiotic niche only, (3) fungal associates only (Abouheif distance, as above), and (4) combined ‘extended’ data (excluding SNP data). We ran RDA and variance partitioning (Peres Neto et al., 2006; Legendre et al., 2011) to quantify the unique and shared contributions among delimitations and test the significance of the adjusted R2 values for each dataset.

**3. Results**

**3.1 ISSRseq and SNP analyses**

The final ISSRseq dataset consisted of 27,117 SNPs (dataset S2), 6,589 of which remained after LD-thinning (dataset S3), and 2,662 of which remained after removal of non-parsimony informative sites with ‘phrynomics’ in R (dataset S4). Phylogenetic analyses based on informative SNPs yielded highly resolved and supported topologies (Fig. 2A-C). Relationships among major clades from RAxML-NG, SVDQuartets, and IQtree2 were congruent, with strong bootstrap support for four major groupings: *C. striata* var. *striata,* *C. striata* var. *vreelandii*, Sierran, and Coast/Cascade accessions. Bootstrap support was >90% for the sister relationships and for the monophyly of each of the four groupings. Topologies from RAxML-NG and IQtree differed slightly within major clades [Robinson-Foulds (RF) symmetric distance = 2], while both topologies differed to a greater degree from the SVDQuartets topology (RF symmetric distance = 74), though these topological differences were largely at mid-levels or towards the tips of the trees. Individuals from the same locality tended to group together in many cases (e.g. var. *striata* population 187 from Lewis and Clark County, Montana, USA; var. *vreelandii* population 163 from Ouray County, Colorado, USA; populations 312 and 432 from Santa Cruz County, California, USA; and population 253 from Fresno County, California, USA).

PCA of LD-thinned ISSRseq data (6,589 SNPs) revealed distinct genomic groupings based on the first three PCs (Fig. 2D), with PC1 differentiating var. *striata* and *vreelandii* from the Sierran and Coast/Cascade populations (12.8% of total variance), PC2 differentiating var. *vreelandii* and *striata* (10.5%), and PC3 differentiating Sierran and Coast/Cascade populations (4.9%). Overall FST was 0.57 (p < 0.001). Population structure analysis with DyStruct yielded the highest Delta K values of 139.8 and 92.3 for K = 3 and K = 4, respectively (Table S3). For K = 3, there were three distinct groupings corresponding to var. *vreelandii*, var. *striata*, and all Californian accessions, while with K = 4, the Californian accessions were distinct between the Sierra Nevada and Coast/Cascades. There is little evidence for admixture among the groupings, except in a few cases, most notably, accession 103e from Otero County, New Mexico, USA, with apparent mixed ancestry between var. *vreelandii* and var. *striata*. A few accessions from the Sierra Nevada show at least some proportion of the genome from the Coast/Cascade cluster, while a few accessions of the latter show at least some proportion of the genome from the var. *striata* cluster.

The SSRgenotyper pipeline recovered 19 polymorphic microsatellite loci from the ISSRseq data (79 alleles total; Dataset S5) after strict filtering. Discriminant analysis of principal components based on microsatellite repeats revealed a similar pattern of differentiation to that based on SNP data (Fig. S1). As in Fig. 2B, PC1 differentiates var. *striata*, var. *vreelandii*, and the Californian accessions, PC2 differentiates var. *striata* from the rest of the groupings, and PC3 differentiates the Sierra Nevadan accessions from the Coast/Cascade accessions. Overall, microsatellites show a somewhat weaker pattern of differentiation relative to that based on SNP data, but recover the same general signal.

Among the four alternative species delimitation models tested with SNAPP-BFD\*, the four-species model had the highest MLE and overwhelming support compared to the alternative species delimitation scenarios based on Bayes Factor comparisons (L = -175,646.5; Table 1). Both of the three-species models tested had lower Bayes Factors compared to the four-species model, with BF = -311.8 for treating the Sierran and Coast/Cascade accessions as a single species, and BF = -5,294.8 for treating vars. *vreelandii* and *striata* as a single species while keeping the Sierran and Coast/Cascade accessions separate. The two-species model also had lower support than the four-species model, with the former treating vars. *vreelandii* and *striata* as a species, while also treating the Sierran and Coast/Cascades accessions as a species (BF = -5,026.6).

**3.2 Morphology**

Analysis of 14 floral measurement characters via PCA and LDA revealed somewhat distinct groupings, but with evidence of overlap among them (Fig. 3A, C, E; Dataset S6). Perianth length/width characters were largely correlated with PC1 (69.77% of total variance, e.g. petal, sepal, and labellum characters; Table S4), while characters associated with the column and callus were correlated with PC2. PC1 reflects overall size differences among the larger-flowered var. *striata*, the small-flowered var. *vreelandii*, and the intermediate-sized Californian accessions. PC3 partially differentiates Sierran and Coast/Cascade individuals based on column and callus characters, but also to some degree on sepal and petal characters, though this axis only explains a small amount of the total variation (6.67%). LDA/MANOVA reveals statistically significant multivariate differences among all four groupings (Wilks’ lambda = 0.062, p < 0.0001), with significant, Bonferroni-corrected pairwise comparisons among all groupings (Fig. 3E, f; p < 0.0001 in all pairwise comparisons). Further, LDA/MANOVA classified 116 of 127 accessions (91.3%) according to their a priori assignment (Table 2). A phylomorphospace representation, combining the RAxML-NG tree with the PCA for morphology, further illustrates the genomic and morphological distinctness of vars. *vreelandii* and *striata*, but also displays overlap among the two Californian entities and vars. *striata* and *vreelandii* (Fig. 3G, H). A closer investigation of column and callus characters reveals that column length, column width (at base, narrowest point, and apex), and callus length/width ratio comprise most of the variation among Sierran and Coast/Cascade individuals. However, there is some overlap in the distributions of all of these characters.

**3.3 Abiotic niche**

Pairwise Niche Identity tests with Ecospat address the null hypothesis that niche overlap is constant when randomly reallocating the occurrences of both groups among their respective ranges. Among the four groupings, pairwise Niche Identity tests revealed that abiotic niche space was less identical than expected by chance in some comparisons but not others, based on the ‘D’ and ‘I’ statistics (Table 3; Dataset S7). *Corallorhiza striata* var. *striata* differed significantly from var. *vreelandii* and Coast/Cascades, but was not significantly different from Sierra Nevadan populations. *Corallorhiza striata* var. *vreelandii* differed from Sierra Nevadan populations, but not those from the Coast Ranges/Cascades, while the Sierran populations differed significantly from the Coast/Cascades populations. Ecospat Background/Similarity Tests further consider information on the geographic availability of environmental niche space. The results revealed that niche similarity was lower than expected by chance in all comparisons, but were overall inconclusive, failing to reject the null hypothesis among any of the comparisons (Table 3).

**3.4 Phenology and reproductive mode**

Flowering time, obtained from 1,743 herbarium records, differs significantly among the lineages (FANOVA = 72.48, p < 0.0001), with Coast/Cascade individuals having earlier flowering times compared to vars. *striata*, *vreelandii*, and Sierran individuals (Fig. 3J; pTukey HSD < 0.0001 for all comparisons; dataset S8). Considering individuals from the Coast Ranges vs. the Cascades separately, those from the California Coast Ranges have a slightly earlier flowering time on average than those from the Cascades (pTukey HSD < 0.001). Individuals of var. *vreelandii* show a higher proportion of swollen ovaries in herbarium records (n = 43 after data filtering; Fig. 3K; dataset S9) than var. *striata* and Californian individuals (FANOVA = 107.6, p < 0.0001; pTukey HSD < 0.0001 for all var. *vreelandii* comparisons), but the other three entities do not vary significantly.

**3.5 Fungal DNA analyses**

The final fungal ITS alignment (130 sequences from the *C. striata* complex + 1,451 *Tomentella* from GenBank), had a total, post-filtered length of 521 bp, with 139 parsimony informative sites (dataset S10). Analysis of *C. striata* fungal associates confirms a high level of specificity on a single clade of ectomycorrhizal *Tomentella*, corresponding to all reference sequences of *T. fuscocinerea* and *T. patagonica* (Fig. 4A). Within the fungal clade of *C. striata* associates, Sierran individuals predominantly associate with a single sub-clade (‘Clade III,’ comprising reference accessions of *T. fuscocinerea* and *T. patagonica*; πsierra = 0.010), while Coast/Cascade individuals associate largely with members of ‘clade II,’ representing accessions of *T. fuscocinerea* (πcoast/cascades = 0.030; Fig. 4B; Table S5). The closest relative to associates of the Sierran individuals is from Sweden, while the closest references to associates from the Coast/Cascades are from Minnesota (USA), Hungary, Iran, Mexico, Montenegro, and the U.K. (Fig. 4B). However, a single associate from the Sierra Nevada (9bR, Nevada County, California, USA) groups instead in ‘Clade II,’ while three associates from Marin County, California, and Josephine County, Oregon, USA are divergent from the remaining Coast/Cascade accession within ‘Clade II.’ Associates of vars. *striata* and *vreelandii* are widely dispersed throughout the tree, occupying Clades I-III (πvreelandii = 0.029, πstriata = 0.032). However, there is evident clustering of associates for each grouping and phylogenetic signal among associates with each of the four major groupings based on analysis with 'phylo.signal.disc' (Fig. S2; p < 0.001).

**3.6 Integrative analyses**

Results from assignPOP were congruent with other clustering methods (Fig. 5). Analysis of SNP-only data revealed a strong pattern of differentiation for LDA, SVM, and RF (Fig. 5A), with posterior probabilities of assignments = 1.0 in all cases (Table S6). Overall, the addition of morphological, abiotic niche, and fungal data did not change any of the analyses (with the exception of the combined data RF analysis for Coast/Cascades and Sierra Nevadan accessions, assignment probability = 0.88; Table S6), suggesting that these data, in combination, hold a similar signal of differentiation and do not controvert the patterns based on SNP-only data (Fig. 5B; Table S6). Analyses of the ‘extended data’ (i.e., including all but the SNP data) recovered similar overall patterns, but with an evidently weaker ability to discriminate the four lineages (Fig. 5C; Table S6). Variance partitioning via RDA, with three- and four-species hypotheses as explanatory variables, showed a preference for a four-species model for all ‘extended’ datasets individually and in combination (Fig. 5D). The models were significant in all comparisons (based on R2-adjusted values, p < 0.01 in all cases; Fig. 5D), but with the four-species model explaining a higher proportion of the overall variance.

**4. Discussion**

**4.1 Distinct lineages and integrative taxonomy in *C. striata***

We analyzed multiple datasets in an integrative approach to species delimitation of the *C. striata* complex. The most significant finding is the existence of four distinct lineages within this widespread species complex. Further, populations from the Coast Ranges/Cascades of California and Oregon, USA, not previously included in studies of *C. striata*, comprise a distinct lineage within *C. striata*, sister to a lineage restricted to the Sierra Nevada in California. Our objective was to balance broad geographic sampling with a focus on contact zones among putative lineages (Fig. 1). Recent studies have emphasized the importance of sampling in contact zones to avoid biases by artificially ‘discretizing’ patterns of variation by lack of sampling (Chambers and Hillis, 2020; Mason et al., 2020; Hillis et al., 2021). To the extent we were able to sample in two putative contact zones (northern Utah and northern California/southern Oregon), we find no evidence for introgression in these regions (Fig. 2D). Instead, we find sharp transitions, indicating distinct lineage boundaries, even for populations sampled within 200 km. The only notable individual of mixed ancestry is from Otero County, New Mexico, USA, but this is far from any putative contact zone, and long-distance dispersal cannot be ruled out (Fig. 2D, E). Additional sampling of contact zones, coupled with analyses based on nuclear SNPs and other data (Derryberry et al., 2014) would be beneficial in testing lineage distinctness. This may be challenging in the *C. striata* complex, occupying rare, clustered habitats in montane areas surrounded by unsuitable or anthropogenically altered habitats.

**4.2 Applying role to lineages of *C. striata***

What constitutes role in the *C. striata* complex? Clearly floral morphology, phenology, and reproductive mode are viable candidates, and could be related if differences among the groupings in floral morphometry are adaptive in pollinator specificity as in many other orchid species (Ackerman, 1983; Van Der Cingel, 2001; Sletvold et al., 2010; van der Kooi et al., 2021). However, so little is known about the pollination biology in this species complex that it is impossible to say at this point, requiring careful future observation and analysis of pollinators and floral morphometry across the broad geographic range of this complex.

The predominant mode of self-pollination hypothesized for *C. striata* var. *vreelandii* could have implications for role, in the sense of shifting away from insect pollination (Darwin, 1876; Ornduff, 1969; Richards, 1986; Sicard and Lenhard, 2011). In fact, vars. *striata* and *vreelandii* appear to be the most morphologically divergent members (Fig. 3A-I), with var. *striata* (large, open flowers; Fig. 1B) showing evidence of outcrossing and var. *vreelandii* showing evidence of self-pollination (drab, less-open, smaller flowers; Fig. 1A; Freudenstein, 1997). Floral morphology in these two varieties is influenced by pollination mode, which may play a role in that these two lineages may be reproductively isolated, with floral morphology serving as a proxy for this shift in ‘role.’ However, there are no clear, fixed differences in morphology (*viz*. non-overlapping floral character values) or reproductive mode other than overall size differences in the features measured, and thus we cannot conclude that there has been a role shift. Further, the intermediate morphological variation observed in both the Sierran and Coast/Cascade lineages casts doubt on whether there is sufficient evidence to recognize separate species based on morphology and reproductive mode. Earlier flowering time of the Coast/Cascade populations could represent a temporal reproductive barrier to gene flow among this and other lineages of *C. striata*. But again, a statistically significant difference in phenology does not equate to a fixed difference, such as is seen in the spring-flowering *Corallorhiza wisteriana* and the fall-flowering *C. odontorhiza* in northern North America, which are sister species (Freudenstein, 1992; 1997; Freudenstein and Barrett, 2014).

Although the abiotic niche cannot be considered part of the extended phenotype *per se*, it can be an agent of divergent selective pressure, modulating adaptations, and serving as a proxy for underlying genetic conditions (e.g. Cicero et al., 2021; Schmidt et al., 2021). Niche identity tests were significant in some cases, indicating less identity than expected by chance among lineages (Table 3), but only at 10 km resolution and not when considering niche data at finer scales (0.5 km resolution). Such tests should be interpreted with caution, however, especially if two lineages being compared are allopatric, in that the habitat of one may not be available to the other, thus making an unfair comparison (Warren, 2008; Cardillo and Warren, 2016; Warren et al., 2021). The background/similarity test attempts to account for this, being more robust in cases of allopatry. However, results of our ‘ecospat’ background/similarity tests were inconclusive at both spatial resolutions (less similarity than expected by chance, but non-significant in all comparisons; Table 3), which we interpret as a lack of decisive evidence for abiotic niche differences playing a clearly defined role in the speciation process.

Lastly, fungal host associations could play a role in speciation, in the contexts of coevolution and diversification by host-switching (Taylor et al., 2002; Waterman and Bidartondo, 2008; Taylor et al., 2013; Wang et al., 2021). Phylogenetic analysis of fungal hosts reveals high specificity, especially considering the predominant associations of Sierran and Coast/Cascade accessions with different fungal clades, with strong phylogenetic signal (Figs. 4b; S2; Barrett et al., 2010). However, the overall pattern of associations within *C. striata* reveals overlap among the four lineages, especially between vars. *vreelandii* and *striata*, and does not fit the hypothesized pattern of non-overlapping associations expected if these orchid lineages had fully co-diversified on their host fungi (Bidartondo and Bruns, 2001; Waterman and Bidartondo, 2008). Instead we interpret this pattern as strong but non-exclusive host preferences among lineages of *C. striata* towards different genotypes of *Tomentella*, potentially following a geographic mosaic pattern of specificity (sensu Thompson 1994; 2005; Barrett et al., 2010). One explanation for this pattern could be that ancestral polymorphisms in host associations are not yet fixed among the four lineages of *C. striata*, if we assume genetic control. Another could be strictly due to geography, in that certain *Tomentella* genotypes may be more regionally abundant, and that our sampling simply recapitulates this. Though collections of *T. fuscocinerea* and *T. patagonia* specimens are few (five records for North America; <https://gbif.org>), their global distribution suggests these fungal taxa are ubiquitous but locally rare or patchily distributed among forested habitats (Larsen, 1965; 1974; Jakucs and Erős-Honti, 2008; Kuhar et al., 2016). Seed burial experiments among sites with known divergent orchid and fungal associates may shed light on what controls this pattern of specificity (Rasmussen and Whigham, 1993; McCormick et al., 2012).

Mycoheterotrophic specificity in orchids and other clades has been studied extensively, and within the orchids, studies reveal frequent major host switches from typical saprotrophic “rhizoctonia” hosts to ectomycorrhizal hosts (Taylor et al., 2002; Leake, 2005; Bidartondo, 2005). These switches to ectomycorrizal hosts most often occur with a significant narrowing of host specificity, in many cases where plant species specifically target specific fungal families, genera, species, or even genotypes. Thus, fungal host specificity is hypothesized to be highly informative as a potential marker of species boundaries for the plants engaged in these parasitic symbioses, and further, these associations may play a role in the speciation process (Hynson and Bruns, 2010; Taylor et al., 2004; 2013). If divergent host associations among lineages somehow disrupt gene flow, then the role of host specificity is clear, as was suggested in the *Corallorhiza maculata* and *odontorhiza-wisteriana* complexes (Taylor and Bruns, 1997; Taylor et al., 2004; Freudenstein and Barrett, 2014). If host associations are genetically controlled, we would expect fixed polymorphisms co-adapted for fungal hosts to be frequent throughout the genome, and in LD within lineages associated with fungal host genotypes (Taylor et al., 2004; Taylor et al., 2013). Host associations could be indicators and drivers of shifts in ‘role.’ Our observation of strong host preferences but some degree of overlap in associations suggests at least at the level of the four *C. striata* lineages, that these associations do not satisfy the strict definition of ‘role,’ which emphasizes fixation (Fig. 4B; Freudenstein et al., 2017). However, finer-scale analyses of fixed differences and fungal specificity may reveal that role is in fact satisfied, if divergence in associations extends below the level of lineage, and instead represents specific orchid-fungal genotype-genotype specificity or ‘host races’ (*sensu* Taylor et al., 2004). Denser sampling, specifically targeted in geographically proximal, putative contact zones, may shed light on specificity in associations, which may not be detectable at range-wide scales.

The strongest case for recognizing species under lineage + role would be genomically distinct lineages, each supported by multiple, fixed traits hypothesized to be ecologically important underlying a role shift (though one fixed trait would be sufficient; Freudenstein et al., 2017). Examples might include different features of feeding apparatus associated with plant hosts, differences in reproductive structures that serve as premating barriers, or adaptations to different abiotic habitats. What if there are no obvious, detectable, fixed differences in any single extended phenotype trait among lineages, but many traits analyzed together, even from different datasets, show clear, emergent patterns of differentiation? Recognizing “clusters in phenotypic [or other] space” as species is best described as a phenetic approach (Sneath and Sokal, 1963; Michener, 1970; Sokal and Crovello, 1970). While phenetics is widely applied at the species level, many have criticized this approach as lacking evolutionary context (Hull, 1988; Donoghue, 1990; de Queiroz and Good, 1997). On one hand, if those features measured, even if not fixed for any single character, are synergistically linked (e.g. morphometric variation in tooth and jaw morphology associated with different food preference, canalized differences in floral dimensions associated with different frequencies of pollinator types), then one could hypothesize that there is an ecologically important role shift (Lockwood, 2006). Combined with lineage evidence, one could justify recognition of separate species, each occupying distinct “hyper-roles,” without detectable, fixed differences in any one feature. On the other hand, the problem of arbitrariness persists: where does one objectively draw the line that determines species boundaries in multivariate (or any single continuous character) space?

Some researchers may view our results, and our interpretation of a lack of sufficient evidence for separate species (requiring fixed differences in expressed organismal features), as overly conservative. An alternative interpretation might be that the similar patterns of divergence among the extended data and lineages, or simply the recovery of distinct lineages themselves, would be sufficient evidence for recognizing separate species. For example, some may be comfortable calling the different lineages species under the Unified Species Concept (*sensu* de Queiroz, 2007), secondarily applying phenetic (i.e. morphology) and phylogenetic species concepts (i.e. monophyly). Further, under another interpretation of the Phylogenetic Species Concept (i.e. diagnosability, *sensu* Nixon and Wheeler, 1990), a single synapomorphy, or suite of variant sites fixed for each lineage, would be sufficient evidence of separate species even without phenotypic evidence of fixation. However, applying this approach to genomic data introduces another level of ambiguity and subjectivity: how does one select the level at which to recognize species, given that there will be fixed synapomorphies at different hierarchical levels in any given dataset? We argue that such interpretations would bring us no closer to objectivity, and instead perpetuate a long-standing problem with species delimitation: a lack of repeatable and consistent criteria.

**5. CONCLUSION**

Here we have taken an integrative approach to species delimitation in the *Corallorhiza striata* complex, employing seven datasets, requiring distinct lineages and fixed differences in features of the extended phenotype as the basis for species recognition. We identified four genomic lineages based on nuclear data, with evidence for divergent roles based on morphology, phenology, reproductive mode, and fungal hosts. However, none of the metrics of extended phenotype (i.e. proxies for role shifts) satisfy the requirement of fixation, and thus we conservatively conclude that there is insufficient evidence for the recognition of distinct species within *C. striata*. Taken together, from the standpoints of biodiversity and functional ecology, our findings reveal that recognizing four distinct lineages here at the species level could equate to taxonomic ‘over-splitting,’ and that these lineages are more appropriately recognized as varieties. Our findings provide support for a previous hypothesis that these lineages represent evolutionarily significant units that may warrant more nuanced conservation approaches (Barrett and Freudenstein, 2011). Future studies should emphasize explicit sampling in zones of potential contact among lineages for all of the above metrics, and in particular, reciprocal seed-baiting experiments may prove to be informative on the nature of orchid-fungal specificity, the heritability of these associations, and potential fitness implications. Finally, the current study illustrates the importance of including the concept of role in integrative species delimitation beyond the recognition of historical lineages, which has important implications for biodiversity assessment.

**6. ACKNOWLEDGEMENTS**

Funding was provided by the US National Science Foundation award DEB CAREER: 2044259 to CFB; the West Virginia University Department of Biology and Eberly College; a NASA Space Grant to MVS; and by the West Virginia Research Challenge Fund through a grant from the Division of Science and Research, HEPC and in part by (i) the WVU Provost’s Office, (ii) the Eberly College of Arts and Sciences, (iii) the Honors College and (iv) the Department of Biology to MVS and NMF. We are grateful to the USDA Forest Service (FS-2400-1) and California State Parks (DPR65) for permission to collect plant tissues, and to Stephen DiFazio, Jun Fan, Ryan Percifield, and Donald Primerano for sequencing support and discussion. We thank the WVU Genomics Core Facility for support provided to help make this publication possible and CTSI Grant #U54 GM104942 which in turn provides financial support to the WVU Core Facility. We further acknowledge WV-INBRE (P20GM103434), a COBRE ACCORD grant (1P20GM121299), and a West Virginia Clinical and Translational Science Institute (WV-CTSI) grant (2U54GM104942) in supporting the Marshall University Genomics Core (Research Citation: Marshall University Genomics Core Facility, RRID:SCR\_018885).

**7. REFERENCES**

Abouheif, E. (1999). A method for testing the assumption of phylogenetic independence in comparative data. *Evolutionary Ecology Research*, 1, 895-909.

Ackerfield, J. R., Keil, D. J., Hodgson, W. C., Simmons, Mark. P., Fehlberg, S. D., & Funk, V. A. (2020). Thistle be a mess: Untangling the taxonomy of *Cirsium* (Cardueae: Compositae) in North America. *Journal of Systematics and Evolution*, *58*(6), 881–912.<https://doi.org/10.1111/jse.12692>

Ackerman, J. D. (1983). Specificity and mutual dependency of the orchid-euglossine bee interaction. *Biological Journal of the Linnean Society*, *20*(3), 301–314.<https://doi.org/10.1111/j.1095-8312.1983.tb01878.x>

Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, *38*(5), 541–545.<https://doi.org/10.1111/ecog.01132>

Baele, G., Li, W. L. S., Drummond, A. J., Suchard, M. A., & Lemey, P. (2013). Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Molecular Biology and Evolution*, *30*(2), 239–243.<https://doi.org/10.1093/molbev/mss243>

Barnard-Kubow, K. B., So, N., & Galloway, L. F. (2016). Cytonuclear incompatibility contributes to the early stages of speciation. *Evolution*, *70*(12), 2752–2766.<https://doi.org/10.1111/evo.13075>

Barrett, C. F., & Freudenstein, J. V. (2009). Patterns of morphological and plastid DNA variation in the *Corallorhiza striata* species complex (Orchidaceae). *Systematic Botany*, *34*(3), 496–504. doi: [10.1600/036364409789271245](https://doi.org/10.1600/036364409789271245)

Barrett, C. F., & Freudenstein, J. V. (2011). An integrative approach to delimiting species in a rare but widespread mycoheterotrophic orchid. *Molecular Ecology*, *20*(13), 2771–2786.<https://doi.org/10.1111/j.1365-294X.2011.05124.x>

Barrett, C. F., Freudenstein, J. V., Lee Taylor, D., & Kõljalg, U. (2010). Rangewide analysis of fungal associations in the fully mycoheterotrophic *Corallorhiza striata* complex (Orchidaceae) reveals extreme specificity on ectomycorrhizal *Tomentella* (Thelephoraceae) across North America. *American Journal of Botany*, *97*(4), 628–643.<https://doi.org/10.3732/ajb.0900230>

Barrett, C. F., Freudenstein, J. V., Li, J., Mayfield-Jones, D. R., Perez, L., Pires, J. C., & Santos, C. (2014). Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. *Molecular Biology and Evolution*, *31*(12), 3095–3112.<https://doi.org/10.1093/molbev/msu252>

Barrett, C. F., Sinn, B. T., & Kennedy, A. H. (2019). Unprecedented parallel photosynthetic losses in a heterotrophic orchid genus. *Molecular Biology and Evolution*, *36*(9), 1884–1901.<https://doi.org/10.1093/molbev/msz111>

Barrett, C. F., Wicke, S., & Sass, C. (2018). Dense infraspecific sampling reveals rapid and independent trajectories of plastome degradation in a heterotrophic orchid complex. *The New Phytologist*, *218*(3), 1192–1204.<https://doi.org/10.1111/nph.15072>

Beier, S., Thiel, T., Münch, T., Scholz, U., & Mascher, M. (2017). MISA-web: A web server for microsatellite prediction. *Bioinformatics*, *33*(16), 2583–2585.<https://doi.org/10.1093/bioinformatics/btx198>

Bidartondo, M. I. (2005). The evolutionary ecology of myco-heterotrophy. *New Phytologist*, *167*(2), 335–352.<https://doi.org/10.1111/j.1469-8137.2005.01429.x>

Bidartondo, M. I., & Bruns, T. D. (2001). Extreme specificity in epiparasitic Monotropoideae (Ericaceae): Widespread phylogenetic and geographical structure. *Molecular Ecology*, *10*(9), 2285–2295.<https://doi.org/10.1046/j.1365-294x.2001.01358.x>

Bolker, R., Butler, M., Cowan, P., Vienne, D. de, Eddelbuettel, D., Holder, M., Jombart, T., Kembel, S., Michonneau, F., Orme, D., O’Meara, B., Paradis, E., Regetz, J., & Zwickl, D. (2020). *phylobase: Base Package for Phylogenetic Structures and Comparative Data* (0.8.10). [https://CRAN.R-project.org/package=phylobase](https://cran.r-project.org/package=phylobase)

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., Maio, N. D., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., Plessis, L. du, Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., … Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, *15*(4), e1006650.<https://doi.org/10.1371/journal.pcbi.1006650>

Braukmann, T. W. A., Broe, M. B., Stefanović, S., & Freudenstein, J. V. (2017). On the brink: The highly reduced plastomes of nonphotosynthetic Ericaceae. *New Phytologist*, *216*(1), 254–266.<https://doi.org/10.1111/nph.14681>

Broe, M. (2014). Phylogenetics of the Monotropoideae (Ericaceae) with special focus on the genus *Hypopitys* Hill, together with a novel approach to phylogenetic inference using lattice theory. *Undefined*.<https://www.semanticscholar.org/paper/Phylogenetics-of-the-Monotropoideae-(Ericaceae)-on-Broe/8f2b05a8f9b268d18a5b221f1a4c11d881905a08>

Broennimann, O., Fitzpatrick, M. C., Pearman, P. B., Petitpierre, B., Pellissier, L., Yoccoz, N. G., et al. (2012). Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography*, *21*(4), 481–497. doi: [10.1111/j.1466-8238.2011.00698.x](https://doi.org/10.1111/j.1466-8238.2011.00698.x)

Bush, S. E., Weckstein, J. D., Gustafsson, D. R., Allen, J., DiBlasi, E., Shreve, S. M., Boldt, R., Skeen, H. R., & Johnson, K. P. (2016). Unlocking the black box of feather louse diversity: A molecular phylogeny of the hyper-diverse genus *Brueelia*. *Molecular Phylogenetics and Evolution*, *94*, 737–751.<https://doi.org/10.1016/j.ympev.2015.09.015>

Camargo, A., & Sites, J. J. (2013). Species delimitation: a decade after the renaissance. In: *The Species Problem—Ongoing Issues*. IntechOpen.<https://doi.org/10.5772/52664>

Cameron, K. M. (2004). Utility of plastid *psaB* gene sequences for investigating intrafamilial relationships within Orchidaceae. *Molecular Phylogenetics and Evolution*, *31*(3), 1157–1180.<https://doi.org/10.1016/j.ympev.2003.10.010>

Cardillo, M., & Warren, D. (2016). Analysing patterns of spatial and niche overlap among species at multiple resolutions: Spatial and niche overlap. *Global Ecology and Biogeography*, *25*.<https://doi.org/10.1111/geb.12455>

Carstens, B. C., Pelletier, T. A., Reid, N. M., & Satler, J. D. (2013). How to fail at species delimitation. *Molecular Ecology*, *22*(17), 4369–4383.<https://doi.org/10.1111/mec.12413>

Caruso, C. M., Case, A. L., & Bailey, M. F. (2012). The evolutionary ecology of cytonuclear interactions in angiosperms. *Trends in Plant Science*, *17*(11), 638–643.<https://doi.org/10.1016/j.tplants.2012.06.006>

Chamberlain S, Barve V, Mcglinn D, Oldoni D, Desmet P, Geffert L, Ram K (2022). rgbif: interface to the global biodiversity information facility API. R package version 3.5.2, [https://CRAN.R-project.org/package=rgbif](https://cran.r-project.org/package=rgbif)

Chambers, E. A., & Hillis, D. M. (2020). The multispecies coalescent over-splits species in the case of geographically widespread taxa. *Systematic Biology*, *69*(1), 184–193.<https://doi.org/10.1093/sysbio/syz042>

Chen, K.-Y., Marschall, E. A., Sovic, M. G., Fries, A. C., Gibbs, H. L., & Ludsin, S. A. (2018). assignPOP: An R package for population assignment using genetic, non-genetic, or integrated data in a machine-learning framework. *Methods in Ecology and Evolution*, *9*(2), 439–446.<https://doi.org/10.1111/2041-210X.12897>

Chifman, J., & Kubatko, L. (2014). Quartet Inference from SNP Data Under the Coalescent Model. *Bioinformatics*, *30*(23), 3317–3324.<https://doi.org/10.1093/bioinformatics/btu530>

Chifman, J., & Kubatko, L. (2015). Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *Journal of Theoretical Biology*, *374*, 35–47.<https://doi.org/10.1016/j.jtbi.2015.03.006>

Cicero, C., Mason, N. A., Jiménez, R. A., Wait, D. R., Wang-Claypool, C. Y., & Bowie, R. C. K. (2021). Integrative taxonomy and geographic sampling underlie successful species delimitation. *Ornithology*, *138*(2), ukab009.<https://doi.org/10.1093/ornithology/ukab009>

Coleman, R. A. (2002). *The Wild Orchids of California*. Cornell University Press.

Coleman, R. A., Wilken, D. H., & Jennings, W. F. (2012). Orchidaceae. In: Baldwin, B. G., Goldman, D., Keil, D. J., Patterson, R., Rosatti, T. J., & Wilken, D. (Eds.). *The Jepson Manual: Vascular Plants of California*, *2nd Edition*. University of California Press, Berkeley.

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer Associates, Sunderland, Massachusetts.

Crotty, S. M., Minh, B. Q., Bean, N. G., Holland, B. R., Tuke, J., Jermiin, L. S., & Haeseler, A. V. (2020). GHOST: recovering historical signal from heterotachously evolved sequence alignments. *Systematic Biology*, *69*(2), 249–264.<https://doi.org/10.1093/sysbio/syz051>

da Cruz, M. de O. R., & Weksler, M. (2018). Impact of tree priors in species delimitation and phylogenetics of the genus *Oligoryzomys* (Rodentia: Cricetidae). *Molecular Phylogenetics and Evolution*, *119*, 1–12.<https://doi.org/10.1016/j.ympev.2017.10.021>

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.<https://doi.org/10.1093/bioinformatics/btr330>

Darriba, D., Weiß, M., & Stamatakis, A. (2016). Prediction of missing sequences and branch lengths in phylogenomic data. *Bioinformatics*, *32*(9), 1331–1337.<https://doi.org/10.1093/bioinformatics/btv768>

Davis, J. I., & Nixon, K. C. (1992). Populations, genetic variation, and the delimitation of phylogenetic species. *Systematic Biology*, *41*(4), 421–435.<https://doi.org/10.2307/2992584>

Dawkins, R. (1982). *The Extended Phenotype: The Long Reach of the Gene*. Oxford, UK: Oxford University Press

de Queiroz, Kevin (1998). The general lineage concept of species, species criteria, and the process of speciation. In D. J. Howard & S. H. Berlocher (eds.), *Endless Forms: Species and Speciation*. Oxford University Press. pp. 57-75.

de Queiroz, K. (2005). Different species problems and their resolution. *BioEssays*, *27*(12), 1263–1269.<https://doi.org/10.1002/bies.20325>

de Queiroz, K. (2007). Species Concepts and Species Delimitation. *Systematic Biology*, *56*(6), 879–886.<https://doi.org/10.1080/10635150701701083>

de Queiroz, K., & Good, D. A. (1997). Phenetic Clustering in Biology: A Critique. *The Quarterly Review of Biology*, *72*(1), 3–30.<https://doi.org/10.1086/419656>

Darwin C. (1876). The effects of cross and self-fertilization in the vegetable kingdom. London: John Murray.

Derryberry, E. P., Derryberry, G. E., Maley, J. M., & Brumfield, R. T. (2014). hzar: Hybrid zone analysis using an R software package. *Molecular Ecology Resources*, *14*(3), 652–663.<https://doi.org/10.1111/1755-0998.12209>

Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F. T., D’Amen, M., Randin, C., Engler, R., Pottier, J., Pio, D., Dubuis, A., Pellissier, L., Mateo, R. G., Hordijk, W., Salamin, N., & Guisan, A. (2017). ecospat: An R package to support spatial analyses and modeling of species niches and distributions. *Ecography*, *40*(6), 774–787.<https://doi.org/10.1111/ecog.02671>

Donoghue, M. J. (1990). Sociology, selection, and success: A critique of David Hull's analysis of science and systematics. *Biology and Philosophy,* 5(4), 459-472.

Doyle, J. J. (2021). Defining coalescent genes: theory meets practice in organelle phylogenomics. *Systematic Biology*, *71*(2), 476–489.<https://doi.org/10.1093/sysbio/syab053>

Doyle, J. J., & Doyle, J. L. (1987). A Rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin, 19: 11-15.

Edgar, R. C. (2021). *MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping* [Preprint]. Bioinformatics.<https://doi.org/10.1101/2021.06.20.449169>

Edwards, D. L., & Knowles, L. L. (2014). Species detection and individual assignment in species delimitation: Can integrative data increase efficacy? *Proceedings of the Royal Society B: Biological Sciences*, *281*(1777), 20132765.<https://doi.org/10.1098/rspb.2013.2765>

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620.<https://doi.org/10.1111/j.1365-294X.2005.02553.x>

Evans, N. M., Lindner, A., Raikova, E. V., Collins, A. G., & Cartwright, P. (2008). Phylogenetic placement of the enigmatic parasite, Polypodium hydriforme, within the Phylum Cnidaria. *BMC Evolutionary Biology*, *8*(1), 139.<https://doi.org/10.1186/1471-2148-8-139>

Firneno Jr., T. J., O’Neill, J. R., Itgen, M. W., Kihneman, T. A., Townsend, J. H., & Fujita, M. K. (2021). Delimitation despite discordance: Evaluating the species limits of a confounding species complex in the face of mitonuclear discordance. *Ecology and Evolution*, *11*(18), 12739–12753.<https://doi.org/10.1002/ece3.8018>

Folk, R. A., Mandel, J. R., & Freudenstein, J. V. (2017). Ancestral Gene Flow and Parallel Organellar Genome Capture Result in Extreme Phylogenomic Discord in a Lineage of Angiosperms. *Systematic Biology*, *66*(3), 320–337.<https://doi.org/10.1093/sysbio/syw083>

Freudenstein, J. V. (1992). Systematics of *Corallorhiza* and the Corallorhizinae (Orchidaceae). Ph.D. thesis. Ithaca, New York: Cornell University.

Freudenstein, J. V. (1997). A MONOGRAPH OF CORALLORHIZA (ORCHIDACEAE). *Harvard Papers in Botany*, *1*(10), 5–51.

Freudenstein, J. V., & Barrett, C. F. (2014). Fungal host utilization helps circumscribe leafless Coralroot orchid species: An integrative analysis of *Corallorhiza odontorhiza* and *C. wisteriana*. *Taxon*, *63*(4), 759–772.<https://doi.org/10.12705/634.3>

Freudenstein, J. V., Broe, M. B., Folk, R. A., & Sinn, B. T. (2017). Biodiversity and the Species Concept—Lineages are not enough. *Systematic Biology*, *66*(4), 644–656.<https://doi.org/10.1093/sysbio/syw098>

Freudenstein, J. V., van den Berg, C., Goldman, D. H., Kores, P. J., Molvray, M., & Chase, M. W. (2004). An expanded plastid DNA phylogeny of Orchidaceae and analysis of jackknife branch support strategy. *American Journal of Botany*, *91*(1), 149–157.<https://doi.org/10.3732/ajb.91.1.149>

Frolov, I., Vondrák, J., Košnar, J., & Arup, U. (2021). Phylogenetic relationships within *Pyrenodesmia* sensu lato and the role of pigments in its taxonomic interpretation. *Journal of Systematics and Evolution*, *59*(3), 454–474.<https://doi.org/10.1111/jse.12717>

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*, *27*(9), 480–488.<https://doi.org/10.1016/j.tree.2012.04.012>

García, N., Folk, R. A., Meerow, A. W., Chamala, S., Gitzendanner, M. A., Oliveira, R. S. de, Soltis, D. E., & Soltis, P. S. (2017). Deep reticulation and incomplete lineage sorting obscure the diploid phylogeny of rain-lilies and allies (Amaryllidaceae tribe Hippeastreae). *Molecular Phylogenetics and Evolution*, *111*, 231–247.<https://doi.org/10.1016/j.ympev.2017.04.003>

Gernandt, D. S., Aguirre Dugua, X., Vázquez-Lobo, A., Willyard, A., Moreno Letelier, A., Pérez de la Rosa, J. A., Piñero, D., & Liston, A. (2018). Multi-locus phylogenetics, lineage sorting, and reticulation in *Pinus* subsection *Australes*. *American Journal of Botany*, *105*(4), 711–725.<https://doi.org/10.1002/ajb2.1052>

Graham, S. W., Lam, V. K. Y., & Merckx, V. S. F. T. (2017). Plastomes on the edge: The evolutionary breakdown of mycoheterotroph plastid genomes. *New Phytologist*, *214*(1), 48–55.<https://doi.org/10.1111/nph.14398>

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, 40(3), 1-25. URL<https://www.jstatsoft.org/v40/i03/>

Grummer, J. A., Bryson, R. W., Jr., & Reeder, T. W. (2014). Species Delimitation Using Bayes Factors: Simulations and Application to the Sceloporus scalaris Species Group (Squamata: Phrynosomatidae). *Systematic Biology*, *63*(2), 119–133.<https://doi.org/10.1093/sysbio/syt069>

Goudet, J. (2005). hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*: 184-186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Haag K. L., James, T. Y., Pombert, F-F., Larsson R., Schaer, T. M. M., Refardt, D., & Dieter, E. (2014). Evolution of a morphological novelty occurred before genome compaction in a lineage of extreme parasites. *Proceedings of the National Academy of Sciences*, *111*(43), 15480–15485. doi: [10.1073/pnas.1410442111](https://doi.org/10.1073/pnas.1410442111)

Hammer, O., Harper, D. A. T., & Ryan, P. D. (n.d.). *PAST: Paleontological Statistics Software Package for Education and Data Analysis*. 9.

Hijmans, R. J., Etten, J. van, Sumner, M., Cheng, J., Baston, D., Bevan, A., Bivand, R., Busetto, L., Canty, M., Fasoli, B., Forrest, D., Ghosh, A., Golicher, D., Gray, J., Greenberg, J. A., Hiemstra, P., Hingee, K., Ilich, A., Geosciences, I. for M. A., … Wueest, R. (2022). *raster: Geographic Data Analysis and Modeling* (3.5-15) [Computer software]. [https://CRAN.R-project.org/package=raster](https://cran.r-project.org/package=raster)

Hillis, D. M., Chambers, E. A., & Devitt, T. J. (2021). Contemporary Methods and Evidence for Species Delimitation. *Ichthyology & Herpetology*, *109*(3), 895–903.<https://doi.org/10.1643/h2021082>

Hull, D. L. (1988). *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*. Chicago, IL: University of Chicago Press. Retrieved from<https://press.uchicago.edu/ucp/books/book/chicago/S/bo3618159.html>

Jackson, N. D., Morales, A. E., Carstens, B. C., & O’Meara, B. C. (2017). PHRAPL: Phylogeographic Inference Using Approximate Likelihoods. *Systematic Biology*, *66*(6), 1045–1053.<https://doi.org/10.1093/sysbio/syx001>

Jakucs, E., & Erős-Honti, Z. (2008). Morphological-anatomical characterization and identification of *Tomentella* ectomycorrhizas. *Mycorrhiza*, *18*(6), 277–285.<https://doi.org/10.1007/s00572-008-0183-4>

Jiao, X., & Yang, Z. (2021). Defining Species When There is Gene Flow. *Systematic Biology*, *70*(1), 108–119.<https://doi.org/10.1093/sysbio/syaa052>

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405.<https://doi.org/10.1093/bioinformatics/btn129>

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071.<https://doi.org/10.1093/bioinformatics/btr521>

Jombart, T., Balloux, F., & Dray, S. (2010). adephylo: New tools for investigating the phylogenetic signal in biological traits. *Bioinformatics*, *26*(15), 1907–1909.<https://doi.org/10.1093/bioinformatics/btq292>

Joseph, T. A., & Pe’er, I. (2019). Inference of population structure from time-series genotype data. *American Journal of Human Genetics*, *105*(2), 317–333.<https://doi.org/10.1016/j.ajhg.2019.06.002>

Kalyaanamoorthy, S., Minh, B., Wong, T. et al. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* *14*, 587–589. https://doi.org/10.1038/nmeth.4285

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773–795.<https://doi.org/10.2307/2291091>

Kassambara, A. (2020). *ggpubr: “ggplot2” Based Publication Ready Plots* (0.4.0) [Computer software]. [https://CRAN.R-project.org/package=ggpubr](https://cran.r-project.org/package=ggpubr)

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780.<https://doi.org/10.1093/molbev/mst010>

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, *35*(21), 4453–4455.<https://doi.org/10.1093/bioinformatics/btz305>

Krug, P. J., Vendetti, J. E., Rodriguez, A. K., Retana, J. N., Hirano, Y. M., & Trowbridge, C. D. (2013). Integrative species delimitation in photosynthetic sea slugs reveals twenty candidate species in three nominal taxa studied for drug discovery, plastid symbiosis or biological control. *Molecular Phylogenetics and Evolution*, *69*(3), 1101–1119.<https://doi.org/10.1016/j.ympev.2013.07.009>

Kuhar, F., Barroetaveña, C., & Rajchenberg, M. (2016). New species of *Tomentella* (Thelephorales) from the Patagonian Andes forests. *Mycologia*, *108*(4), 780–790.<https://doi.org/10.3852/15-244>

Lam, V. K. Y., Darby, H., Merckx, V. S. F. T., Lim, G., Yukawa, T., Neubig, K. M., Abbott, J. R., Beatty, G. E., Provan, J., Soto Gomez, M., & Graham, S. W. (2018). Phylogenomic inference *in extremis*: A case study with mycoheterotroph plastomes. *American Journal of Botany*, *105*(3), 480–494.<https://doi.org/10.1002/ajb2.1070>

Larsen, M. J. (1965). *Tomentella* and related genera in North America: I. Studies of nomenclatural types of species of *Hypochnus* described by Burt. *Canadian Journal of Botany*, *43*(12), 1485–1510.<https://doi.org/10.1139/b65-159>

Larsen, M. J. (1974). A contribution to the taxonomy of the genus *Tomentella*. *Mycologia Memoirs,* 4: 1–145.

Leaché, A. D., Fujita, M. K., Minin, V. N., & Bouckaert, R. R. (2014). Species delimitation using genome-wide SNP data. *Systematic Biology*, 63(4), 534–542. doi: 10.1093/sysbio/syu018

Leake, J. R. (1994). Tansley Review No. 69. The biology of myco-heterotrophic ('saprophytic’) plants. *The New Phytologist*, *127*(2), 171–216.

Legendre, P., & Anderson, M. J. (1999). Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs*, *69*(1), 24.

Legendre, P., Oksanen, J., & ter Braak, C. J. F. (2011). Testing the significance of canonical axes in redundancy analysis. *Methods in Ecology and Evolution*, *2*(3), 269–277.<https://doi.org/10.1111/j.2041-210X.2010.00078.x>

Levin, D. A. (2000). The origin, expansion, and demise of plant species. Oxford University Press, Oxford.<https://www.cabdirect.org/cabdirect/abstract/20003032923>

Leigh, JW, Bryant D (2015). PopART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9),1110–1116.<https://doi.org/10.1111/2041-210X.12410>

Lewis, D. H., Jarvis, D. E., & Maughan, P. J. (2020). SSRgenotyper: A simple sequence repeat genotyping application for whole-genome resequencing and reduced representational sequencing projects. *Applications in Plant Sciences*, *8*(12), e11402.<https://doi.org/10.1002/aps3.11402>

Lewis, P. (2001). A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data. *Systematic Biology*, *50*, 913–925.<https://doi.org/10.1080/106351501753462876>

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760.<https://doi.org/10.1093/bioinformatics/btp324>

Lockwood, C. (2007). Adaptation and functional integration in primate phylogeny. *Journal of Human Evolution*, *52*, 490–503.<https://doi.org/10.1016/j.jhevol.2006.11.013>

Lowry, D. B., Modliszewski, J. L., Wright, K. M., Wu, C. A., & Willis, J. H. (2008). The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1506), 3009–3021.<https://doi.org/10.1098/rstb.2008.0064>

Magrath, L. K., & Freudenstein, J.V. (2002). *Corollorhiza*. In: *Flora of North America Editorial Committee*, (Eds.). *Flora of North America*. 26: 633-638. Oxford University Press., New York and Oxford.

Mason, N. A., Fletcher, N. K., Gill, B. A., Funk, W. C., & Zamudio, K. R. (2020). Coalescent-based species delimitation is sensitive to geographic sampling and isolation by distance. *Systematics and Biodiversity*, *18*(3), 269–280.<https://doi.org/10.1080/14772000.2020.1730475>

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 diversity*, Chapman & Hall. pp. 381–423.

Mayr, E. (1982). Speciation and Macroevolution. *Evolution*, *36*(6), 1119–1132.<https://doi.org/10.1111/j.1558-5646.1982.tb05483.x>

McCormick, M., Whigham, D., Neill, Becker, J., Werner, S., Rasmussen, H., Bruns, T., & Taylor, D. L. (2009). Abundance and distribution of Corallorhiza odontorhiza reflect variations in climate and ectomycorrhizae. *Ecological Monographs*, *79*, 619–635.<https://doi.org/10.1890/08-0729.1>

McCormick, M. K., & Jacquemyn, H. (2014). What constrains the distribution of orchid populations? *New Phytologist*, *202*(2), 392–400.<https://doi.org/10.1111/nph.12639>

McNeal, J. R., Bennett, J. R., Wolfe, A. D., & Mathews, S. (2013). Phylogeny and origins of holoparasitism in Orobanchaceae. *American Journal of Botany*, *100*(5), 971–983.<https://doi.org/10.3732/ajb.1200448>

Meleshko, O., Martin, M. D., Korneliussen, T. S., Schröck, C., Lamkowski, P., Schmutz, J., Healey, A., Piatkowski, B. T., Shaw, A. J., Weston, D. J., Flatberg, K. I., Szövényi, P., Hassel, K., & Stenøien, H. K. (2021). Extensive genome-wide phylogenetic discordance is due to incomplete lineage sorting and not ongoing introgression in a rapidly radiated bryophyte genus. *Molecular Biology and Evolution*, *38*(7), 2750–2766.<https://doi.org/10.1093/molbev/msab063>

Merckx, V. (Ed.). (2013). *Mycoheterotrophy*. Springer New York.<https://doi.org/10.1007/978-1-4614-5209-6>

Merckx, V., & Freudenstein, J. V. (2010). Evolution of mycoheterotrophy in plants: A phylogenetic perspective. *New Phytologist*, *185*(3), 605–609. doi: [10.1111/j.1469-8137.2009.03155.x](https://doi.org/10.1111/j.1469-8137.2009.03155.x)

Michener, C. D. (1970). Diverse approaches to systematics. *Evolutionary Biology*. 4: 1-38.

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.<https://doi.org/10.1093/molbev/msaa015>

Molvray M., Kores P., Chase M.W. (2000). Polyphyly of mycoheterotrophic orchids and functional influences on floral and molecular characters In: K.L. Wilson, D.A. Morrison (Eds.), *Monocots: Systematics and Evolution*, CSIRO, Melbourne, pp. 441-448.

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. *Molecular Biology and Evolution*, 32:268-274.<https://doi.org/10.1093/molbev/msu300>

Nixon, K. C., & Wheeler, Q. D. (1990). An Amplification of the Phylogenetic Species Concept. *Cladistics*, *6*(3), 211–223.<https://doi.org/10.1111/j.1096-0031.1990.tb00541.x>

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P., O’Hara, R., Simpson, G., Solymos, P., Stevens, M., Szöcs, E., & Wagner, H. (2020). *Vegan community ecology package version 2.5-7 November 2020*.

Ornduff, R. (1969). Reproductive biology in relation to systematics. *Taxon*, *18*:121–133.

Padial, J. M., & de la Riva, I. (2010). A response to recent proposals for integrative taxonomy. *Biological Journal of the Linnean Society*, *101*(3), 747–756.<https://doi.org/10.1111/j.1095-8312.2010.01528.x>

Padial, J. M., & De la Riva, I. (2021). A paradigm shift in our view of species drives current trends in biological classification. *Biological Reviews*, *96*(2), 731–751.<https://doi.org/10.1111/brv.12676>

Padial, J. M., Miralles, A., De la Riva, I., & Vences, M. (2010). The integrative future of taxonomy. *Frontiers in Zoology*, *7*(1), 16.<https://doi.org/10.1186/1742-9994-7-16>

Pagel, M., & Meade, A. (2008). Modelling heterotachy in phylogenetic inference by reversible-jump Markov chain Monte Carlo. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1512), 3955–3964.<https://doi.org/10.1098/rstb.2008.0178>

Papakostas, S., Michaloudi, E., Proios, K., Brehm, M., Verhage, L., Rota, J., Peña, C., Stamou, G., Pritchard, V. L., Fontaneto, D., & Declerck, S. A. J. (2016). integrative taxonomy recognizes evolutionary units despite widespread mitonuclear discordance: evidence from a rotifer cryptic species complex. *Systematic Biology*, *65*(3), 508–524.<https://doi.org/10.1093/sysbio/syw016>

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528.<https://doi.org/10.1093/bioinformatics/bty633>

Pedraza-Marrón, C. del R., Silva, R., Deeds, J., Van Belleghem, S. M., Mastretta-Yanes, A., Domínguez-Domínguez, O., Rivero-Vega, R. A., Lutackas, L., Murie, D., Parkyn, D., Bullock, L. H., Foss, K., Ortiz-Zuazaga, H., Narváez-Barandica, J., Acero, A., Gomes, G., & Betancur-R, R. (2019). Genomics overrules mitochondrial DNA, siding with morphology on a controversial case of species delimitation. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1900), 20182924.<https://doi.org/10.1098/rspb.2018.2924>

Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn, R. G., Alfaro, M. E., & Harmon, L. J. (2014). geiger v2.0: An expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics*, *30*(15), 2216–2218.<https://doi.org/10.1093/bioinformatics/btu181>

Peres-Neto, P. R., Legendre, P., Dray, S., & Borcard, D. (2006). Variation Partitioning of Species Data Matrices: Estimation and Comparison of Fractions. *Ecology*, *87*(10), 2614–2625. [https://doi.org/10.1890/0012-9658(2006)87[2614:VPOSDM]2.0.CO;2](https://doi.org/10.1890/0012-9658(2006)87%5b2614:VPOSDM%5d2.0.CO;2)

Philippe, H., Zhou, Y., Brinkmann, H., Rodrigue, N., & Delsuc, F. (2005). Heterotachy and long-branch attraction in phylogenetics. *BMC Evolutionary Biology*, *5*(1), 50.<https://doi.org/10.1186/1471-2148-5-50>

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE*, *5*(3), e9490.<https://doi.org/10.1371/journal.pone.0009490>

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904.<https://doi.org/10.1093/sysbio/syy032>

Rannala, B; Yang, Z; (2020) Species delimitation. In: Scornavacca, C and Delsuc, F and Galtier, N, (eds.) *Phylogenetics in the Genomic Era*. (5.5:1-5.5:18). Self-published.

Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. Sankhyā: *The Indian Journal of Statistics*, *Series A,* 26, 329–358.

Rasmussen, H. N., & Whigham, D. F. (1993). Seed ecology of dust seeds in situ: a new study technique and its application in terrestrial orchids. *American Journal of Botany*, *80*(12), 1374–1378.<https://doi.org/10.2307/2445665>

Reeves, P. A., & Richards, C. M. (2011). Species delimitation under the general lineage concept: an empirical example using wild North American hops (Cannabaceae: *Humulus lupulus*). *Systematic Biology*, *60*(1), 45–59.<https://doi.org/10.1093/sysbio/syq056>

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3, 217–223.<https://cran.r-project.org/web/packages/phytools/index.html>

Richards, A. J. (1986). Plant breeding systems. London: George Allen & Unwin.

Rose, J. P., Toledo, C. A. P., Lemmon, E. M., Lemmon, A. R., & Sytsma, K. J. (2021). Out of sight, out of mind: widespread nuclear and plastid-nuclear discordance in the flowering plant genus *polemonium* (polemoniaceae) suggests widespread historical gene flow despite limited nuclear signal. *Systematic Biology*, *70*(1), 162–180.<https://doi.org/10.1093/sysbio/syaa049>

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, *34*(12), 3299–3302.<https://doi.org/10.1093/molbev/msx248>

Rueden, C. T., Schindelin, J., Hiner, M. C., DeZonia, B. E., Walter, A. E., Arena, E. T., & Eliceiri, K. W. (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*, *18*(1). [doi:10.1186/s12859-017-1934-z](https://doi.org/10.1186/s12859-017-1934-z)

Schlick-Steiner, B. C., Steiner, F. M., Seifert, B., Stauffer, C., Christian, E., & Crozier, R. H. (2010). Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual Review of Entomology*, *55*(1), 421–438.<https://doi.org/10.1146/annurev-ento-112408-085432>

Schliep, K. P. (2011). phangorn: Phylogenetic analysis in R. *Bioinformatics (Oxford, England)*, *27*(4), 592–593.<https://doi.org/10.1093/bioinformatics/btq706>

Schmitt, S., Tysklind, N., Derroire, G., Heuertz, M., & Hérault, B. (2021). Topography shapes the local coexistence of tree species within species complexes of Neotropical forests. *Oecologia*, *196*(2), 389–398.<https://doi.org/10.1007/s00442-021-04939-2>

Schoener, T. W. (1968). The anolis lizards of Bimini: resource partitioning in a complex fauna. *Ecology*, *49*(4), 704–726.<https://doi.org/10.2307/1935534>

Sicard, A., & Lenhard, M. (2011). The selfing syndrome: a model for studying the genetic and evolutionary basis of morphological adaptation in plants. *Annals of botany*, *107(9)*, 1433–1443. https://doi.org/10.1093/aob/mcr023

Simpson, G. G. (1951). The Species Concept. *Evolution*, *5*(4), 285–298.<https://doi.org/10.2307/2405675>

Simpson, G. G. (1961). Principles of Animal Taxonomy. In *Principles of Animal Taxonomy*. Columbia University Press.<https://doi.org/10.7312/simp92414>

Sinn, B. (2017). *Asarum rosei* (Aristolochiaceae), a new species from the Blue Ridge Escarpment of North Carolina, USA. *Phytotaxa*, *296*, 53.<https://doi.org/10.11646/phytotaxa.296.1.3>

Sinn, B. T., Simon, S. J., Santee, M. V., DiFazio, S. P., Fama, N. M., & Barrett, C. F. (2021). ISSRseq: An extensible method for reduced representation sequencing. *Methods in Ecology and Evolution*, 00, 1– 14.<https://doi.org/10.1111/2041-210X.13784>

Sites, J. W., & Marshall, J. C. (2004). Operational criteria for delimiting species. *Annual Review of Ecology, Evolution, and Systematics*, *35*(1), 199–227.<https://doi.org/10.1146/annurev.ecolsys.35.112202.130128>

Sletvold, N., Grindeland, J. M., & Ågren, J. (2010). Pollinator-mediated selection on floral display, spur length and flowering phenology in the deceptive orchid *Dactylorhiza lapponica*. *New Phytologist*, *188*(2), 385–392.<https://doi.org/10.1111/j.1469-8137.2010.03296.x>

Sokal, R. R., & Sneath, P. H. A. 1963. *Principles of numerical taxonomy*. W. H. Freeman, San Francisco.

Sokal, R. R., & Crovello, T. J. (1970). The Biological Species Concept: a critical evaluation. *The American Naturalist*, *104*(936), 127–153.<https://doi.org/10.1086/282646>

Solís-Lemus, C., Knowles, L. L., & Ané, C. (2015). Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution*, *69*(2), 492–507.<https://doi.org/10.1111/evo.12582>

Sterelny, K. (1999). Species as ecological mosaics. In Robert A. Wilson (Ed.), *Species:* *New Interdisciplinary* *Essays*. Bradford, Denver.

Su Huei-Jiun, Barkman Todd J., Hao Weilong, Jones Samuel S., Naumann Julia, Skippington Elizabeth, et al. (2019). Novel genetic code and record-setting AT-richness in the highly reduced plastid genome of the holoparasitic plant Balanophora. *Proceedings of the National Academy of Sciences*, *116*(3), 934–943. doi: [10.1073/pnas.1816822116](https://doi.org/10.1073/pnas.1816822116)

Sukumaran, J., Holder, M. T., & Knowles, L. L. (2021). Incorporating the speciation process into species delimitation. *PLOS Computational Biology*, *17*(5), e1008924.<https://doi.org/10.1371/journal.pcbi.1008924>

Sukumaran, J., & Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(7), 1607–1612.<https://doi.org/10.1073/pnas.1607921114>

Swofford, D. (2002). PAUP\*. Phylogenetic Analysis Using Parsimony (\*and other methods). Version 4.0b10.<https://doi.org/10.1111/j.0014-3820.2002.tb00191.x>

Taylor, D. L., & Bruns, T. D. (1997). Independent, specialized invasions of ectomycorrhizal mutualism by two nonphotosynthetic orchids. *Proceedings of the National Academy of Sciences*, *94*(9), 4510–4515.<https://doi.org/10.1073/pnas.94.9.4510>

Taylor, D. L., Bruns, T., Leake, J. R., & Read, D. (2002). Mycorrhizal specificity and function in myco-heterotrophic plants. *Mycorrhizal Ecology*, 375–413.

Taylor, D. L., Bruns, T. D., & Hodges, S. A. (2004). Evidence for mycorrhizal races in a cheating orchid. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(1534), 35–43.<https://doi.org/10.1098/rspb.2003.2557>

Taylor, D. L., Barrett, C. F., Beatty, G. E., Hopkins, S. E., Kennedy, A. H., & Klooster, M. R. (2013). Progress and prospects for the ecological genetics of mycoheterotrophs. In V. Merckx (Ed.), *Mycoheterotrophy* (pp. 245–266). Springer New York.<https://doi.org/10.1007/978-1-4614-5209-6_6>

Thompson, J. N. (1994). *The Coevolutionary Process*. University of Chicago Press.<https://press.uchicago.edu/ucp/books/book/chicago/C/bo3637930.html>

Thompson, J. N. (2005). *The Geographic Mosaic of Coevolution*. University of Chicago Press.<https://press.uchicago.edu/ucp/books/book/chicago/G/bo3533766.html>

Tiffin, P., Olson, M. S., & Moyle, L. C. (2001). Asymmetrical Crossing Barriers in Angiosperms. *Proceedings: Biological Sciences*, *268*(1469), 861–867.<https://royalsocietypublishing.org/doi/10.1098/rspb.2000.1578>

Tsukaya, H. (2018). How leaves of mycoheterotrophic plants evolved – from the view point of a developmental biologist. *New Phytologist*, *217*(4), 1401–1406.<https://doi.org/10.1111/nph.14994>

Van der Cingel, N.A. (2001). An atlas of orchid pollination: America, Africa, Asia and Australia. AA Balkema Publishers: Rotterdam.

van der Kooi, C. J., Vallejo-Marín, M., & Leonhardt, S. D. (2021). Mutualisms and (A)symmetry in Plant–Pollinator Interactions. *Current Biology*, *31*(2), R91–R99.<https://doi.org/10.1016/j.cub.2020.11.020>

Van Valen, L. (1976). Ecological Species, Multispecies, and Oaks. *TAXON*, *25*(2–3), 233–239.<https://doi.org/10.2307/1219444>

Wang, D., Jacquemyn, H., Gomes, S. I. F., Vos, R. A., & Merckx, V. S. F. T. (2021). Symbiont switching and trophic mode shifts in Orchidaceae. *New Phytologist*, *231*(2), 791–800.<https://doi.org/10.1111/nph.17414>

Warren, D. L., Glor, R. E., & Turelli, M. (2008). Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution*, *62*(11), 2868–2883.<https://doi.org/10.1111/j.1558-5646.2008.00482.x>

Warren, D. L., Matzke, N. J., Cardillo, M., Baumgartner, J. B., Beaumont, L. J., Turelli, M., Glor, R. E., Huron, N. A., Simões, M., Iglesias, T. L., Piquet, J. C., & Dinnage, R. (2021). ENMTools 1.0: An R package for comparative ecological biogeography. *Ecography*, *44*(4), 504–511.<https://doi.org/10.1111/ecog.05485>

Waterman, R. J., & Bidartondo, M. I. (2008). Deception above, deception below: Linking pollination and mycorrhizal biology of orchids. *Journal of Experimental Botany*, *59*(5), 1085–1096.<https://doi.org/10.1093/jxb/erm366>

Wells, T., Carruthers, T., Muñoz-Rodríguez, P., Sumadijaya, A., Wood, J. R. I., & Scotland, R. W. (2021). Species as a heuristic: reconciling theory and practice. *Systematic Biology*, syab087.<https://doi.org/10.1093/sysbio/syab087>

White, T. J., Bruns, T. D., Lee, S., & Taylor, J. W. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M. A., Gelfand, D. H., Sninsky, J. J., White, T. J., (Eds.). *PCR protocols: a guide to methods and applications*. New York, N.Y: Academic Press, Inc., pp. 315–322.

Wicke, S., Müller, K. F., de Pamphilis, C. W., Quandt, D., Wickett, N. J., Zhang, Y., Renner, S. S., & Schneeweiss, G. M. (2013). Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *The Plant Cell*, *25*(10), 3711–3725.<https://doi.org/10.1105/tpc.113.113373>

Wickham, H. (2016). ggplot2: elegant graphics for data analysis. Springer-Verlag, New York.

Wiens, J. J. (2007). Species delimitation: new approaches for discovering diversity. *Systematic Biology*, *56*(6), 875–878.<https://doi.org/10.1080/10635150701748506>

Wiley, E. O. (1978). The Evolutionary Species Concept reconsidered. *Systematic Biology*, *27*(1), 17–26.<https://doi.org/10.2307/2412809>

Willyard, A., Cronn, R., & Liston, A. (2009). Reticulate evolution and incomplete lineage sorting among the ponderosa pines. *Molecular Phylogenetics and Evolution*, *52*(2), 498–511.<https://doi.org/10.1016/j.ympev.2009.02.011>

Wilson, E. O. (1999). *The Diversity of Life*. W. W. Norton & Company.

Wilson, E. O. (2017). Biodiversity research requires more boots on the ground. *Nature Ecology & Evolution*, *1*(11), 1590–1591.<https://doi.org/10.1038/s41559-017-0360-y>

Yang Z, Wafula EK, Kim G, Shahid S, McNeal JR, Ralph PE, Timilsena PR, Yu WB, Kelly EA, Zhang H, Person TN, Altman NS, Axtell MJ, Westwood JH, dePamphilis CW. (2019). Convergent horizontal gene transfer and cross-talk of mobile nucleic acids in parasitic plants. *Nature Plants*, *5*(9), 991-1001.<https://doi.org/10.1038/s41477-019-0458-0>

Yang, Z., & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(20), 9264–9269.<https://doi.org/10.1073/pnas.0913022107>

Yeates, D. K., Seago, A., Nelson, L., Cameron, S. L., Joseph, L., & Trueman, J. W. H. (2011). Integrative taxonomy, or iterative taxonomy? *Systematic Entomology*, *36*(2), 209–217.<https://doi.org/10.1111/j.1365-3113.2010.00558.x>

Young, A. D., & Gillung, J. P. (2020). Phylogenomics—Principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology*, *45*(2), 225–247.<https://doi.org/10.1111/syen.12406>

Yuan, Y., Jin, X., Liu, J., Zhao, X., Zhou, J., Wang, X., et al. (2018). The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. *Nature Communications*, *9*(1), 1615.<https://doi.org/10.1038/s41467-018-03423-5>

Zachos, F. E. (2018). (New) Species concepts, species delimitation and the inherent limitations of taxonomy. *Journal of Genetics*, *97*(4), 811–815.<https://doi.org/10.1007/s12041-018-0965-1>

Zapata, F., & Jiménez, I. (2012). Species Delimitation: Inferring Gaps in Morphology across Geography. *Systematic Biology*, *61*(2), 179.<https://doi.org/10.1093/sysbio/syr084>

Zietkiewicz, E., A. Rafalski, & Labuda, D. 1994. Genome fingerprinting by Simple Sequence Repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* *20*:176–183.

CrossRefPubMedWeb of ScienceGoogle Scholar

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V., & Antonelli, A. (2019). CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*, *10*(5), 744–751.<https://doi.org/10.1111/2041-210X.13152>

**8. DATA ACCESSIBILITY**

Sequences of the *rbcL* pseudogene and fungal ITS region are available via NCBI GenBank accession numbers: FJ445519–FJ445595, GU223916-GU223976, JF319708–

JF31977 and OM334812-OM334825 (*rbcL*); GU224038-GU220711 and OM282086-OM282096 (fungal ITS). Genomic data from ISSRseq are available via the NCBI Sequence Read Archive (BioProject PRJNA771539). Morphological, genomic (SNP), phenological, abiotic niche, reproductive mode, and Fungal ITS data are available as supplementary materials via Zenodo (https://doi.org/10.5281/zenodo.6341174). Code for bioinformatic processing of ISSRseq data is available at GitHub ([www.github.com/btsinn/ISSRseq](http://www.github.com/btsinn/ISSRseq)).

**9. AUTHOR CONTRIBUTIONS**

CFB, JVF, and MVS conceived the study. MVS, NMF, BTS, SJS, and CFB designed the laboratory methodology and protocols. BTS, SJS, MVS, NMF, and CFB conducted the laboratory work. CFB conducted field work. BTS, SJS, CFB and MVS

analyzed data. MVS and CFB led the writing of the manuscript. All authors contributed to manuscript drafts and provided consent for publication.

**10. TABLES**

**Table 1**. Results of SNAPP-BFD\* species delimitation analyses. ‘# spp’ = the number of species in a particular scenario, ‘MLE’ = Maximum likelihood estimate for a particular scenario, ‘rank’ = the rank of each scenario based on MLE values, ‘BF’ = Bayes Factor comparison of a particular scenario to that with the lowest MLE value. Under ‘Delimitation,’ ‘str’ = var. *striata*, ‘vre’ = var. *vreelandii*, ‘snv’ = Sierra Nevada, and ‘crc’ = Coast Ranges/Cascades.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| # spp | Delimitation | MLE | rank | BF |
| 4 | (str) (vre) (snv) (crc) | -175646.5 | 1 | -- |
| 3 | (str) (vre) (snv + crc) | -175802.4 | 2 | -311.8 |
| 2 | (str + vre) (snv + crc) | -178159.8 | 3 | -5026.6 |
| 3 | (str + vre) (snv) (crc) | -178293.9 | 4 | -5294.8 |

**Table 2.** Confusion matrix from Linear Discriminant Analysis analysis of 14 morphological characters, showing the number of positive classifications based on *a priori* assignments. Overall, 91.34% of accessions were classified according to their *a priori* grouping. Rows are the given, *a priori* groups and columns are the classified groups. The diagonal values (boldface) indicate the numbers of successful classifications to *a priori* groupings.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | var. *striata* | var. *vreelandii* | Sierra Nevada | Coast/ Cascades | Total |
| var. *striata* | **36** | 0 | 0 | 0 | 36 |
| var. *vreelandii* | 0 | **29** | 2 | 1 | 32 |
| Sierra Nevada | 3 | 2 | **29** | 1 | 35 |
| Coast/Cascades | 1 | 1 | 0 | **22** | 24 |
| Total | 40 | 32 | 31 | 24 | 127 |

**Table 3**. Ecospat identity and background/similarity tests for 10 km and 0.5 km resolutions. Both the ‘D’ and ‘I’ statistics are shown, and cell values represent identity (ID, left value) and background/similarity (BG, right value) test statistics. \*p <0.05, \*\*p < 0.01.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Comparison** | **D, ID/BG,**  **10 km** | **I, ID/BG,**  **10 km** | **D, ID/BG,**  **0.5 km** | **I, ID/BG,**  **0.5 km** |
| *striata* - *vreelandii* | 0.24\*\*/0.31 | 0.33\*\*/0.40 | 0.28/0.27 | 0.37\*\*/0.37 |
| *striata* - Sierra Nevada | 0.05/0.04 | 0.09/0.10 | 0.18/0.15 | 0.22/0.10 |
| *striata* - Coast/Cascades | 0.10\*\*/0.11 | 0.14\*\*/0.13 | 0.23/0.19 | 0.29/0.24 |
| *vreelandii* - Sierra Nevada | 0.06\*\*/0.05 | 0.23\*\*/0.22 | 0.18/0.15 | 0.57/0.64 |
| *vreelandii* - Coast/Cascades | 0.37/0.37 | 0.62\*/0.62 | 0.31/0.29 | 0.48/0.46 |
| Sierra Nevada - Coast/Cascades | 0.24\*\*/0.22 | 0.41\*\*/0.45 | 0.30/0.29 | 0.53/0.52 |

**11. FIGURE CAPTIONS**

**Figure 1. A.** *Corallorhiza striata* var. *vreelandii* (green). **B.** *C. striata* var. *striata* (red). **C.** *C. striata* from the Sierra Nevada of California, USA (blue). **D.** *C. striata* from the Coast Ranges/Cascades of California and Oregon, USA (magenta). **E.** Map showing the geographic range of the *C. striata* complex. Black empty circles are GBIF records based on herbarium collections. Colored, filled circles are sampling localities included in the ISSRseq analysis (see Fig. 2) for each of the four groupings in A.

**Figure 2.** Relationships and genetic differentiation of the *C. striata* complex based on SNP data from ISSRseq. **A.** Relationships based on maximum likelihood analysis in RAxML-NG with an ascertainment bias model. **B.** Relationships based on analysis in SVDQuartets. Total weight of compatible quartets = 0.927. **C.** Relationships based on maximum likelihood analysis in IQtree2 with an ascertainment bias model. **D.** Principal Components Analysis of linkage disequilibrium-thinned SNP data, showing PCs 1-3. **E.** Population structure analysis in DyStruct for K=2-4. Colors correspond to legend in d).

**Figure 3**. Morphological, phenological, and reproductive mode analyses of the *C. striata* complex. **A.** Principal Components Analysis (PCA) of 14 log-transformed floral morphological characters, showing PCs 1-2. **B.** Biplot showing the loading scores of each character on PCs 1-2. **C.** PCA of PC axes 1 and 3. **D.** Biplot of character loadings on PCs 1-3. **E.** Linear Discriminant Analysis (LDA) of the same data, showing LDA axes 1 and 2. **F.** Biplot of the morphological characters on LDA axes 1 and 2. **G.** Phylomorphospace representation of PC scores for PC axis 1, with relationships from the RAxML-NG tree superimposed. **H.** Two-dimensional phylomorphospace representation with the RAxML-NG tree on PCs 1 and 2. **I.** Density plots of the six most informative characters differentiating Sierra Nevadan and Coast/Cascade accessions. Scale is in mm. **J.** Violin plots of specimen records by flowering date. Note that Cascades accessions were split from Coast Range accessions here to investigate differences in flowering time between accessions from these two regions. **K.** Boxplots of the proportion of flowers in a raceme with clear evidence of swollen ovaries. Letters ‘a’ and ‘b’ are Tukey *post-hoc* comparison values that are significantly different (p < 0.0001).

**Figure 4. A.** Phylogenetic analysis of >1,400 *Tomentella* fungal accessions, including GenBank reference sequences and members of the *C. striata* complex. Scale bar = 0.2 substitutions/site. **B.** Closeup of the clade occupied by all accessions of the *C. striata* complex, with closely related reference sequences from *Tomentella fuscocinerea* and *T. patagonia*. Branch lengths are scaled proportionally. Blue and magenta arrows point to associates of the Sierra Nevadan *C. striata* and the Coast/Cascade *C. striata*, respectively. Roman numerals indicate three principal sub-clades.

**Figure 5.** Integrative analyses of SNP and ‘extended’ data. **A.** SNPs only analyzed in assignPop. **B**. Combined SNP and ‘extended’ data (morphology, biotic niche, and fungal hosts). **C**. ‘Extended’ data only. ‘LDA’ = linear discriminant analysis; ‘SVM’ = support vector machine; ‘RF’ = random forest. **D**. Redundancy Analysis (RDA) and variance partitioning using four- and three-species delimitations as explanatory variables with ‘extended’ data as response variables (\* = *Corallorhiza striata* var. *striata*, *C*. *striata* var. *vreelandii*, Sierra Nevada + Coast/Cascades).

**Figure S1**. Discriminant Analysis of Principal Components in ‘adegenet.’ **A.** Analysis of SNP data from ISSRseq, axes 1 and 2. **B.** Analysis of SNP data from ISSRseq, axes 1 and 3. **C.** Analysis of microsatellite data from ISSRseq, axes 1 and 2. **D.** Analysis of microsatellite data from ISSRseq, axes 1 and 2.

**Figure S2.** Results from ‘phylo.signal.disc’ analysis of phylogenetic signal among groupings for fungal associates of *C. striata*. **A.** Null distribution of traits (groupings) based on random permutations of states on the fungal ITS tree. X-axis = the number of randomized character state transitions in the null distribution; Y-axis = the frequency of each number of transitions. Red arrow = the number of observed transitions. **B.** Observed distribution of character states (i.e. groupings) on the fungal ITS tree.