

Lineage and role in integrative taxonomy of a heterotrophic orchid complex

Craig F. Barrett¹  | Mathilda V. Santee¹  | Nicole M. Fama¹ | John V. Freudenstein²  | Sandra J. Simon³  | Brandon T. Sinn^{4,5} 

¹Department of Biology, West Virginia University, West, USA

²Department of Evolution, Ecology, and Organismal Biology, Ohio State University, Columbus, Ohio, USA

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

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

⁵Department of Botany and Ecology, University of Latvia, Riga, Latvia

Correspondence

Craig F. Barrett, Department of Biology, West Virginia University, 53 Campus Drive, Morgantown, West Virginia 26506, USA.

Email: craig.barrett@mail.wvu.edu

Funding information

USDA Forest Service, Grant/Award Number: DPR65; Eberly College of Arts and Sciences, West Virginia University; National Science Foundation

Handling Editor: SUHUA SHI

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, analysing 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 data sets 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, Orchidaceae, SNP data, species boundaries, species concept

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., Camargo & Sites, 2013; Dobzhansky, 1950; Donoghue, 1985; Hennig, 1966; Hillis et al., 2021; Mayr, 1942; Pedraza-Marrón et al., 2019; Sites &

Marshall, 2004; Sneath & Sokal, 1973; Sterelny, 1999; Wiens, 2007; Wiley, 1978; Wilson, 1999; Wilson, 2017; Wright & Huxley, 1940). 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, focusing on identifying such lineages has become a key aspect of species delimitation (de Queiroz, 1998, 2005, 2007; Reeves & Richards, 2011)

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 [Carstens & Knowles, 2007; Jackson et al., 2017; Jiao & Yang, 2021; Leaché & Fujita, 2010; O'Meara, 2010]). Other species criteria 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 (Rannala & Yang, 2020; Yang & Rannala, 2010). Recent studies have questioned the relevance of the lineage as the sole criterion for recognizing species, however, and some methods designed to delimit species, for example, the MSC (Freudenstein et al., 2017; Sukumaran et al., 2021; Sukumaran & Knowles, 2017; Wells et al., 2021), because they may not circumscribe the units we care about. Arguments against lineage-only approaches centre around observations that such methods may tend to delimit population structure rather than species and are sensitive to biases in range-wide sampling (Chambers & Hillis, 2020; Hillis et al., 2021; Mason et al., 2020; Sukumaran et al., 2021; Sukumaran & Knowles, 2017). 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; Carstens et al., 2013; Cicero et al., 2021; Edwards & Knowles, 2014; Krug et al., 2013; Leaché et al., 2014; Padial et al., 2010; Padial & de la Riva, 2010; Padial & de la Riva, 2021; Schlick-Steiner et al., 2010; Solís-Lemus et al., 2015; 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, behaviour, metabolites, and ecological interactions (Carstens et al., 2013; Fujita et al., 2012). A further distinction can be made between iterative and integrative taxonomy, where in the former, data sets are analysed separately or in a specific order (Padial et al., 2010; Padial & de la Riva, 2010), and in the latter they are analysed simultaneously in a single framework (Edwards & Knowles, 2014; Solís-Lemus et al., 2015; Sukumaran et al., 2021; Yeates et al., 2011).

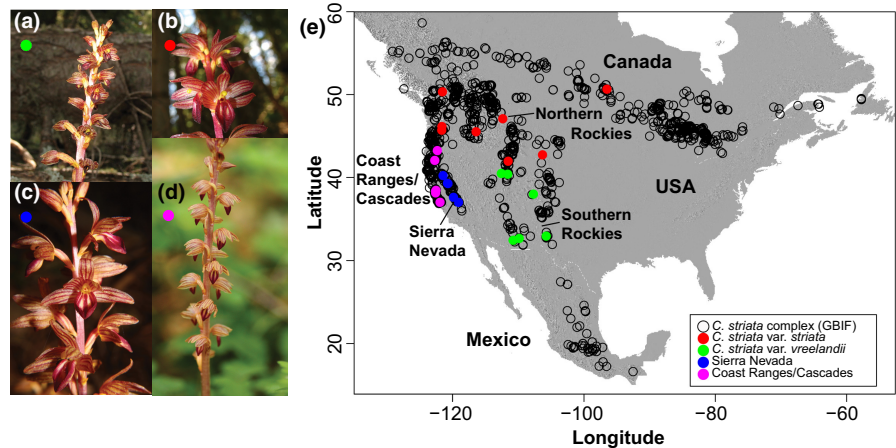
Freudenstein et al. (2017) emphasized the coequal importance of lineage and phenotype in their explication of Simpson's ESC. Simpson's use of the term role in his statement of the concept embodied the phenotypic aspect, which Freudenstein et al. (2017) viewed broadly as the extended phenotype (Dawkins, 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 (Levin, 2000; Mayr, 1982;

Simpson, 1951, 1961; Van Valen, 1976). Extended phenotype is an intentionally broad category, encompassing any fixed, observable differences among lineages in phenotype, ecological interactions (including symbionts), behaviour, 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., Davis & Nixon, 1992; Nixon & Wheeler, 1990), 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 (Ackerfield et al., 2020; da Cruz & Weksler, 2018; Folk et al., 2017; Frolov et al., 2021; Sinn, 2017; Zachos, 2018), but a comprehensive application is lacking.

Some organisms are inherently taxonomically challenging. Parasites, with reduced morphology and genomes, are especially problematic (Cameron, 2004; Evans et al., 2008; Freudenstein et al., 2004; Haag et al., 2014; McNeal et al., 2013; Molvray et al., 2000; Wicke et al., 2013). Mycoheterotrophic plants are parasites specializing on and deriving nutrients from their fungal hosts (Bidartondo, 2005; Leake, 1994; Merckx, 2013). Mycoheterotrophs convergently display morphological reductions, lacking leaves or other features that contain character information for species delimitation (Merckx & 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 (Barrett et al., 2014, 2019; Braukmann et al., 2017; Graham et al., 2017; Merckx, 2013; Su et al., 2019; Yang et al., 2019; Yuan et al., 2018). 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 & Meade, 2008; Darriba et al., 2016; Givnish et al., 2018; Lam et al., 2018; Crotty et al., 2020; Young & Gillung, 2020). An integrative approach to taxonomy is essential for such groups, as it maximizes potential information for their placement, including species delimitation (Barrett & Freudenstein, 2011; Broe, 2014; Freudenstein & Barrett, 2014).

One problematic group is the *Corallorhiza striata* complex, a wide-ranging, variable orchid of uncertain status (Figure 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

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. Coloured, filled circles are sampling localities included in the ISSRseq analysis (see Figure 2) for each of the four groupings in (a).



distributed from Mexico to Canada (Barrett & Freudenstein, 2011; Freudenstein, 1997; Magrath & Freudenstein, 2002). *Corallorhiza bentleyi* and *C. involuta* share a close relationship yet are divergent from *C. striata* based on nuclear markers and plastid genomes (Barrett et al., 2018; Barrett & Freudenstein, 2011). 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 (British Columbia to Newfoundland).

Taxonomic delimitation within *C. striata* is challenging. Freudenstein (1997) noted clinal patterns of morphological variation across North America, with *C. striata* var. *vreelandii* (Figure 1a) having smaller, less-open flowers, suggesting a reproductive mode of self-pollination. *Corallorhiza striata* var. *striata* (Figure 1b) has large, open flowers, and is 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 & Freudenstein, 2009, 2011). Adding to the confusion, populations in California, USA (Figures 1c,d), are morphologically intermediate between the two recognized varieties (Barrett & Freudenstein, 2009, 2011), having been proposed as introgressants by some researchers (e.g., Magrath & 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 & 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 (Jakucs & Erős-Honti, 2008; Larsen, 1974).

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 therefore subject to potential confounding effects of ILS and past/present gene flow relative to representations based on a broad sampling of nuclear loci (Doyle, 2021; Folk et al., 2017; García et al., 2017; Gernandt et al., 2018; Rose et al., 2021; Willyard et al., 2009). 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, 2002), 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 analyse data from genome-wide single nucleotide polymorphisms, floral morphometry, phenology, reproductive mode, abiotic niche, and patterns of fungal host specificity to quantify and analyse variation in *Corallorhiza striata* within the framework of “lineage + role”. The rationale for including each of these data sets is as follows: genomic data can provide a powerful source of evidence for distinct lineages (de Queiroz, 2007; Leaché 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 &

Bruns, 1997). This potentially equates to novel and highly specific interactions that have been hypothesized to drive speciation, analogous to the situation in many plant-feeding insects (e.g., Drès & Mallett, 2002; 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 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 Simpson, 1951; 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 centre of diversity in western North America, but also sampled in putative contact zones in northern Utah and California/Oregon (Figure 1e). A CTAB protocol was performed to isolate shoot (floral) and rhizome (plant+fungal) DNA (Doyle & 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 et al. (2021). Briefly, this method involves PCR amplification and high-throughput sequencing of intersimple 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 a genomic-scale data set 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 et al. (2021). Biallelic SNPs were filtered using phrynomics in R to remove nonbinary and noninformative SNPs (Leaché et al., 2015). SNPs were used to infer relationships among individuals with RAXML-NG using 1000 bootstraps with the best-fit model: GTR + G4m + ASC (Kozlov et al., 2019). Bootstrap support was assessed with 1000 standard bootstrap pseudoreplicates. Another tree was built using IQtree2 (Minh et al., 2020; Nguyen et al., 2015) with best fit model TVM + F + ASC + R3 as determined by ModelFinder (Kalyaanamoorthy et al., 2017). The “ASC” option accounts for ascertainment bias in SNP-only data (Leaché et al., 2015; Lewis, 2001). Support was assessed with 1000 ultrafast bootstrap pseudoreplicates (Hoang et al., 2018). A coalescent-based species tree was estimated with SVDquartets in PAUP version 4 with 1000 bootstraps (coalescent model, sampling all quartets, Erik+2 normalization; Swofford, 2002; Chifman & Kubatko, 2014, 2015).

To test among alternative species delimitation scenarios using genomic data alone, Bayes factor delimitation (Leaché et al., 2014) was conducted via SNAPP-BFD* in BEAST2 (Bouckaert et al., 2019). Path sampling was conducted via stepping stone analysis (Baele et al., 2013; Grummer et al., 2014; Kass & Raftery, 1995) 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 data set 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% burnin, after preburnin of 50,000 generations. We verified stationarity and convergence across runs in Tracer version 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: $BF = 2 \times (MLE_1 - MLE_0)$, in which model₀ had the highest MLE. Bayes Factors < -10 indicate “decisive” support for model₀ over model₁.

Population structure was characterized using principal component analysis (PCA) in ADEGENET version 2.1.5 (Jombart, 2008; Jombart & Ahmed, 2011), after linkage disequilibrium (LD) thinning to one SNP per locus in VCFTOOLS (Danecek et al., 2011); thinning was set to 3 kb, greater than the length of the longest assembled contig. Global F_{ST} among lineages was assessed with hierfstat (Goudet, 2005). Analysis of admixture was conducted in DyStruct version 1.1.0 (Joseph & Pe'er, 2019), with K = 1–8, to determine the most likely number of ancestral clusters. Runs were replicated 10 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/>

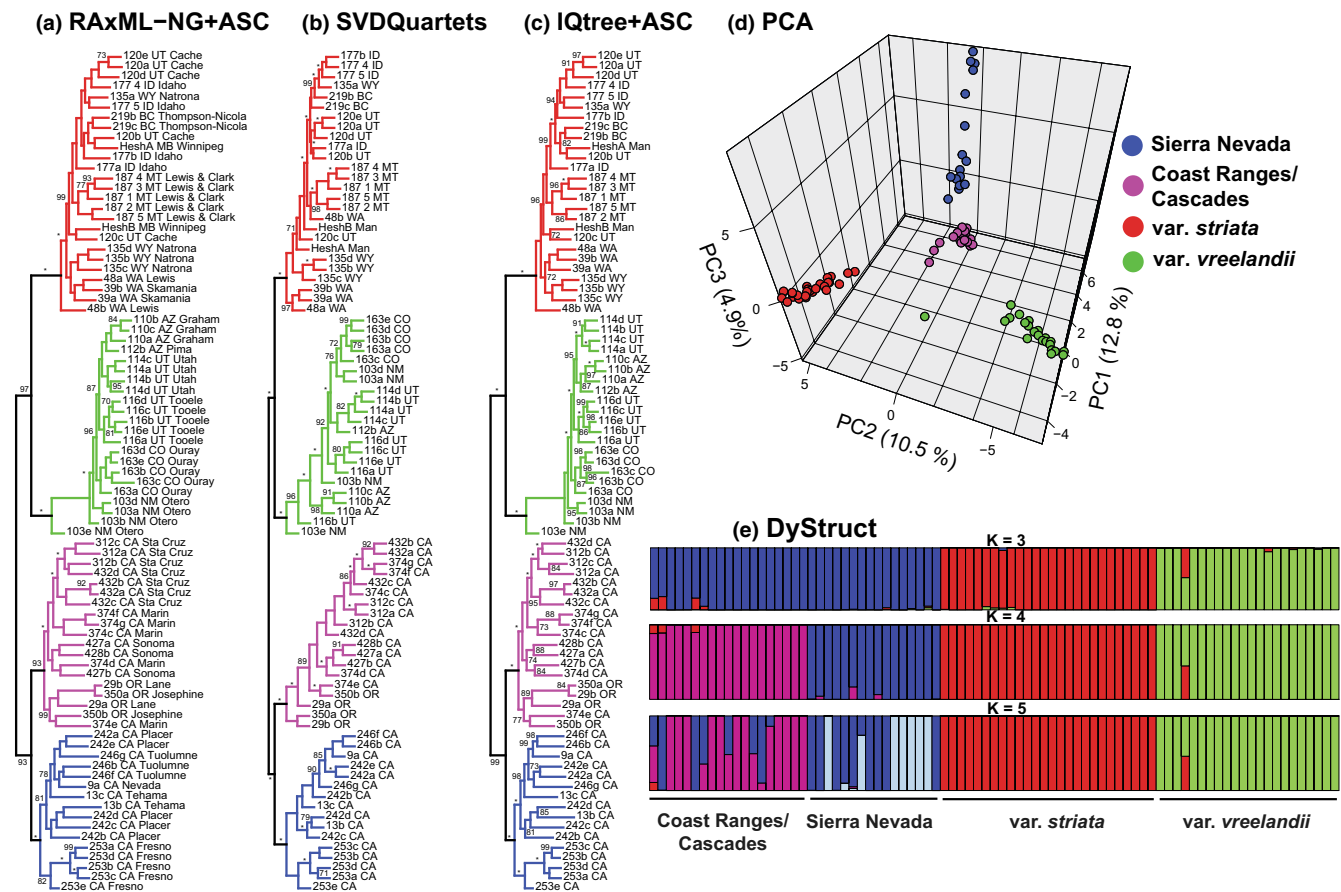


FIGURE 2 Relationships and genetic differentiation of the *C. striata* complex based on SNP data from ISSRseq. (a) Relationships based on maximum likelihood analysis in RAXML-NG with an ascertainment bias model. (b) Relationships based on analysis in SVDQuartets. Total weight of compatible quartets = 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. Colours correspond to legend in (d).

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 et al., 2021 for details) with BWA-MEM (Li & Durbin, 2009), and processed with SAMTOOLS (sorting, removing duplicates, filtering on mapping quality of $-q$ 45). GENEPOP output was read into adegenet version 2.1.3 (Jombart, 2008) in R, as was the LD-thinned SNP data set. We used discriminant analysis of principal components (DAPC; Jombart, Devillard, & Balloux, 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, we measured 14 floral characters for a single flower from each of 127 individual plants (as in Barrett & Freudenstein, 2011) using a stereomicroscope (Fisher Scientific). 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

analysed with Image J version 1.53e (Rueden et al., 2017). PAST version 4.09 software (Hammer et al., 2001) was used to conduct PCA of $\log_{10}$ -transformed data with a variance-covariance matrix and multivariate analysis of variance (MANOVA). Plots were constructed using ggplot2 (Wickham, 2016) and ggpubr version 0.4.0 in R (Kassambara, 2020). We further constructed a phylomorphospace with the R package phytools version 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 difference occupancy of niche space constitutes 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 version 3.6.0 (Chamberlain et al., 2022). Occurrences were then filtered for coordinate uncertainty with

CoordinateCleaner version 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 both 10 and 0.5 km resolutions in R via the raster version 3.5.2 package (Hijmans et al., 2022). The full BIOCLIM data set was subjected to principal components analysis in Past version 4 (Hammer et al., 2001), specifying a correlation matrix to account for differences in the scale of the variables.

The full BIOCLIM data set was then subjected to pairwise Pearson's correlation analysis using ENMTools version 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 version 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 1000 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 version 1.7.10 (Grolemund & 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, 2002; Coleman et al., 2012), we included specimens collected after 1 April 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's comparisons in R.

2.6 | Fungal DNA analyses

In order to quantify patterns of specificity and divergence in fungal associations among lineages of *C. striata*, we extracted DNAs from rhizome tissues (mixed orchid and fungal DNAs) of 11 accessions from 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 version 10 (Drummond et al., 2011) and further aligned using the MAFFT version 7.490 plugin (auto algorithm; Katoh & Standley, 2013). In addition, we included >1400 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 version 2.1.10 under a GTR model with 1000 bootstrap pseudoreplicates (Price et al., 2010).

To test for phylogenetic signal in associations among the four groupings (i.e., nonrandom associations), we used the subclade from the fungal ITS tree containing all *C. striata* accessions, pruning reference *Tomentella* and associates of *C. bentleyi*+*involuta* with ape version 5.5 in R (Paradis & 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 version 2.0.7, phangorn version 2.5.5, and phylobase version 0.8.10 (Schliep, 2011; Pennell et al., 2014; Bolker et al., 2020; 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 data sets to quantify their contribution to divergence among lineages, and to identify evidence for shifts in role. In order to analyse multiple data sets integratively, population assignment analyses were conducted with the R package assignPOP version 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 data set 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 data set 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 version 1.1.11 (Jombart, Balloux, & Dray, 2010). 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 $\log_{10}$ -transformed.

Analyses were conducted using the assign.MC (Markov Chain Monte Carlo [MCMC]) 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 F_{ST} 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 data sets 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 & Anderson, 1999) in the R package vegan version 2.5.7 (Oksanen et al., 2020), to address the hypothesis that extended data sets may form the basis for hypotheses of role shifts with respect to the genomic lineages identified. We tested four species versus 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 & Jiménez, 2012; Papakostas et al., 2016; Firreno et al., 2021), and were used as potential explanatory variables in four RDA scenarios with different data sets 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 (Legendre et al., 2011; Peres-Neto et al., 2006) to quantify the unique and shared contributions among delimitations and test the significance of the adjusted R^2 values for each data set.

3 | RESULTS

3.1 | ISSRseq and SNP analyses

The final ISSRseq data set consisted of 27,117 SNPs (data set S2), 6589 of which remained after LD-thinning (data set S3), and 2662 of which remained after removal of nonparsimony informative sites

with phrynomics in R (data set S4). Phylogenetic analyses based on informative SNPs yielded highly resolved and supported topologies (Figures 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 (6589 SNPs) revealed distinct genomic groupings based on the first three PCs (Figure 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 F_{ST} was 0.57 ($p < .001$). 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 S2). 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; data set 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 (Figure S1). As in Figure 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 Nevada 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 $BF = -5294.8$ 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 ($BF = -5026.6$).

3.2 | Morphology

Analysis of 14 floral measurement characters via PCA and LDA revealed somewhat distinct groupings, but with evidence of overlap among them (Figures 3a,c,e; data set S6). Perianth length/width characters were largely correlated with PC1 (69.77% of total variance, e.g. petal, sepal, and labellum characters; Table S3), 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, $p < .0001$), with significant, Bonferroni-corrected pairwise comparisons among all groupings (Figure 3e, f; $p < .0001$ 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* (Figure 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.

TABLE 1 Results of SNAPP-BFD* species delimitation analyses

# 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

Abbreviations: # 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; str, var. *striata*; vre, var. *vreelandii*; snv, Sierra Nevada; crc, Coast Ranges/Cascades.

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; data set S7). *Corallorhiza striata* var. *striata* differed significantly from var. *vreelandii* and Coast/Cascades, but was not significantly different from Sierra Nevada populations. *Corallorhiza striata* var. *vreelandii* differed from Sierra Nevada 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 1743 herbarium records, differs significantly among the lineages ($F_{ANOVA} = 72.48$, $p < .0001$), with Coast/Cascade individuals having earlier flowering times compared to vars. *striata*, *vreelandii*, and Sierran individuals (Figure 3j; $p_{Tukey\ HSD} < 0.0001$ for all comparisons; data set S8). Considering individuals from the Coast Ranges versus the Cascades separately, those from the California Coast Ranges have a slightly earlier flowering time on average than those from the Cascades ($p_{Tukey\ HSD} < 0.001$). Individuals of var. *vreelandii* show a higher proportion of swollen ovaries in herbarium records ($n = 43$ after data filtering; Figure 3k; data set S9) than var. *striata* and Californian individuals ($F_{ANOVA} = 107.6$, $p < .0001$; $p_{Tukey\ HSD} < 0.0001$ 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 +1451 *Tomentella* from GenBank), had a total, post-filtered length of 521 bp, with 139 parsimony informative sites (data set 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* (Figure 4a). Within the fungal clade of *C. striata* associates, Sierran individuals predominantly associate with a single subclade (clade III, comprising reference accessions of *T. fuscocinerea* and *T. patagonica*; $\pi_{Sierra} = 0.010$), while Coast/Cascade individuals associate largely with members of clade II, representing accessions of *T. fuscocinerea* ($\pi_{Coast/Cascades} = 0.030$; Figure 4b; Table S4). The closest relative to associates of the Sierran individuals is from Sweden,

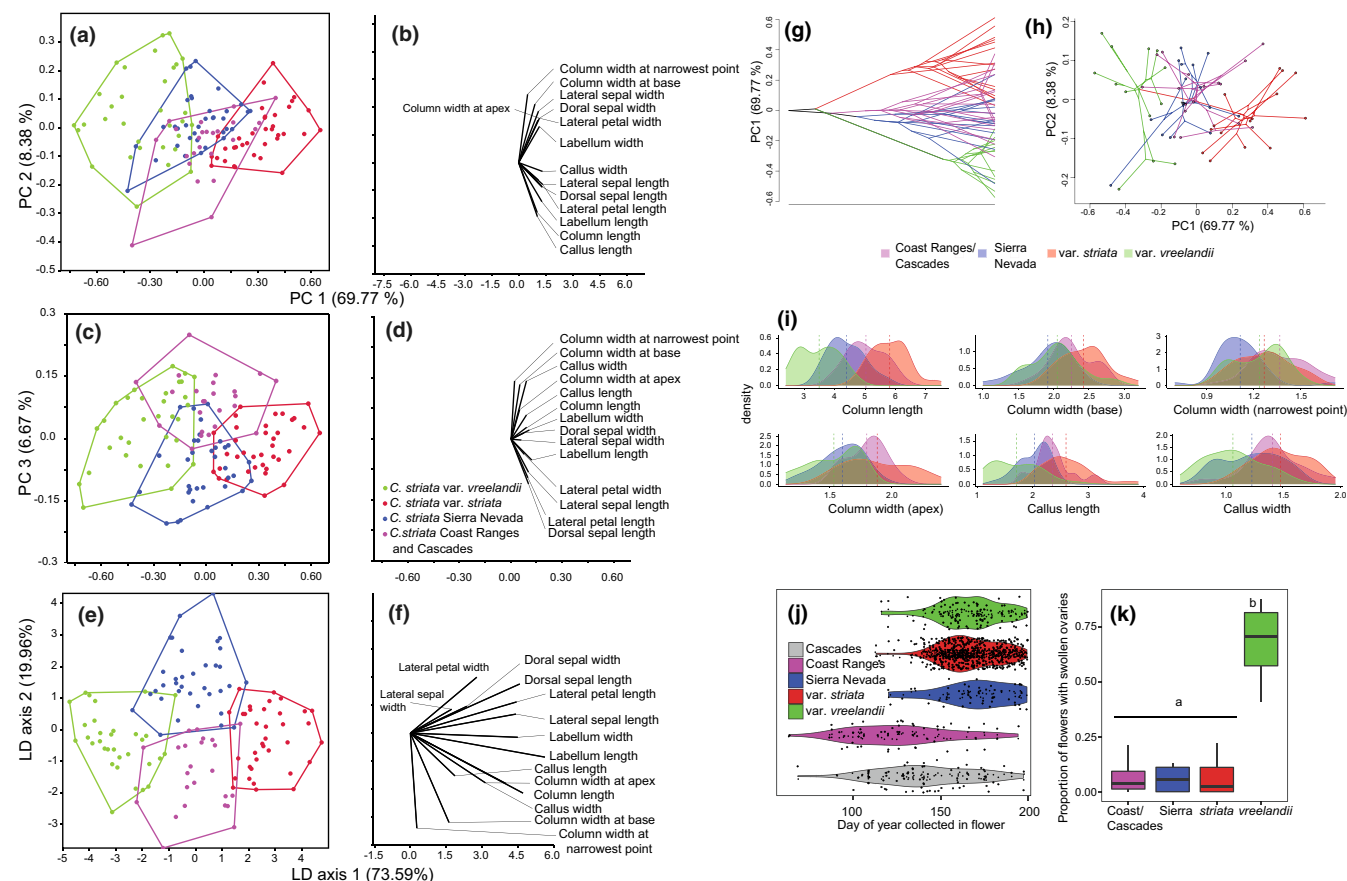


FIGURE 3 Morphological, phenological, and reproductive mode analyses of the *C. striata* complex. (a) Principal components analysis (PCA) of 14 log-transformed floral morphological characters, showing PCs 1–2. (b) Biplot showing the loading scores of each character on PCs 1–2. (c) PCA of PC axes 1 and 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 Nevada 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's post hoc comparison values that are significantly different ($p < .0001$).

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 (bold type) indicate the numbers of successful classifications to a priori groupings

	<i>Var. striata</i>	<i>Var. vreelandii</i>	Sierra Nevada	Coast/ cascades	Total
<i>var. striata</i>	36	0	0	0	36
<i>var. vreelandii</i>	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

while the closest references to associates from the Coast/Cascades are from Minnesota (USA), Hungary, Iran, Mexico, Montenegro, and the UK. (Figure 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 ($\pi_{vreelandii} = 0.029$, $\pi_{striata} = 0.032$). However, there is evident clustering of associates for each grouping and phylogenetic

TABLE 3 Ecospat identity and background/similarity tests for 10 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

Comparison	D, ID/BG, 10 km	I, ID/BG, 10 km	D, ID/BG, 0.5 km	I, ID/BG, 0.5 km
<i>striata</i> - <i>vreelandii</i>	0.24**/0.31	0.33**/0.40	0.28/0.27	0.37**/0.37
<i>striata</i> - Sierra Nevada	0.05/0.04	0.09/0.10	0.18/0.15	0.22/0.10
<i>striata</i> - Coast/Cascades	0.10**/0.11	0.14**/0.13	0.23/0.19	0.29/0.24
<i>vreelandii</i> - Sierra Nevada	0.06**/0.05	0.23**/0.22	0.18/0.15	0.57/0.64
<i>vreelandii</i> - 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

Note: * $p < .05$, ** $p < .01$.

signal among associates with each of the four major groupings based on analysis with phylo.signal.disc (Figure S2; $p < .001$).

3.6 | Integrative analyses

Results from assignPOP were congruent with other clustering methods (Figure 5). Analysis of SNP-only data revealed a strong pattern of differentiation for LDA, SVM, and RF (Figure 5a), with posterior probabilities of assignments = 1.0 in all cases (Table S5). 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 Nevada accessions, assignment probability = 0.88; Table S5), suggesting that these data, in combination, hold a similar signal of differentiation and do not controvert the patterns based on SNP-only data (Figure 5b; Table S5). 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 (Figure 5c; Table S5). Variance partitioning via RDA, with three- and four-species hypotheses as explanatory variables, showed a preference for a four-species model for all extended data sets individually and in combination (Figure 5d). The models were significant in all comparisons (based on R^2 -adjusted values, $p < .01$ in all cases; Figure 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 analysed multiple data sets 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 (Figure 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 & Hillis, 2020; Hillis et al., 2021; Mason et al., 2020). 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 (Figure 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 (Figure 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; Sletvold et al., 2010; Van Der Cingel, 2001; 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 & Lenhard, 2011). In fact, vars. *striata* and *vreelandii* appear to be the most morphologically divergent members (Figures 3a–i), with var. *striata* (large, open flowers; Figure 1b) showing evidence of outcrossing and var. *vreelandii* showing evidence of self-pollination (drab, less-open, smaller flowers; Figure 1a; Freudenstein, 1997). Floral morphology in these two varieties is

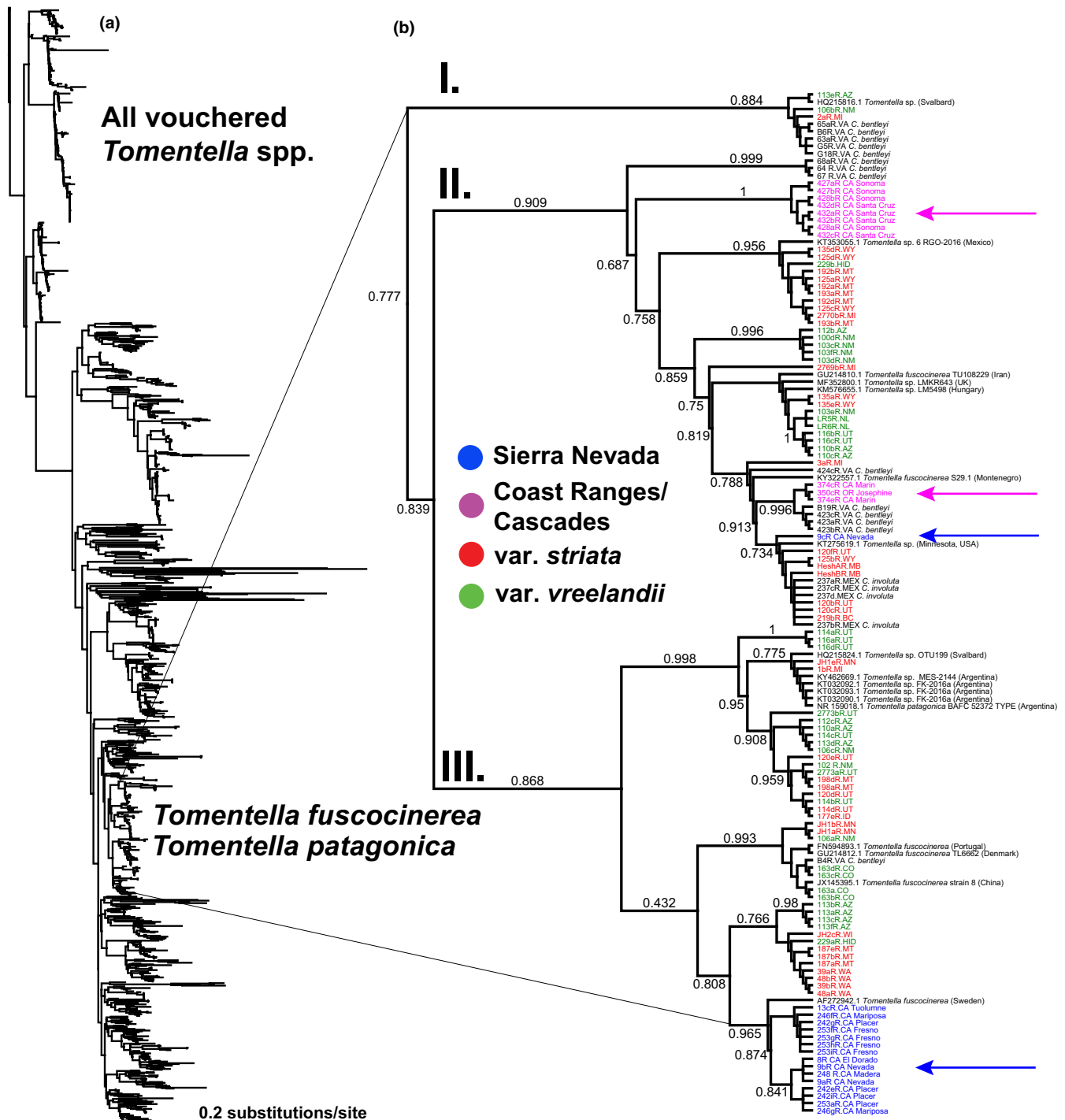


FIGURE 4 (a) Phylogenetic analysis of >1400 *Tomentella* fungal accessions, including GenBank reference sequences and members of the *C. striata* complex. Scale bar = 0.2 substitutions/site. (b) Closeup of the clade occupied by all accessions of the *C. striata* complex, with closely related reference sequences from *Tomentella fuscocinerea* and *T. Patagonia*. Branch lengths are scaled proportionally. Blue and magenta arrows point to associates of the sierra Nevada *C. striata* and the coast/Cascade *C. striata*, respectively. Roman numerals indicate three principal subclades.

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. nonoverlapping 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

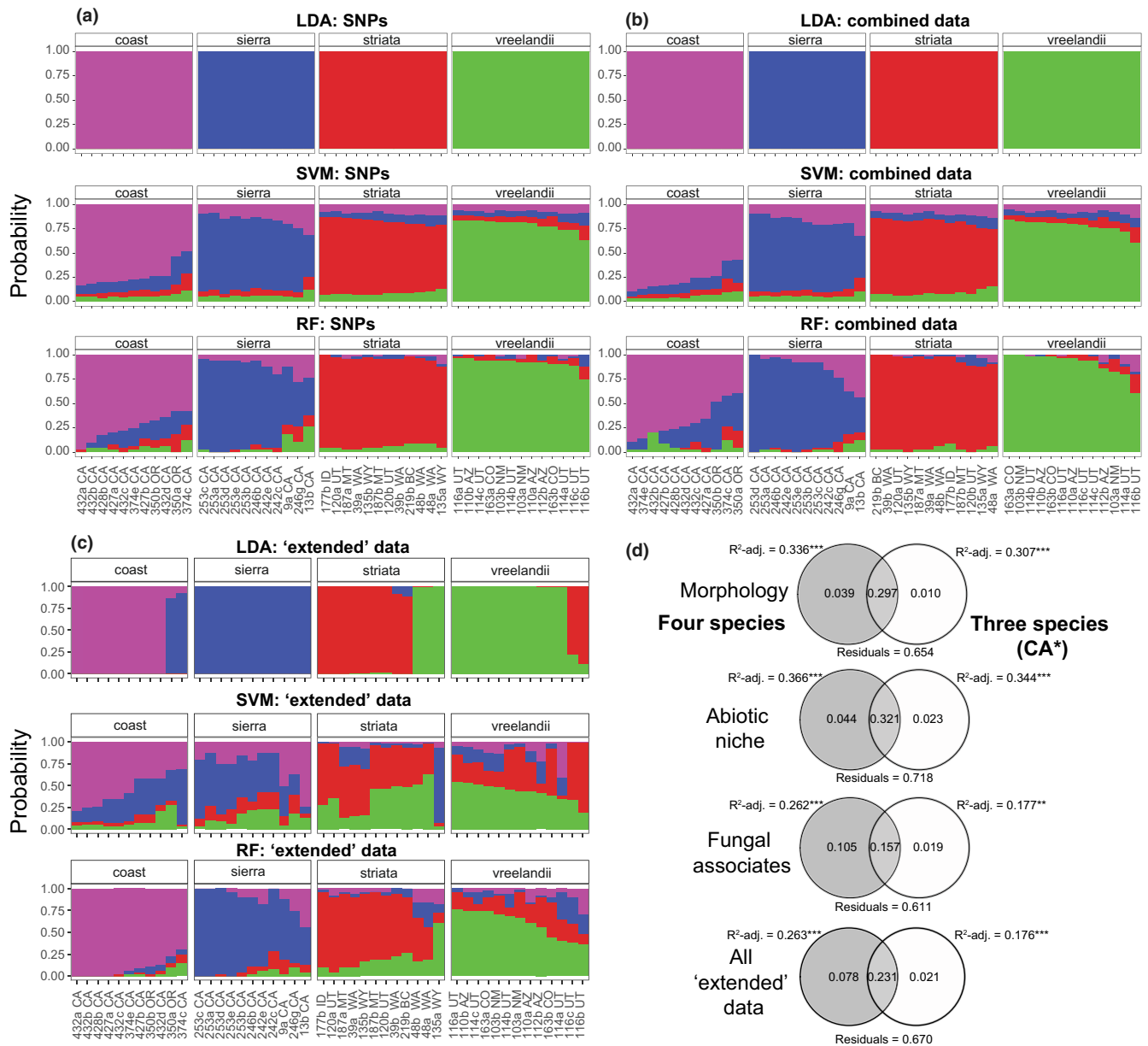


FIGURE 5 Integrative analyses of SNP and extended data. (a) SNPs only analysed in assignPop. (b) Combined SNP and extended data (morphology, biotic niche, and fungal hosts). (c) Extended data only. LDA, linear discriminant analysis; SVM, support vector machine; RF, random forest. (d) Redundancy analysis (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).

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. odorathiza* in northern North America, which are sister species (Freudenstein, 1992, 1997; Freudenstein & 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 (Cardillo & Warren, 2016; Warren et al., 2008; Warren et al., 2021). The background/similarity test attempts to account for this, being more robust in cases of allopatry. However, results of our background/similarity tests were inconclusive at both spatial resolutions (less similarity than expected by chance, but nonsignificant 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, 2013; Wang et al., 2021; Waterman & Bidartondo, 2008). 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 (Figures 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 nonoverlapping associations expected if these orchid lineages had fully codiversified on their host fungi (Bidartondo & Bruns, 2001; Waterman & Bidartondo, 2008). Instead we interpret this pattern as strong but nonexclusive 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 (Jakucs & Erős-Honti, 2008; Kuhar et al., 2016; Larsen, 1965, 1974). Seed burial experiments among sites with known divergent orchid and fungal associates may shed light on what controls this pattern of specificity (Rasmussen & Whigham, 1993; McCormick et al., 2009).

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 (Leake, 1994; Taylor et al., 2002, 2013). These switches to ectomycorrhizal 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 & 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 (Freudenstein & Barrett, 2014; Taylor et al., 2004; Taylor & Bruns, 1997). If host associations are genetically controlled, we would expect fixed polymorphisms coadapted for fungal hosts to be frequent throughout the genome, and in LD within lineages associated with fungal host genotypes (Taylor et al., 2004, 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 (Figure 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 analysed together, even from different data sets, show clear, emergent patterns of differentiation? Recognizing “clusters in phenotypic (or other) space” as species is best described as a phenetic approach (Michener, 1970; Sneath & Sokal, 1973; Sokal & Crovello, 1970; Sokal & Sneath, 1963). While phenetics is widely applied at the species level, many have criticized this approach as lacking evolutionary context (de Queiroz & Good, 1997; Donoghue, 1990; Hull, 1988). 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, 2007). 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 the interpretation of our results—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 observed among the extended data and lineages, or simply the recovery of distinct lineages themselves, is sufficient evidence for recognizing separate species. For example, some may be comfortable calling the different lineages species under the general lineage or unified species concept (sensu de Queiroz, 1998, 2005, 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 & Wheeler, 1990), a single molecular synapomorphy, or suite of SNPs 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 may be fixed synapomorphies at different hierarchical levels in any given data set? We argue that such interpretations would bring us no closer to objectivity, and instead perpetuate some long-standing problems with species delimitation: a lack of repeatable and consistent criteria, and ignoring the role of species in their environments in favour of relying entirely upon genetic data.

5 | CONCLUSION

Here, we have taken an integrative approach to species delimitation in the *Corallorhiza striata* complex, employing seven data sets, 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 “oversplitting”, 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 & 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.

AUTHOR CONTRIBUTIONS

Craig F. Barrett, John V. Freudenstein, and Mathilda V. Santee conceived the study. Mathilda V. Santee, Nicole M. Fama, Brandon T. Sinn, Sandra J. Simon, and Craig F. Barrett designed the laboratory methodology and protocols. Brandon T. Sinn, Sandra J. Simon, Mathilda V. Santee, Nicole M. Fama, and Craig F. Barrett conducted the laboratory work. Craig F. Barrett conducted fieldwork. Brandon T. Sinn, Sandra J. Simon, Craig F. Barrett and Mathilda V. Santee analysed data. Mathilda V. Santee and Craig F. Barrett led the writing of the manuscript. All authors contributed to manuscript drafts and provided consent for publication.

ACKNOWLEDGEMENTS

Funding was provided by the US National Science Foundation awards DEB CAREER 2044259 and OIA 1920858 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 (1) the WVU Provost's Office, (2) the Eberly College of Arts and Sciences, (3) the Honours 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 no. 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:SCR018885](https://doi.org/10.1111/rrid.12885)).

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Sequences of the fungal ITS region have been made available via NCBI GenBank accession numbers GU224038-GU220711 and OM282086-OM282096. Genomic data from ISSRseq have been made available via the NCBI Sequence Read Archive (BioProject PRJNA771539). Morphological, genomic (SNP), phenological, abiotic niche, reproductive mode, and fungal ITS data are available as Supporting Information 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).

ORCID

Craig F. Barrett  <https://orcid.org/0000-0001-8870-3672>

Mathilda V. Santee  <https://orcid.org/0000-0001-5349-724X>

John V. Freudenstein  <https://orcid.org/0000-0001-9410-0624>

Sandra J. Simon  <https://orcid.org/0000-0002-7928-3503>

Brandon T. Sinn  <https://orcid.org/0000-0002-5596-6895>

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, M. 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>
- 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. <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 mycoheterotrophy. *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>
- 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., du Plessis, L., Poppinga, 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. Available from: [https://www.semanticscholar.org/paper/Phylogenetics-of-the-Monotropoideae-\(Ericaceae\)-on-Broe/8f2b05a8f9b268d18a5b221f1a4c11d881905a08](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., Thuiller, W., Fortin, M. J., Randin, C., Zimmermann, N. E., Graham, C. H., & Guisan, A. (2012). Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography*, 21(4), 481–497. <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, 951–963. <https://doi.org/10.1111/geb.12455>
- Carstens, B. C., & Knowles, L. L. (2007). Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: An example from *Melanoplus* grasshoppers. *Systematic Biology*, 56, 400–411. <https://doi.org/10.1080/10635150701405560>
- 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>
- 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>
- 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., & Ludsins, 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 B. G. Baldwin, D. Goldman, D. J. Keil, R. Patterson, T. J. Rosatti, & D. Wilken (Eds.), *The Jepson manual: Vascular plants of California* (2nd ed.). University of California Press.
- Crotty, S. M., Minh, B. Q., Bean, N. G., Holland, B. R., Tuke, J., Jermin, 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. D. O., & 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>
- Darwin, C. (1876). *The effects of cross and self-fertilization in the vegetable kingdom*. John Murray.
- 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 University Press.
- 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>
- de Queiroz, K. (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* (pp. 57–75). Oxford University Press.
- 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>
- Dobzhansky, T. (1950). Mendelian populations and their evolution. *American Naturalist*, 84, 401–418.
- Donoghue, M. J. (1985). A critique of the biological species concept and recommendations for a phylogenetic alternative. *Bryologist*, 88, 172–181.
- 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.
- Drès, M., & Mallet, J. (2002). Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 357, 471–492.
- 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.
- 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>
- Firreno, T. J., Jr., 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. 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., de Oliveira, R. S., 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>
- 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>
- 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. <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>
- 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. <https://doi.org/10.1073/pnas.1410442111>
- Hammer, O., Harper, D. A. T., & Ryan, P. D. (2001). PAST: Paleontological Statistics Software Package for Education and Data Analysis. 9.
- Hennig, W. (1966). *Phylogenetic systematics*. University of Illinois Press.
- 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). <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 and Herpetology*, 109(3), 895–903. <https://doi.org/10.1643/h2021082>
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35, 518–522. <https://doi.org/10.1093/molbev/msx281>
- Hull, D. L. (1988). *Science as a process: An evolutionary account of the social and conceptual development of science*. 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). Adegnet: 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). Adegnet 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>
- Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, 11, 94. <https://doi.org/10.1186/1471-2156-11-94>
- 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>
- Kalyanamoorthy, 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>
- 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 *Hypocnys* 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., Banbury, B. L., Felsenstein, J., de Oca, A. N., & Stamatakis, A. (2015). Short tree, long tree, right tree, wrong tree: New acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology*, 64(6), 1032–1047. <https://doi.org/10.1093/sysbio/syv053>
- Leaché, A. D., & Fujita, M. K. (2010). Bayesian species delimitation in west African forest geckos (*Hemidactylus fasciatus*). *Proceeding of the Royal Society, Series B*, 277, 3071–3077. <https://doi.org/10.1098/rspb.2010.0662>
- 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. <https://doi.org/10.1093/sysbio/syu018>
- Leake, J. R. (1994). Tansley review no. 69. The biology of mycoheterotrophic ('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. <https://www.cabdirect.org/cabdirect/abstract/20003032923>
- 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*, 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>
- Magrath, L. K., & Freudenstein, J. V. (2002). Corollorhiza. In Flora of North America editorial committee (Ed.), *Flora of North America* (Vol. 26, pp. 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 M. F. Claridge, H. A. Dawah, & M. R. Wilson (Eds.), *Species: The units of diversity* (pp. 381–423). Chapman & Hall.
- Mayr, E. (1942). *Systematics and the origin of species*. Columbia University Press.
- Mayr, E. (1982). Speciation and macroevolution. *Evolution*, 36(6), 1119–1132. <https://doi.org/10.1111/j.1558-5646.1982.tb05483.x>
- McCormick, M. K., Whigham, D. F., O'Neill, J. P., Becker, J. J., Werner, S., Rasmussen, H. N., Bruns, T. D., & 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>
- 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>
- Merckx, V. (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. <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* (pp. 441–448). CSIRO.
- 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.
- O'Meara, B. C. (2010). New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology*, 59(1), 59–73. <https://doi.org/10.1093/sysbio/syp077>
- 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. D. 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. (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[2614:VPOSDM]2.0.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 C. Scornavacca, F. Delsuc, & N. Galtier (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*. 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>
- 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, 529. <https://doi.org/10.1186/s12859-017-1934-z>
- Schlick-Steiner, B. C., Steiner, F. M., Seifert, B., Staufer, 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>
- 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/10.1093/10.2307/2405675>
- Simpson, G. G. (1961). *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>
- Sneath, P. H., & Sokal, R. R. (1973). *Numerical taxonomy: The principles and practice of numerical classification* (1st ed.). Freeman.
- 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>
- Sokal, R. R., & Sneath, P. H. A. (1963). *Principles of numerical taxonomy*. W. H.
- 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 R. A. Wilson (Ed.), *Species: New interdisciplinary essays*. Bradford.
- Su, H.-J., Barkman, T. J., Hao, W., Jones, S. S., Naumann, J., Skippington, E., Wafula, E. K., Hu, J. M., Palmer, J. D., & de Pamphilis, C. W. (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. <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 its 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., 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
- Taylor, D. L., Bruns, T., Leake, J. R., & Read, D. (2002). Mycorrhizal specificity and function in myco-heterotrophic plants. In *Mycorrhizal Ecology* (pp. 375–413). Springer.
- 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. 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>
- 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>
- 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.
- van der Kooij, 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. Epub ahead of print. <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 M. A. Innis, D. H. Gelfand, J. J. Sninsky, & T. J. White (Eds.), *PCR protocols: A guide to methods and applications* (pp. 315–322). Academic Press, Inc..
- 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.113.13373>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.
- 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*. 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>
- Wright, S., & Huxley, J. (1940). The statistical consequences of mendelian heredity in relation to speciation. In *The new systematics* (pp. 161–183). Oxford University Press.
- Yang, Z., & Rannala, B. (2010). Bayesian species delimitation using multi-locus 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>
- Yang, Z., Wafula, E. K., Kim, G., Shahid, S., McNeal, J. R., Ralph, P. E., Timilsena, P. R., Yu, W. B., Kelly, E. A., Zhang, H., Person, T. N., Altman, N. S., Axtell, M. J., Westwood, J. H., & de Pamphilis, C. W. (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>
- 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., Wang, D., Lai, C., Xu, W., Huang, J., Zha, L., Liu, D., Ma, X., Wang, L., Zhou, M., Jiang, Z., Meng, H., Peng, H., Liang, Y., ... Huang, L. (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), 179–194. <https://doi.org/10.1093/sysbio/syr084>
- Zietkiewicz, E., Rafalski, A., & Labuda, D. (1994). Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics*, 20, 176–183.
- 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>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Barrett, C. F., Santee, M. V., Fama, N. M., Freudenstein, J. V., Simon, S. J., & Sinn, B. T. (2022). Lineage and role in integrative taxonomy of a heterotrophic orchid complex. *Molecular Ecology*, 00, 1–20. <https://doi.org/10.1111/mec.16617>