MS. SARAH J JACOBS (Orcid ID : 0000-0002-3125-6266)

Article type : Original Article

RUNNING HEAD: Molecular Species Delimitation Incongruence

Incongruence in molecular species delimitation schemes: what to do when adding more data is difficult.

Sarah J. Jacobs[1,2,3,*], Casey Kristoferson[1,2,3], Simon Uribe-Convers[1,2,3,§], Maribeth Latvis[1,2,3,¶], and David C. Tank[1,2,3]

[1] University of Idaho, Department of Biological Sciences, 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051

[2] Stillinger Herbarium, University of Idaho, 875 Perimeter Drive MS 1133, Moscow, ID 83843-1133

[3] Institute for Bioinformatics and Evolutionary Studies (IBEST), University of Idaho, 875 Perimeter Drive MS 3051, Moscow, ID 83843-3051

[§] University of Michigan, Department of Ecology and Evolutionary Biology, 830 North University, Kraus Natural Science Building, Ann Arbor, MI 48109

[¶] South Dakota State University, Department of Natural Resource Management, 1390 College Ave., Brookings, SD 57007

[*] *Corresponding Author*: sarahjjacobs@gmail.com

ABSTRACT

Using multiple, independent approaches to molecular species delimitation is advocated to accommodate limitations and assumptions of a single approach. Incongruence in delimitation schemes is a potential byproduct of employing multiple methods on the same data, and little attention has been paid to its reconciliation. Instead, a particular scheme is prioritized and/or molecular delimitations are coupled with additional, independent lines of evidence that mitigate incongruence. We advocate that incongruence *within* a line of evidence should be accounted for before comparing *across* lines of evidence, that can themselves be incongruent. Additionally, it is not uncommon for empiricists working in non-model systems to be data-limited, generating some concern for the adequacy of available data to address the question of interest. With conservation and management decisions often hinging on the status of *species*, it seems prudent to understand the capabilities of approaches we use given the data we have. Here we apply two molecular species delimitation approaches, spedeSTEM and BPP, to the *Castilleja ambigua* (Orobanchaceae) species complex, a relatively young plant lineage in western North America. Upon finding incongruence in our delimitation, we employed a post-hoc simulation study to examine the power of these approaches to delimit species. Given the data we collected, we find that spedeSTEM lacks the power to delimit while BPP is capable, thus allowing us to address incongruence before proceeding in delimitation. We suggest post-hoc simulation studies like this compliment empirical delimitation and serve as a means of exploring conflict within a line of evidence and dealing with it appropriately.

Keywords: species delimitation, incongruent delimitation scheme, spedeSTEM, BPP, simulation, *Castilleja*

INTRODUCTION

Species are one of the basic units of scientific inquiry, and the way we define species can have far-reaching impact – e.g., our understanding of biodiversity (Agapow *et al.* 2004; Pimm *et al.* 2014; Adams *et al.* 2014), our approaches to conservation (Myers *et al.* 2000; Hedrick 2001; Costello *et al.* 2013), and our understanding of evolutionary processes (Ruane *et al.* 2014; Morales *et al.* 2016). Because of this, species delimitation is central to the biodiversity sciences (e.g., Sites & Marshall 2003; Wiens 2007; Leaché & Fujita 2010; Camargo & Sites 2013; Carstens *et al.* 2013; Rannala 2015; Flot 2015). The advancement of molecular-based delimitation approaches through the incorporation of coalescent theory (e.g., Pons *et al.* 2006; Knowles & Carstens 2007; O'Meara 2010; Yang & Rannala 2010), has represented a huge step forward in our ability to robustly delimit species, especially at recent timescales. The past ten years have seen an explosion in molecular species delimitation approaches (e.g., Pons *et al.* 2006; Knowles & Carstens 2007; O'Meara 2010; Yang & Rannala 2010; Ence & Carstens 2010; Camargo *et al.* 2012; Grummer *et al.* 2014;

Solís-Lemus *et al.* 2015), empirical examples (e.g., Reeves & Richards 2010; Goldberg *et al.* 2011; Satler *et al.* 2013; Singh *et al.* 2015), and critical reviews (e.g., Leaché & Fujita 2010; Camargo *et al.* 2012; Carstens *et al.* 2013). Most authors agree that the use of multiple lines of evidence (Schlick-Steiner *et al.* 2010; Yeates *et al.* 2010), multiple approaches in conjunction (Fujita 2012; Aguilar *et al.* 2013; Andújar *et al.* 2014), and when possible, integrated analyses (Padial *et al.* 2010; Zapata & Jiménez 2012; Guillot *et al.* 2012; Edwards & Knowles 2014), are necessary to be objective in our delimitations.

However, despite the amount of work in this area, few studies have specifically addressed how to handle conflict. Conflict occurs when independent approaches result in incongruent delimitations—i.e., the delimitation scheme of one approach differs from that of another. Possible explanations of incongruent delimitations might include different signals across different lines of evidence (e.g., morphological delimitation differs from molecular delimitation) or violation of assumptions and/or different degrees of statistical power of an analysis. Incongruence in delimitation *across* lines of evidence can be mediated by evaluating delimitation with each line of evidence independently and then determining which data source to rely on given biological and/or evolutionary explanations for disagreement across datasets (e.g., Schlick-Steiner *et al.* 2010; Yeates *et al.* 2010). The integration of multiple lines of evidence into unified species delimitation analyses—i.e., where all data are used simultaneously—may help alleviate this subjectivity (e.g., Edwards & Knowles 2014; Solís-Lemus *et al.* 2015). However, results of multiple analyses on the *same* dataset (for example, applying several molecular species delimitation methods on the same molecular dataset) can also differ, highlighting when the limitations of a particular approach may impact delimitation (e.g., Satler *et al.* 2013).

For example, consider spedeSTEM (Ence & Carstens 2010) and BPP (Yang & Rannala 2010), two commonly applied delimitation methods utilizing the multispecies coalescent that can disagree in practice; the likelihood-based approach spedeSTEM relies on highly informative gene trees to build a species tree, which is then used to test and rank all possible permutations of lineage composition, and the Bayesian approach BPP estimates the posterior probability of bifurcations on a guide tree that are collapsed to examine all

possible combinations of putative lineages. The largely conservative spedeSTEM has been shown to under-delimit species (Ence & Carstens 2010), while BPP may over-delimit (Leaché & Fujita 2010), especially in the case of inaccurate guide trees (but see Zhang *et al.* 2014) and/or misspecified priors (Giarla *et al.* 2014). Therefore, if conflict occurs between these two approaches, it could mean that uninformative gene trees may be limiting spedeSTEM, and/or misinformed analytical parameters may be limiting BPP (e.g., Camargo *et al*. 2012, Carstens & Satler 2013, Pelletier *et al*. 2014, Giarla *et al*. 2014). Improvements to BPP have addressed this possibility by incorporating the estimation of the species tree topology in conjunction with species delimitation (Yang & Rannala 2014a). Recent theoretical work has highlighted the sensitivity of the multispecies coalescent and its use by BPP, highlighting the potential for detecting population structure, rather than what many delimitation analyses are aiming for, i.e., species boundaries (Sukumaran & Knowles 2017). Other methods employing the coalescent potentially risk this as well. It is apparent that now, more than ever, we should be addressing the capability of the methods we employ to perform the tasks that we expect they do.

If we find incongruent delimitation schemes from analyses that use the same input data, it may suggest differing degrees of statistical power in the approaches we use. Additionally, because the parameter space associated with any question of species delimitation is complex and intractable, simplifying assumptions must be made to minimize the number of parameters considered; each analytical approach will simplify in different ways, and thus, each approach will have different implicit assumptions (Carstens *et al*. 2013). Statistical power is a topic explored in methodological papers, and most often includes simulations and an empirical example to understand the limitations of the method. How the approach behaves in *other* systems is left to the exploration of the user. Incongruence across delimitations using the same input data is not uncommon and has been shown to be particularly problematic in studies with small sample sizes (Carstens *et al.* 2013). When working with small or limited datasets, a knee-jerk reaction might be to increase sampling (loci or individuals). Several studies have documented the impact of small sample sizes on delimitation, and general 'good practices' of species delimitation

143 suggest at least 10 individuals per putative lineage and as many loci as possible (Carstens *et*
144 *al.* 2013). Increasing the number of loci in a dataset has become easier to do (e.g.,
145 McCormack *et al.* 2013; Lemmon & Lemmon 2013), and there is a general consensus in the
146 phylogenetics community that more loci typically result in increased resolution (Ruane *et*
147 *al.* 2015; Blaimer *et al.* 2015). However, genome-scale data are still time consuming and
148 expensive to generate, particularly for non-model organisms, and there can be
149 computational disadvantages to using hundreds of loci (Ruane *et al.* 2015). Furthermore,
150 for rare taxa—e.g., those known from only a few, often small, populations, and/or those that
151 are spatially restricted—the incorporation of 10 individuals per putative lineage may not be
152 possible (Lim *et al.* 2012). For these reasons, empirical studies, especially those dealing
153 with rare or spatially restricted taxa, often begin with existing datasets (often Sanger
154 sequenced data or data obtained from GenBank) that, in terms of individuals and loci
155 sampled, are often smaller in size.
156 When a researcher recovers conflicting delimitation schemes across approaches
157 using a dataset that is limited in size, an alternative analytical tactic is an assessment of the
158 data already at hand (i.e., less than ideal datasets). In other words, an assessment of the
159 capability of each methodological approach to detect the signal of independent lineages in
160 the data collected. This can be directly tested in empirical studies using post-hoc
161 simulations. While this has been implied as an appropriate and important step in empirical
162 delimitation (Carstens *et al.* 2013), and some studies have simulated data in order to
163 compare methodological approaches (e.g., Camargo *et al.* 2012; Barley *et al.* 2017) or to
164 specifically address sample size (e.g., Giarla *et al.* 2014; Hime *et al.* 2016), to our knowledge
165 an assessment of inferential error has not been specifically done in any empirical study.
166 In this study, we apply species delimitation approaches to a species complex in the
167 plant genus *Castilleja*, a widespread and iconic wildflower that is most diverse in western
168 North America. A recent, rapid radiation (Tank & Olmstead 2008), *Castilleja* is an important
169 target for species delimitation, both theoretically and practically. Theoretically, the young
170 age of this lineage affords us the opportunity to test the limits and capabilities of
171 delimitation approaches in a group where molecular, morphological, ecological, and

geographic boundaries between species are often 'fuzzy'. Furthermore, *Castilleja* is known to have a rich history of hybridization and genome duplication events that have complicated the taxonomy and systematics of the genus (Heckard & Chuang 1977; Chuang & Heckard 1991; Tank & Olmstead 2008). Practically speaking, recent advances in sequence generation (e.g., Uribe-Convers *et al.* 2016) and analytical approaches (e.g., Morales *et al.* 2016), combined with focused delimitation efforts, provide an opportunity to refine what we know about the evolutionary history and species composition of *Castilleja*. However, as is the case with many empiricists working in non-model systems, we are working towards becoming 'data-rich' in *Castilleja*, but to some degree we are still currently data-limited (i.e., we do not have tens to hundreds of loci). This is important from a conservation standpoint. Many species of *Castilleja* (including two taxa studied here) are only known from narrow ranges that are vulnerable to extirpation. Knowledge of their evolutionary relationships, and, if warranted, status as a species, will impact conservation and management efforts.

Here, we propose a strategy to species delimitation when data is limited. By simulating data comparable to the empirical data and under a known species tree topology, we can directly test the capability of molecular species delimitation approaches to delimit the known number of distinct evolutionary lineages. Given this information, we can address conflicting delimitations from an informed position using the data at hand. We think it is important to consider what can (and cannot) be done with small, non-genomic datasets. We suggest an approach that allows us to address the assumption that a given species delimitation method is capable of delimiting species with the data that we currently have available to us.

METHODS

Study System

We focus our attention on two annual, diploid lineages of *Castilleja*: the polymorphic *Castilleja ambigua* Hook. & Arn. and a close relative, *Castilleja victoriae* Fairbarns and J.M. Egger (Fig. 1). Generally occurring in maritime locations, members of *C. ambigua* typically inhabit coastal bluffs, salt marshes, and grasslands of the western coast of North America, and are united by vegetative morphology and reproductive similarities (Egger *et al.* 2012;

Wetherwax *et al.* 2016). There is, however, variability within the species that has led to the description of multiple intraspecific varieties that are primarily distinguished from one another by ecological preferences and geographic ranges, but also differ in some morphological characters (Fairbarns & Egger 2007; Egger *et al.* 2012).

The typic and most widespread of these varieties, *C. ambigua* var. *ambigua*, has white and yellow flowers and occurs on coastal bluffs and grasslands along the Pacific coast from southern California north, into British Columbia (Fig. 1). *C. ambigua* var. *humboldtiensis* (D.D. Keck) J.M. Egger, is a fleshy, less-branched variety and has primarily pink to rose-purple flowers and a much narrower distribution. It occurs in salt marshes along the northern coast of California in Mendocino and Humboldt counties. Another narrow-ranged variety, *C. ambigua* var. *insalutata* (Jeps.) J.M. Egger is non-fleshy and its stems are highly branched. It, too, has pink-purple flower coloration and occurs in grassy coastal bluffs along the central California coast, between San Mateo and San Luis Obispo counties. More recently, Egger *et al.* (2012) described the variety *C. ambigua* var. *meadii* J.M. Egger & Ruygt. Vegetative morphology, restricted range, and ecological preferences readily distinguish *C. ambigua* var. *meadii* from the other varieties; variety *meadii* is typically erect, with un-branched stems, and leaves and bracts with narrow, linear lobes. In addition, it is restricted to the Atlas Peak Plateau district of Napa County, California, where it occurs in seasonally wet places associated with freshwater and is known from only four extant populations (a fifth being recently documented as extirpated (Egger *et al.* 2012)).

Another member of this complex described in 2007 (Fairbarns & Egger 2007), *Castilleja victoriae*, has been allied to *C. ambigua*. Both species share a coastal range, but *C. victoriae* is associated with edge habitat of fresh water seeps and vernal pools, and is restricted to southwestern British Columbia, Canada, and a single island in the San Juan Archipelago of extreme northwestern Washington State, USA. This species is formally known from only three extant populations (a fourth being recently documented as extirpated (Fairbarns & Egger 2007). Morphologically, *C. victoriae* tends toward a compact, single-stemmed habit and lacks the distinctive contrasting floral coloration of *C. ambigua*. A

difference in stigma position at peak flowering time between *C. ambigua* (exserted) and *C. victoriae* (inserted) is also diagnostic.

Because of the morphological and ecological variation outlined above, in addition to the conservation and management implications of species status of the two range-restricted taxa, we focus on testing the distinctiveness of the following three taxa: *Castilleja ambigua*, *C. ambigua* var. *meadii*, and *C. victoriae*. For the purposes of this work we treat *Castilleja ambigua* varieties *ambigua*, *insalutata*, and *humboldtiensis* as part of *Castilleja ambigua*.

Molecular Methods

Taxon sampling and DNA extraction.—Thirteen accessions of *Castilleja ambigua* (including two accessions of var. *insalutata* and one of var. *humboldtiensis*), three accessions of *C. ambigua* var. *meadii*, and three accessions of *C. victoriae* were sampled throughout their ranges, and the closely related *C. lacera* (Tank & Olmstead 2008; Tank *et al.* 2009) was chosen to serve as outgroup for phylogenetic analyses (Fig. 1; Supplementary Table S1). Total genomic DNA was extracted from either silica-gel dried tissue or tissue sampled from herbarium specimens using a modified CTAB method (Doyle and Doyle 1987).

Chloroplast dataset.—We used a set of *Castilleja*-specific chloroplast primers designed to amplify the most variable regions of the chloroplast genome (Latvis *et al.* 2017; Supplementary Table S2). Following Uribe-Convers *et al.* (2016), microfluidic PCR was performed on 45 primer pairs on the Fluidigm Access Array System (Fluidigm Co., San Francisco, California, USA). The resulting amplicons were sequenced on an Illumina MiSeq platform using the Reagent Kit v.3 (300 bp paired-end reads; Illumina Inc., San Diego, California, USA). Microfluidic PCR, downstream quality control and assurance, and Illumina sequencing was performed in the University of Idaho Institute for Bioinformatics and Evolutionary Studies (IBEST) Genomics Resources Core Facility.

Nuclear dataset.— The nuclear ribosomal sequences from the internal and external transcribed spacers (ITS and ETS, respectively) used here were collected in two ways—

first, following traditional Sanger sequencing approaches, and second, using a targeted amplicon sequencing (TAS) strategy modified from (Bybee *et al.* 2011). Both approaches used ITS2, ITS3, ITS4, and ITS5 primers from (Baldwin 1992) to amplify the entire ITS region, as well as the ETS-B (Beardsley & Olmstead 2002) and 18S-IGS primers (Baldwin & Markos 1998) to amplify a portion of the 3' end of the ETS region. For Sanger sequenced products (Supplementary Table S1), PCR was performed following Tank and Olmstead (2008), and prior to sequencing, amplified PCR products were cleaned and purified by precipitation from 20% polyethylene glycol solution and washed in 70% ethanol. Both strands of the cleaned PCR products were sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, California, USA) with the same primers used during amplification on an ABI 3130xl Genetic Analyzer (Applied Biosystems, Foster City, California, USA). For TAS, the ITS and ETS regions were amplified using a two-round PCR strategy (Supplementary Table S1). Following Uribe-Convers *et al.* (2016), each target-specific primer sequence contained a conserved sequence tag that was added to the 5' end at the time of oligonucleotide synthesis (CS1 for forward primers and CS2 for reverse primers), to provide an annealing site for the second pair of primers. After an initial round of PCR using the CS-tagged, target specific primers (PCR1), a second round of PCR was used to add sample-specific barcodes and high-throughput sequencing adapters to both the 5' and 3' ends of each PCR amplicon (PCR2). From 5' to 3', the PCR2 primers included either Illumina P5 (CS1-tagged forward primers) or P7 (CS2-tagged reverse primers) sequencing adapters, 8 bp sample-specific barcodes, and the reverse complement of the conserved sequence tags. Sequences for the CS1 and CS2 conserved sequence tags, barcodes, and sequencing adapters were taken from Uribe-Convers *et al.* (2016). Following PCR2, the resulting amplicons were pooled together and sequenced on an Illumina MiSeq platform using 300 bp paired-end reads, as with chloroplast sequencing. PCR conditions were as follows: PCR1—25 uL reactions included 2.5 uL of 10x PCR buffer, 3 uL of 25 mM $MgCl_2$, 0.30 uL of 20 mg/mL BSA, 1 uL of 10 mM dNTP mix, 0.125 uL 10 uM CS1-tagged target specific forward primer, 0.125 uL 10 uM CS2-tagged target specific reverse primer, 0.125 uL of 5000 U/ml Taq DNA polymerase, 1 uL template DNA, and PCR-grade H_2O to volume;

PCR1 cycling conditions - 95°C for 2 min. followed by 20 cycles of 95°C for 2 min., 50°C for 1 min., 68°C for 1 min., followed by a final extension of 68°C for 10 min.; PCR2 – 20 uL reactions included 2 uL of 10x PCR buffer, 3.6 uL of 25 mM $MgCl_2$, 0.60 uL of 20 mg/mL BSA, 0.40 uL of 10 mM dNTP mix, 0.75 uL of 2 uM barcoded primer mix, 0.125 uL of 5000 U/mL Taq DNA polymerase, 1 uL of PCR1 product as template, and PCR-grade H_2O to volume; PCR2 cycling conditions—95°C for 1 min. followed by 15 cycles of 95°C for 30 sec., 60°C for 30 sec., 68°C for 1 min., followed by a final extension of 68°C for 5 min.

Dataset preparation—For the chloroplast and TAS-generated nuclear ribosomal datasets, pooled reads from Illumina MiSeq runs were demultiplexed using the dbcAmplicons pipeline, and consensus sequences were generated using the R script reduce_amplicons.R (https://github.com/msettles/dbcAmplicons) following the workflow detailed in Uribe-Convers *et al.* (2016). Briefly, for each sample, read-pairs were identified, sample-specific dual barcodes and target specific primers were identified and removed, and each read was annotated to include the species name and read number for each gene region. To eliminate fungal contamination that may have been amplified for ITS, each read was screened against a reference file of annotated sequences retrieved from GenBank (using the "-screen" option in dbcAmplicons). Reads that mapped with default sensitivity settings were kept. Each read was reduced to the most frequent length variant, paired reads that overlapped by at least 10bp (default) were merged into a single continuous sequence, and a consensus sequences without ambiguities were produced ("-p consensus" in the R script reduce_amplicons.R from dbcAmplicons). Paired reads that did not overlap were concatenated together using Phyutility v.2.2.6 (Smith & Dunn 2008), and any merged segments were added to the concatenated reads (Supplementary Table S2). The resulting chromatograms from Sanger sequencing were edited and contigs were assembled using Sequencher v.4.7 (Gene Codes Corp., Ann Arbor, Michigan, USA).

Phylogenetic Analyses

Alignment and model selection—Each chloroplast (cp) and nuclear ribosomal (nr) DNA region was aligned separately using Muscle v.3.8.31 (Edgar 2004). Sequences from individual chloroplast regions were concatenated into a single dataset with Phyutility v.2.2.6 (Smith & Dunn 2008) and treated as a single locus. Likewise, the ITS and ETS regions are tightly linked in the nrDNA repeat and were also treated as a single locus. The best-fit partitioning schemes and models of molecular evolution for nucleotide alignments were selected using PartitionFinder (Lanfear *et al.* 2012), where predefined data blocks corresponded to each region of the chloroplast dataset (i.e., single-end reads or merged reads; Supplemental Table S2), and ITS and ETS, in the case of the nuclear dataset. The Bayesian information criterion (BIC), as implemented in PartitionFinder, was used to identify the highest-ranking models of molecular evolution. All downstream phylogenetic analyses used these partitioning schemes and models.

Gene trees—Maximum likelihood gene trees were estimated with cpDNA and nrDNA as implemented in the program Garli v.2.0 (Zwickl 2006). Twenty-five search replicates were performed, and subsequent log files were examined to ensure that each replicate search resulted in similar trees and log likelihood scores, thus indicating that the analyses consistently found the same topology. A bootstrap run of 1,000 replicates was conducted to assess nodal support. The SumTrees function of the DendroPy package v.4.0 (Sukumaran & Holder 2010) was used to summarize bootstrap results.

Bayesian phylogenetic analyses were conducted on cpDNA and nrDNA datasets using MrBayes v.3.2.1 (Ronquist *et al.* 2012). Each analysis consisted of four Markov chains (using default heating schemes), sampled every 10,000 generations for a total of 5,000,000 generations. To avoid false stationarity at local optima, we conducted four independent runs of each analysis. Stationarity of the chains and convergence of parameter estimates were determined by plotting the likelihood score and all other parameter values against the generation time using the computer program Tracer v.1.5 (Drummond *et al.* 2012). Stationarity was assumed when all parameter estimates and the likelihood had stabilized. Additionally, the likelihoods of the independent runs were considered indistinguishable

when the average standard deviation of split frequencies was <0.01. Burn-in positions were visually assessed and a conservative initial 25% of trees were discarded, and the remaining trees and their associated values saved. The *sump* and *sumt* commands in MrBayes were used to summarize the estimated posterior distributions of both the parameter values and the trees across runs. A majority rule consensus tree showing all compatible partitions from the resulting posterior distribution of topologies was used to recover the posterior probabilities of nodes.

Species tree—We performed a *BEAST analysis with BEAST v.2.0 (Bouckaert *et al.* 2014) via the CIPRES Science Gateway (Miller *et al.* 2010) using the nrDNA and the cpDNA dataset and previously identified partitioning schemes and nucleotide substitution models. Individuals were mapped to species according to taxonomic identification. We employed a strict molecular clock to estimate relative times of diversification events and a constant population size prior. Five independent analyses were conducted for 500 million generations each, sampling the posterior every 10,000 generations. In addition, a run without data was performed to examine the influence of the priors on posterior parameter estimates. Convergence and stationarity of the chains was assessed the same way as with the mrBayes analyses. Burn-in was estimated from each trace file separately, the trees discarded, and then all analyses were combined using LogCombiner v.2.2.0 and a maximum clade credibility tree was summarized with TreeAnnotator v.2.2.0 (Drummond *et al.* 2012).

Molecular Species Delimitation

Here we aim to test the delimitation of our focal taxa (*C. ambigua*, *C. ambigua* var. *meadii*, and *C. victoriae*) as distinct evolutionary lineages. We apply two independent coalescent-based species delimitation methods – the maximum-likelihood approach spedeSTEM (Ence & Carstens 2010), and the Bayesian approach BPP v.3.1 (Yang & Rannala 2014b). We use these methods in a *validation* context (as opposed to *discovery* (*sensu* Ence & Carstens 2010), as the assignment of individuals to a taxonomic group is done prior to the delimitation analysis. When referring to topological relationships in the following sections,

we use the following acronyms for simplification: *C. ambigua* (AMB), *C. ambigua* var. *meadii* (MEA), *C. victoriae* (VIC), and *C. lacera* (LAC).

Estimating theta and tau—Both molecular species delimitation approaches used here require an estimate of population size parameters, encompassed in the variable theta (θ); BPP also requires an estimate of divergence time, tau (τ). We used the program MIGRATE-N v.3.6 (Beerli & Felsenstein 2001) to estimate a value of θ appropriate for our dataset. Sequences were organized into populations corresponding to their taxonomic identification; each taxon was treated as one population. Three independent analyses were conducted to ensure convergence on the same parameter estimates, each consisting of one long chain and 10 short chains (four of which were statically heated). We used analysis A00 (part of the BPP program, this analysis estimates both θ and τ parameters) of the program BPP to estimate τ. We modeled this parameter on the species tree topology from our *BEAST analysis and loosely informed the prior with our MIGRATE-N results. Multiple independent analyses were conducted to confirm results were stable across runs. This analysis also estimates θ, affording us the opportunity to compare our MIGRATE-N and BPP estimates of this parameter. Further details of both approaches can be found in the Supplementary Data S3.

spedeSTEM—The maximum likelihood (ML) delimitation approach spedeSTEM (Ence & Carstens 2010) calculates the ML species tree for all possible models of lineage-composition, given a set of gene trees and an estimate of θ. In our case, this corresponds to five models that reflect all possible combinations of our focal, *a priori* defined taxa: one model with three distinct lineages (AMB, VIC, MEA), three models with two distinct lineages (where the '_' between acronyms indicates a combined lineage) [AMB_VIC, MEA], [AMB_MEA, VIC], and [MEA_VIC, AMB], and a final model of one distinct lineage [AMB_MEA_VIC]. Post likelihood calculations, the competing lineage-composition models are ranked and scored using information theory to identify the best model (further detail below). Because our sampling efforts were disproportionately weighted towards *Castilleja*

ambigua, we used the replicated subsampling approach in STEM (Hird *et al.* 2010) to generate 100 sets of gene trees (a set composed of one chloroplast and one nuclear gene tree) with three alleles subsampled from our dataset per focal lineage (except *C. lacera*, which is represented in our dataset with a single allele only and is therefore present once in each gene tree). Our subsampling was constrained to three per focal lineage, given that we had three alleles only from *C. victoriae* and *C. ambigua* var. *meadii* from which to subsample. Hird *et al*. (2010) demonstrated that as few as three to five alleles could produce accurate estimates of the species tree, provided enough loci. These subsampled gene trees were then used as input in 100 separate spedeSTEM analyses. At the end of the analysis, we are left with 100 likelihoods for each model of lineage composition. Following Ence and Carstens (2010), we then calculated the average likelihood for each model and used the Akaike Information Criterion (AIC) to calculate model differences (Δ_i) and weights (w_i). This series of calculations describes the amount of information lost between a given model *i* and the next best model and describes the probability that this model *i* is the best model (Anderson 2008).

BPP—The Bayesian approach BPP v.3.1 (Yang & Rannala 2014b), when provided with sequence data and parameter estimates (that include θ, τ), examines support for various delimitation schemes by collapsing internal nodes of a species tree and calculating probabilities of those nodes. Previous versions of BPP (Rannala & Yang 2013) required the user to provide the species tree (called the guide tree). Simulations and empirical studies have suggested that incorrect guide-trees could lead to strongly supported, over-split lineages (e.g., Leaché & Fujita 2010; but see (Zhang *et al.* 2014). The version used here retains the user-provided guide tree (called analysis A10, which can be beneficial when the species phylogeny is known because it is computationally more tractable), but also includes an analysis of delimitation that does not require an estimate of the species tree (called analysis A11). This analysis performs species delimitation and estimates the species phylogeny simultaneously.

Here, we applied both approaches. In the guided analysis (A10) we provided a guide tree representing our best estimate of the species tree from our *BEAST analysis ((AMB, VIC), MEA) (following (Leaché & Fujita 2010), in addition to our taxonomic hypothesis, ((AMB, MEA), VIC) and the alternative topology, ((MEA, VIC), AMB). In both analyses (A10 (guided) and A11 (unguided)), we performed a series of multiple replicates to ensure convergence across rjMCMC algorithms, species tree topology (the guide trees in A10; the starting trees in A11), and species model priors (in analysis A11). The guided analysis in BPP reports probabilities of distinction at each node of the guide tree (i.e., probability of speciation at each node of the user-provided guide tree topology). The unguided analysis in BPP reports posterior probabilities for the number of species in the dataset and their probability of species delimitation (i.e., probability that an *a priori* defined taxon is a distinct lineage), and estimates a posterior distribution of species tree topologies.

Post-hoc Simulation Study

To test the capability of these approaches to delimit species in our dataset, we used a simulation approach (Fig. 2). We first simulated one genealogy per locus with the same number of tips and species designations as our empirical gene trees using the program *ms* (Hudson 2002). Next, using scaled versions of these genealogies as guide topologies, we simulated the evolution of nucleotide sequences along the genealogy to generate sequence alignments that are comparable to our empirical dataset using the program *seq-gen* (Rambaut & Grass 1997). The subsequent sequence alignments then become the input datasets for species delimitation with a known topology (i.e., a 'known topology' that we simulated data on), thus allowing us to directly test the capability of each delimitation approach to recover the 'true' delimitation (i.e., the known number of lineages that the data were simulated under). Furthermore, we performed this series of simulations on multiple topologies: the species tree topology (((AMB, VIC), MEA), LAC), the taxonomic topology (((AMB, MEA), VIC), LAC), the alternative of these two topologies (((MEA, VIC), AMB), LAC), and a 'one lineage' topology ((AMB_MEA_VIC), LAC). In this way, we can confirm the capability of each analysis to delimit, regardless of the biological or evolutionary reality of

459 the underlying topology. Because a failure to delimit could be due to limitations of the
460 analysis, or because the relationship among the tips in the simulation is incorrect, by
461 modeling on several topologies, we can test the true capability of each analysis to delimit.
462 We have outlined these simulation steps in further detail in the supplementary materials
463 (Supplementary Data S4).
464
465 *Set up and expectations of the simulations*—We simulated 100 datasets to test the
466 capability of each delimitation approach to delimit correctly. If the delimitation approach
467 correctly delimits (i.e., identifies the same number of lineages as simulated), we can assume
468 that the approach is sensitive enough to delimit given a dataset with the size and amount of
469 variability that we have collected. If the delimitation incorrectly delimits (i.e., identifies a
470 number of lineages different from what we simulated), we conclude that the approach is not
471 sensitive enough to delimit given the data we have collected.
472
473 *Post-hoc simulation study of molecular delimitation approaches*—We have developed our
474 own code that combines the simulation steps described above with the spedeSTEM analysis
475 (available on Dryad). For each topology, this code simulates one genealogy per locus,
476 simulates sequences on the genealogy, and then performs all steps of the spedeSTEM
477 approach (including the 100 subsampled replicates) using the same values of θ used in the
478 empirical delimitation. We performed this simulation-plus-analysis procedure 100
479 independent times and report the proportion of models that are ranked in each position
480 (first through fifth) across simulations.
481 For BPP, we randomly sampled 10 datasets from the 100 simulated datasets made
482 during the spedeSTEM simulation study using R (R Development Core Team 2016), and
483 performed the unguided delimitation analysis using the same prior settings for θ and
484 divergence times used in our empirical analyses. We used species model prior '1' in each
485 analysis, which assigns equal probabilities across all rooted topologies. For each randomly
486 sampled dataset, we performed two replicates to ensure convergence across independent
487 analyses using different rjMCMC algorithms. We summarize the results by reporting the

posterior probability of lineage distinction and the component models of the 95% credibility set of models.

RESULTS

Phylogenetic reconstructions

Gene trees and species trees—Maximum likelihood and Bayesian reconstructions of chloroplast and nuclear phylogenies were largely similar, varying mostly in the amount of topological support, with one primary exception. Bayesian nuclear reconstructions recovered *Castilleja ambigua* var. *meadii* as sister to the remaining taxa, while maximum likelihood reconstructions recovered it within *C. ambigua* + *C. victoriae* clade (Supplementary Data S5). To keep things simple, we refer only to the Bayesian reconstruction from here forward, noting that with the exception of the previous relationship, all results mentioned here apply to the ML reconstructions as well.

In both gene tree reconstructions, we recovered a monophyletic *C. ambigua* var. *meadii* and a monophyletic *C. victoriae* (Fig. 3a). Furthermore, in our nuclear reconstruction, both *C. victoriae* and *C. ambigua* var. *meadii* were placed on long branches relative to other taxa. The chloroplast reconstruction recovered *C. ambigua* as paraphyletic with respect to *C. ambigua* var. *meadii* and *C. victoriae*, while the nuclear reconstruction supported *C. ambigua* var. *meadii* as sister to a paraphyletic *C. ambigua* and *C. victoriae*. This paraphyletic relationship was also recovered in our estimate of the species tree (Fig. 3b), where *C. ambigua* var. *meadii* is sister to a clade composed of both *C. victoriae* and *C. ambigua*. Taken together, *C. ambigua* var. *meadii* and *C. victoriae* are each monophyletic, and their relationship to *C. ambigua* is difficult to place with certainty.

Molecular Species Delimitation

Estimate of theta—Given the three independent MIGRATE-N analyses, we estimated an average nuclear θ of 0.0146, an average chloroplast θ of 0.0064, and a genome-wide average θ of 0.0105 (Supplementary Table S3.1). After a series of preliminary tests to ensure the priors suited this dataset (see Supplementary Data S3 for details), four

independent BPP A00 analyses estimated an averaged θ of 0.0326 for *C. ambigua*, 0.0055 for *C. ambigua* var. *meadii*, and 0.0054 for *C. victoriae* (Supplementary Table S3.1). We take these separate estimates of θ as corroborative of each other. While these estimates were not identical, they did fall within the same order of magnitude and locus-wide averages were similarly close.

Molecular delimitation with spedeSTEM and BPP—Results of spedeSTEM analyses, averaged over 100 subsampled replicate analyses, strongly supported only one of five possible models of lineage composition (Table 1). This highest ranked model considers our three focal taxa as a single evolutionary lineage, (AMB_MEA_VIC). An extremely large *Δi* separated this best model from that of the next best. Therefore, this model composes all of the total model probabilities, indicating no support for other models of lineage composition.

Results of the guided delimitation (analysis A10) with BPP recovered high probabilities of lineage divergence at each node in each of our guide topologies (Fig. 4a). The unguided delimitation in BPP (Analysis A11) reports high posterior probability for the presence of three distinct lineages (four, including the outgroup *C. lacera*, (Table 2)) and recovers high posterior probabilities for all taxonomic species. Across all replicates, the 95% credibility set of species tree topologies was composed of four topologies (Fig. 4b; Table 2). Among these, a sister relationship of *C. ambigua* and *C. ambigua* var. *meadii* was consistently the most highly supported model; however, it was rarely recovered with strong probability (6 of 22 replicates with probability of 0.95 or greater (Table 2)).

It has been suggested that lineages be declared distinct only if posterior probabilities exceed thresholds of 95% or greater (Rannala & Yang 2013). The results of our independent molecular species delimitation approaches are in conflict; spedeSTEM supports a single-lineage model while BPP finds evidence of three distinct lineages.

Post-hoc Simulation Study

Delimitation with simulated data—Here we present the results of our simulation study of spedeSTEM and BPP, using 100 and 10 simulated datasets respectively, from four

alternative topologies: our estimate of the species tree, ((AMB, VIC), MEA); the taxonomic hypothesis, ((AMB, MEA), VIC); the alternative three-lineage topology, ((MEA, VIC), AMB); the one-lineage topology (AMB_MEA_VIC). We expect that an analysis will have sufficient power to delimit if it identifies the same number of lineages as modeled in the simulations. spedeSTEM reports results as support for lineage composition (i.e., how many lineages are present, and which taxa make up those lineages, with no comment on relationship of those lineages) and unguided BPP reports probabilities of lineage distinction, with an additional estimate of species phylogeny.

spedeSTEM—In two of our three, three-lineage simulations spedeSTEM did not recover the correct number of lineages (Fig. 5, rows 1-2). In all simulations based on the species tree and taxonomic hypotheses, the highest ranked model was composed of a single lineage. In the alternative three-lineage simulations, spedeSTEM most often ranked a one-lineage model as highest, therefore failing the majority of the time to identify the correct number of lineages (Fig. 5, row 3); however, in six of the 100 simulations, spedeSTEM ranked the three-lineage model as the highest (Supplemental Table S6). In our one-lineage simulations, spedeSTEM delimited the correct number of lineages 20 times out of 100. Most often it ranked a two-lineage model first (71 times), but also ranked a three-lineage model as first 9 times (Fig. 5, row 4; Supplemental Table S6).

BPP—In two of our three, three-lineage simulations BPP correctly delimited (Fig. 6, rows 1-2). In simulations of the species tree and taxonomic hypotheses, BPP recovered very strong support for the delimitation of taxonomic species corresponding to our focal taxa. Furthermore, in all simulations, the 95% clade credibility set contained models corresponding to the simulated topology, indicating that BPP was reconstructing the topology correctly (Fig. 6, rows 1-2; Supplemental Table S7.1 and S7.2). In simulations of the alternative three-lineage topology, BPP incorrectly delimited a single species. This corresponds to no posterior support for taxonomic species and an incorrect topological reconstruction (Fig. 6, row 3; Supplemental Table S7.3). In our one-lineage simulations, BPP

correctly delimits a single species, recovered very strong support for the delimitation of one species, and reconstructed the correct topology (Fig. 6, row 4).

DISCUSSION

Initial phylogenetic analyses often hint at the conflict between taxonomy and phylogeny that may be present in a system, as we see here in the *Castilleja ambigua* species complex (Fig. 3). In cases such as these, where there is a need for species delimitation with *limited* data, it is important to explore the capability of the data and analyses at hand to address the question of interest. In our case, when individual gene trees are considered alongside the results of our species tree reconstruction, we have reason to suspect 1) that we may have signal of distinct lineages that do not correspond with taxonomy, and 2) that the relationship between these lineages is poorly understood. The application of two independent molecular delimitation approaches results in incongruent delimitations (Table 1 and 2); spedeSTEM ranks highest a one-lineage model, while BPP supports three distinct lineages. BPP results are further complicated by strong support for different topologies (guided analysis (A10) recovers high support for all three topologies tested (average over all replicates > 0.95, Fig. 4, a); unguided analysis (A11) moderately supports the taxonomic hypothesis (average over all replicates between 0.75 and 0.95; Fig. 4, b)).

Had we stopped here, we would be faced with a subjective decision about which delimitation to prioritize. We would have attempted to explain the conflict in a biological context to arrive at a delimitation decision. However, knowing that each approach has its own set of limitations casts doubt on the interpretations of the results. spedeSTEM is known to be more conservative; it is highly reliant on the phylogenetic certainty of gene trees and simulations have shown that the validity of shallower nodes is most difficult to establish (Ence & Carstens 2010). Guided BPP can over-delimit, given an incorrect guide tree (Leaché & Fujita 2010) (but see Zhang *et al.* 2014) or misspecified prior settings (Giarla *et al.* 2014). In addition to testing the impact of the prior settings on results, we also provided BPP with alternative topologies and found each was strongly supported with high probability, suggesting one or more may be incorrect. The unguided delimitation is

intended to eliminate the need for a guide tree. We find this analysis strongly supports distinct lineages (for our focal taxa) and most often recovers a topology consistent with taxonomy—a hypothesis that is in conflict with one of our gene trees, as well as our species tree, and is only recovered six of 22 times with strong probability (Table 2). Furthermore, two of the 22 replicate unguided analyses recovered the species tree topology with noteworthy support, though moderate (*pp* = 0.86, results not shown). With such striking contrasts between delimitations, we find ourselves back at the starting point—how many lineages do we have? Is it lack of signal in the data that causes spedeSTEM to fail to delimit, or are we somehow biasing our delimitation, resulting in over-delimitation with guided BPP?

Pertinent to this conversation are the quality of the data we are using and the particular characteristics of the study system. Despite having many base pairs of data (25,351 bp of the most variable regions of *Castilleja* plastome, and 1,139 bp nrDNA totaling 26,490bp; Table S4.5), we are effectively delimiting with only two loci. In addition, the sampling of two of our focal taxa is small (three individuals for both *Castilleja ambigua* var. *meadii* and *Castilleja victoriae*). These small sample sizes could be impacting our results. If that is the case, an easy fix is to increase sample size, but generating more data by adding loci and/or increasing individuals sampled is difficult and expensive. Furthermore, two of our focal taxa are extremely rare and known from only a few populations that are very spatially restricted (Fairbarns & Egger 2007; Egger *et al.* 2012) (Fig. 1). As such, incorporating additional individuals that will represent additional, currently unsampled molecular variation is unlikely, not to mention practically difficult. This is a common position for empiricists, especially those working in non-model systems with rare and/or spatially restricted taxa. While many of us are focused on gathering more data, it is important to remember that we do have other tools available to assess the suitability of the data *already at hand*. Post-hoc simulation studies can help us evaluate the adequacy of our data for addressing our question of interest.

Simulations are useful in cases such as these—By simulating data on a known topology (i.e., a topology that we know for certain because we simulated it (rather than estimating it)) with variation similar to what we observe in our dataset, we can specifically test if there is signal in our data to delimit species, and if that signal is detectable with these analyses. In addition, by simulating data on multiple topologies (including our estimated species tree topology, as well as alternative relationships, therefore accommodating uncertainty in the underlying species level relationships), we can assess the sensitivity of these analyses to different topological relationships, therefore testing the ability of each approach to delimit, regardless of our knowledge of the true underlying species relationships.

In our simulation study, spedeSTEM fails to delimit in three of four cases where we see dominating support for a one-lineage model in our three, three-lineage simulations (Supplemental Table S6; Fig. 5). In the fourth case, the one-lineage simulation, spedeSTEM accurately delimits a single lineage 20 times, but also delimits a two or three lineage model 80 times (71 and 9, respectively). Unguided delimitation with BPP, on the other hand, correctly delimits in three of four cases (Fig. 6, Supplementary Table S7.1, S7.2, and S7.4), and fails when we simulate the alternative three-lineage topology (Fig. 6, Supplementary Table S7.3). Given the results of these simulations, we conclude that spedeSTEM is not suitable for delimitation with the dataset that we have collected here. BPP, on the other hand, appears to be sensitive enough to delimit the number of lineages, but perhaps not the evolutionary relationship of these lineages.

Other reasons for conflict in delimitation—There are, of course, other explanations for conflicting delimitations, other than the limitations of the approaches as we have described them here. For example, we may have violated assumptions implicit in both approaches. Probably the assumption most in jeopardy of violation is that polymorphism present in the data are the result of incomplete lineage sorting (ILS) and not gene flow (Ence & Carstens 2010; Yang & Rannala 2014b). Breaking this particular assumption has been shown to impact both approaches by homogenizing allele frequencies across lineage boundaries, thus impeding delimitation (e.g., Ence & Carstens 2011, Camargo *et al*. 2012, Pelletier *et al*.

2014). In this system, there are distinct floral differences that exists between *C. victoriae* and *C. ambigua* (including *C. ambigua* var. *meadii*) that suggests the possibility that contemporary gene flow between these taxa is unlikely. In *C. victoriae,* stigmas are inserted at anthesis (i.e., female reproductive organs enclosed within the flower at peak flowering time), suggesting the possibility of self-pollination as a reproductive strategy. This is in direct contrast with all of *C. ambigua* where stigmas are exserted at anthesis (i.e., female reproductive organs held up and out of the flower at peak flowering times), which is the typical placement for an outcrossing mode of pollination. These differences are likely to be a strong functional barrier to cross-pollination.

While floral morphological distinction between *C. ambigua* and *C. ambigua* var. *meadii* is less apparent, vegetative morphological variation is apparent and may reflect the ecological differentiation of these taxa. *C. ambigua* var. *meadii* is found further inland than most other *C. ambigua* (which are typically coastal) and is associated with freshwater (as opposed to salt water habitats where other members of *C. ambigua* occur) (Fig. 1). For these reasons, we consider contemporary gene flow unlikely in this particular complex of species; however, historical gene flow is something we cannot rule out and, given the young age of this lineage, something that may be relatively recent.

Hybridization has played, and may continue to play, a big role in the history of *Castilleja*, both at recent and deep time scales (e.g., Heckard 1968; Heckard & Chuang 1977; Tank & Olmstead 2009; Hersch-Green 2012; Clay *et al.* 2012). We have evidence of ongoing hybridization that we can observe in the field (e.g., Anderson & Taylor 1983; Hersch-Green & Cronn 2009), as well as signatures of hybridization deep in the history of the lineage (Hersch-Green & Cronn 2009; Tank & Olmstead 2009; Hersch-Green 2012). Furthermore, there is reason to expect gene flow at relatively shallow nodes in the phylogeny. Between the uplift of the Cascades and the Sierras between 2 – 5 million years ago, and the last glacial maximum (LGM) that peaked around 20,000 years ago, western North America has seen many geographic changes and there are many examples of geologic impact on flora and fauna, including diversification (e.g., Hewitt 1996; Brunsfeld *et al.* 2001; Shafer *et al.* 2010; Espíndola *et al.* 2012; Folk *et al.* 2016; 2017). Therefore, it is not unreasonable to

690 suggest that diversification of this species complex happened within this timeframe. Indeed,
691 major north-south post-glacial re-colonization routes pass through extreme southwestern
692 British Columbia and northwestern Washington state (Shafer *et al.* 2010) where current
693 day *C. victoriae* occurs (Fig. 1). As such, expecting a shallow node of divergence of both *C.*
694 *victoriae* and *C. ambigua* var. *meadii* from *C. ambigua* is perhaps realistic— this would
695 explain the low amount of variation we recover in our sequence data and the difficulty
696 spedeSTEM has detecting it.
697 While we consider the results of this work to confirm the distinction of three
698 lineages corresponding to our focal taxa, there is still evidence wanting with respect to
699 species delimitation. First, a robust delimitation must include additional lines of evidence
700 that corroborate (or refute) the evidence presented here. For example, given the distinctive
701 habitats of *C. victoriae* and *C. ambigua* var. *meadii*, we expect a signature of ecological
702 differentiation in these lineages. This is especially important given recent criticism about
703 the nature of what BPP— and coalescent-based, molecular species delimitation approaches,
704 in general—is delimiting (i.e., population structure or species, (Sukumaran & Knowles
705 2017). Second, recent advances in modeling the complex history of lineages (including gene
706 flow, alongside that of population subdivision, and/or population size differences) (e.g.,
707 Morales *et al.* 2016; Jackson *et al.* 2016) provide us with opportunities to examine the
708 possibility of historical and contemporary gene flow in this system, and possibly rule out
709 (or identify) potential causes of incongruence in our delimitation. Future work in the
710 *Castilleja ambigua* species complex will address additional lines of evidence, and include
711 more holistic species delimitation analyses (e.g., Solís-Lemus *et al.* 2015), and any formal
712 changes to species limits will follow accordingly.
713 Carstens *et al*. (2013) report that only 30% of species delimitation studies make
714 taxonomic recommendations and only 25% describe new species and suggest that this
715 could indicate a lack of confidence in the study, an inability to resolve incongruence across
716 approaches, or acknowledgement of inadequacy of the data. Formal simulation studies, like
717 ours, provide an avenue for researchers to address these concerns. Ultimately, empiricists
718 have an obligation to use species delimitation approaches carefully and according to

'manufacturer instructions.' By carefully considering the assumptions and limitations of the approaches we use, we are off to a good start; by keeping abreast of both empirical and theoretical studies that refine our understanding of the limitations of these approaches, we are in an even better position to appropriately use the methods we employ. Finally, by performing simulation studies, such as those shown here, we have the opportunity to test if our approach is appropriate given our specific study system and the data at hand. This will be particularly important and useful in systems that are in the process of becoming data-rich (but currently have smaller, non-genomic datasets) and have pressing need for formal delimitations. Regardless, post-hoc simulation studies such as this can be important to success in species delimitation, especially at recent time scales where the depth of the nodes we are examining may be very shallow. It is likely that in many systems, such as this one, where we are interested in distinguishing incipient lineages, incongruence across delimitations will be common.

ACKNOWLEDGEMENTS

We thank past and present members of the Tank Lab, Bryan Carstens, Luke Harmon, Jack Sullivan, Daniel Caetano, Megan Ruffley, Anahí Espindola, Josef Uyeda, and Mark Egger for helpful and insightful conversations. Additionally, we thank four anonymous reviewers for insightful and helpful comments on the manuscript. This research was supported by resources at the Institute for Bioinformatics and Evolutionary Studies (IBEST- NIH/NCRR P20RR16448 and P20RR016454), NSF DEB-1253463 (to DCT), and NSF DEB-1502061 (to DCT for SJJ).

REFERENCES

Adams M, Raadik TA, Burridge CP, Georges A (2014) Global biodiversity assessment and hyper-cryptic species complexes: more than one species of elephant in the room? *Systematic Biology*, **63**, 518–533.

Agapow PM, Bininda-Emonds ORP, Crandall KA, Gittleman JL, Mace GM, Marshall JC, and A Purvis (2004) The impact of species concept on biodiversity studies. *Quarterly Review*

748 *of Biology*, **79**, 161–179.
749 Aguilar C, Wood P, Cusi JC, Guzman A, Huari F, Lundberg M, Mortensen E, Ramirez C, Robles
750 D, Suarez J, Ticona A, Vargas VJ, Venegas PJ and JW Sites (2013) Integrative taxonomy
751 and preliminary assessment of species limits in Liolaemus walker complex (Squamata,
752 Liolaemidae) with descriptions of three new species from Peru. *ZooKeys*, **364**, 47–91.
753 Anderson AVR, Taylor RJ (1983) Patterns of morphological variation in a population of
754 mixed species of Castilleja (Scrophulariaceae). *Systematic Botany*, **8**, 225–232.
755 Andújar C, Arribas P, Ruiz C, Serrano J, Gómez-Zurita J (2014) Integration of conflict into
756 integrative taxonomy: fitting hybridization in species delimitation of Mesocarabus
757 (Coleoptera: Carabidae). *Molecular Ecology*, **23**, 4344–4361.
758 Baldwin BG (1992) Phylogenetic utility of the internal transcribed spacers of nuclear
759 ribosomal DNA in plants: an example from Compositae. *Molecular Phylogenetics and*
760 *Evolution*, **1**, 3–16.
761 Baldwin BG, Markos S (1998) Phylogenetic utility of the external transcribed spacer (ETS)
762 of 18S–26S rDNA: congruence of ETS and ITS trees of Calycadenia (Compositae).
763 *Molecular Phylogenetics and Evolution*, **10**, 449–463.
764 Barley AJ, Brown JM, Thomson RC (2018) Impact of model violations on the inference of
765 species boundaries under the multispecies coalescent. *Systematic Biology*, **67**, 269-284.
766 Beardsley PM, Olmstead RG (2002) Redefining Phrymaceae: the placement of Mimulus,
767 tribe Mimuleae and Phryma. *American Journal of Botany*, **89**, 1093–1102.
768 Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and
769 effective population sizes in n subpopulations by using a coalescent approach.
770 *Proceedings of the National Academy of Sciences*, **98**, 4563–4568.
771 Blaimer BB, Brady SG, Schultz TR (2015) Phylogenomic methods outperform traditional
772 multi-locus approaches in resolving deep evolutionary history: a case study of formicine
773 ants. *BMC Evolutionary Biology*, **15**, 271.
774 Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C, Xi D, Suchard M, Rambaut A and AJ
775 Drummond (2014) BEAST 2: A software platform for bayesian evolutionary analysis.
776 *PLoS Computational Biology*, **10**, e1003537.

Brunsfeld SJ, Sullivan J, Soltis DE, Soltis PS (2001) Comparative phylogeography of north-western North America: a synthesis. *Special Publication-British Ecological Society*, **14**, 319–340.

Bybee SM, Bracken-Grissom H, Haynes BD, Hermansen RA, Byers RL, Clement MJ, Udall JA, Wilcox ER and KA Crandall (2011) Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. *Genome Biology and Evolution*, **3**, 1312–1323.

Camargo A, Sites J Jr. (2013) Species delimitation: a decade after the renaissance. *The Species Problem – Ongoing Issues*. Ed. Igor Y. Pavlinov. InTech. 2013.

Camargo A, Morando M, Avila LJ, Sites JW Jr. (2012) Species delimitation with ABC and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the Liolaemus darwinii complex (Squamata: Liolaemidae). *Evolution*, **66**, 2834–2849.

Carstens BC, Pelletier TA, Reid NM, Satler JD (2013) How to fail at species delimitation. *Molecular Ecology*, **22**, 4369–4383.

Chuang TI, Heckard LR (1991) Generic realignment and synopsis of subtribe Castillejinae (Scrophulariaceae-tribe Pediculareae). *Systematic Botany*, **16**, 644–666.

Clay DL, Novak SJ, Serpe MD, Tank DC, Smith JF (2012) Homoploid hybrid speciation in a rare endemic Castilleja from Idaho (Castilleja christii, Orobanchaceae). *American Journal of Botany*, **99**, 1976–1990.

Costello MJ, May RM, Stork NE (2013) Can we name earth's species before they go extinct? *Science*, **339**, 413–416.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Edwards DL, Knowles LL (2014) Species detection and individual assignment in species delimitation: can integrative data increase efficacy? *Proceedings of the Royal Society B*, **281**, 20132765.

Egger JM, Ruygt JA, Tank DC (2012) Castilleja ambigua var. meadii (OROBANCHACEAE): A new variety from Napa County, California. *Phytoneuron,* **68,** 1–12.

Ence DD, Carstens BC (2010) SpedeSTEM: a rapid and accurate method for species delimitation. *Molecular Ecology Resources*, **11**, 473–480.

Espíndola A, Pellissier L, Maiorano L, Hordijk W, Guisan A and N Alvarez (2012) Predicting present and future intra-specific genetic structure through niche hindcasting across 24 millennia. *Ecology Letters*, **15**, 649–657.

Fairbarns M, Egger JM (2007) Castilleja victoriae (Orobanchaceae): a new rare species from southeastern Vancouver Island, British Columbia, Canada, and the adjacent San Juan Islands, Washington, USA. *Madroño*, **54**, 334–342.

Flot J-F (2015) Species delimitation's coming of age. *Systematic Biology*, **64**, 897–899.

Folk RA, Mandel JR, Freudenstein JV (2017) Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. *Systematic Biology*, **66**, 320-337.

Folk RA, Visger CJ, Soltis PS, Soltis DE, Guralnick RP (2017) Geographic range dynamics drove ancient hybridization in a lineage of angiosperms. *bioRxiv,* 129189.

Fujita MK, Leaché AD, Burbrink FT, McGuire JA and C Moritz (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution*, **27**, 480–488.

Giarla TC, Voss RS, Jansa SA (2014) Hidden diversity in the Andes: comparison of species delimitation methods on montane marsupials. *Molecular Phylogenetics and Evolution*, **70**, 137–151.

Goldberg CS, Tank DC, Uribe-Convers S, Bosworth WR, Marx HE and LP Waits (2011) Species designation of the Bruneau Dune tiger beetle (Cicindela waynei) is supported by phylogenetic analysis of mitochondrial DNA sequence data. *Conservation Genetics*, **13**, 373–380.

Grummer JA, Bryson RW, Reeder TW (2014) Species delimitation using bayes factors: simulations and application to the Sceloporus scalaris species group (Squamata: Phrynosomatidae). *Systematic Biology*, **63**, 119–133.

Guillot G, Renaud S, Ledevin R, Michaux J, Claude J (2012) A unifying model for the analysis of phenotypic, genetic, and geographic data. *Systematic Biology*, **61**, 897–911.

Heckard LR (1968) Chromosome numbers and polyploidy in Castilleja (Scrophulariaceae). *Brittonia*, **20**, 212–226.

Heckard LR, Chuang TI (1977) Chromosome-numbers, polyploidy, and hybridization in Castilleja (Scrophulariaceae) of Great Basin and Rocky Mountains. *Brittonia*, **29**, 159–172.

Hedrick PW (2001) Conservation genetics: where are we now? *Trends in Ecology & Evolution*, **16**, 629-636.

Hersch-Green EI (2012) Polyploidy in Indian paintbrush (Castilleja; Orobanchaceae) species shapes but does not prevent gene flow across species boundaries. *American Journal of Botany*, **99**, 1680–1690.

Hersch-Green EI, Cronn R (2009) Tangled trios?: characterizing a hybrid zone in Castilleja (Orobanchaceae). *American Journal of Botany*, **96**, 1519–1531.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hime PM, Hotaling S, Grewelle RE, O'Neill EM, Voss SR, Shaffer HB and DW Weisrock (2016) The influence of locus number and information content on species delimitation: an empirical test case in an endangered Mexican salamander. *Molecular Ecology*, **25,** 5959-5974.

Hird S, Kubatko L, Carstens B (2010) Rapid and accurate species tree estimation for phylogeographic investigations using replicated subsampling. *Molecular Phylogenetics and Evolution,* **57**, 888-898.

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, **18**, 337–338.

Jackson ND, Carstens BC, Morales AE, O'Meara BC (2017) Species delimitation with gene flow. *Systematic Biology*, **66**, 799-812.

Knowles LL, Carstens B (2007) Delimiting species without monophyletic gene trees. *Systematic Biology*, **56**, 887–895.

Lanfear R, Calcott B, Ho SYW, Guindon S (2012) PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–1701.

Latvis ML, Mortimer SME, Morales-Briones DF, Torpey S, Uribe-Convers S, Jacobs SJ, Mathews S, Tank DC (2017). Primers for *Castilleja* and their utility across Orobanchaceae: I. Chloroplast Primers. *Applications in Plant Sciences*. **5**, 1-7.

Leaché AD, Fujita MK (2010) Bayesian species delimitation in West African forest geckos (Hemidactylus fasciatus). *Proceedings of the Royal Society B*, **277**, 3071–3077.

Lemmon EM, Lemmon AR (2013) High-Throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 99–121.

Lim GS, Balke M, Meier R (2012) Determining species boundaries in a world full of rarity: singletons, species delimitation methods. *Systematic Biology*, **61**, 165–169.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526–538.

Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gateway Computing Environments Workshop (GCE)*, 1-8.

Morales AE, Jackson ND, Dewey TA, O'Meara BC, Carstens BC (2017) Speciation with gene flow in North American Myotis bats. *Systematic Biology*, **66**, 440-452.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca G, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

O'Meara BC (2010) New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology*, **59**, 59–73.

Padial JM, Miralles A, la Riva De I, Vences M (2010) The integrative future of taxonomy. *Frontiers in Zoology*, **7**, 16.

Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN and PH Raven (2014) The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, **344**, 1246792.

Pons J, Barraclough T, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD

and AP Vogler (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**, 595–609.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rambaut A, Grass NC (1997) Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Bioinformatics*, **13**, 235–238.

Rannala B (2015) The art and science of species delimitation. *Current Zoology*, **61**, 846–853.

Rannala B, Yang Z (2013) Improved reversible jump algorithms for bayesian species delimitation. *Genetics,* **194**, 245-253.

Reeves PA, Richards CM (2010) Species delimitation under the general lineage concept: an empirical example using wild North American hops (Cannabaceae: Humulus lupulus). *Systematic Biology*, **60**, 45–59.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA and JP Huelsenbeck (2012) MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.

Ruane S, Bryson RW, Pyron RA, Burbrink FT (2014) Coalescent species delimitation in milksnakes (Genus Lampropeltis) and impacts on phylogenetic comparative analyses. *Systematic Biology*, **63**, 231–250.

Ruane S, Raxworthy CJ, Lemmon AR, Lemmon EM, Burbrink FT (2015) Comparing species tree estimation with large anchored phylogenomic and small Sanger-sequenced molecular datasets: an empirical study on Malagasy pseudoxyrhophiine snakes. *BMC Evolutionary Biology*, **15,** 221.

Satler JD, Carstens BC, Hedin M (2013) Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, Aliatypus). *Systematic Biology*, **62**, 805–823.

Schlick-Steiner BC, Steiner FM, Seifert B, Stauffer C, Christian E and RH Crozier (2010) Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual Review of Entomology*, **55**, 421–438.

Shafer ABA, Cullingham CI, Cote SD, Coltman DW (2010) Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology*, **19**, 4589–4621.

Singh G, Dal Grande F, Divakar PK, Otte J, Leavitt SD, Szczepanska K, Crespo A, Rico VJ, Aptroot A, da Silva Caceres ME, Lunbsch HT and I Schmitt (2015) Coalescent-based species delimitation approach uncovers high cryptic diversity in the cosmopolitan lichen-forming fungal genus Protoparmelia (Lecanorales, Ascomycota). *PLoS One*, **10**, e0124625.

Sites JW Jr., Marshall JC (2003) Delimiting species: a renaissance issue in systematic biology. *Trends in Ecology & Evolution*, **18**, 462–470.

Smith SA, Dunn CW (2008) Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics*, **24**, 715–716.

Solís-Lemus C, Knowles LL, Ané C (2015) Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution*, **69**, 492–507.

Sukumaran J, Holder MT (2010) DendroPy: a Python library for phylogenetic computing. *Bioinformatics*, **26**, 1569–1571.

Sukumaran J, Knowles LL (2017) Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences*, **114**, 1607–1612.

Tank DC, Olmstead RG (2008) From annuals to perennials: phylogeny of subtribe Castillejinae (Orobanchaceae). *American Journal of Botany*, **95**, 608–625.

Tank DC, Olmstead RG (2009) The evolutionary origin of a second radiation of annual Castilleja (Orobanchaceae) species in South America: the role of long distance dispersal and allopolyploidy. *American Journal of Botany*, **96**, 1907–1921.

Tank DC, Egger JM, Olmstead RG (2009) Phylogenetic classification of subtribe Castillejinae (Orobanchaceae). *Systematic Botany*, **34**, 182–197.

Uribe-Convers S, Settles ML, Tank DC (2016) A phylogenomic approach based on PCR target enrichment and high throughput sequencing: resolving the diversity within the South American species of Bartsia L. (Orobanchaceae) *PLoS ONE*, **11**, e0148203.

Wetherwax M, Chuang TI, Heckard L (2017). *Castilleja ambigua*, *Jepson eFlora*,

http://ucjeps.berkeley.edu/cgi-bin/get_IJM.pl?tid=18158, accessed on September 23, 2017.
Wiens J (2007) Species delimitation: new approaches for discovering diversity. *Systematic Biology*, **56**, 875–878.
Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, **107**, 9264–9269.
Yang Z, Rannala B (2014a) Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution,* **31**, 3125-3135.
Yeates DK, Seago A, Nelson L, Cameron SL, Joseph L and Trueman JWH (2010) Integrative taxonomy, or iterative taxonomy? *Systematic Entomology*, **36**, 209–217.
Zapata F, Jiménez I (2012) Species delimitation: inferring gaps in morphology across geography. *Systematic Biology*, **61**, 179–194.
Zhang C, Rannala B, Yang Z (2014) Bayesian species delimitation can be robust to guide-tree inference errors. *Systematic Biology*, **63**, 993–1004.
Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. University of Texas, Austin, PhD dissertation.

DATA ACCESSIBILITY

DNA Sequences are deposited in GENBANK; alignments, gene trees, species trees, R code are deposited in Dryad (doi:10.5061/dryad.97d83g0).

AUTHOR CONTRIBUTIONS

SJJ and DCT designed and performed research, contributed new analytical scripts, analyzed data, and wrote paper; CK, SUC, and ML collected data and edited manuscript.

TABLES AND FIGURES

TABLE 1. Results of empirical species delimitation using spedeSTEM. Information-theoretic metrics from 100 subsampled replicates (replicates 3-99 omitted for simplicity).

Models of lineage composition	Average likelihood for each subsampled replicate					k	AICc	Δ_i	Model likelihood	w
	Rep. 1	Rep. 2	…	Rep. 100	*ln*L (avg)					
AMB_MEA_VIC	-11935.99	-12682.72	…	-12682.72	-12223.46	1	24450.25	0.00	1.00	1.
MEA, AMB_VIC	-12193.18	-12905.89	…	-12913.24	-12478.08	2	24966.16	515.91	0.00	0.
VIC, AMB_MEA	-13416.58	-14163.11	…	-14163.11	-13511.09	2	27032.18	2581.93	0.00	0.
MEA, VIC, AMB	-13668.46	-14371.97	…	-14379.33	-13713.90	3	27457.80	3007.55	0.00	0.
MEA_VIC, AMB	-14149.41	-14886.75	…	-14886.75	-14372.91	2	28755.82	4305.57	0.00	0.

AMB – *Castilleja ambigua* (including varieties *ambigua, humboldtiensis*, and *insalutata*)

MEA – *Castilleja ambigua* var. *meadii*

VIC – *Castilleja victoriae*

*ln*L (avg) – Log likelihood of the model, averaged across all replicates

k – the number of free parameters in the model

AICc – Akaike information criterion, corrected for small sample sizes

Δ_i – Akaike differences between current and best model

w_i – Model weights

TABLE 2. Results of empirical molecular species delimitation using BPP, analysis A11, averaged across 22 independent runs. Each panel represents a portion of the output of this analysis; the probability of the taxonomic species (first and second panel) and the best models found in the 95% credibility set of species tree topologies (third panel).

	Posterior probabilities			
Best Model (ignoring species tree phylogeny)	mean	Min	Max	Number reps above 0.95 (out of 22)
4 distinct lineages: A, M, V, L	0.9610	0.74	0.99	18

Posterior probability of taxonomic species	mean	Min	Max	Number reps above 0.95 (out of 22)
Castilleja ambigua	0.9713	0.79	0.99	18
Castilleja ambigua var. *meadii*	0.9845	0.79	0.99	21
Castilleja victoriae	0.9645	0.75	0.99	18
Castilleja lacera	0.9780	0.74	0.99	20

Best models in 95% credibility set	mean	Min	Max	Number reps above 0.95 (out of 22)	Number reps occurred in (out of 22)
(((A, M), V), L)	0.7434	0.59	0.99	6	22
(((A, V), M), L)	0.1838	0.02	0.87	0	6
(((M, V), A), L)	0.2401	0.01	0.34	0	6
((A, M), (V, L))	0.1413	0.02	0.63	0	3

A – *Castilleja ambigua* (including varieties *ambigua, humboldtiensis*, and *insalutata*)
M – *Castilleja ambigua* var. *meadii*
V – *Castilleja victoriae*
L – *Castilleja lacera*

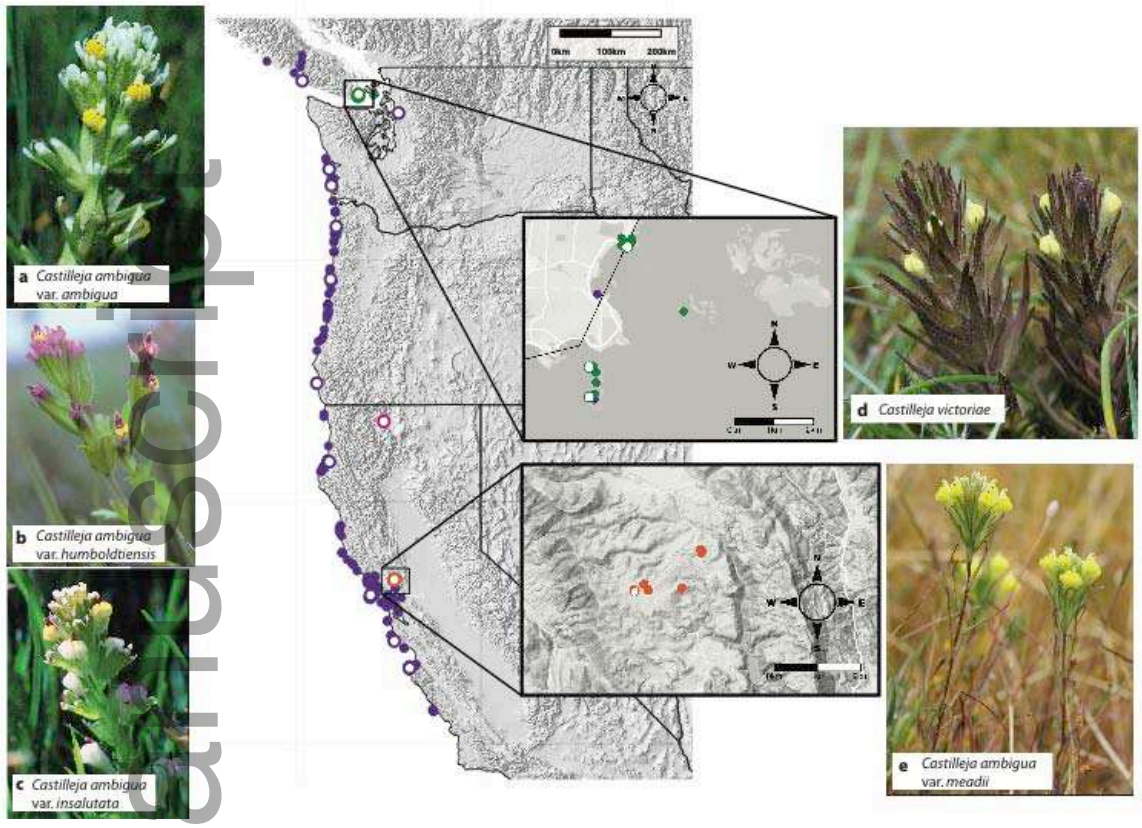


FIGURE 1. Distributions and location of sampled individuals for focal taxa considered here; (a) the polymorphic *Castilleja ambigua* (purple) (which we treat as including varieties *C. ambigua* var. *ambigua*, (b) *C. ambigua* var. *insalutata*, and (c) *C. ambigua* var. *humboldtiensis*), (d; green) *C. victoriae*, and (e; orange) *C. ambigua* var. *meadii*. Filled circles are known localities of each taxon; empty circles represent sampled localities. Photographs by J. Mark Egger.

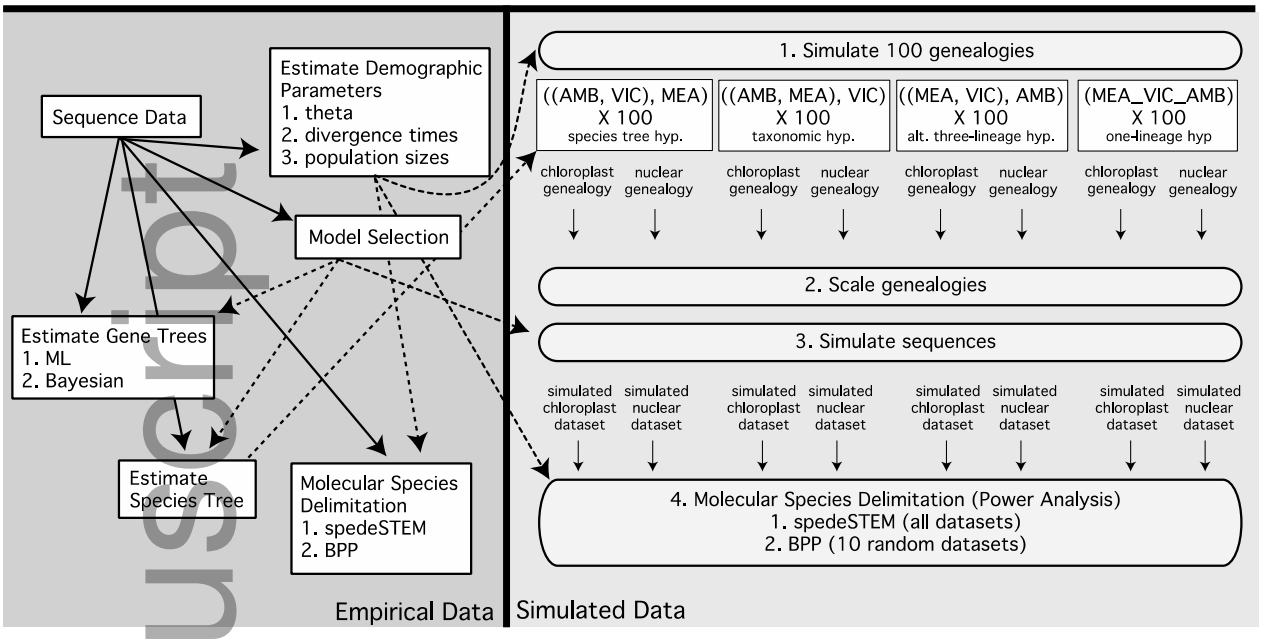


FIGURE 2. Schematic illustrating components of our empirical analyses (left) and simulations (right), highlighting the use of estimated models of nucleotide evolution, demographic parameters, and inferred species tree topology from empirical data in our simulations (dashed lines connecting the left side to the right). Solid arrows represent use of sequence data in each step of phylogenetic, species tree, and molecular species delimitation inference; dashed arrows indicate estimated models of nucleotide evolution and demographic parameters necessary for phylogenetic, species tree, and molecular species delimitation analyses. AMB = *Castilleja ambigua*, MEA = *Castilleja ambigua* var. *meadii*, VIC = *Castilleja victoriae*.

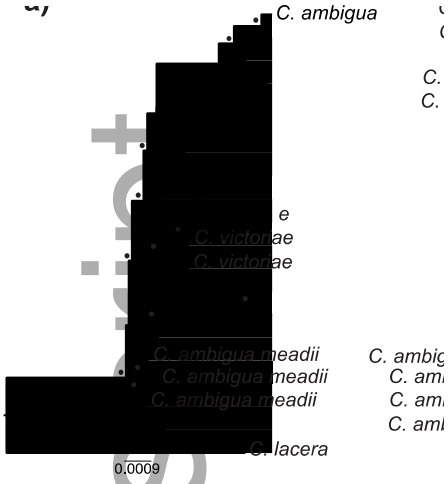


FIGURE 3. (a) Results of Bayesian gene tree inference (chloroplast reconstruction at left, nuclear reconstruction at right). Dots above branches indicate support > 0.95. Branch lengths are proportional to the number of substitutions per site, as measured by the scale bar. (b) Species tree estimation with posterior probabilities indicated at nodes. Dashed lines indicate median node heights used to inform timing of population splits in simulated genealogies.

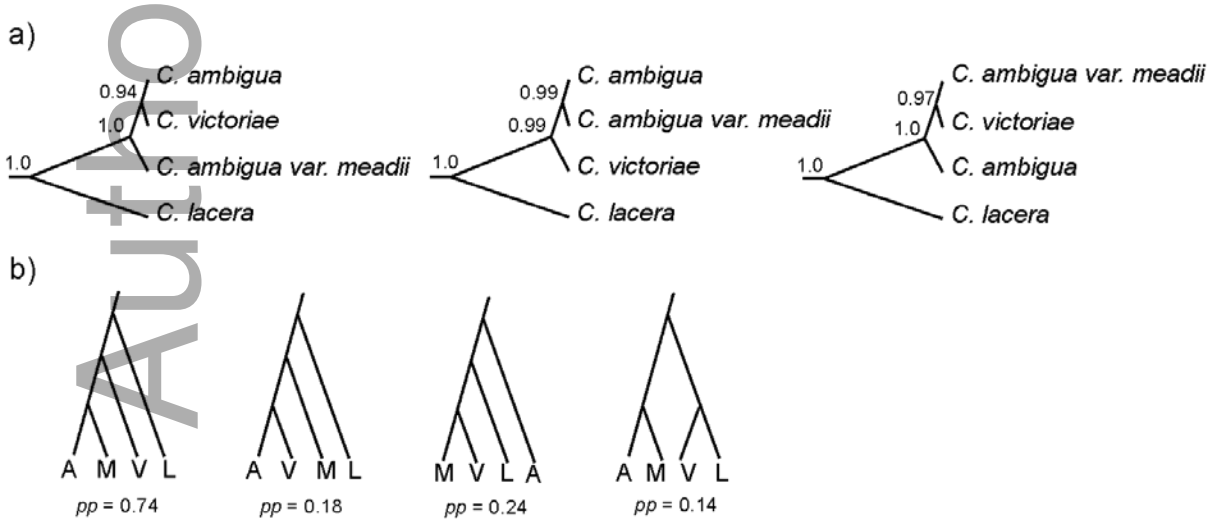


FIGURE 4. (a) Results of empirical molecular species delimitation using guided BPP. The three topologies correspond to the species tree hypothesis (left), and its two alternative topologies. Values at nodes represent lineage distinctiveness. (b) The set of models included in the 95% credibility set of trees from unguided delimitation with BPP. Posterior probability for each topology is reported beneath the tree. A = *Castilleja ambigua* ; M = *Castilleja ambigua* var. *meadii* ; V = *Castilleja victoriae* ; L = *Castilleja lacera* (outgroup).

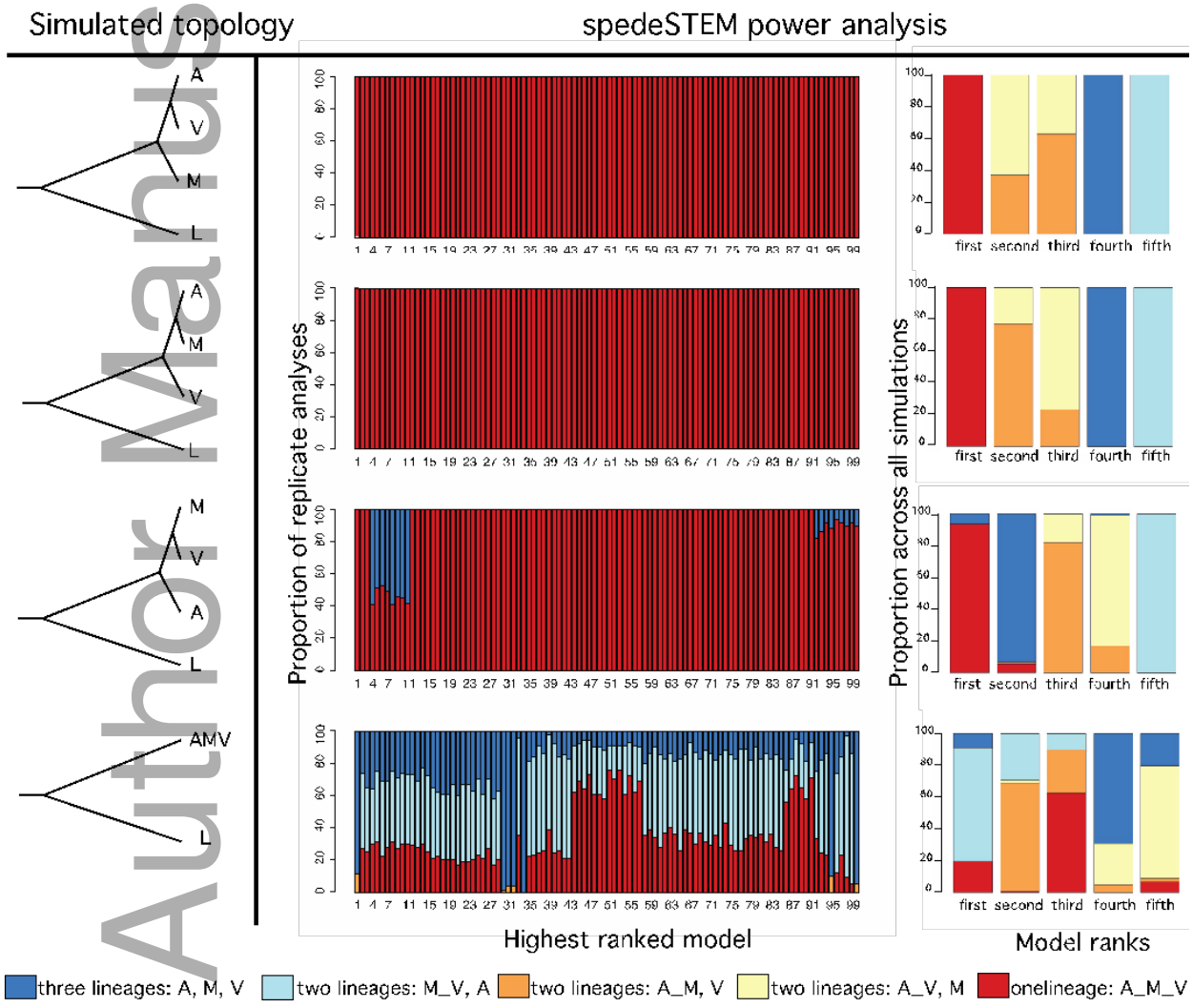


FIGURE 5. Results of simulation study of spedeSTEM. Here we report for each simulated topology (left column), the model ranked highest across 100 independent simulations (middle column) and the proportion of models at each rank position (right column) across the 100 simulations. Lineage models are color coded according to their composition (linear key along bottom of figure). A = *Castilleja ambigua* ; M = *Castilleja ambigua* var. *meadii* ; V = *Castilleja victoriae* ; L = *Castilleja lacera* (outgroup). AMV = a single lineage composed of *Castilleja ambigua* + *Castilleja ambigua* var. *meadii* + *Castilleja victoriae.*

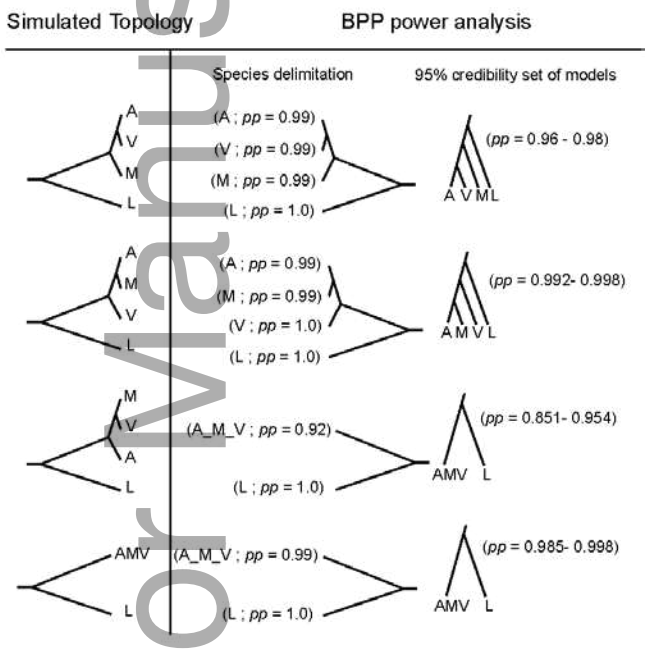
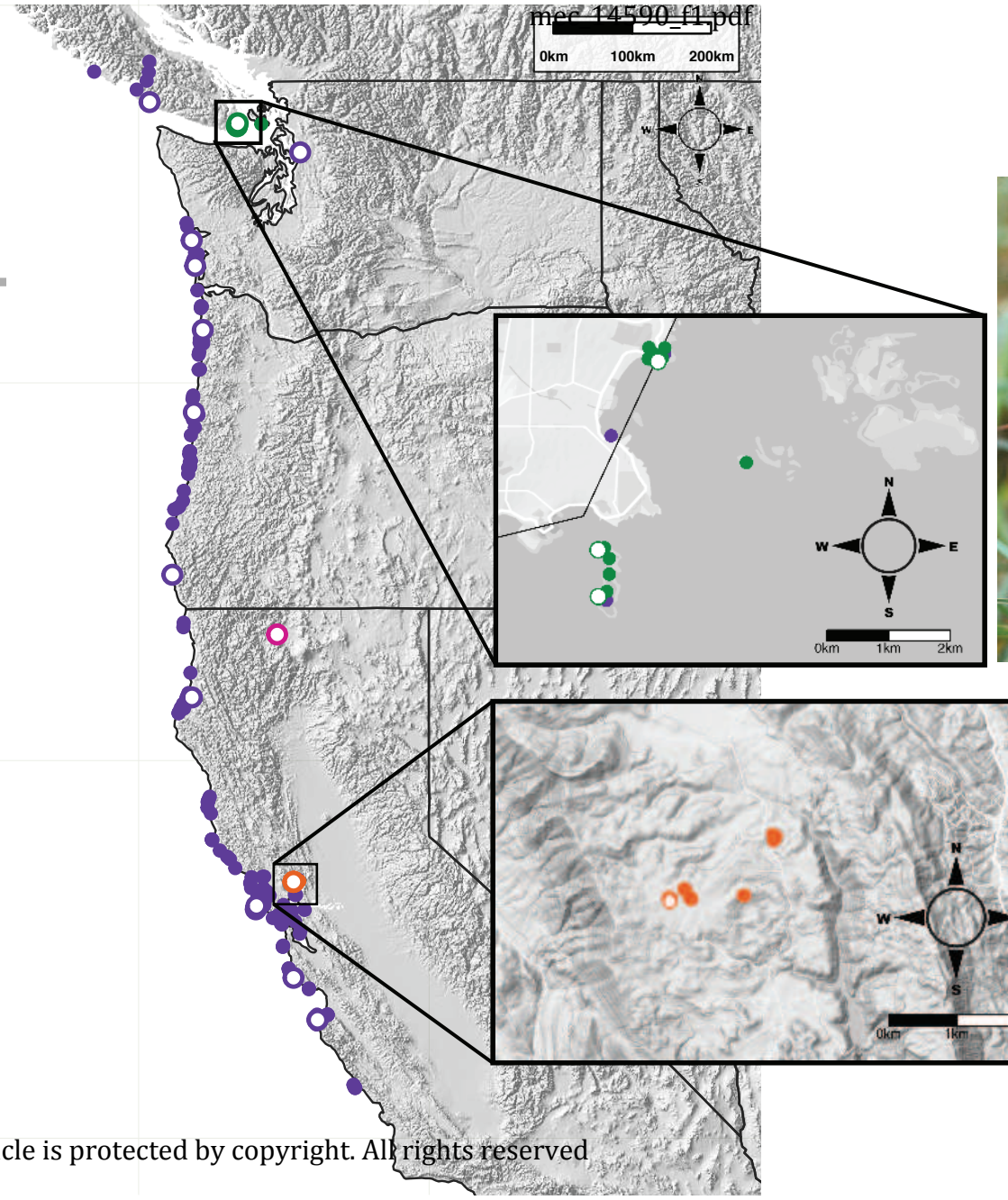


FIGURE 6. Results of simulation study of unguided BPP, averaged across 10 separate simulations. For each simulated topology (left column), we report the posterior probability for lineage distinctiveness (middle column), and the component models recovered in the 95% credibility set of models (right column). Ranges of probabilities reported under the latter represent the range of support across 10 independent simulations.

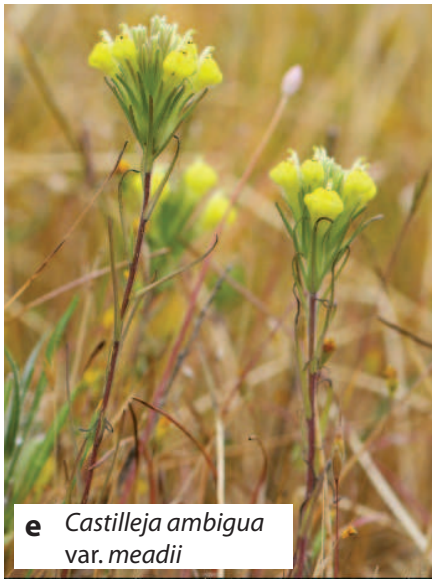
a *Castilleja ambigua* var. *ambigua*

b *Castilleja ambigua* var. *humboldtiensis*

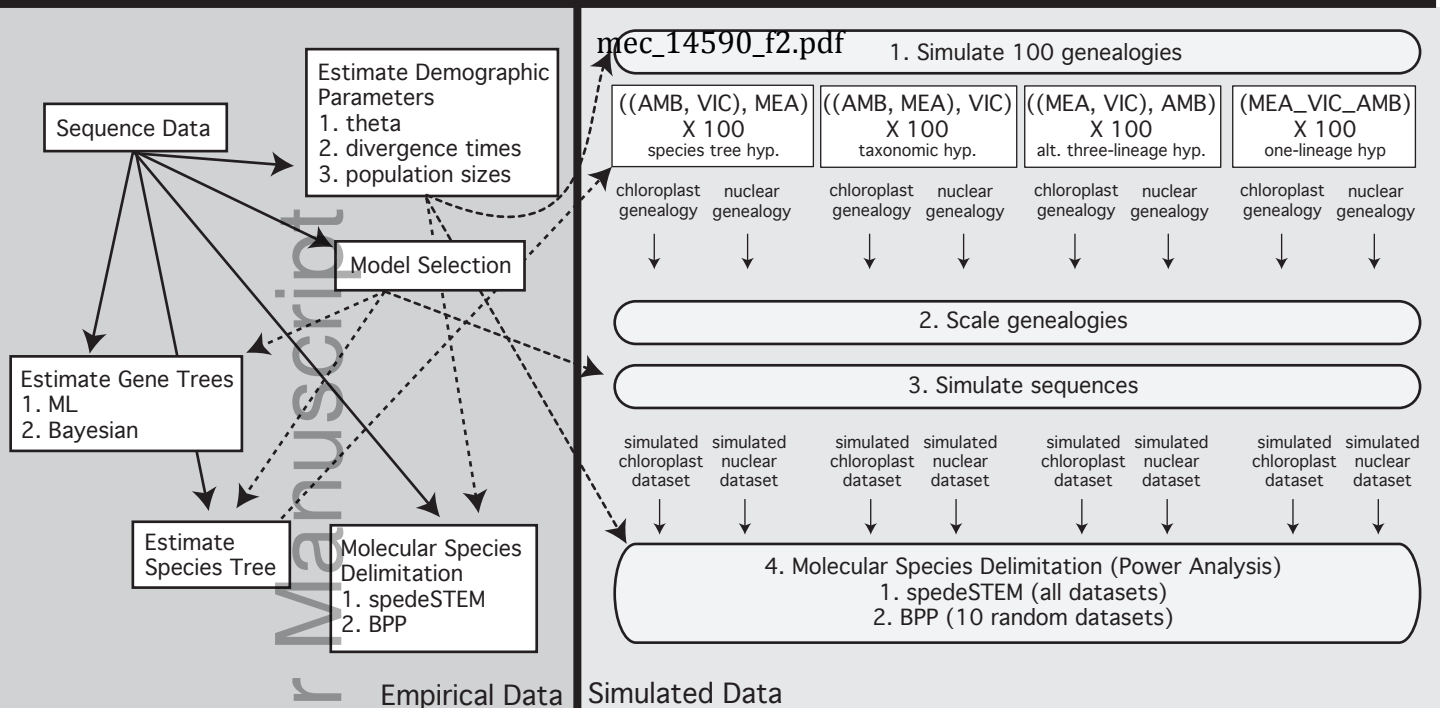
c *Castilleja ambigua* var. *insalutata*

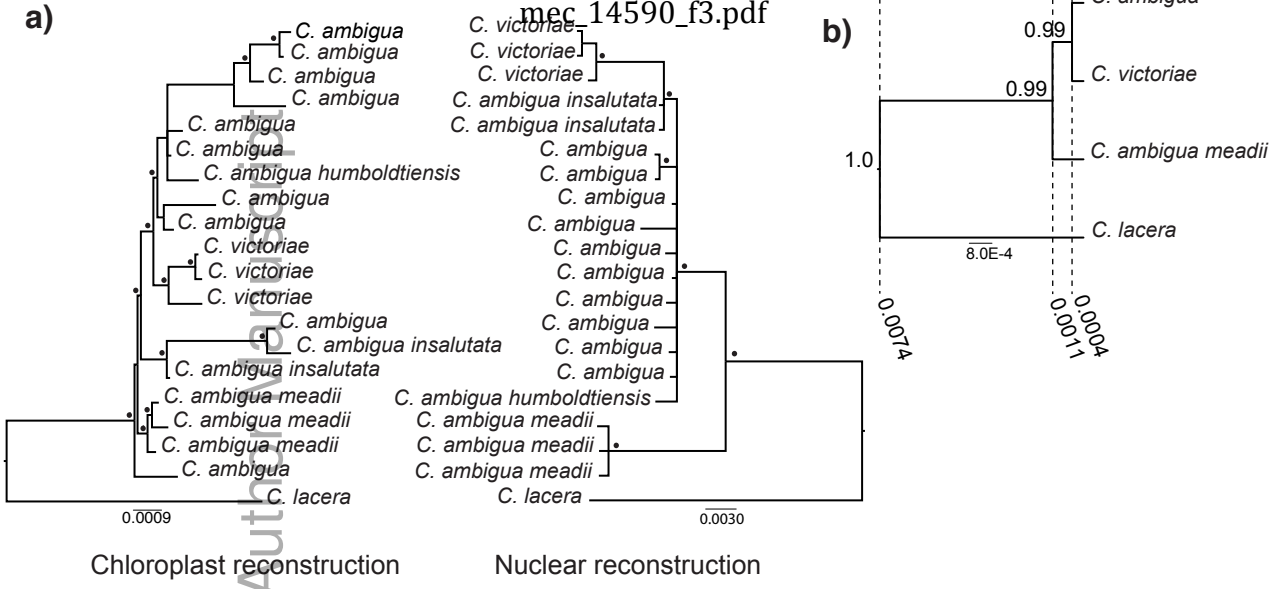


d *Castilleja victoriae*



e *Castilleja ambigua* var. *meadii*





Chloroplast reconstruction Nuclear reconstruction