The challenge of defining species boundaries in incipient plant lineages: a case study in the paintbrushes (*Castilleja*; Orobanchaceae)

A Dissertation
Presented in Partial Fulfillment of the Requirements for the
Degree of Doctorate of Philosophy
with a
Major in Biology
in the
College of Graduate Studies
University of Idaho
by
Sarah J. Jacobs

Major Professor: David C. Tank, Ph.D.
Committee Members: Jack Sullivan, Ph.D.; Luke J. Harmon, Ph.D.; Bryan Carstens, Ph.D.;
Christine Parent, Ph.D.
Department Administrator: James J. Nagler, Ph.D.

August 2018

AUTHORIZATION TO SUBMIT DISSERTATION

This dissertation of Sarah J. Jacobs, submitted for the degree of Doctorate of Philosophy with a Major in Biology and titled "The challenge of defining species boundaries in incipient plant lineages: a case study in the paintbrushes (*Castilleja*; Orobanchaceae) has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor: _____ Date: _____
David. C. Tank, Ph.D.

Committee Members: _____ Date: _____
Jack Sullivan, Ph.D.

_____ Date: _____
Luke J. Harmon, Ph.D.

_____ Date: _____
Bryan Carstens, Ph.D.

_____ Date: _____
Christine Parent, Ph.D.

Department Administrator: _____ Date: _____
James L. Nagler, Ph.D.

ABSTRACT

The genus *Castilleja* (also known as 'the paintbrushes') is an iconic and widespread group of plants. Infrageneric classifications in this young and rapid radiation have long been fraught with difficulty, mostly attributed to high incidence of polyploidy and interspecific gene flow when species co-occur. Subsequently, taxonomies have nearly continuous morphological and ecological diagnostic characters across species boundaries. This has resulted in the genus *Castilleja* being notoriously taxonomically difficult. During the course of my dissertation I have applied species delimitation to small species complexes to explore the capability of current approaches to delimit species in the face of tremendous amounts of interspecific similarity. Each of these complexes has required the use of suites of methods and analytical tools, some applied in new and novel ways, to delimit species. Recently, the use of multiple, independent approaches to molecular delimitation has been advocated as a means of accommodating the limitations of a single approach; however, this can result in incongruent delimitation schemes (i.e., where one approach delimits differently than another) with no widely used objective way to mitigate incongruence. The first chapter discusses the application of *post-hoc* simulations to address the capability of each approach to correctly delimit, particularly in the face of small sample sizes. Chapter two examines the application of environmental variables to the question of species boundaries. Given robustly estimated species ranges (using occurrence data from museum collections), I estimated niche models and extracted climatic variables associated with focal taxa to corroborate molecular species boundaries. The final chapter considers morphology as a line of evidence to define species boundaries. By quantifying the amount of morphological similarity in the *pilosa* species complex, I show that morphological characters do not distinguish taxonomic entities, suggesting little to no morphological distinction among species in this particular group. My work shows that when validating species boundaries in incipient lineages, multiple lines of evidence should be carefully scrutinized. Additionally, it is still unclear how best to reconcile incongruent delimitation schemes across multiple lines of evidence. Currently, fully integrated analyses are advocated, but in some cases, these reduce data to transformed variables that are difficult to interpret biologically.

ACKNOWLEDGEMENTS

DEDICATION

I dedicate this dissertation to Mary Ann Jacobs and Mildred Haerther Homan, and to their daughters Kathy, Becky, Ellen, Kathy, Susan, Cindy, Celeste, and Lynne. These women have raised and supported me, inspired me and guided me, never doubted me and have always, always loved me. I am extremely fortunate.

I also dedicate this dissertation to my amazing parents. I am truly grateful, and so very lucky, to have been raised in such a supportive and loving family.

Finally, to KMM. I dedicate this dissertation to us. I know I could have done this without you, but I'm very, very glad that I didn't have to.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INCONGRUENCE IN MOLECULAR SPECIES DELIMITATION SCHEMES: WHAT TO DO WHEN ADDING MORE DATA IS DIFFICULT.

with Casey Kristofferson, Simon Uribe-Convers, Maribeth Latvis, and David C. Tank

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.

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; Leache & 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., Leache & 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 (Leache & 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 on the part of the method 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 suggest at least 10 individuals per putative lineage and as many loci as possible (Carstens *et al.* 2013). Increasing the number of loci in a dataset has become easier to do (e.g., McCormack *et al.* 2013; Lemmon & Lemmon 2013), and there is a general consensus in the phylogenetics community that more loci typically result in increased resolution (Ruane *et al.* 2015; Blaimer *et al.* 2015). However, genome-scale data are still time consuming and expensive to generate, particularly for non-model organisms, and there can be computational disadvantages to using hundreds of loci (Ruane *et al.* 2015).

Furthermore, for rare taxa—e.g., those known from only a few, often small, populations, and/or those that are spatially restricted—the incorporation of 10 individuals per putative lineage may not be possible (Lim *et al.* 2012). For these reasons, empirical studies, especially those dealing with rare or spatially restricted taxa, often begin with existing datasets (often Sanger sequenced data or data obtained from GenBank) that, in terms of individuals and loci sampled, are often smaller in size.

When a researcher recovers conflicting delimitation schemes across approaches using a dataset that is limited in size, an alternative analytical tactic is an assessment of the data already at hand (i.e., less than ideal datasets). In other words, an assessment of the capability of each methodological approach to detect the signal of independent lineages in the data collected. This can be directly tested in empirical studies using post-hoc simulations. While this has been implied as an appropriate and important step in empirical delimitation (Carstens *et al.* 2013), and some studies have simulated data in order to compare methodological approaches (e.g., Camargo *et al.* 2012; Barley *et al.* 2017) or to specifically address sample size (e.g., Giarla *et al.* 2014; Hime *et al.* 2016), to our knowledge an assessment of inferential error has not been specifically done in any empirical study.

In this study, we apply species delimitation approaches to a species complex in the plant genus *Castilleja*, a widespread and iconic wildflower that is most diverse in western North America. A recent, rapid radiation (Tank & Olmstead 2008), *Castilleja* is an important target for species delimitation, both theoretically and practically. Theoretically, the young age of this lineage affords us the opportunity to test the limits and capabilities of 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 can not) 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., Leache and 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 (Leache & 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 the underlying topology. Because a failure to delimit could be due to limitations of the analysis, or because the relationship among the tips in the simulation is incorrect, by modeling on several topologies, we can test the true capability of each analysis to delimit. We have outlined these simulation steps in further detail in the supplementary materials (Supplementary Data S4).

Set up and expectations of the simulations—We simulated 100 datasets to test the capability of each delimitation approach to delimit correctly. If the delimitation approach correctly delimits (i.e., identifies the same number of lineages as simulated), we can assume that the approach is sensitive enough to delimit given a dataset with the size and amount of variability that we have collected. If the delimitation incorrectly delimits (i.e., identifies a number of lineages different form what we simulated), we conclude that the approach is not sensitive enough to delimit given the data we have collected.

Post-hoc simulation study of molecular delimitation approaches—We have developed our own code that combines the simulation steps described above with the spedeSTEM analysis (available on Dryad). For each topology, this code simulates one genealogy per locus, simulates sequences on the

genealogy, and then performs all steps of the spedeSTEM approach (including the 100 subsampled replicates) using the same values of θ used in the empirical delimitation. We performed this simulation-plus-analysis procedure 100 independent times and report the proportion of models that are ranked in each position (first through fifth) across simulations.

For BPP, we randomly sampled 10 datasets from the 100 simulated datasets made during the spedeSTEM simulation study using R (R Development Core Team 2016), and performed the unguided delimitation analysis using the same prior settings for θ and divergence times used in our empirical analyses. We used species model prior '1' in each analysis, which assigns equal probabilities across all rooted topologies. For each randomly sampled dataset, we performed two replicates to ensure convergence across independent 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 (Leache & 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 suggest that diversification of this species complex happened within this timeframe. Indeed, major north-south post-glacial re-colonization routes pass through extreme southwestern British Columbia and northwestern Washington state (Shafer *et al.* 2010) where current day *C. victoriae* occurs (Fig. 1). As such, expecting a shallow node of divergence of both *C. victoriae* and *C. ambigua* var. *meadii*

from *C. ambigua* is perhaps realistic— this would explain the low amount of variation we recover in our sequence data and the difficulty spedeSTEM has detecting it.

While we consider the results of this work to confirm the distinction of three lineages corresponding to our focal taxa, there is still evidence wanting with respect to species delimitation. First, a robust delimitation must include additional lines of evidence that corroborate (or refute) the evidence presented here. For example, given the distinctive habitats of *C. victoriae* and *C. ambigua* var. *meadii*, we expect a signature of ecological differentiation in these lineages. This is especially important given recent criticism about the nature of what BPP— and coalescent-based, molecular species delimitation approaches, in general—is delimiting (i.e., population structure or species, (Sukumaran & Knowles 2017). Second, recent advances in modeling the complex history of lineages (including gene flow, alongside that of population subdivision, and/or population size differences) (e.g., Morales *et al.* 2016; Jackson *et al.* 2016) provide us with opportunities to examine the possibility of historical and contemporary gene flow in this system, and possibly rule out (or identify) potential causes of incongruence in our delimitation. Future work in the *Castilleja ambigua* species complex will address additional lines of evidence, and include more holistic species delimitation analyses (e.g., Solís-Lemus *et al.* 2015), and any formal changes to species limits will follow accordingly.

Carstens et al (2013) report that only 30% of species delimitation studies make taxonomic recommendations and only 25% describe new species, and suggest that this could indicate a lack of confidence in the study, an inability to resolve incongruence across approaches, or acknowledgement of inadequacy of the data. Formal simulation studies, like ours, provide an avenue for researchers to address these concerns. Ultimately, empiricists 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.
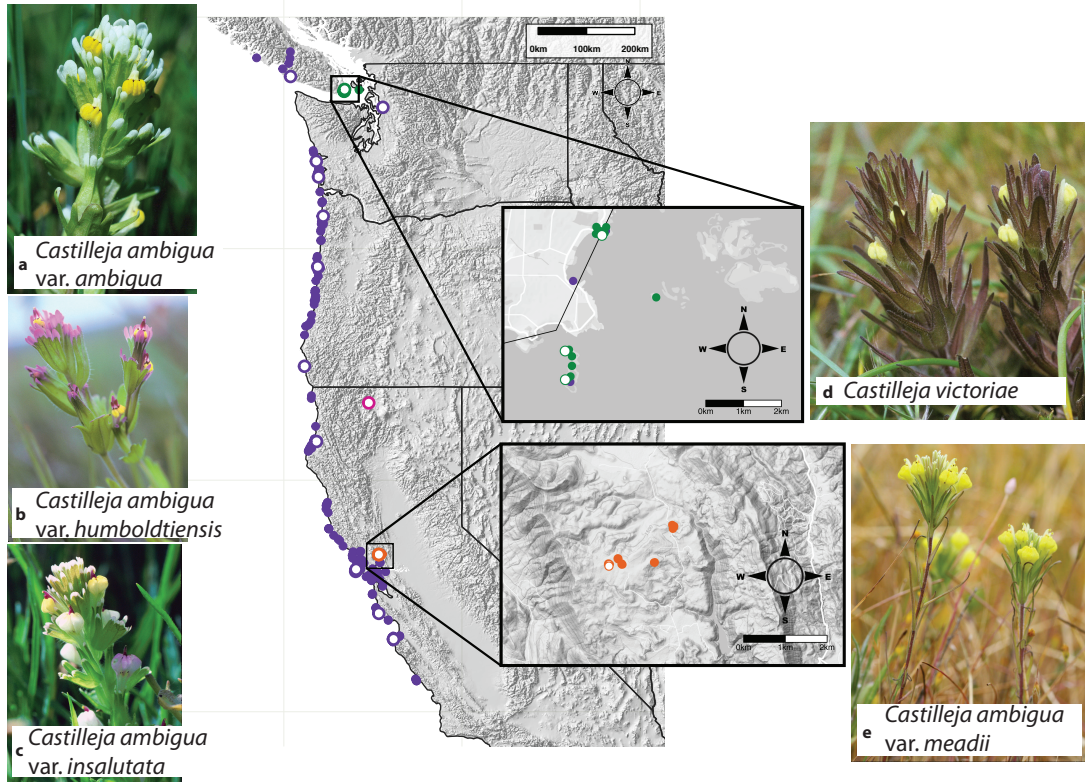
FIGURE 1.1. Distributions and locations of sampled individuals for focal taxa considered here; (a) the polymorphic *Castilleja ambigua* (purple) (which we treat as including varieties *C. ambigua* var. *ambigua*, (b) *C. ambigua* var. *insalutata*, and (c) *C. ambigua* var. *humboldtiensis*), (d; green) *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.
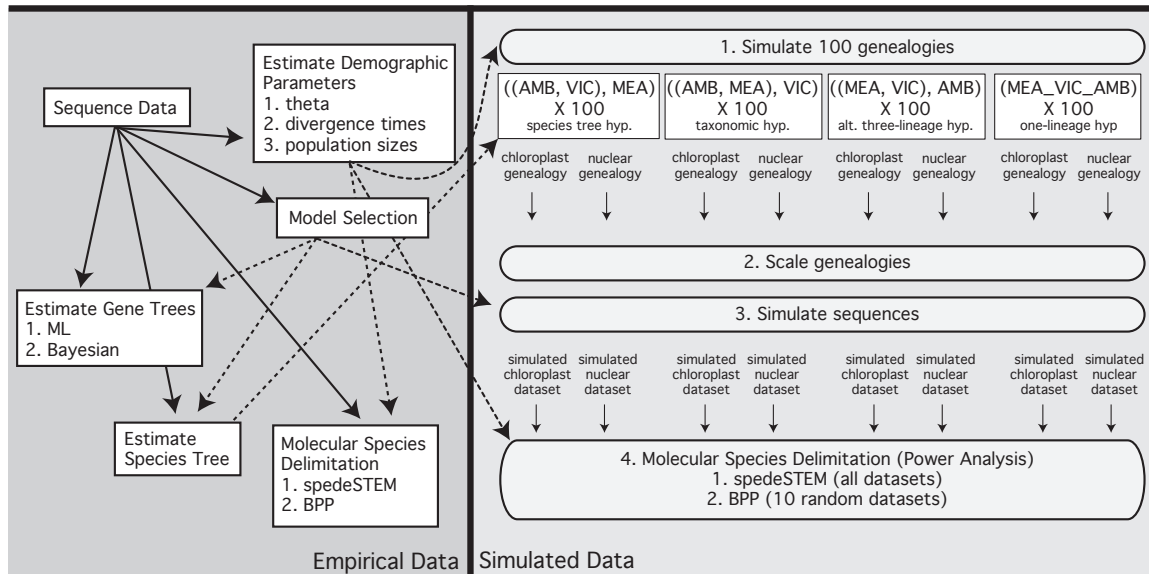
FIGURE 1.2. Schematic illustrating components of our empirical analyses (left) and simulations (right), highlighting the use of estimated models of nucleotide evolution, demographic parameters, and inferred species tree topology from empirical data in our simulations (dashed lines connecting the left 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 victoria*.
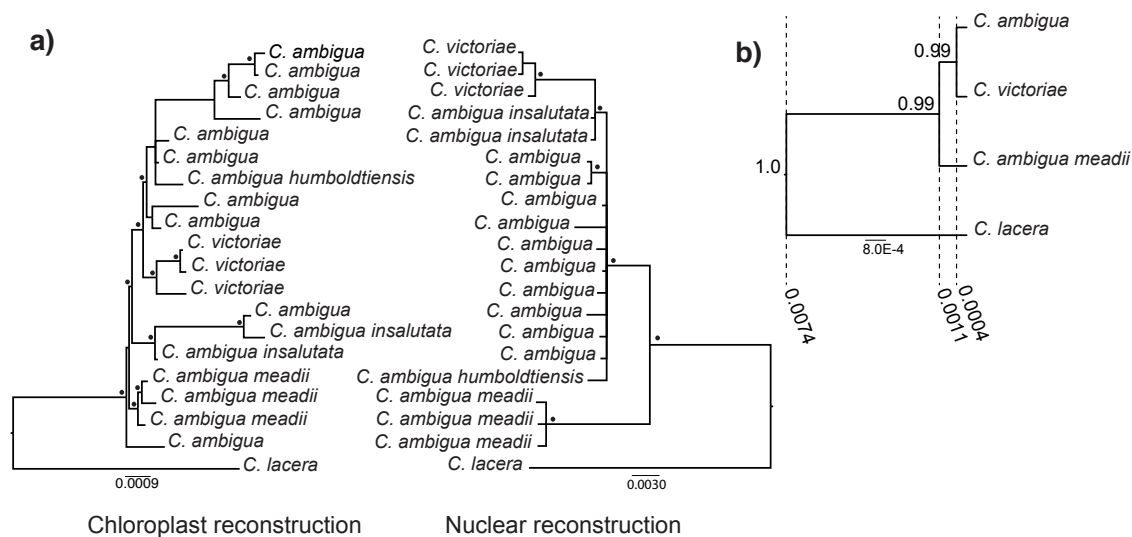
FIGURE 1.3. (a) Results of Bayesian gene tree inference (chloroplast reconstruction at left, nuclear reconstruction at right). Dots above branches indicate support > 0.95. Branch lengths are proportional to the number of substitutions per site, as measured by the scale bar. (b) Species tree estimation with posterior probabilities indicated at nodes. Dashed lines indicate median node heights used to inform timing of population splits in simulated genealogies.

a)

b)

A M V L
pp = 0.74

A V M L
pp = 0.18

M V L A
pp = 0.24

A M V L
pp = 0.14
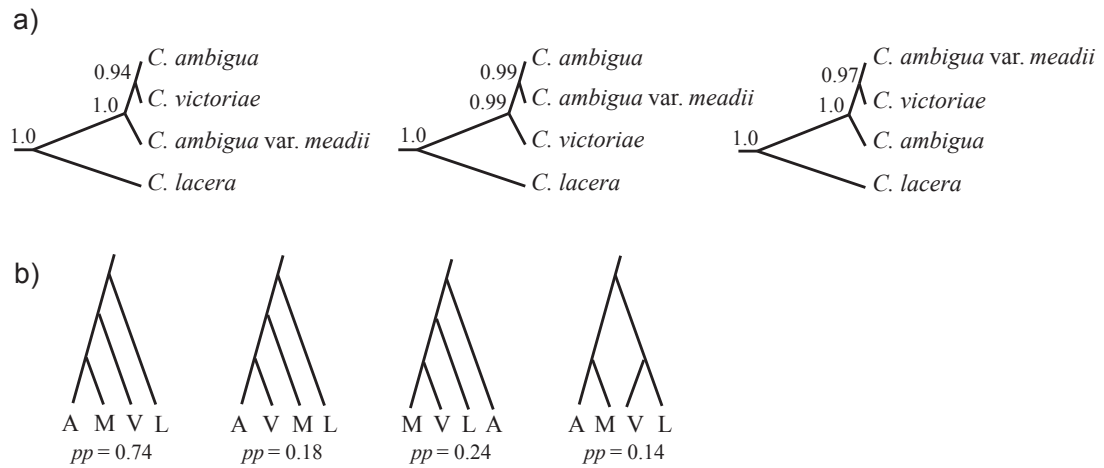
FIGURE 1.4. (a) Results of empirical molecular species delimitation using guided BPP. The three topologies correspond to the species tree hypothesis (left), and its two alternative topologies. Values at nodes represent lineage distinctiveness. (b) The set of models included in the 95% credibility set 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).
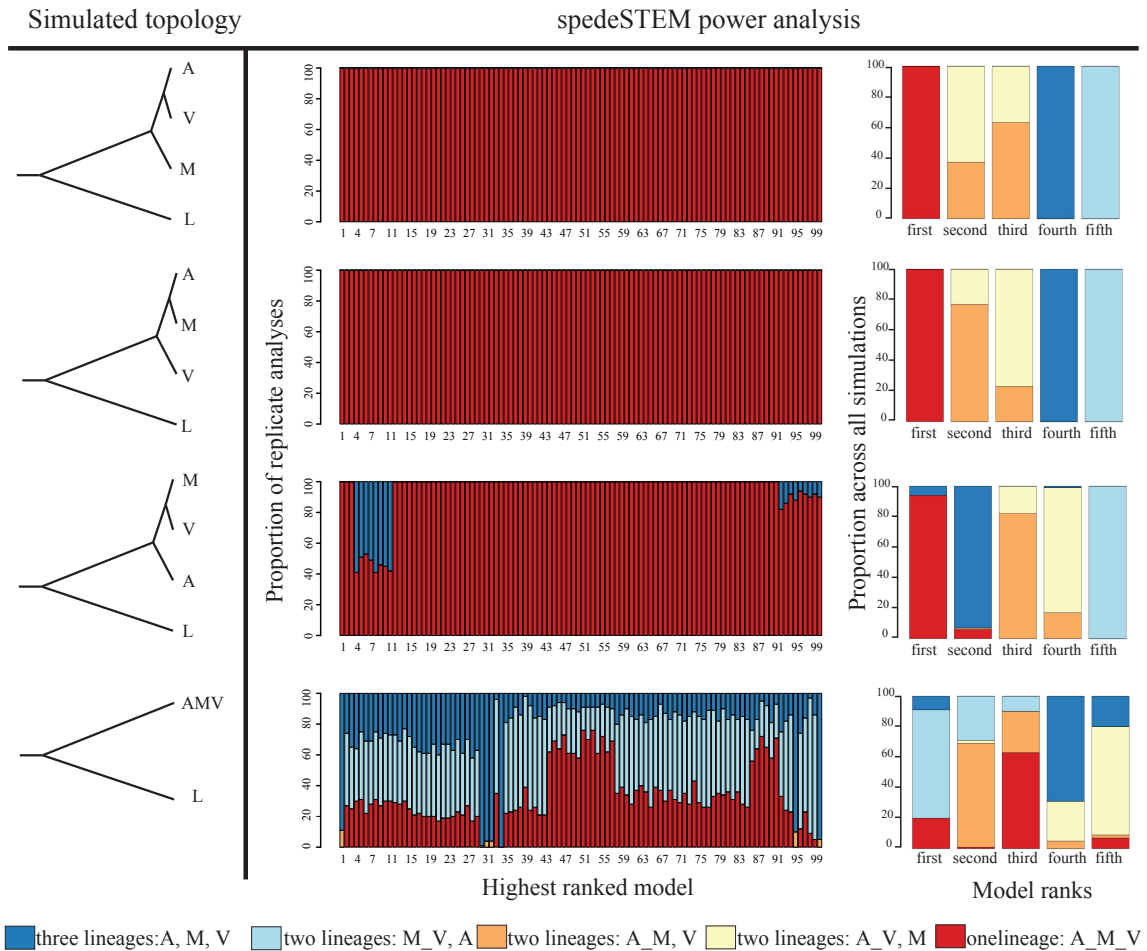
FIGURE 1.5. Results of simulation study of spedeSTEM. Here we report for each simulated topology (left column), the model ranked highest across 100 independent simulations (middle column) and the proportion of models at each rank position (right column) across the 100 simulations. lineage models are color coded according to their composition (linear key along the 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*.
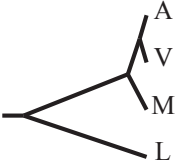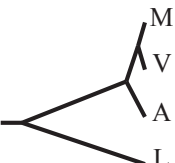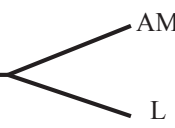
FIGURE 1.6. Results of simulation study of unguided BPP, averaged across 10 separate simulations. For each simulated topology (left column), we report the posterior probability for lineage distinctiveness (middle column), and the component models recovered in the 95% credibility set of models (right column). Ranges of probabilities reported under the latter represent the range of support across 10 independent simulations.

TABLE 1.1. 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_i
	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	1.00
MEA, AMB_VIC	-12193.18	-12905.89	...	-12913.24	-12478.08	2	24966.16	515.91	0.00	0.00
VIC, AMB_MEA	-13416.58	-14163.11	...	-14163.11	-13511.09	2	27032.18	2581.93	0.00	0.00
MEA, VIC, AMB	-13668.46	-14371.97	...	-14379.33	-13713.90	3	27457.80	3007.55	0.00	0.00
MEA_VIC, AMB	-14149.41	-14886.75	...	-14886.75	-14372.91	2	28755.82	4305.57	0.00	0.00

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 1.2. 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*

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 *et al.* (2004) The impact of species concept on biodiversity studies. *Quarterly Review of Biology*, **79**, 161–179.

Aguilar C, Wood P, Cusi JC *et al.* (2013) Integrative taxonomy and preliminary assessment of species limits in Liolaemus walker complex (Squamata, Liolaemidae) with descriptions of three new species from Peru. *ZooKeys*, **364**, 47–91.

Anderson AVR, Taylor RJ (1983) Patterns of morphological variation in a population of mixed species of Castilleja (Scrophulariaceae). *Systematic Botany*, **8**, 225–232.

Andújar C, Arribas P, Ruiz C, Serrano J, Gómez-Zurita J (2014) Integration of conflict into integrative taxonomy: fitting hybridization in species delimitation of Mesocarabus (Coleoptera: Carabidae). *Molecular Ecology*, **23**, 4344–4361.

Baldwin BG (1992) Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from Compositae. *Molecular Phylogenetics and Evolution*, **1**, 3–16.

Baldwin BG, Markos S (1998) Phylogenetic utility of the external transcribed spacer (ETS) of 18S–26S rDNA: congruence of ETS and ITS trees of Calycadenia (Compositae). *Molecular Phylogenetics and Evolution*, **10**, 449–463.

Barley AJ, Brown JM, Thomson RC (2018) Impact of model violations on the inference of species boundaries under the multispecies coalescent. *Systematic Biology*, **67**, 269-284.

Beardsley PM, Olmstead RG (2002) Redefining Phrymaceae: the placement of Mimulus, tribe Mimuleae and Phryma. *American Journal of Botany*, **89**, 1093–1102.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences*, **98**, 4563–4568.

Blaimer BB, Brady SG, Schultz TR *et al.* (2015) Phylogenomic methods outperform traditional multi-locus approaches in resolving deep evolutionary history: a case study of formicine ants. *BMC Evolutionary Biology*, **15**, 271.

Bouckaert R, Heled J, Kühnert D *et al.* (2014) BEAST 2: A software platform for bayesian evolutionary analysis. *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 *et al.* (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 *et al.* (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, Leache AD, Burbrink FT, *et al*. (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 *et al.* (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 *et al.* (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.

Leache 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 *et al.* (2014) The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, **344**, 1246792.

Pons J, Barraclough T, Gomez-Zurita J *et al.* (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 *et al.* (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 *et al.* (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 *et al.* (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 *et al.* (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

SUPPLEMENTAL DATA S1

SUPPLEMENTAL TABLE S1.1. Collection information and molecular sampling for individuals used in this study. Column 1) taxonomic identification, column 2) collector and collection number followed by the herbarium housing the collection voucher (represented by its acronym), and columns 3 through 5) molecules sampled, indicated by an 'X'. Individuals with an asterisk superscript indicate those sequences derived from dbc-amplicon approach.

Scientific name	Collector and Voucher location	cpDNA	ETS	ITS
Castilleja ambigua var. *ambigua*	Egger 567 WTU	X	X	X
	Egger 1463 WTU	X	X	
	Egger 337 WTU	X	X	X
	Gage and Rodman 375 WTU	X	X	X
	Holmgren 2643 UC	X	X*	X*
	Egger 578 WTU	X	X	X
	Avis. s.n. WTU	X	X	
	Stansell s.n. OSC	X	X*	
	Halse 4905 WTU	X	X*	X*
	Frenkel 1654 OSC	X		X
Castilleja ambigua var. *humboldtiensis*	Egger 409 WTU	X	X*	
Castilleja ambigua var. *insalutata*	Egger 528 WTU	X	X	X
	Egger 523 WTU	X	X	X
Castilleja ambigua var. *meadii*	Egger 1468 (#1) WTU	X	X	X
	Ruygt 5575 (#1) WTU	X	X	
	Ruygt 5575 (#2) WTU	X	X	
Castilleja victoriae	Egger s.n. WTU	X	X	X
	Egger s.n. WTU	X	X	X
	Calder and MacKay 29531 WTU	X	X*	
Castilleja lacera	Egger 400 WTU	X	X	X

WTU – University of Washington, Burke Museum
UC – University of California, Jepson Herbarium
OSC – Oregon State University Herbarium

SUPPLEMENTAL DATA S2

SUPPLEMENTAL TABLE S1.2. Primer pairs used in this study (from Latvis et al 2017, referenced in main text). Because our reads are paired-end (i.e., they were sequenced from either end of the target region), reads that overlap in the middle of the sequence were merged into continuous reads (indicated in bold); reads without overlap were utilized as single-end reads.

	Forward		Reverse	
	Primer Name	Primer Sequence	Primer Name	Primer Sequence
Pair 1	Cas_120561_F	GTCCAAACGATCCCATACCA	Cas_121371_R	TTTAGGTCGGTTACCGGTGT
Pair 2	Cas_111970_F	GGTGGAAAGTGAGGAAGAAAGA	Cas_112789_R	TCAAGAAGGAACAGGTTTGGA
Pair 3	Cas_129331_F	TGAGTTTAATCAACCGGAGA	Cas_130126_R	GACCCTTTCCTGAACAAATCA
Pair 4	Cas_112854_F	ACATAGTATTGTCCGATTCATAAGGA	Cas_113746_R	GGAGGGACCCACTCCTATTT
Pair 5	Cas_59866_F	TTGCCGTCAAAGACATTCG	Cas_60624_R	GCCTGTTTGAACAGCCTCAG
Pair 6	Cas_127891_F	GGATTCCTTGATAGTGAAGAACAGA	Cas_128420_R	GAAGGATCTGGACGATCGAA
Pair 7	Cas_130168_F	ACAACCGAGTCCTTGTTTCAA	Cas_130760_R	GGTGGAAAGTGAGGAAGAAAGA
Pair 8	**Cas_126110_F**	**TTCTAATCGATAATTAGGCCAAAGA**	**Cas_126868_R**	**GGATCCGTTCTATCACAACCA**
Pair 9	Cas_32159_F	AATCGGATCAATATCATGAATAACAA	Cas_32745_R	ATTCGCCAATCTACCACGAG
Pair 10	Cas_77140_F	TGTCGCAATGGCTCTATTTG	Cas_78034_R	TCATCTCGTACAGCTCAAGCA
Pair 11	**Cas_10778_F**	**TCAGTTTGATGATCCTTTGATGA**	**Cas_11525_R**	**CATTCGGCTCCTTTATGGAA**
Pair 12	Cas_46472_F	GGGAACTATTCCGATTTCATTG	Cas_47162_R	CTAACTGGTGGAATAAAGGTCTCC
Pair 13	Cas_3185_F	TTGCGGTCAATAAGGTAGGG	Cas_3987_R	AGGCTCAGAGTTGTTGGAAGA
Pair 14	Cas_17609_F	ACACTCTCGCAGAGCCGTAT	Cas_18412_R	CCACGATAGACCAGAACAATCA
Pair 15	Cas_33546_F	GACTCGTTTGGGAATTAAATCAA	Cas_34499_R	CTTCAACCATTTCCGAGCAC
Pair 16	Cas_67504_F	TTGTACCGAGGGCATCTTTAG	Cas_68343_R	AACCGAAATAACTCGTTATAGTAAGCA
Pair 17	Cas_72399_F	GGGCTGGTTTAGATTGATCCT	Cas_73245_R	TTTCATTGGATTATGTATCGAGAGAG
Pair 18	Cas_62840_F	ATTCCGTTGACCCGTACTGA	Cas_63772_R	AAGAGAGAAATCCACCAAGGTAAA
Pair 19	Cas_65707_F	CTCGGGAAATCCCTTGTACC	Cas_66634_R	TCCGGATTAGGTTCATCCCTA
Pair 20	Cas_69456_F	CAACTCTAAGCGACCCTTAAAATACA	Cas_70174_R	TACTCGCGCATCTTTCCTCT
Pair 21	Cas_4537_F	GGTTCCTTGACCAACCACAG	Cas_5319_R	ATCCCAACAACAGACTTCC

Pair 22	Cas_48611_F	TGTAATTTCGATTTCTTGATCACAAT	Cas_49520_R	CAGATACAGATTTGGGCCATC
Pair 23	Cas_29425_F	TTGAAGCGAAGTAGGATAATTTGA	Cas_30291_R	TCCTACCAGAGGCTACAATCTGA
Pair 24	Cas_125001_F	AAATAATCCCAACGCGTTACA	Cas_125859_R	AATGTTTCAATTAGCTCTCGAAATG
Pair 25	Cas_21290_F	TGTTCTGATTCTACATATTGATCGTTT	Cas_22036_R	CGTGAAGGGCTTTCTTTAACA
Pair 26	Cas_20851_F	CTGGTAGAGAGTGGTCGGATCT	Cas_21307_R	GGTTGAATTGGGAGAAGCTG
Pair 27	Cas_11589_F	AGCCTTCCAAGCTAACGATG	Cas_12461_R	CTGGAATCAGACCCGCTATT
Pair 28	Cas_47139_F	GGAGACCTTTATTCCACCAGTTAG	Cas_47689_R	TTTCGATTGGGTATGGCTTC
Pair 29	Cas_14073_F	TTCGATTCATTTGGCTCTCA	Cas_14724_R	TGGAAAGGGAGTGTGTGTGA
Pair 30	Cas_122476_F	ACGGCTCCTCATAGGTCACA	Cas_123331_R	TGCTGTTAAAGGAATTCAATCTCA
Pair 31	**Cas_73947_F**	**TCTTGTTCCTGAATGGGTCTC**	**Cas_74498_R**	**GTTACGTTTCCCACATCAAAGTGA**
Pair 32	Cas_123306_F	AATGAGATTGAAATTCCTTTAAACAGC	Cas_124104_R	TGAAATTGGCTGATATTATGACG
Pair 33	Cas_24256_F	ATAAACCGCGTAATCGCAAG	Cas_25037_R	TGTCATCCCCAGTCAATCCAA
Pair 34	**Cas_85769_F**	**CATCAGGATATACCATAGTTGCCTTT**	**Cas_86417_R**	**CCATACGATTGCCGTTCATA**
Pair 35	Cas_36699_F	GCGGTCCGCAGAATATATGA	Cas_37444_R	TTATTTCACAAATGGGAATCCTG
Pair 36	Cas_61880_F	GCAATGGCTTCTTTATTTCTTCA	Cas_62831_R	GGCCTCGGATGTCCATATAA
Pair 37	Cas_71554_F	TCCAATGGCTTCGGCTACTA	Cas_72431_R	AATCATCCGGTTAGGATCAATCT
Pair 38	Cas_63749_F	TTTACCTTGGTGGATTTCTCTCTT	Cas_64669_R	AGCGATACGGAAATAGATCGAG
Pair 39	Cas_5508_F	CCCATTCATTTCCTTTAATTCG	Cas_6230_R	TTAGCTCAACAGTTTGATTAGCTTG
Pair 40	Cas_13394_F	CGAGCAATACCATCGCCTAC	Cas_14062_R	TTGGTTCGGGAAGGGATTAT
Pair 41	Cas_19198_F	TCCTGGAGTTGGCCAAATAAG	Cas_19976_R	CCTTTGTTGAAATAAGGGCAAA
Pair 42	Cas_124082_F	CGTCATAATATCAGCCAATTTCA	Cas_124968_R	ATGGACCGAACGACTAGG
Pair 43	Cas_52327_F	TCGATAAATACAGATACACCCAATACA	Cas_52920_R	GCAAGAAATCCTAGGCGAAGA
Pair 44	Cas_50548_F	TCGGTTCAGATACAAATAAATCCA	Cas_51414_R	AGGGTCATTTGTCTGCTTGG
Pair 45	Cas_126859_F	GAAACGGATCCAAGATCTCCTC	Cas_127713_R	GGCGAAATGCGCTTCATA

Estimating theta and tau

We used an estimate of θ in this study in several analyses, each having its own assumptions about what theta represents. We used an estimate of θ in our empirical species delimitation as 1) an estimate of θ for all loci to be used by STEM (as part of spedeSTEM) to construct species trees, and 2) by BPP as a prior estimate of the population size of modern and ancestral species used in the implementation of the multispecies coalescent model. We also used θ in our simulations in the program *ms* where we simulated genealogies under the multispecies coalescent. spedeSTEM and BPP expect the mutation rate component of θ to be representative across loci; *ms*, expects the mutation rate to be representative of the locus the genealogy is being modeled after (i.e., μ is the mutation rate for the specific locus being modeled). Furthermore, STEM (used by spedeSTEM to build species trees) requires a conversion of theta to a per-site mutation rate estimate.

We used the program MIGRATE-N v.3.6 (Beerli and Felsenstein 2001) to estimate a value of θ appropriate for our dataset. Three independent Bayesian MIGRATE-N analyses were conducted to ensure convergence on the same parameter estimates. Sequences were organized into populations corresponding to their taxonomic identification; each taxon was treated as one population. Each analysis consisted of one long chain and 10 short chains (four of which were statically heated), replicated twice within each analysis. The results of these analyses are an estimate of theta for each taxon and locus (Table S3.1). At the end of our MIGRATE-N analyses, we averaged the estimates of θ across each run, and then averaged the estimates of θ for each locus (to be used in *ms*) and across loci (to be used in spedeSTEM and as priors for BPP) (Table S3.2).

We used analysis A00 of BPP v.3.1 (Yang and Rannala 2014) to estimate the divergence time parameter, τ, and the population size parameters for modern and ancestral species, θ. We modeled these parameters on the species tree topology from our *BEAST analysis, loosely informed by our MIGRATE-N results. Because we had little information available to guide the prior settings, we applied a diffuse setting to the alpha shape parameter (α=2) and then tested multiple β values such that the posterior estimate of gamma (calculated as α/β) was both above and below the estimate of theta from our MIGRATE-N analyses by multiple orders of magnitude (following (Yang 2015)). To do so, we had to make some general assumptions about time of divergence (we used a conservative age of 1,000,000 years ago) and an estimate of the overall mutation rate in *Castilleja* (locus-wide estimate of theta 0.0105 / total number of base pairs 26490 = 0.000000396). During this exploratory period, we further confirmed that the posterior was insensitive to the prior.

Our exploratory runs sampled every two generations for a total of 20,000 samples, after an initial burn-in of 2,000 samples had been achieved. Once we had determined appropriate settings for the prior, we conducted four independent analyses from which to draw our posterior estimates of these parameters. These analyses sampled every ten generations for a total 200,000 samples, after a burn-in of 2,000 samples. Analyses were conducted multiple times to confirm that results were stable across runs (Table S3.1 and Fig S3.1).

SUPPLEMENTAL TABLE S1.3.1. Results of MIGRATE-N and BPP estimates of theta and tau. Migrate-n results are an average of three independent analyses; BPP results are an average of four independent analyses. The table below compares estimates of theta between MIGRATE-N and BPP A00 analyses.

	Migrate-n results			BPP results
	estimated theta		averaged theta	averaged theta
	nuclear	*chloroplast*	*across loci*	
C. ambigua	0.0233	0.0102	0.0167	0.0326
C. ambigua v. *meadii*	0.0156	0.0009	0.0082	0.0055
C. victoriae	0.0051	0.0084	0.0067	0.0054
averaged theta *across species*	0.0146	0.0064	0.0105	0.0145

SUPPLEMENTAL TABLE S1.3.2. Theta estimates and locus specific length and variability used to calculate dataset-wide, per site mutation rate (bold box) and scaling factors for spedeSTEM (last column).

	estimated theta	length; site patterns	per site mutation rate	per locus variation	scaling factor for spedeSTEM
			(theta / length)	(site patterns/ locus length)	(nuc variation/chl variation)
nuclear	0.0146	1139; 107	0.00001281	0.093942054	1
chloroplast	0.0064	25351; 350	0.000000252	0.013806161	6.8043
		26490	0.000000396		
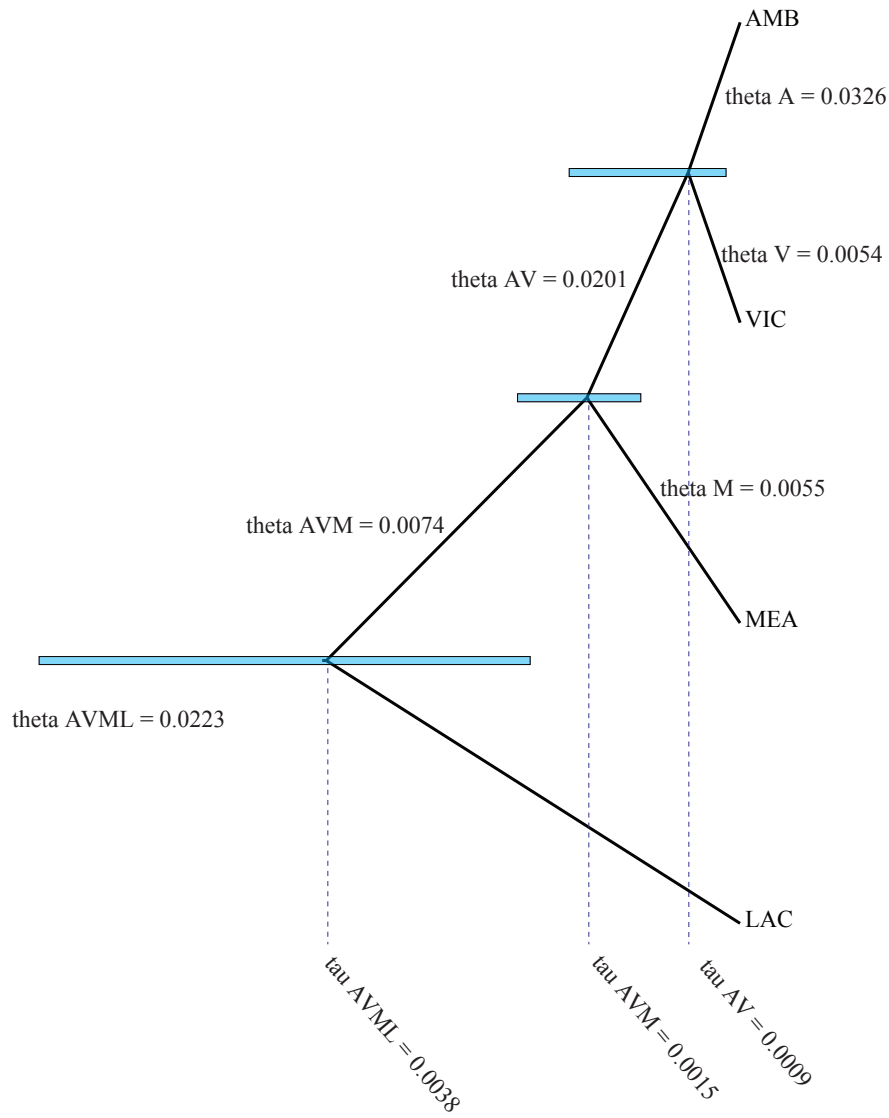
SUPPLEMENTAL FIGURE S1.3. Estimates (posterior ranges and means) of parameters obtained from BPP A00 analysis. The node bars represent the 95% HPD intervals for divergence times, with mean values represented at ends of dashed lines, and resulting theta values are placed along each branch. There is only one sequence for LAC, and therefore no estimate of theta for this lineage. AMB and A - *Castilleja ambigua* (including varieties *ambigua*, *humboldtiensis*, and *insalutata*); VIC and V - *Castilleja victoriae*; MEA and M - *Castilleja ambigua* var. *meadii*.

Further detail of simulation procedure

        To test the capability of these approaches to delimit species given our dataset, we used a simulation approach (Fig. 2, main text). 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 to generate sequence alignments that were comparable to our empirical dataset using the program *seq-gen* (Rambaut and Grass 1997). The subsequent sequence alignments then became the input for species delimitation of simulated data with a known topology.

        In the following outline, we specify demographic components modeled during genealogy simulation (Step 1, Fig S4.1), the way in which we scaled our genealogies (Step 2), and the methods used to evolve sequences along our genealogies (Step 3). Commands used to simulate genealogies and evolve sequences are included below. Detailed values for genealogy simulation (Supplemental Table S4.1 – Table S4.4) and molecular evolution (Supplemental Table S4.5) follow.

Step 1. Components of demographic modeling for our genealogy simulation

1. Number of populations and how many individuals within each population (Table S4.1).

    - Modeled after the dataset we collected: twenty total individuals from four populations where one population has 13 members (corresponds to AMB), two populations have three members (corresponds to VIC and MEA), and one population has one member (corresponds to LAC). In our 'one lineage' model, we identify 19 individuals belonging to lineage AMB_MEA_VIC, and one individual belonging to lineage LAC.

2. The modern and ancestral sizes of the populations (corresponding to N0, N1, N2, N3, and N4 in Fig. S4.1) (Table S4.3).

    - Population sizes were set relative to the ancestral population common to all lineages (N0). We made the assumption that the AMB and LAC populations were both half the size of population N0; further, we assumed that the MEA and VIC populations were both 0.001 the size of the N0. We made a conservative estimate of the ancestral population size N0 to be 100,000 individuals. Therefore, AMB and LAC were each 50,000 individuals and MEA and VIC were each 100 individuals.

3. Timing of the merging of populations (corresponding to T1, T2, and T3 in Fig. S4.1) (Table S4.4).

- We used the median node heights of our preliminary empirical species tree (from *BEAST), which were estimated relative to one another, to inform the timing of these events; from oldest to most recent, these values were 0.0074, 0.0011, 0.0004 (Fig 3, b, from main text). Population merging times were scaled by the ploidy of the locus and the ancestral population size.

The following commands were used to simulate genealogies:

- (((AMB, VIC), MEA), LAC) (species tree topology)
    i. chl genealogy: ./ms 20 100 -T -t 1232.5 -I 4 13 3 3 1 -n 1 0.5 -n 2 0.001 -n 3 0.001 -n 4 0.5 -ej 0.2 3 1 -ej 0.55 2 1 -ej 3.7 4 1
    ii. nuc genealogy: ./ms 20 100 -T -t 5786.12 -I 4 13 3 3 1 -n 1 0.5 -n 2 0.001 -n 3 0.001 -n 4 0.5 -ej 0.1 3 1 -ej 0.275 2 1 -ej 1.85 4 1
- (((AMB, MEA), VIC), LAC) (taxonomic hypothesis)
    i. chl genealogy: ./ms 20 1 -T -t 1232.5 -I 4 13 3 3 1 -n 1 0.5 -n 2 0.001 -n 3 0.001 -n 4 0.5 -ej 0.2 2 1 -ej 0.55 3 1 -ej 3.7 4 1
    ii. nuc genealogy: ./ms 20 1 -T -t 5786.12 -I 4 13 3 3 1 -n 1 0.5 -n 2 0.001 -n 3 0.001 -n 4 0.5 -ej 0.1 2 1 -ej 0.275 3 1 -ej 1.85 4 1
- (((MEA, VIC), AMB), LAC) (alternative three lineage topology)
    i. chl genealogy: ./ms 20 1 -T -t 1232.5 -I 4 3 3 13 1 -n 1 0.001 -n 2 0.001 -n 3 0.5 -n 4 0.5 -ej 0.2 2 1 -ej 0.55 3 1 -ej 3.7 4 1
    ii. nuc genealogy: ./ms 20 1 -T -t 5786.12 -I 4 3 3 13 1 -n 1 0.001 -n 2 0.001 -n 3 0.5 -n 4 0.5 -ej 0.1 2 1 -ej 0.275 3 1 -ej 1.85 4 1
- ((AMB_MEA_VIC), LAC) (one lineage topology)
    i. chl genealogy: ./ms 20 1 -T -t 1232.5 -I 2 19 1 -n 1 0.5 -n 2 0.5 -ej 3.7 2 1
    ii. nuc genealogy: ./ms 20 1 -T -t 5786.12 -I 2 19 1 -n 1 0.5 -n 2 0.5 -ej 1.85 2 1

Step 2. Genealogy Scaling
1. Scaling factors for simulated genealogies.
- The genealogies were simulated under the coalescent. Therefore, branch lengths are in coalescent units. Seq-gen, used in the next step to simulate sequences, expects branch lengths in units of substitutions per site. As a proxy for converting from coalescent units to substitutions per site, we calculated a scaling factor to be used as part of the sequence evolution step. Using the ape package in R (Paradis et al 2004; R Core Team 2016), we calculated the mean tree height of all trees from the burned

posterior distribution of empirical gene trees (mrbayes runs) and the mean tree height of simulated genealogies for each locus. The ratio of these lengths provided the scaling factor for the simulated sequences using the –s flag (see below).

Step 3. Evolving Sequences

1. Models of nucleotide evolution.
   - Nucleotide sequences were evolved along the simulated genealogies using the program *seq-gen* (Rambaut and Grass 1997). Parameters of nucleotide evolution estimated from our empirical model selection process (partitionfinder), and the branch length scaling factor mentioned above, served as simulation parameters. For each locus, partitions recovered during the model selection process were simulated separately and then concatenated to create sequence alignments with the same number and variability of nucleotides found in our empirical datasets. The following commands were used to simulate sequences on all simulated genealogies (regardless of topology).

The following commands were used to evolve sequences on the simulated genealogies. Scaling of the branches occurs in this step with the –s flag.

- Chloroplast sequences:

  par.1 -> ./seq-gen -mHKY -l4386 -i0.939 -f0.3033,0.1286,0.158,0.4101 -t0.5 -s0.003

  par.2 -> ./seq-gen -mHKY -l11921 -i0.92 -f0.3496,0.177,0.1676,0.3059 -t0.5 -s0.003

  par.3 -> ./seq-gen -mHKY -l1643 —a0.768 -g4 -i0.947 -f0.3408,0.1523,0.13,0.3769 -t0.5 -s0.003

  par.4 -> ./seq-gen -mHKY -l6339 -i0.95 -f0.2738,0.1983,0.1837,0.3441 -s0.003

  par.5 -> ./seq-gen -mHKY -l508 —a0.272 -g4 -i0.252 -f0.3409,0.1491,0.1503,0.3597 -t0.5 -s0.003

  par.6 -> ./seq-gen -mHKY -l554 -f0.3779,0.1557,0.1888,0.2776 -t0.5 -s0.003

- Nuclear sequences:

  par.1 -> ./seq-gen -mGTR -l450 -a15.988 -g4 -f0.1579,0.2779,0.2859,0.2783 -r0.91238,1.63328,2.78612,0.32558,1.63328,1 -s0.015

  par.2 -> ./seq-gen -mGTR -l689 —i0.763 -f0.1888,0.3222,0.3051,0.1839 -r0.6289,1.0201,0.9926,0.0134,2.2502,1 -s0.015
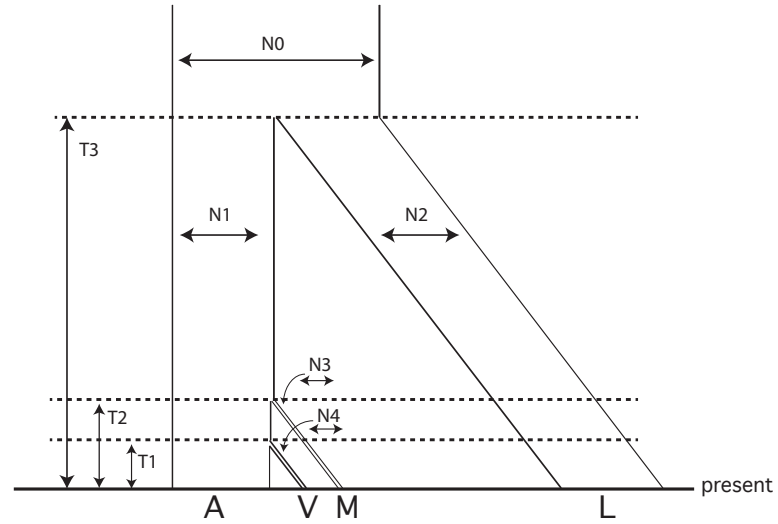
SUPPLEMENTAL FIGURE S1.4. Visual representation of simulated parameters including population sizes (N0, N1, N2, N3, and N4) and divergence times (T1, T2, and T3) modeled on the species tree topology. A – *Castilleja ambigua* (including varieties *ambigua, humboldtiensis*, and *insalutata*); V – *Castilleja victoriae*; M – *Castilleja ambigua* var. *meadii*; L – *Castilleja lacera*.

SUPPLEMENTAL TABLE S1.4.1. Number of populations and associated individuals simulated for each topology.

	Populations	Number of individuals per lineage
Species tree hypothesis:		
(((AMB, VIC), MEA), LAC)	four	13 (AMB), 3 (VIC), 3 (MEA), 1 (LAC)
Taxonomic hypothesis:		
(((AMB, MEA), VIC), LAC)	four	13 (AMB), 3 (MEA), 3 (VIC), 1 (LAC)
Alternative three lineage hypothesis:		
(((MEA, VIC), AMB), LAC)	four	3 (MEA), 3 (VIC), 13 (AMB), 1 (LAC)
One lineage hypothesis:		
((AMB_MEA_VIC), LAC)	one	19 (AMB_VIC_MEA), 1 (LAC)

AMB – *Castilleja ambigua* (including varieties *ambigua*, *humboldtiensis*, and *insalutata*)
VIC – *Castilleja victoriae*
MEA – *Castilleja ambigua* var. *meadii*.
LAC – *Castilleja lacera*

SUPPLEMENTAL TABLE S1.4.2. Preliminary mutation rate estimates, locus length, ploidy values, and our estimate of ancestral population size used to calculate theta value used in genealogy simulation.

	Estimated mutation rate	length of locus	ploidy	Ancestral population size (N0)	theta for *ms*
chloroplast	0.000000243	25351	2	100,000	1232.5
nuclear	0.0000127	1139	4	100,000	5786.12

S<small>UPPLEMENTAL</small> T<small>ABLE</small> S1.4.3. Population sizes of each lineage, relative to ancestral population.

	With respect to ancestral population size (N0=100,000)	Implied size
AMB	0.5	50,000
MEA	0.001	100
VIC	0.001	100
LAC	0.5	50,000

AMB – *Castilleja ambigua* (including varieties *ambigua*, *humboldtiensis*, and *insalutata*)
VIC – *Castilleja victoriae*
MEA – *Castilleja ambigua* var. *meadii*.
LAC – *Castilleja lacera*

SUPPLEMENTAL TABLE S1.4.4. Timing of population merging for each topology tested

		Scaled by ploidy and N0	
		chloroplast	nuclear
Species tree hypothesis: (((AMB, VIC), MEA), LAC)			
VIC - AMB	40,000 generations ago	0.2	0.1
VIC_AMB - MEA	110,000 generations ago	0.55	0.275
VIC_AMB_MEA - LAC	740,000 generations ago	3.7	1.85
Taxonomic hypothesis: (((AMB, MEA), VIC), LAC)			
MEA - AMB	40,000 generations ago	0.2	0.1
MEA_AMB - VIC	110,000 generations ago	0.55	0.275
MEA_AMB_VIC - LAC	740,000 generations ago	3.7	1.85
Alternative three lineage hypothesis: (((MEA, VIC), AMB), LAC)			
MEA - VIC	40,000 generations ago	0.2	0.1
MEA_VIC - AMB	110,000 generations ago	0.55	0.275
MEA_VIC_AMB - LAC	740,000 generations ago	3.7	1.85
One lineage hypothesis: ((AMB_MEA_VIC), LAC)			
MEA_VIC_AMB - LAC	740,000 generations ago	3.7	1.85

AMB – *Castilleja ambigua* (including varieties *ambigua*, *humboldtiensis*, and *insalutata*)
VIC – *Castilleja victoriae*
MEA – *Castilleja ambigua* var. *meadii*.
LAC – *Castilleja lacera*

SUPPLEMENTAL TABLE S1.4.5. Results of estimation of nucleotide evolution for our dataset. These values were used directly to evolve sequences on our simulated genealogies (refer to command lines noted above).

				frequency			
locus	partition	bp	model of selection	A	C	G	T
cp	1	4386	F81 + I	0.3033	0.1286	0.158	0.4101
	2	11921	F81 + I	0.3496	0.177	0.1676	0.3059
	3	1643	F81 + I + G	0.3408	0.1523	0.13	0.3769
	4	6339	HKY + I	0.2738	0.1983	0.1837	0.3441
	5	508	F81 + I + G	0.3409	0.1491	0.1503	0.3597
	6	554	F81	0.3779	0.1557	0.1888	0.2776
	Total bp	25351					
nuc	1	450	TVM + G	0.1579	0.2779	0.2859	0.2783
	2	689	GTR + I	0.1888	0.3222	0.3051	0.1839
	Total bp	1139					

				rate matrix					
locus	partition	p-inv	gamma; # cat.	a	b	c	d	e	f
cp	1	0.939							
	2	0.92							
	3	0.947	0.768; 4						
	4	0.95							
	5	0.252	0.272; 4						
	6								
	Total bp	25351							
nuc	1		15.988; 4	0.91238	1.63328	2.78612	0.32558	1.63328	1
	2	0.763		0.6289	1.0201	0.9926	0.0134	2.2502	1
	Total bp	1139							

SUPPLEMENTAL DATA S5
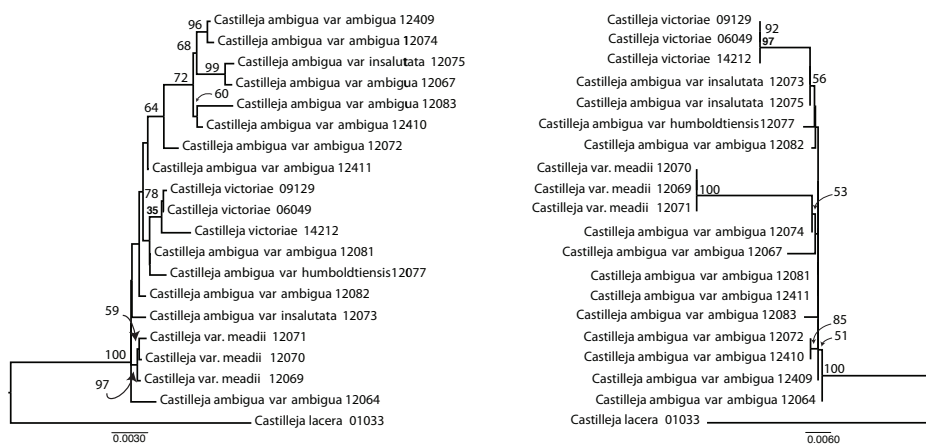
Maximum likelihood reconstructions (chloroplast (left) and nuclear (right))
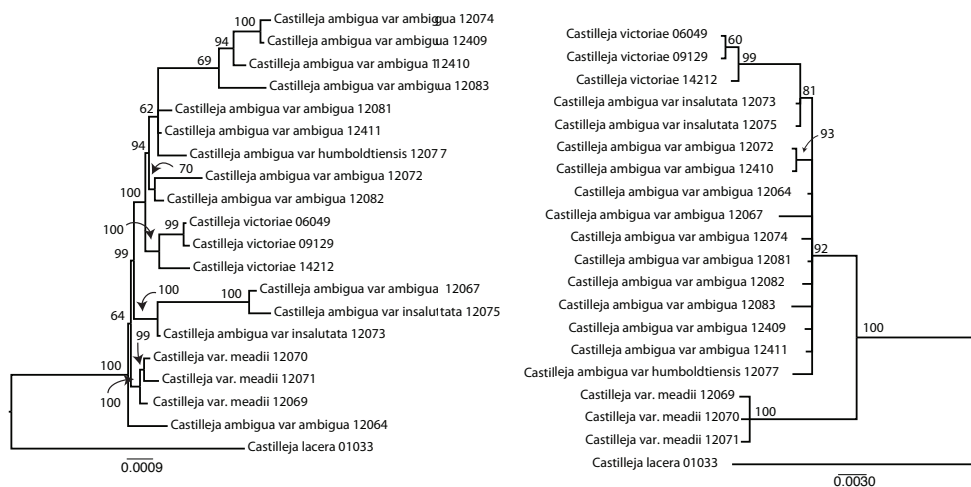
Bayesian reconstructions (chloroplast (left) and nuclear (right))

SUPPLEMENTAL FIGURE S1.5. Results of maximum likelihood (top) and Bayesian (bottom) reconstructions of chloroplast (left) and nuclear (right) datasets. Values at nodes reflect bootstraps (in the case of the maximum likelihood inference) and posterior probabilities (in the case of Bayesian inference).

SUPPLEMENTAL DATA S6

SUPPLEMENTAL TABLE S1.6. Results of delimitation with spedeSTEM using simulated data. Here, for brevity, we report the final rankings of each simulation organized by topology modeled. Rows represent the lineage composition model and columns represent the rank position, first through fifth. For each topology modeled, we performed 100 simulations.

**Taxonomic Hypothesis
((AMB, MEA), VIC)**

	First	Second	Third	Fourth	Fifth
AMB_MEA_VIC	100	0	0	0	0
AMB_MEA, VIC	0	77	23	0	0
AMB_VIC, MEA	0	23	77	0	0
MEA_VIC, AMB	0	0	0	0	100
AMB, MEA, VIC	0	0	0	100	0

**Species Tree Hypothesis
((AMB, VIC), MEA)**

	First	Second	Third	Fourth	Fifth
AMB_MEA_VIC	100	0	0	0	0
AMB_MEA, VIC	0	37	63	0	0
AMB_VIC, MEA	0	63	37	0	0
MEA_VIC, AMB	0	0	0	0	100
AMB, MEA, VIC	0	0	0	100	0

**Alternative Three-Lineage Hypothesis
((MEA, VIC), AMB)**

	First	Second	Third	Fourth	Fifth
AMB_MEA_VIC	94	6	0	0	0
AMB_MEA, VIC	0	1	82	17	0
AMB_VIC, MEA	0	0	18	82	0
MEA_VIC, AMB	0	0	0	0	100
AMB, MEA, VIC	6	93	0	1	0

**One Lineage Hypothesis
(AMB_MEA_VIC)**

	First	Second	Third	Fourth	Fifth
AMB_MEA_VIC	20	1	63	0	7
AMB_MEA, VIC	0	68	27	5	2
AMB_VIC, MEA	0	2	0	26	71
MEA_VIC, AMB	71	29	10	0	0
AMB, MEA, VIC	9	0	0	69	20

AMB – *Castilleja ambigua* (including varieties *ambigua*, *humboldtiensis*, and *insalutata*)
VIC – *Castilleja victoriae*
MEA – *Castilleja ambigua* var. *meadii*.

SUPPLEMENTAL DATA S7

Results of delimitation with BPP Analysis A11 (unguided delimitation), using simulated data. Each table below represents one of four topologies simulated. For each topology examined, ten independently simulated datasets were analyzed, each run twice to confirm stability across rjMCMC algorithms, for a total of 20 separate analyses per topology. For each topology, we report the best model recovered (the number of distinct lineages, ignoring phylogeny), the probability of taxonomic species, and all models included in the 95% credibility set of models. For each of these categories, we report the mean posterior probability across analyses, the minimum and maximum posterior probabilities recovered, and the number of independent analyses for which the posterior probability recovered was greater than 0.95. In the case of the alternative three-lineage and the one lineage topology, results of species delimitation recovered an alternative taxonomic species with high posterior probability. The last two tables (corresponding to those topological simulations) include information for these highly supported alternative species.

SUPPLEMENTAL TABLE S1.7.1. Results of delimitation with BPP Analysis A11 using data simulated under our species tree hypothesis, (((AMB, VIC), MEA), LAC).

	Posterior probabilities			
Best Model (ignoring species tree phylogeny)	mean	min	max	number reps (in 19) $pp > 0.95$
4 distinct lineages: A, M, V, L	0.9997	0.9994	0.9998	19
Posterior probability of taxonomic species	mean	min	max	number reps (in 19) $pp > 0.95$
C.ambigua.ambigua	0.9999	0.9998	0.9999	19
C.ambigua.meadii	0.9998	0.9996	0.9999	19
C.victoriae	0.9997	0.9994	0.9998	19
C.lacera	1.0000	1.0000	1.0000	19
Best models in the 95% credibility set	mean	min	max	number reps (in 19) $pp > 0.95$
(((A, V), M), L)	0.9745	0.9652	0.9810	19

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

SUPPLEMENTAL TABLE S1.7.2. Results of delimitation with BPP Analysis A11 using data simulated under the taxonomic hypothesis, (((AMB, MEA), VIC), LAC).

	Posterior probabilities			
Best Model (ignoring species tree phylogeny)	mean	min	max	number reps (in 20) $pp > 0.95$
4 distinct lineages: A, M, V, L	0.9999	0.9997	1.0000	20
Posterior probability of taxonomic species	mean	min	max	number reps (in 20) $pp > 0.95$
C.ambigua.ambigua	0.9998	0.9997	1.0000	20
C.ambigua.meadii	0.9998	0.9997	1.0000	20
C.victoriae	1.0000	1.0000	1.0000	20
C.lacera	1.0000	1.0000	1.0000	20
Best models in the 95% credibility set	mean	min	max	number reps (in 20) $pp > 0.95$
(((A, M), V), L)	0.9964	0.9929	0.9989	20

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

SUPPLEMENTAL TABLE S1.7.3. Results of delimitation with BPP Analysis A11 using data simulated under the alternative three-lineage topology, (((MEA, VIC), AMB), LAC).

	Posterior probabilities				
Best Model (ignoring species tree phylogeny)	mean	min	max	number reps (in 20) *pp* > 0.95	
2 distinct lineages: AMV, L	0.9300	0.8511	0.9545	4	
Posterior probability of taxonomic species	mean	min	max	number reps (in 20) *pp* > 0.95	
C.ambigua.ambigua	0.0503	0.0134	0.1340	0	
C.ambigua.meadii	0.0175	0.0029	0.0429	0	
C.victoriae	0.0035	0.0024	0.0067	0	
C.lacera	1.0000	0.9999	1.0000	20	
Posterior probability of *alternative* species	mean	min	max	number reps (in 20) *pp* > 0.95	occurrence
C.ambigua.ambigua + *C.ambigua.meadii* + *C.victoriae*	0.9283	0.8418	0.9519	4	20
C.ambigua.meadii + C.victoriae	0.0644	0.0367	0.1323	0	13
C.ambigua.ambigua + C.victoriae	0.0412	0.0392	0.0424	0	3
Best models in the 95% credibility set	mean	min	max	number reps (in 20) *pp* > 0.95	occurrence
(AMV, L)	0.9301	0.8512	0.9545	4	20
((MV, A), L)	0.0644	0.0367	0.1323	0	13
((AV, M), L)	0.0412	0.0392	0.0424	0	3

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

SUPPLEMENTAL TABLE S1.7.4. Results of delimitation with BPP Analysis A11 using data simulated under the one lineage topology, (MEA_VIC_AMB), LAC).

	Posterior probabilities			
Best Model (ignoring species tree phylogeny)	mean	min	max	number reps (in 20) $pp > 0.95$
2 distinct lineages: AMV, L	0.9960	0.9857	0.9986	20
Posterior probability of taxonomic species	mean	min	max	number reps (in 20) $pp > 0.95$
C.ambigua.ambigua	0.0004	0.0000	0.0010	0
C.ambigua.meadii	0.0033	0.0004	0.0132	0
C.victoriae	0.0003	0.0000	0.0007	0
C.lacera	1.0000	1.0000	1.0000	20
Posterior probability of *alternative* species	mean	min	max	number reps (in 20) $pp > 0.95$
C.ambigua.ambigua + *C.ambigua.meadii* + *C.victoriae*	0.9960185 0.9960185	0.98576 0.98576	0.99869 0.99869	20
Best models in the 95% credibility set	mean	min	max	number reps (in 20) $pp > 0.95$
(AMV, L)	0.9960185	0.98576	0.99869	20

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

SUPPLEMENTAL REFERENCES

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences*. **98**, 4563–4568.

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

Paradis E, Claude J, Strimmer K (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*. **20**, 289-290.

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.

Yang Z (2015) The BPP program for species tree estimation and species delimitation. *Current Zoology*. **61**, 854–865.

Yang Z, Rannala B (2014) Unguided species selimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution*. **31**, 3125–3135.

CHAPTER 2: QUANTIFYING MORPHOLOGICAL VARIATION IN THE CASTILLEJA PILOSA SPECIES COMPLEX

with Sarah Herzog and David C. Tank

Abstract

Robustly delimited species are of paramount importance, the identification of which relies on our ability to discern boundaries between one species and the next. This is not difficult to do when *species* are very distinct from one another. However, in recently evolved lineages where putative species may have relatively few diagnostic features (e.g., species complexes composed of very similar species, the boundaries between which are often unclear), defining species boundaries can be more challenging. Hence the field of species delimitation has widely advocated the use of multiple lines of evidence to delimit species, particularly in species complexes. Excessive taxonomic confusion, often the result of species descriptions that shift through time (e.g. during revisionary work and regional treatments), can further complicate the search for diagnostic features in species complexes. Here, as a first step in robustly delimiting species boundaries, we quantify and describe morphological variation in the *Castilleja pilosa* species complex. We first infer the morphospace of the complex and use fuzzy-clustering techniques to explore the morphological variation in the system. Next, we hypothesize the position of type specimens within morphospace. In so doing, we visualize the impact that regional treatments have had on the conceptualization of taxa through time. We find that there is limited morphological variation among members of this complex, and we determine that current species boundaries are no longer accurately represented by type specimens.

Introduction

Because they provide the basis for the recognition of one of the primary units of biodiversity, the *species,* classifications are the cornerstone of the biodiversity sciences. As such, classifications are vital to our understanding of biodiversity and the process of speciation. Therefore, the careful and robust delimitation of *species* is imperative. Species delimitation relies on our ability to define boundaries between one population and the next. Historically, this has been done using morphological evidence (Sneath and Sokal 1974), ecological evidence (Van Valen 1976), and more recently, in light of technical and analytical advances, molecular evidence (Baum and Shaw 1995). Each criterion has limitations for being widely applied across the tree of life (De Queiroz 2007), and no one criterion has been universally applied to defining species boundaries (De Queiroz 2009).

Instead, there has been a movement to include multiple lines of evidence in the delimitation of species (e.g., Padial et al. 2010; Schlick-Steiner et al. 2010; Carstens et al. 2013; Dejaco et al. 2016; Freudenstein et al. 2016). Species delimitation methods are often applied to existing classifications where species boundaries are poorly defined and/or sample assignment to species is difficult (e.g. Barley et al. 2013; Giarla et al. 2014). In these cases, species delimitation is used in a *validation* context (where taxonomic boundaries are validated—i.e., individuals are assigned to a group *a priori* (Ence and Carstens 2010)) and attempts to clarify species boundaries and which lines of evidence (morphological, ecological, molecular) do and do not describe *species*.

Traditionally, species have been described and anchored by a type specimen and its corresponding morphological and ecological traits, providing a central point around which some amount of variation occurs (Fig. 1). However, the characteristics of this variation (the amount, the direction, etc.) are not static, and additional collections assigned to a species can shift the conceptual boundaries of the species, in particular how this is applied on-the-ground. For example, regional and floristic studies can result in treatments and species descriptions that incorporate variation observed in the field on a local scale. Revisionary work, typically happening at a broader scale (e.g. the Flora of North America), often recognizes overlapping variation between similar species and synonymizes names where appropriate. As a result, there can be a shift of species boundaries and known variation. In essence, these shifts can inflate or deflate the taxonomic conceptualization outside of the realm of its original description, and in some cases this can result in a type specimen that no longer serves as a central, anchoring point within the range of known variation for a species, and instead only represents a portion of that variation (Fig. 1).

Often, the species involved in these taxonomic fluctuations are characterized as species complexes (i.e., groups of species that are difficult to distinguish from one another),and are already known to have overlapping variation that is difficult to classify. The shifting of recognized and ascribed variation through time and across treatments can further increase the fuzziness of species boundaries, making the identification of unknown individuals (and therefore the usefulness of the classification) even more difficult. This is further complicated when an unknown comes from a geographic boundary (or, conversely, one that is widespread but has varieties that occur in geographically restricted areas), served by two or more regional or localized treatments that have varying interpretations of variation within a taxon. This requires choices to be made by the identifier in preferring one treatment to another when treatments are in conflict (e.g., one treatment recognizes varieties, while another does not). Cases such as these—species complexes with a great deal of taxonomic confusion—are good targets for robust species delimitation. By clarifying and

determining which lines of evidence distinguish species, classifications can be updated to reflect more accurate estimates of species boundaries.

Recognizing species in the genus *Castilleja* Mutis ex L. f. (Orobanchaceae Vent.) is notoriously difficult, particularly in the field. These difficulties largely stem from a nearly continuous range of variation both within and across taxonomic boundaries (Cronquist et al. 1984). The source of this morphological continuity is likely a combination of the age of the lineage, the widespread and highly variable instances of polyploidy, and interspecific gene flow when species co-occur (Heckard and Chuang 1977; Tank et al. 2008; Tank et al. 2009). This means that most often the characters that diagnose species are slight and often overlapping. A good example of these difficulties is the *Castilleja pilosa* species complex, composed of two taxonomic species, *Castilleja nana* Eastwood and *Castilleja pilosa* (S. Watson) Rydberg (with three named varieties), that are geographically distinct and morphologically very similar (Fig. 2a). All members of this complex belong to a wider group of *Castilleja* species whose tubular flowers have less-showy corollas with short beaks and a pouchy lower corolla lip, with somewhat petaloid teeth. The calyx lobes of these species are often subequal and the depth of the sinuses, and the corresponding size and shape of the calyx lobe segments, are often diagnostic (Fig. 2b) (Cronquist et al. 1984; Hitchcock et al. 1984; Wetherwax et al. 2012).

Castilleja pilosa is composed of three taxonomically recognized varieties, distinguished primarily by geography, in addition to slight variations in a suite of morphological characters. *Castilleja pilosa* (S. Watson) Rydberg var. *pilosa* is found in the Sierra Nevada, north and east into Oregon; *Castilleja pilosa* (S.Watson) Rydb. var. *steenensis* (Pennell) N.H.Holmgren is endemic to the high ridges of Steens Mountain in southeastern Oregon; *Castilleja pilosa* (S.Watson) Rydb. var. *longispica* (A.Nelson) N.H.Holmgren occurs in the southern half of Idaho, east into western Wyoming and Montana, and has a disjunct population in northern Idaho. These varieties are distinguished by calyx length, herbage pubescence, elevation, and geographic position (Cronquist et al. 1984). *Castilleja nana* occurs throughout the central and southern Sierra Nevada range of eastern California and extends eastward on the high ridges of Nevada's basin and range topography. *Castilleja nana* is primarily distinguished from *C. pilosa* by elevation; *C. nana* occurs between 2400 and 4200 meters, while *C. pilosa* is found primarily at lower elevations, between 1200 and 3400 meters. Additionally, *C. nana* is often a smaller plant with decumbent branches and smaller features.

When *Castilleja pilosa* and *Castilleja nana* occur in sympatry and at the same elevation, it is often quite difficult to distinguish the two species. Additionally, many of the members of this complex occur across geographic and political boundaries and are represented in multiple, overlapping regional and floristic treatments (Cronquist et al. 1984; Hitchcock et al. 1984;

Wetherwax et al. 2012). Subsequently, there has been a great deal of taxonomic confusion, demonstrated by the number of synonyms associated with *C. pilosa* and *C. nana*. Several of these incorporations are centered in the Sierra Nevada where both species occur in sympatry, as well as northern California at the border with Oregon and Nevada. These regions also lie at the boundary between the Great Basin, the Pacific Northwest, and the Sierra Nevada and California floristic province, where a great deal of taxonomic work has been done. The taxonomic confusion in this group could be the result of any of the following factors: the young age of the lineage, the propensity for gene flow when species are sympatric, little to no morphological distinction between species, and/or the absence of species in the complex (i.e., the entire complex is actually a single lineage). As such, this complex is in great need for robust species delimitation.

Here we begin this process by quantifying morphological variation in the complex and assessing its correlation (or not) with the current taxonomy. By sampling many populations across the known ranges of these entities, identifying them using regional treatments, and measuring and analyzing a suite of morphological traits, we test the assumption that there are morphological clusters that correspond to taxonomic entities. We perform principal coordinate analyses to understand the position of individuals in morphospace, and then apply a non-hierarchical clustering method to assess the signal of morphological similarity that exists among these entities. In this way, we aim to quantify and begin to characterize the morphological variation in this species complex, information that will ultimately become part of a robust delimitation of species boundaries in this group.

Methods

Sampling and Range Estimation—Both mounted and unmounted collections of *Castilleja pilosa* var. *pilosa*, *C. pilosa* var. *longispica, C. pilosa* var. *steenensis*, and *C. nana* were examined for this study, with emphasis placed on representing the known distributional ranges of these taxa. Prior to measurement, all collections were identified using the primary literature currently available— regional floras and treatments ((Wetherwax et al. 2012) California; (Hitchcock et al. 1984), Oregon, Idaho, and adjacent Montana and Wyoming; (Cronquist et al. 1984), Great Basin). Species ranges were estimated based on loan material and specimen label data accessed through regional databases (Consortium of Pacific Northwest Herbaria (pnwherbaria.org); Southwest Environmental Information Network (SEINet; swbiodiversity.org); University and Jepson Herbaria Specimen Portal (webapps.cspace.berkeley.edu); New York Botanical Garden (NYBG; nybg.org); Rocky Mountain Herbarium (RM; rmh.uwyo.edu)). Latitude and longitude were taken directly from collection labels, when available. In some cases, coordinates were not provided on the collection label, in which case

they were estimated by hand based on locality information provided by the collector. For specimens whose identification we did not confirm (i.e., specimens not on loan), we only considered collections or identifications determined by collectors that we considered to have extensive expertise in *Castilleja* identification. All sampling information, including coordinates and voucher locations, can be found in Supplemental Table S1.

Morphological Measurements—We used a combination of continuous and categorical traits to characterize morphology. These traits are known to be taxonomically informative and are widely used to identify and distinguish *Castilleja* species (Cronquist et al. 1984; Hitchcock et al. 1984; Chuang and Heckard 1991; Hersch-Green and Cronn 2009) (Table 1). Specimens were chosen for data collection based on the overall condition of the collection and maturity of the plant when it was collected, preferring specimens as close to peak maturity as possible. Multiple stems within each collection were measured in order to record a complete set of measurements for each collection. Floral measurements were taken from dissected flowers rehydrated with Pohl's solution (Pohl 1965). Flowers at peak maturity were identified, removed from the indeterminate inflorescence, and saturated with Pohl's solution for five minutes. The bract, calyx, and corolla were separated from one another, and measurements taken from the dissected tissues (Fig. 2b). Habit, inflorescence, and leaf characters were taken from the specimen without further dissection; surface textures were taken from stem midway between the inflorescence and the base of the plant.

Nineteen continuous characters were measured from a total of 171 collections: *Castilleja nana* (n=50), *C. pilosa* var. *longispica* (n=34)*, C. pilosa* var. *pilosa* (n=76)*,* and *C. pilosa* var. *steenensis* (n=11). Several continuous variables were used in the auto-calculation of additional continuous variables, thus creating a composite variable (Table 1, characters 18, 27, and 28). To avoid pseudo-replication of traits in the dataset, we removed the component traits (Table 1, characters 16, 17, 20, 21, 22, and 27), leaving only the composite variables in the dataset, resulting in thirteen quantitative characters. Nine categorical characters were recorded from the same 171 collections (Table 1). Three of these characters did not vary across individuals and were removed from the dataset (Table 1, characters 24, 25, and 26), leaving a total of six qualitative characters included in the analyses.

Data Preparation and Quantification of Morphological Variation—When present, raw measurements from different stems of the same collection were combined to produce an average measurement for each individual for each trait examined. Individuals with missing data for any of the traits measured (indicating the tissue was unavailable for sampling, a total of 60 collections) were

removed from downstream analyses. We identified possible outliers in the dataset by calculating the multivariate normal density function of all continuous variables using the *stats* package in R (R Core Team 2016), resulting in the pruning of 10 collections. Continuous variables were log transformed, and presence/absence data were coded as binary variables. In order to quantify morphological measurements for each taxonomic entity, kernel density estimates for each continuous trait were generated with the density function in the R package *stats* (R core Team 2016) using a "gaussian" kernel and default bandwidth parameters. Categorical traits were summed across taxonomic groups.

Principal Coordinate Analysis—To represent the morphological similarity in our dataset, we applied a metric, multidimensional scaling approach that positions each individual in a reduced dimension morphospace, preserving the distance relationship between individuals as well as possible (Gower 1966). Because the categorical variables that we measured are taxonomically diagnostic, it was important to include them in a quantification of morpholospace in this species complex. We performed a principal coordinate analysis (PCoA), which can handle both quantitative and qualitative data by using measures of (dis)similarity calculated from mixed variables (Gower 1966; Legendre and Legendre 1998). We calculated a dissimilarity matrix based on our log-transformed continuous variables, our nominal categorical variables, and our symmetric dichotomous variables, using Gower's dissimilarity coefficient (Gower 1971), as implemented using the daisy function in the R package *cluster* (Maechler et al. 2016). We then performed PCoA on the dissimilarity matrix using the function pcoa in the R package *ape* (Paradis et al. 2004). PCoA can sometimes result in negative eigenvalues when dealing with non-Euclidean distance measures (as we are doing here). As such, we used the Cailliez correction (Cailliez 1983), where a constant is added to each original measure of dissimilarity (except the diagonals). Because PCoA is based on a pairwise distance matrix, there are approximately as many dimensions as there are pairwise comparisons, and they are ordered by their eigenvalues. By plotting each individual at the first two to three principal coordinates, we can represent the best possible Euclidean approximation of the morphological distance between them (Gower 1982).

Fuzzy Clustering—To explore and describe the signal of morphological similarity that we have quantified, we apply a clustering technique that can accommodate situations where cluster boundaries may not be clear-cut. Fuzzy clustering (Dunn 1976; Kaufman and Rousseeuw 2005) is a 'soft' approach to clustering where individuals are assigned a probability of membership (the coefficient of membership) to each recovered cluster; this is in contrast to 'hard' clustering where an individual is assigned to a single cluster only. The benefit of this type of clustering approach is that it

can accommodate ambiguity in cluster assignments and provide more detailed information about the structure of the dataset.

The objective of the fuzzy clustering algorithm is to minimize the within cluster variance and maximize between cluster variance; put another way, the objective is to minimize the distance between two objects belonging to the same cluster. This is accomplished through an iterative procedure where cluster membership is initiated and a coefficient of membership is calculated for each individual based on the distance of the individual to the centroid of each cluster. The process is repeated until new clustering iterations fail to maximize the objective. After clustering, a final coefficient of membership to each cluster is calculated for each individual. When an individual is assigned equal coefficients to all clusters, it is described as having 'complete fuzziness' and can be imagined as falling in the 'middle ground' between all clusters; when an individual has a membership close to 1 to a particular cluster, the clustering is essentially hard (i.e., it is a partition). Dunn's normalized partition coefficient (1976) can be used to describe the overall fuzziness of an analysis, regardless of the number of clusters considered, where values close to 0 indicate high levels of fuzziness (near equal membership to all clusters) and values close to 1 indicate very low levels of fuzziness (i.e., hard partitions). After generating the coefficients of membership, one can find the hard partitioning scheme that most closely approximates the fuzzy clustering by assigning each individual to the cluster in which it has the largest membership.

One way to visualize the results of fuzzy clustering is by examining silhouette plots of the hard clusters. These plots are constructed of horizontal bars representing the silhouette coefficient ($s(i)$ – a measure of that individual's similarity to other members of the same cluster) of each individual in the analysis, organized by hard cluster assignment. When $s(i)$ is at its largest for an individual (close to 1), that means that the individual is much more similar to other members of its cluster than it is to individuals outside of the cluster. When $s(i)$ is low for an individual (closer to 0), it means that the individual is equally similar to both members of its cluster and members of other clusters. When an individual has an $s(i)$ value that is negative, the within cluster similarity is much smaller than the between cluster similarity. Finally, we can calculate the mean silhouette coefficient (i.e., the mean silhouette coefficient of all samples in the analysis) as a way of interpreting and validating the clustering. Kaufman and Rousseeuw (2005) suggest that datasets with silhouette coefficients less than or equal to 0.25 have no substantial structure, values between 0.26 and 0.50 indicate weak structure that could be artificial and require additional methods to corroborate, values between 0.51 and 0.70 suggest reasonable structure, and values between 0.71 and 1.0 suggest strong structure has been found.

Fuzzy clustering analyses were run using the function fanny in the R package *cluster* (Maechler et al. 2016), and the same dissimilarity matrix for fuzzy clustering used for PCoA. Fuzzy clustering requires the user to define the number of clusters (k) to optimize. We chose to examine clustering of k = 4, 3, and 2 clusters. We begin at four because this corresponds with the number of named taxonomic entities focal to this study; three and two clusters were also examined to explore the morphological signature of the data. We further examined the effect of the membership exponent (a variable in the cluster optimization process) on our clustering results. It has been shown that higher values (near two) lead to greater fuzziness while lower values (near one) yield less fuzzy clustering (Kaufman and Rousseeuw 2005). We examined the effect of this variable on clustering results by adjusting its value between 1.1 and 1.7, by increments of 0.1. We ran all fuzzy clustering analysis for 100,000 iterations, to assure convergence.

Estimating Position of Type Specimen in Morphospace—To explore the position of type specimens in morphospace, we took the geographic position of each type specimen and found the nearest population of the same species from which we took morphological measurements. We make the assumption that these populations would have similar morphologies.

Results

Sampling—A total of 171 individuals were examined for this study. While normality is not a strict assumption of the approaches used here, extremely non-normal traits may affect results in unpredictable ways. As a conservative measure, we eliminated from downstream analyses approximately the top 10% of individuals that deviated extremely from the natural variability in the data. The impact of outlier removal on downstream analyses was examined and found to have minimal influence (results not shown). After data cleaning and outlier removal, our final dataset consisted of *Castilleja nana* (n=29), *C. pilosa* var. *longispica* (n=23), *C. pilosa* var. *pilosa* (n=52), and *C. pilosa* var. *steenensis* (n=4), and covered the known ranges of each focal taxon (Fig. 2, open circles). Individuals measured, the housing herbarium of each collection, and associated raw data are available on Dryad.

Quantifying Morphological Variation—Kernel density estimates of quantitative trait values grouped by taxonomic identity revealed a great deal of overlap in trait values for each taxon across many traits. In some cases, this overlap occurs across all focal taxa, as in bract width and leaf width (Fig. 3), where all taxa have widely overlapping trait distributions. In other cases, the distribution of trait

values distinguishes one of the focal taxa from the remaining three. For example, *C. pilosa* var. *steenensis* has a larger beak to tube ratio than the remaining taxa (meaning that the difference in length between the tube and the beak is greater); *C. nana* has a longer bract than all varieties of *C. pilosa*; most *C. pilosa* var. *longispica* have shorter calyces than other varieties of *C. pilosa* and *C. nana*. There are also cases of interspecific overlapping trait distributions, as in plant height where *C. nana* and *C. pilosa* var. *steenensis* are generally shorter in height than *C. pilosa* var. *longispica* and *C. pilosa* var *pilosa.* We see a similar pattern of overlap in traits across taxa in our qualitative data (Fig. 3). With the exception of the decumbent habit, no one qualitative trait is found primarily in one taxon, let alone exclusively (Fig. 4). In general, pubescence traits were equally variable across taxa, *C. nana* was the only taxon that occasionally lacked lobes on the leaves, and *C. pilosa* var. *pilosa* and *C. pilosa* var. *steenensis* were the only focal taxa that were never scored as having broader, deltoid shaped calyx lobes. Summary statistics for raw values of continuous traits and raw counts of categorical traits can be found in the supplemental data (Table S1 and S2, respectively).

Principal Coordinate Analysis—A Cailliez correction, equal to D' = -0.5 * (D + 0.57237) ^ 2, was applied to all negative eigenvalues. The position of each individual in the first two and three principal coordinates are shown in Fig. 5, with 95% confidence ellipses around the mean position of each focal taxon in morphospace. The first 10 principal coordinate axes are required to account for 50% of the corrected, relative eigenvalues. An examination of axes 4 through 10 does not change the interpretation of results presented here; the first two principal coordinate axes represent the maximum morphological distance among individuals sampled and the third axis reveals no further distinction (Fig. 5).

In general, and considering all three principal coordinate axes, individuals identified as *Castilleja nana* (yellow) occupy a different part of the scatterplot than those identified as *C. pilosa*, including its named varieties (blue (var. *pilosa*), orange (var. *longispica*), and red (var. *steenensis*)). Considering only those individuals identified as *Castilleja pilosa*, there is a large amount of overlap with no discernible position in morphospace unique to any variety (Fig. 5). Confidence ellipses lend support to this conclusion and further suggests a greater distinction of *C. pilosa* var. *steenensis* (in red) from any other focal taxon. The variation in distances of these individuals lies along a different axis than the rest of the focal taxa; however, the effect of sample size (n=4) cannot be discounted.

Fuzzy Clustering—We performed seven fuzzy clustering analyses (corresponding to seven different values of the membership exponent variable; values between 1.1 and 1.7, in increments of 0.1) for each of three possible numbers of clusters (k=4, 3, and 2). Different values of the membership

exponent produced consistent results within each "k=X number" of clusters. For simplicity, we present the results from all clustering scenarios with a membership exponent of 1.3.

Fuzzy clustering analyses, regardless of number of clusters considered, resulted in clusters with small silhouette coefficients (both within and across clusters), and low values for the normalized Dunn coefficient (Fig. 6, Table 2). As cluster number was reduced, there appeared to be some small improvement in these measures (average silhouette coefficient increased from 0.2 (k=4) to 0.22 (k=3), and to 0.25 (k=2) and normalized Dunn coefficient increased from 0.37 (k=4), to 0.38 (k=3), and 0.44 (k=2)); however, overall these values are extremely low.

A somewhat subjective approach to quantifying the structure in a dataset is to calculate the silhouette coefficient (SC) of the dataset (Kaufman and Rousseeuw 2005). This value is the maximum, average silhouette coefficient of all possible numbers of clusters, from k = 2 as a minimum, to k = n as a maximum (k = 108, in this study). At k=53, our standard 100,000 iterations of clustering were not enough to satisfy fuzzy clustering objectives, and we ran into convergence issues. However, considering k=2 through k=53 clusters, the average silhouette coefficients were highest at k=2 (average s(i) = 0.25), and steadily dropped as values of k increased.

To visualize the taxonomic composition of clusters, we painted the silhouettes with colors corresponding to the taxonomic identity of each individual. Across all three clustering schemes, one cluster is consistently composed of mostly *C. nana* individuals with the remaining clusters being variously composed of all three varieties of *C. pilosa*. When we restrict the cluster number to two, the *C. nana* cluster begins to be more heavily composed of *C. pilosa* individuals (Fig. 6).

Discussion

Classifications are useful when they organize objects based on relationships, when they reflect similarities and differences among the constituent parts, and when they aid in the identification and placement of unknowns within the classification (Sokal 1974, de Queiroz and Donoghue 2011, de Queiroz and Donoghue 2013). The type specimen, as a central point of the species description, plays an important role in the creation and implementation of classifications, but with a reliance on it comes the challenge of tracing and managing type collections and species descriptions through time—a problem that we are still dealing with (Hitchcock 1905; Dayrat 2005). In addition, when objects are discrete and discontinuous, classifications are easy to build and use; however, when there is continuous variation in characters used in the classification, this becomes more difficult.

In this study we have closely examined morphology—a commonly used character for describing taxonomic boundaries—for four named taxa, from across their ranges, in a species complex known to be taxonomically difficult to diagnose. Here we have quantified a great deal of overlap in character traits that are typically used to diagnose species in *Castilleja* (Fig. 3, 4). In some cases, these traits are continuous across taxonomic boundaries (Fig. 3), emphasizing the extreme morphological similarity among these named entities. This is what we observe within *C. pilosa*, for example, where we are essentially incapable of distinguishing taxonomic varieties using morphology alone (Fig. 5), even in *C. pilosa* var. *steenensis,* considered the most distinctive of the three varieties due to its isolation on Steens Mountain in SE Oregon (Hitchcock et al. 1984).

In other cases, the distinction between taxa is apparent despite the overall high levels of similarity, indicating some morphological distinction between *C. nana* and *C. pilosa* (Fig. 5). This is also supported by the results of fuzzy clustering analyses that, regardless of the number of clusters considered, recover a cluster composed primarily of *C. nana*, with *C. pilosa* individuals variously scattered among the remaining clusters (Fig. 6). Several continuous traits distinguish *C. nana* from *C. pilosa* (Fig.3; see also Supplemental Fig. S1), however, the overlapping tails of these distributions, and the nature of these distinguishing traits (i.e.—size and length traits that could be environmentally plastic), goes a long way towards explaining the morphological confusion that has plagued this complex historically.

It is clear that geographic and ecological characters must have played a dominant role in shaping the species descriptions in this complex. This is apparent from the species descriptions included both in regional and genus-wide treatments (Cronquist et al. 1984; Hitchcock et al. 1984; Wetherwax et al. 2012), as well as the inferred species ranges (Fig. 2). For example, *C. nana* does not occur in the northern limits of the *C. pilosa* range. So, if you encounter a relatively small individual in Idaho, there is no way to confuse it with *C. nana* (a California and Nevada species), as the ranges do not overlap and the regional treatment does not consider *C. nana* (Hitchcock et al. 1984). Similarly, *C. pilosa* var. *steenensis* only occurs on Steens Mountain in Eastern Oregon. If you found a relatively small individual in central Oregon, you could only classify it as *C. pilosa* var. *pilosa*, using these regional treatments.

When species occur sympatrically, however, the distinction between named entities becomes much more difficult to parse. In the Sierra Nevada, *C. pilosa* var. *pilosa* (a moderate elevation taxon) and *C. nana* (a high elevation taxon) can co-occur at the limits of their elevational ranges (high and low, respectively) where environments are heterogeneous. Similarly, *C. pilosa* var. *pilosa* and *C. pilosa* var. *steenensis* can co-occur on the western slopes of Steens Mountain in the transition area between the high, exposed ridge and the surrounding lower elevation steppe. In heterogeneous

habitats and at ecological boundaries, phenotypes can be accentuated and variable (Van Kleunen and Fischer 2005), potentially in response to local microhabitat conditions such as light availability and precipitation (Schlichting 1986; Dorn et al. 2000; van Kleunen et al. 2000; Nicotra et al. 2010). As a result, it is possible that in these areas of sympatry that correspond with environmental transitions, individuals could experience extreme conditions that may affect the morphological traits that we examine when we try to identify unknowns. We see this in several individuals from the Sierra Nevada that have extreme values in the traits that distinguish *C. nana* and *C. pilosa* (Fig. 7). Furthermore, these are the individuals that occur in the region of overlap in morphospace between these two taxa (Fig. 7).

In some cases, these regions of sympatry also correspond with hotspots of taxonomic synonymy historically—i.e., these sympatric areas are places where synonyms of currently accepted taxa were described (Fig. 8). For example, the area surrounding Lake Tahoe has seen the description of four distinct taxa (*Castilleja jusselii* Eastw. (Eastwood 1940)*, Orthocarpus pilosus* S. Wats. (Watson 1871)*, Castilleja inconspicua* A. Nels & Kennedy (Nelson and Kennedy 1906)*, Castilleja nana* Eastw. (Eastwood 1902)), two of which (*O. pilosus* and *C. nana*) are the type specimens for *Castilleja nana* and *Castilleja pilosa* (Fig. 8). The remaining two taxa were incorporated into *C. nana* (*C. inconspicua* and *C. pilosa* (*C. jusselii*), effectively meaning that these entities are no different from *C. nana* and *C. pilosa*. However, when we place our best approximation of *Castilleja inconspicua* in morphospace (i.e., a specimen of the same taxon (*C. inconspicua* is a synonym of *C. nana*) that was measured by us that is as geographically close to the type collection of *C. inconspicua* as possible), we find that this collection occupies a region of morphospace very different from that of the type collection of *C. nana* (Fig. 8). By including this species into the concept of *C. nana* through synonymization in the Intermountain Flora (Cronquist et al. 1984), the amount of variation attributed to *C. nana* likely expanded.

Areas of sympatry are not the only source of potential confusion in the taxonomic history of either taxon. For example, the synonymization of *Castilleja lapidicola* A.A. Heller (Heller 1912) in eastern Nevada with *C. nana* also expanded the region of morphospace attributed to *C. nana* ((Cronquist et al. 1984), Fig. 8). Similarly, in northern California the inclusion of *C. ochracea* Eastw. (Eastwood 1941) and *C. pisttacinus* increased the area of morphospace occupied by *C. pilosa* ((Cronquist et al. 1984); Fig. 8). Ultimately, the qualitative decisions made about species boundaries based on regional treatments has extended and inflated the morphological concepts of both taxa. By going through this procedure of quantifying morphological variation, we can visualize what morphological variation the taxonomy currently embodies. It is apparent that the morphological concept of both *C. nana* and *C. pilosa* have expanded through the incorporation of additional taxa as

synonyms, and it is possible that the type collections of both taxa may no longer be centralizing or anchoring the features of either taxon.

The inflation of morphological variation attributed to *C. nana* and *C. pilosa* during species level revisions, much of them regionally based, in addition to an apparent reliance on potentially plastic morphological characters to distinguish species in sympatry, has resulted in a great deal of morphological confusion in this complex. This likely contributes to the tumultuous taxonomic history of these taxa, and suggests that relying on morphology alone to define species boundaries in this complex is problematic. This is where molecular and ecological lines of evidence will be incredibly important to delimit species (e.g., Jacobs *et al* 2018 (Chapter 1), Jacobs *et al* 2018 *in prep* (Chapter 3)). In a robust and integrated delimitation of species, we may find that taxa that have been synonymized are not truly part of their corresponding taxa, or vice versa. Subsequent classifications should reflect these boundaries and highlight the similarities and differences between them.

Here we have begun that process by quantifying morphological variation in this species complex and we have estimated the position of type specimens in that space. The next steps in this group will be to gather molecular and ecological evidence to contribute to a robust species delimitation that is based on multiple lines of evidence. With all data in hand, we can more confidently apply names, whether that is applying an old name, a new name, or combining them all in one.
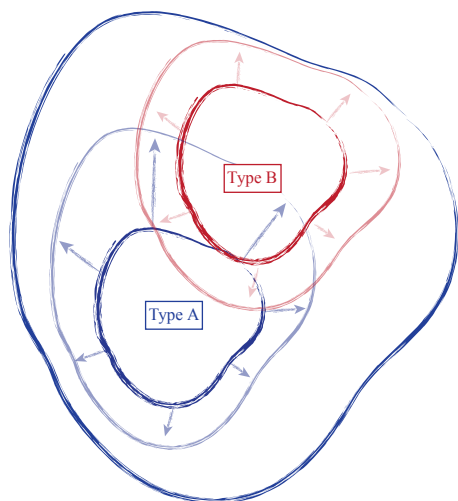
Species "A" and "B" described, anchored by type collections "A" and "B", and including variation around the type
(smaller, dark continuous lines)

Species descriptions of "A" and "B" are updated in regional treatments and floristic studies to incorporate variation observed in the field
(arrows and lighter, continuous lines)

Revisionary work recognizes overlap in variation in species "A" and "B" and synonymizes species "B" with species "A".
(dark, most inclusive continuous line)

FIGURE 2.1. Schematic representing the amount of variation attributed to a species through time.
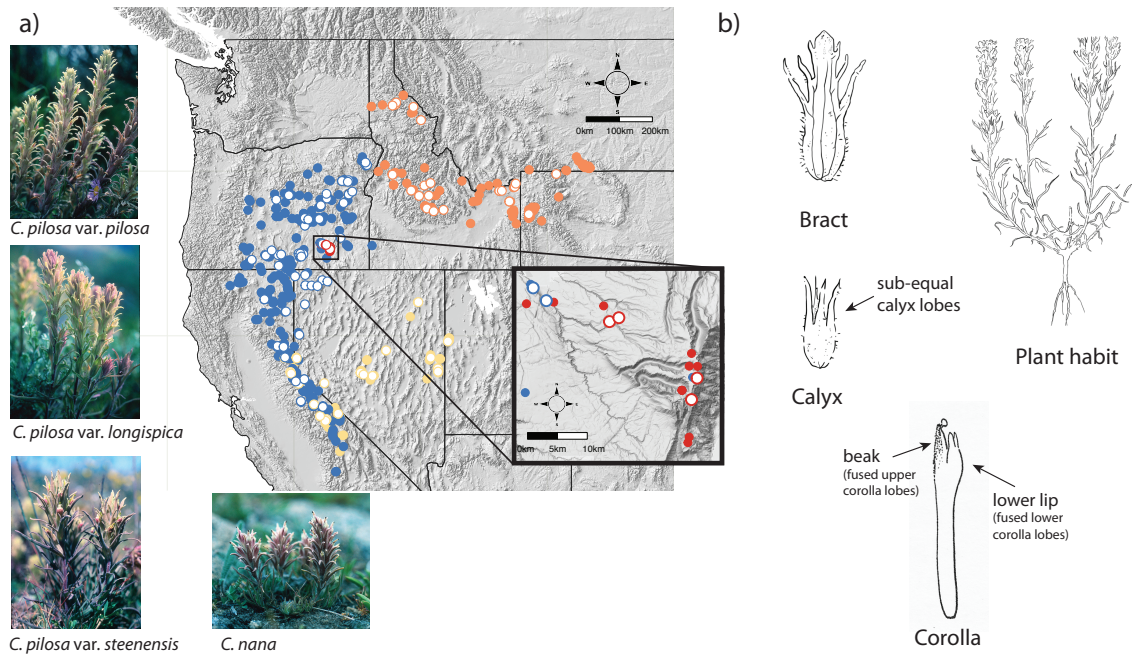
FIGURE 2.2. Distribution of focal taxa (a) and diagrams of species morphology (b). Filled circles represent specimens used to estimate ranges only (accessed through regional databases Consortium of Pacific Northwest Herbaria (pnwherbaria.org); Southwest Environmental Information Network (SEINet; swbiodiversity.org); University and Jepson Herbaria Specimen Portal (webapps.cspace.berkeley.edu); New York Botanical Garden (NYBG; nybg.org); Rocky Mountain Herbarium (RM; rmh.uwyo.edu)); open circles are individuals measured in this study. *Castilleja pilosa* var. *pilosa* (blue), *C. pilosa* var. *longispica* (orange), *C. pilosa* var. *steenensis* (red), *C. nana* (yellow). Photos by J.M. Egger.
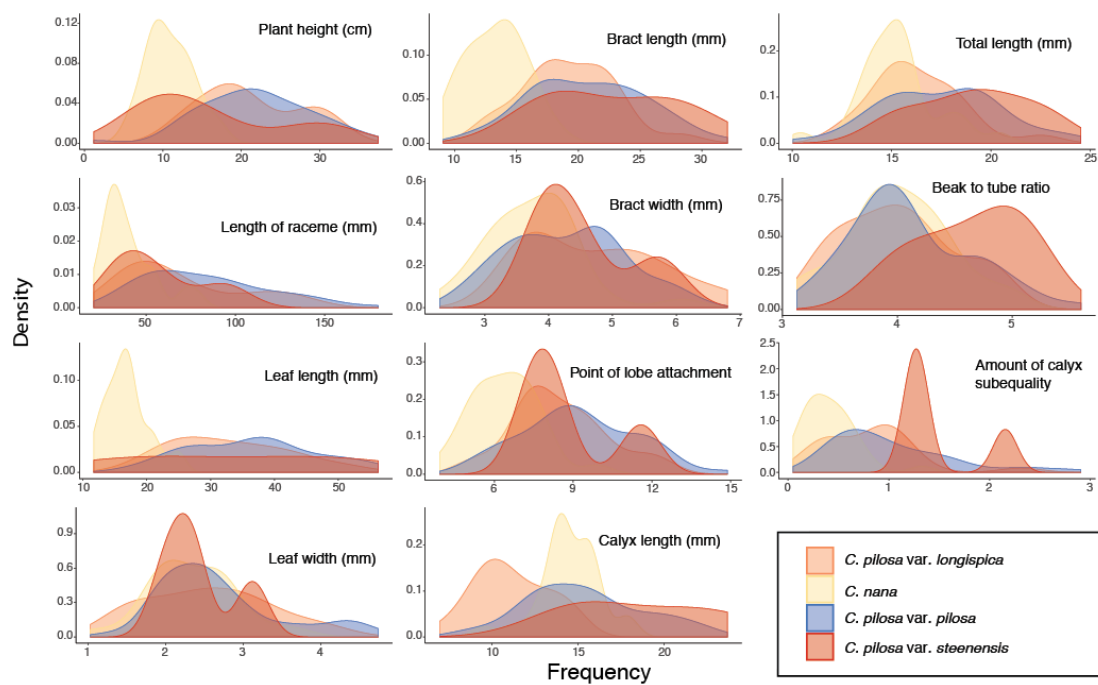
FIGURE 2.3. Kernel density estimates of raw trait values for the continuous traits measured in this study. *C. pilosa* var. *longispica* (orange), *C. pilosa* var. *pilosa* (blue), *C. pilosa* var. *steenensis* (red), and *C. nana* (yellow).
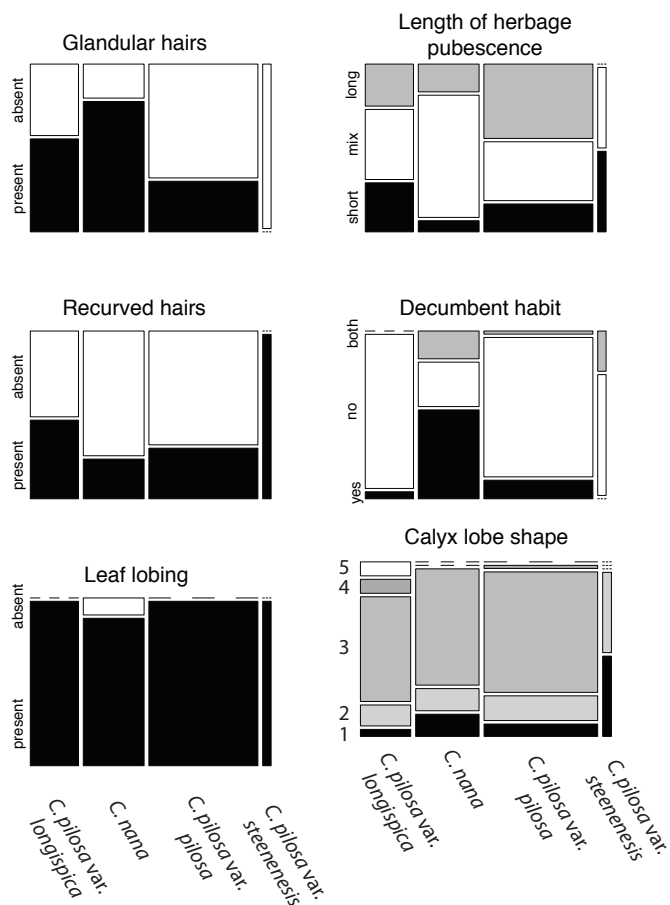
FIGURE 2.4. Summary of counts for categorical characters measured here. Columns represent focal taxa whose area represents all individuals identified to that taxon in our dataset. Shading represents different character states scored for each individual. Dashes represent a character state unobserved in a particular taxon. For calyx lobe shapes, numbers are used in place of trait descriptions for simplicity. These correspond to: 1) linear, 2) lanceolate/linear, 3) lanceolate, 4) deltoid/lanceolate, and 5) deltoid.
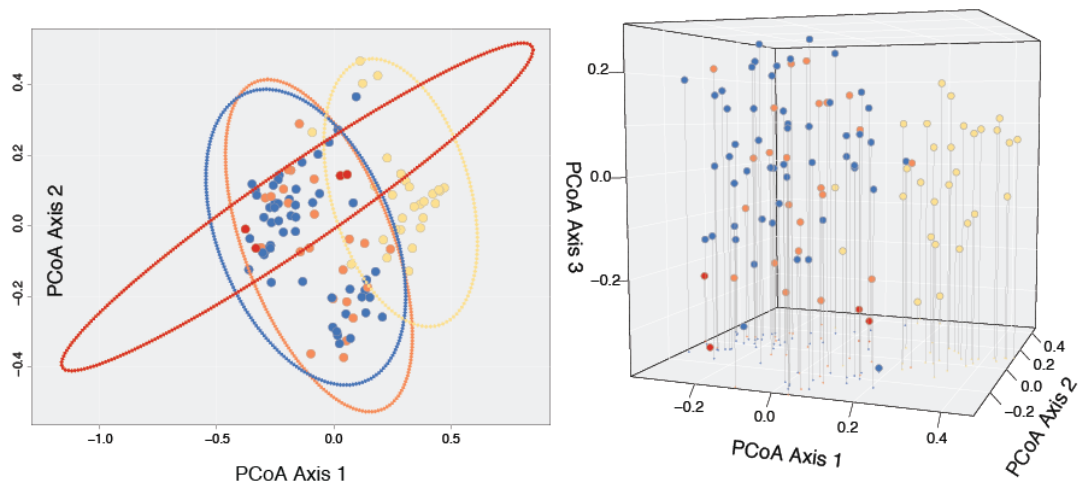
FIGURE 2.5. Results of Principle Coordinate Analysis (PCoA) considering the first two axes of variation (left), and including a third axis (right). Individuals are represented by points in morphospace, and colored according to species identification: *Castilleja pilosa* var. *pilosa* (blue), *C. pilosa* var. *longispica* (orange), *C. pilosa* var. *steenensis* (red), and *C. nana* (yellow).
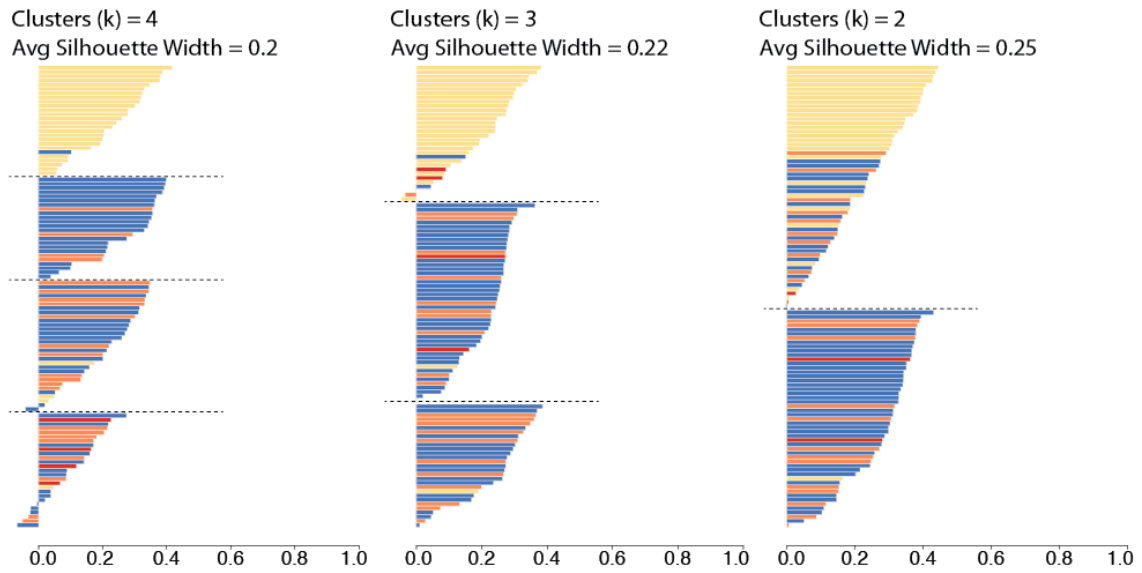
FIGURE 2.6. Results of fuzzy clustering for k=4 clusters (left), k=3 clusters (center), and k=2 clusters (right). For each set of silhouettes, the width of each bar corresponds to the silhouette coefficient for that individual in the analysis; average silhouette coefficient for each analysis (k=4, 3, 2) is reported. Bars are painted with colors corresponding to species identification: *Castilleja pilosa* var. *pilosa* (blue), *C. pilosa* var. *longispica* (orange), *C. pilosa* var. *steenensis* (red), and *C. nana* (yellow).
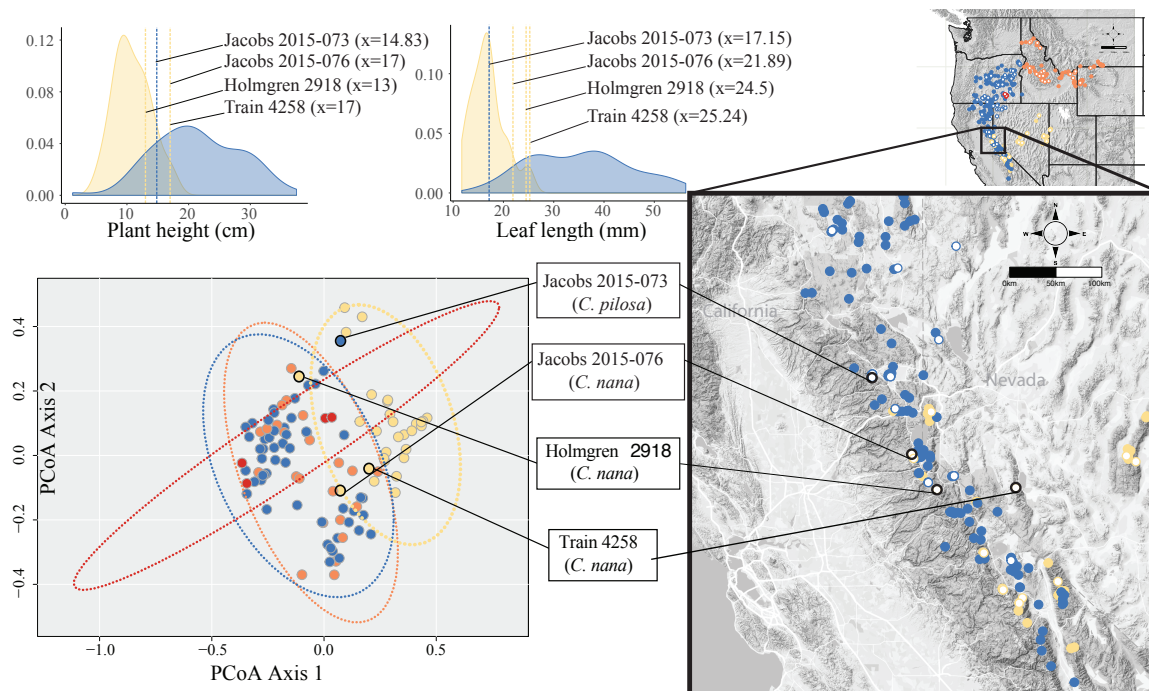
FIGURE 2.7. Position of individuals with extreme trait values in morphospace (left) and in geographic space (right). Individuals are color-coded according to taxonomic identification: Castilleja pilosa var. pilosa (blue), C. pilosa var. longispica (orange), C. pilosa var. steenensis (red), and C. nana (yellow). Histograms at the top of the diagram show trait distributions for C. nana (yellow) and C. pilosa (including all varieties, for simplification; blue). Vertical lines represent raw trait values and are color-coded corresponding to taxonomic identification.
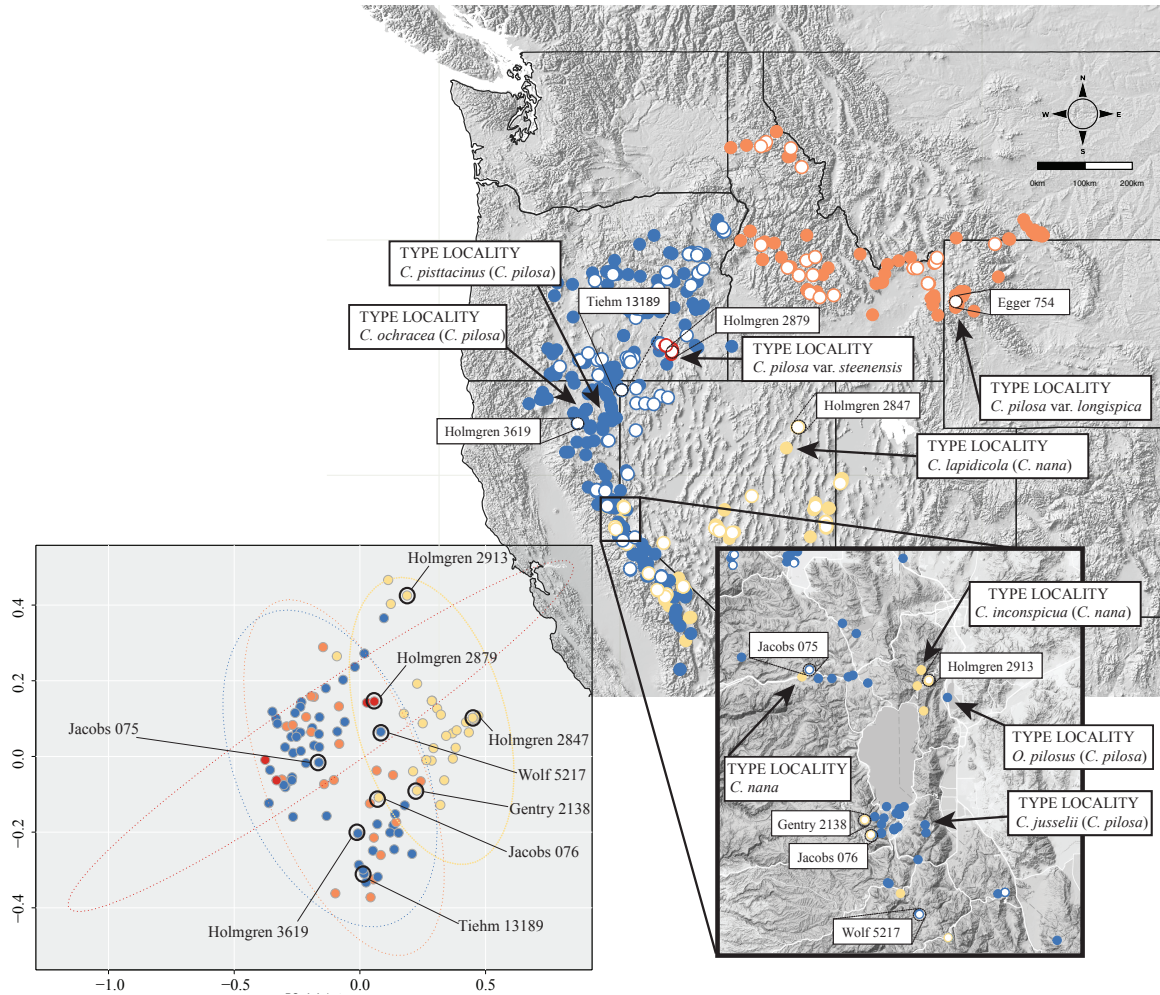
FIGURE 2.8. Position of type collections of focal taxa and associated synonyms, within the known ranges of each taxon (right) and the corresponding position of the nearest geographic individual that we have measurements for in our dataset is identified in morphospace (left). Individuals are color-coded according to taxonomic identification: *Castilleja pilosa* var. *pilosa* (blue), *C. pilosa* var. *longispica* (orange), *C. pilosa* var. *steenensis* (red), and *C. nana* (yellow).

TABLE 2.1. Morphological characters measured in *Castilleja*. The first column following the character name reflects the type of character measured: continuous (C), nominal (N), or dichotomous (D); the second column provides the unit of measurement, the number of levels for nominal or ordinal data, and (when necessary) the formula for character calculation. Asterisks (*) indicate characters were not directly included in analyses, but used to calculate composite variables.

| Character	Data type	
Habit:		
1 Plant height	C	cm
2 Decumbent at base	N	3
Surface textures:		
3 Length of herbage pubescence	N	3
4 Recurved hairs present	D	2
5 Glandular hairs present	D	2
Inflorescence:		
6 Number of racemes per stem	C	8
7 Length of raceme	C	mm
Leaf:		
8 Length of leaf	C	mm
9 Width of leaf	C	mm
10 Leaf lobing	D	2
Bract:		
11 Length of bract	C	mm
12 Width of bract	C	mm
13 Number of secondary lobe pairs	C	4
14 Point of lobe attachment	C	mm
Calyx:		
15 Length of calyx	C	mm
16 Tip of calyx to sinus 1	C	mm*
17 Tip of calyx to sinus 2	C	mm*
18 Calyx lobe subequality	C	mm; | #16 - #17 |
19 Shape of tip of calyx segments	N	5
Corolla:		
20 Total length	C	mm
21 Teeth to bottom of corolla	C	mm*
22 Sinus of beak and lower lip to bottom	C	mm*
23 Tube length	C	mm*
24 Lower lip pouchy	D	2*
25 Teeth petaloid	D	2*
26 Stigmas exserted	D	2*
27 Length of beak	C	mm; #20 - #22*
28 Beak length to tube length ratio	C	ratio; #20 / #27

TABLE 2.2. Results of fuzzy clustering analyses with k = 4, 3, and 2 clusters. Here we report average silhouette coefficients within and across clusters in analyses, as well as normalized Dunn coefficients for each analysis. Silhouette coefficients close to 0 represent less similarity, those close to 1 represent high similarity, and negative silhouette coefficients indicate likely misassignment to a cluster. The normalized Dunn coefficient is a measure of the overall fuzziness of an analysis. Values close to 0 indicate high levels of fuzziness (near equal membership of individuals to all clusters) and values close to 1 indicate very low levels of fuzziness (i.e., hard partitions).

	k = 4			*k = 3*			*k = 2*		
	n	Avg s(i)	stdev s(i)	n	Avg s(i)	stdev s(i)	n	Avg s(i)	stdev s(i)
Cluster 1	26	0.24	0.11	32	0.2	0.11	57	0.23	0.13
Cluster 2	24	0.28	0.11	47	0.21	0.08	51	0.27	0.1
Cluster 3	31	0.21	0.11	29	0.24	0.11			
Cluster 4	27	0.1	0.09						
Avg s(i) across analysis		0.2			0.22			0.25	
normalized Dunn Coefficient		0.3768			0.3879			0.4424	

Table XX. Results of fuzzy clustering analyses with k = 4, 3 and 2 clusters. Here we report average silhouet and across clusters in analyses, as well as normalized Dunn coefficients for each analysis. Silhouette width for each individual in the analysis, can range between -1 and 1, and are a measure of the similarity of the other members of its cluster as well as the closest individual of a different cluster. Silhouette widths clos less similarity and indicate individuals who are equally similar to members of their cluster and a non-clust those close to 1 represent high similarity, and negative silhouette widths indicate likely misassignment to normalized Dunn coefficient is a measure of the overall fuzziness of an analysis. Values close to 0 indicate fuzziness (near equal membership of individuals to all clusters) and values close to 1 indicate very low lev (i.e. hard partitions).

REFERENCES

Barley, A. J., White J., Diesmos A.C., and R.M. Brown. 2013. The challenge of species delimitation at the extremes: diversification without morphological change in Philippine sun skinks. *Evolution*. 67: 3556–3572.

Baum, D. A., and K. L. Shaw. 1995. Genealogical perspectives on the species problem. *Experimental and molecular approaches to plant biosystematics*. 53: 289–303.

Cailliez, F. 1983. The analytical solution of the additive constant problem. *Psychometrika*. 48: 305–308.

Carstens, B. C., Pelletier T.A., Reid, N.M. and J. D. Satler. 2013. How to fail at species delimitation. *Molecular Ecology*. 22: 4369–4383.

Chuang, T. I., and L. R. Heckard. 1991. Generic realignment and synopsis of subtribe Castillejinae (Scrophulariaceae-tribe Pediculareae). *Systematic Botany*. 644–666.

Cronquist, A. J., Holmgren, A.H., Holmgren, N.H. and J. L. Reveal. 1984. Castilleja. Pp. 476-496 in *Vascular Plants of the Intermountain West, U.S.A.* eds. A. J. Cronquist, A. H. Holmgren, N. H. Holmgren, J. L. Reveal, & P. K. Holmgren. Hafner Pub. Co., New York.

Dayrat, B. 2005. Towards integrative taxonomy. *Biological Journal of the Linnean Society*. 85: 407–415.

de Queiroz, K. 2007. Species concepts and species delimitation. *Systematic Biology*. 56: 879–886.

de Queiroz, K. 2009. The General Lineage Concept of Species and the Defining Properties of the Species Category. Pp. 49-89. in *Species: New interdisciplinary essays*. ed. R.A. Wilson. Cambridge, MIT Press.

de Queiroz, K. and M. Donoghue. 2011. Phylogenetic Nomenclature, Three-Taxon Statements, and Unnecessary Name Changes. *Systematic Biology*. 60 : 887-892.

de Queiroz, K. and M. Donoghue. 2013. Phylogenetic Nomenclature, Hierarchical Information, and Testability. *Systematic Biology*. 62: 167-174.

Dejaco, T., Gassner, M., Arthofer, W., Schlick-Steiner, B.C. and F. M. Steiner. 2016. Taxonomist's Nightmare … Evolutionist's Delight : An Integrative Approach Resolves Species Limits in Jumping Bristletails Despite Widespread Hybridization and Parthenogenesis. *Systematic Biology*. 65: 947–974.

Dorn, L. A., Pyle, E.H. and J. Schmitt. 2000. Plasticity to light cues and resources in Arabidopsis thaliana: Testing for adaptive value and costs. *Evolution*. 54: 1982–1994.

Dunn, J. C. 1976. Indices of partition fuzziness and the detection of clusters in large data sets. in *Fuzzy Automata and Decision Processes*. ed. M. Gupta. New York: Elsevier.

Eastwood, A. 1902. *Castilleja nana*. Proceedings of the California Academy of Sciences, Series 3. 2: 289-290.

Eastwood, A. 1940. *Castilleja jussellii.* Leaflets of Western Botany. 2: 243-244.

Eastwood, A. 1941. *Castilleja ochracea*. Leaflets of Western Botany. 3: 91.

Ence, D. D., and B. C. Carstens. 2010. SpedeSTEM: a rapid and accurate method for species delimitation. *Molecular Ecology Resources*. 11: 473–480.

Freudenstein, J. V., Broe, M.B., Folk, R.A. and B. T. Sinn. 2016. Biodiversity and the Species Concept – Lineages are not Enough. *Systematic Biology*. 66: 644-656.

Giarla, T. C., Voss, R.S. and S. A. Jansa. 2014. *Molecular Phylogenetics and Evolution*. 70: 137–151.

Gower, J. C. 1966. Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis. *Biometrika.* 53: 325.

Gower, J. C. 1971. A General Coefficient of Similarity and Some of Its Properties. *Biometrics*. 27: 857–871.

Gower, J. C. 1982. Euclidean distance geometry. *Math. Scientist*. 7: 1–14.

Heckard, L. R., and T. I. Chuang. 1977. Chromosome-Numbers, Polyploidy, and Hybridization in Castilleja (Scrophulariaceae) of Great Basin and Rocky Mountains. *Brittonia*. 29: 159–172.

Heller, A. 1912. *Castilleja lapidicola*. Muhlenbergia; a journal of botany. 8:49-50.

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

Hitchcock, A. S. 1905. Nomenclatorial type specimens of plant species. *Science.* 21: 0828–0832.

Hitchcock, C. H., Cronquist, A.J., Ownbey, F.M., and J. W. Thompson. 1984. Castilleja. Pp. 295-326 in *Vascular Plants of the Pacific Northwest*. Seattle: University of Washington Press.

Kaufman, L., and P. J. Rousseeuw. 2005. *Finding Groups in Data*. John Wiley & Sons, Inc.

Legendre, P., and L. Legendre. 1998. *Numerical ecology: second English edition.* Elsevier.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. and K. Hornik. 2016. cluster: Cluster Analysis Basics and Extensions.

Nelson, A. and P. B. Kennedy. 1906. *Castilleja inconspicua.* Proceedings of the Biological Society of Washington. 19: 38.

Nicotra, A. B., Atkin, O.K., Bonser, S.P., Davidson, A.M., Finnegan, E. J., Mathesius,U. P. Poot, et al. 2010. Plant phenotypic plasticity in a changing climate. *Trends in plant science.* 15: 684–692.

Padial, J. M., A. Miralles, I. De la Riva, and M. Vences. 2010. The integrative future of taxonomy. *Frontiers in Zoology.* 7: 16.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics.* 20: 289–290.

Pohl, R. W. 1965. Dissecting equipment and materials for the study of minute plant structures. *Rhodora*. 67: 95-96.

R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: https://www.R-project.org/.

Schlichting, C. D. 1986. The Evolution of Phenotypic Plasticity in Plants. *Annual Review of Ecology and Systematics.* 17: 667–693.

Schlick-Steiner, B. C., F. M. Steiner, B. Seifert, C. Stauffer, E. Christian, and R. H. Crozier. 2010. Integrative Taxonomy: A Multisource Approach to Exploring Biodiversity. *Annual Review of Entomology.* 55: 421–438.

Sneath, P. H. A., and R. R. Sokal. 1973. *Numerical Taxonomy: The Principles and Practice of Numerical Classification*.

Sokal, R. R. 1974. Classification - Purposes, Principles, Progress, Prospects. *Science.* 185: 1115–1123.

Tank, D. C., and R. G. Olmstead. 2008. From annuals to perennials: phylogeny of subtribe Castillejinae (Orobanchaceae). *American Journal of Botany.* 95: 608-625.

Tank, D. C., J. M. Egger, and R. G. Olmstead. 2009. Phylogenetic Classification of Subtribe Castillejinae (Orobanchaceae). *Systematic Botany.* 34: 182–197.

Van Kleunen, M., and M. Fischer. 2005. Constraints on the evolution of adaptive phenotypic plasticity in plants. *New Phytologist.* 166: 49–60.

van Kleunen, M., M. Fischer, and B. Schmid. 2000. Clonal integration in Ranunculus reptans: by-product or adaptation? *Journal of Evolutionary Biology.* 13: 237–248.

Van Valen, L. 1976. Ecological Species, Multispecies, and Oaks. *Taxon.* 25: 233–239.

Watson, S. 1871. *Orthocarpus pilosus.* United States Geological Exploration of the Fortieth Parallel. Vol. 5, Botany.

Wetherwax, M., T. I. Chuang, and L. R. Heckard. 2012. Castilleja. Jepson Flora Project. Available from: http://www.ucjeps.berkeley.edu/eflora.

SUPPLEMENTAL DATA

SUPPLEMENTAL TABLE S2.1. Sampling information of individuals examined in this study, including where specimen is housed, the corresponding accession number, collector and collection number, and georeferenced coordinates.

Herbarium	Accession Number	Collector/Collection number	Latitude	Longitude
ID	119927	Atwood 20851	39.55826	-116.361459
WTU	278959	Bafus 222	44.724404	-117.840999
WTU	21167	Bradley 156	44.170637	-118.700144
WTU	380570	Brainerd 1402	42.973617	-118.158967
WTU	74153	Brunsfeld 1525	44.562468	-114.850573
ID	73416	Brunsfeld 1626	44.638249	-114.610082
WTU	174223	Chisaki 895	39.662819	-120.411063
ID	99127	Cholewa 7244	47.08291	-115.966847
WTU	335479	Colwell JM213B	42.4942	-119.7473
WTU	258484	Cronquist 11001	38.87834	-117.37347
IDS	1999.1.252	Cronquist 1214	44.402091	-111.792505
IDS	1999.1.249	Cronquist 1593	44.402136	-111.893862
ID	30654	Cronquist 2556	43.866995	-114.751467
WTU	152443	Cronquist 7067	43.959874	-118.938616
WTU	162435	Cronquist 7817	44.3798	-117.6981
WTU	209154	Cronquist 8218	42.438501	-121.10985
IDS	61166	Davis 148 32	44.078803	-111.496346
IDS	1999.1.253	Davis 216	44.621132	-111.243147
IDS	1999.1.250	Davis 556	43.82491	-114.096432
WTU	335727	Denton ALD 01	42.4797	-119.6291
WTU	268085	Denton 3886	39.688241	-120.649019
WTU	348839	Egger 1225	45.1845	-117.1164
WTU	348837	Egger 1232	45.258597	-117.17627
WTU	384656	Egger 1495	44.31238	-118.716611

WTU	331634	Egger 754	43.693363	-110.728604
WTU	331633	Egger 755	44.930027	-109.72614
WTU	335531	Egger 959a	44.676926	-117.87124
WTU	335540	Egger 969a	44.699166	-118.101544
WTU	335539	Egger 969b	44.699166	-118.101544
WTU	335500	Egger 980d	43.782544	-114.484922
WTU	335508	Egger 984a	44.254663	-114.673863
OSC	320832	Ertter 5728	43.3455	-119.5668
ID	101339	Ertter 8808	44.8974	-116.0962
WTU	103860	Ferris 11084	37.425594	-118.752799
ID	120919	Fox 689	46.992255	-116.108595
ID	54448	Gentry 2138	38.896474	-120.136851
ID	in curation	Gilman 2015 036	39.01257	-114.319
ID	in curation	Gilman 2015 038	39.00898	-114.32142
WTU	386085	Gross 717	41.688519	-118.877582
WTU	286756	Halse 1831	43.532117	-119.433468
WTU	56553	Hitchcock 5613	38.820664	-117.251792
WTU	176670	Hitchcock s.n.	43.543687	-119.551075
WTU	160066	Holmgren 10760	38.943519	-114.295626
WTU	230347	Holmgren 1233	41.651488	-118.677257
WTU	230595	Holmgren 1447	38.78018	-116.89771
WTU	230630	Holmgren 1569	38.9881	-114.3151
WTU	230633	Holmgren 1652	38.91183	-114.309944
WTU	230640	Holmgren 2220	38.887057	-114.300574
WTU	230343	Holmgren 2235	39.35013	-114.6038
WTU	233890	Holmgren 2812	38.98767	-114.31367
WTU	233889	Holmgren 2847	41.021822	-115.08902
WTU	233957	Holmgren 2879	42.6324	-118.5768
WTU	233874	Holmgren 2913	39.313734	-119.888309
WTU	233866	Holmgren 2918	38.544999	-119.812862
WTU	233867	Holmgren 2924	37.91705	-119.20943
WTU	233950	Holmgren 2926	37.841297	-118.859258

WTU	233951	Holmgren 2936	37.471739	-118.712628
ID	56238	Holmgren 3619	41.109412	-121.164038
WTU	94776	Holmgren 3625	39.82831	-113.91981
WTU	275648	Holmgren 4119	41.510036	-119.070733
WTU	256063	Holmgren 5347	44.40222	-111.89306
WTU	285160	Holmgren 8861	44.0328	-118.04
WTU	305897	Holmgren 9484	40.73795	-120.317349
WTU	160729	Holmgren 9618	43.547328	-119.51255
ID	349961	Ionta 00 14	44.266667	-120.2
ID	*in curation*	Jacobs 2015 072	39.68792	-120.64284
ID	*in curation*	Jacobs 2015 073	39.65645	-120.64812
ID	*in curation*	Jacobs 2015 075	39.34468	-120.35135
ID	*in curation*	Jacobs 2015 076	38.852557	-120.113117
ID	*in curation*	Jacobs 2015 084	38.68021	-119.59237
ID	*in curation*	Jacobs 2015 093	37.564337	-118.969637
ID	*in curation*	Jacobs 2015 096	37.551945	-118.961407
ID	*in curation*	Jacobs 2015 102	37.419073	-118.755334
ID	*in curation*	Jacobs 2015 106	42.666218	-118.565355
OSC	150400	Johnson 850289	44.892117	-116.10818
WTU	143805	Jones 295	46.5483	-114.9841
ID	128680	Kemper 143	47.018594	-115.998474
WTU	367767	Knoke 472	42.59365	-120.8771
WTU	365965	Knoke 577	42.452567	-120.625933
WTU	368420	Knoke 611	42.468317	-120.505333
WTU	189100	Kruckeburg 4167	44.411086	-115.372201
WTU	228033	Maguire 21128	39.004811	-114.3195
WTU	168646	Maguire 25800	38.77045	-116.93213
WTU	111723	Maguire 26487	43.575814	-119.57339
ID	104936	Moseley 591	46.949266	-115.282459
WTU	315673	Olmstead 717	42.788085	-118.873845
WTU	315928	Olmstead 752	42.4069	-119.7079
WTU	335334	Olmstead RGO96 54	42.5	-119.816667
WTU	177720	Ownbey 3412	44.91425	-109.64321

WTU	283942	Packard 78 89	44.2934	-117.8399
WTU	9558	Peck 18941	43.545277	-118.997542
ID	51128	Peck 21406	42.808191	-118.900715
WTU	155020	Pennell 22950	39.336482	-114.600028
WTU	297774	Reveal 2428	42.761243	-118.725451
WTU	297700	Reveal 2450	42.1268	-120.5973
WTU	360722	Rodman 681	45.26246	-117.17709
WTU	276444	Rogers 1006	41.845608	-119.555372
IDS	1999.1.462	Tiehm 10589	40.95485	-119.56708
ID	123943	Tiehm 13189	41.8071	-119.9558
WTU	352396	Tiehm 13885	40.029279	-119.789141
ID	85852	Tiehm 8790	41.532305	-119.549094
IDS	1999.1.461	Tiehm 9598	41.52593	-119.30607
WTU	133925	Train 4258	38.562339	-118.798346
WTU	140726	Vollmer 229	37.633927	-118.255726
IDS	2012.012	Whitehead 487	44.52329	-111.287024
WTU	398549	Wilson s.n.	44.133539	-120.69683
WTU	109153	Wolf 5217	38.614361	-119.924651
WTU	370131	Zika 11219	42.7551	-118.7445

SUPPLEMENTAL TABLE S2.2. Mean and standard deviation of raw, continuous trait values for measured individuals, organized by taxon.

	C. nana (n=29)	*C. pilosa* (all varieties, n=79)		
		var. *longispica* (n=23)	var. *pilosa* (n=52)	var. *steenensis* (n=4)
Habit				
Plant height (cm)	10.87 (2.95)	21.37 (6.84)	21.27 (6.32)	21.82 (7.02)	16.13 (9.80)
Inflorescence				
Number racemes per stem	1.10 (0.41)	2.32 (1.70)	2.00 (1.28)	2.52 (1.89)	1.50 (1.00)
Length of raceme (mm)	40.29 (15.08)	78.15 (33.39)	69.96 (33.34)	83.40 (34.80)	56.98 (26.23)
Leaf				
Leaf length (mm)	16.95 (3.39)	35.27 (9.67)	32.82 (9.08)	36.50 (9.83)	33.39 (17.19)
Leaf width (mm)	2.35 (0.56)	2.63 (0.77)	2.55 (0.79)	2.68 (0.79)	2.43 (0.49)
Bract				
Bract length (mm)	13.33 (2.41)	20.40 (4.47)	19.22 (3.74)	20.75 (4.69)	22.68 (5.39)
Bract width (mm)	3.75 (0.74)	4.41 (0.93)	4.70 (1.03)	4.27 (0.90)	4.55 (0.83)
Number secondary lobe pairs	1.21 (0.41)	2.11 (0.71)	2.52 (0.73)	1.98 (0.70)	1.50 (0.58)
Point of lobe attachment (mm)	6.40 (1.29)	8.97 (2.03)	8.47 (1.78)	9.21 (2.16)	8.77 (1.93)
Calyx				
Calyx length (mm)	14.47 (1.77)	14.29 (3.79)	11.14 (2.16)	15.36 (3.36)	18.60 (4.23)
Calyx lobe subequality (mm)	0.42 (0.27)	0.98 (0.57)	0.81 (0.47)	1.02 (0.61)	1.49 (0.44)
Corolla				
Total corolla length (mm)	15.27 (1.97)	17.20 (2.81)	16.28 (2.28)	17.46 (3.02)	19.14 (3.06)
Beak to tube ratio	4.01 (0.46)	4.11 (0.52)	4.01 (0.49)	4.11 (0.53)	4.64 (0.50)

SUPPLEMENTAL TABLE S2.3. Raw counts of categorical traits scored for each individual, organized by taxon.

		C. nana	C. pilosa (all varieties)			
				var. *pilosa*	var. *longispica*	var. *steenensis*
Habit						
Decumbent base:	both	5	2	1	0	1
	no	8	70	45	22	3
	yes	16	7	6	1	0
Surface textures						
Recurved hair:	absent	22	48	36	12	0
	present	7	31	16	11	4
Glandular hair:	absent	6	50	36	10	4
	present	23	29	16	13	0
Leaf						
Leaf lobing:	absent	3	0	0	0	0
	present	26	79	52	23	4
Calyx						
Calyx lobe shape:	deltoid	0	2	0	2	0
	deltoid/lanceolate	0	3	1	2	0
	lanceolate	21	54	39	15	0
	lanceolate/linear	4	13	8	3	2
	linear	4	7	4	1	2
Corolla						
Lower lip pouchy:	absent	0	0	0	0	0
	present	29	79	52	23	4
Teeth petaloid:	absent	0	0	0	0	0
	present	29	79	52	23	4
Stigmas exserted:	absent	0	0	0	0	0
	present	29	79	52	23	4

SUPPLEMENTAL TABLE S2.4. Raw results of fuzzy clustering analyses for k=4, 3, and 2 clusters. For each collection, and for each clustering analysis, the cluster assignment and silhouette coefficient are reported. Additionally, for each cluster in each analysis, a membership coefficient is reported for each individual. Finally, the taxonomic identification of each individual is provided.

| Collector/number | Taxon ID | k=4 Clusters | | | | | | | k=3 Clusters | | | | | | k=2 Clusters | | | |
| | | Cluster assign. | Silhouette width | membership coefficients | | | | | | Cluster assign. | Silhouette width | membership coefficients | | | Cluster assign. | Silhouette width | membership coefficients | |
				Clust. 1	Clust. 2	Clust. 3	Clust. 4					Clust. 1	Clust. 2	Clust. 3			Clust. 1	Clust. 2
Atwood 20851	nana	1	0.2523	0.86	0.02	0.1	0.02		1	0.2429	0.88	0.02	0.1	1	0.4081	0.95	0.05
Bafus 222	pilosa	2	0.3878	0.01	0.88	0.03	0.08		2	0.2825	0.03	0.86	0.11	2	0.3369	0.07	0.93
Bradley 156	pilosa	2	0.3336	0.05	0.74	0.06	0.14		2	0.2734	0.11	0.74	0.16	2	0.2593	0.18	0.82
Brainerd 1402	pilosa	3	0.0551	0.34	0.07	0.39	0.2		3	0.047	0.43	0.11	0.46	1	0.2375	0.79	0.21
Brunsfeld 1525	longispica	3	0.3348	0.08	0.08	0.76	0.09		3	0.3449	0.12	0.1	0.78	1	0.1481	0.68	0.32
Brunsfeld 1626	longispica	4	-0.0371	0.06	0.33	0.26	0.34		2	0.0022	0.09	0.5	0.41	2	0.1687	0.25	0.75
Chisaki 895	pilosa	3	0.2603	0.11	0.12	0.68	0.1		3	0.2993	0.16	0.14	0.71	1	0.1368	0.67	0.33
Cholewa 7244	longispica	3	0.0675	0.38	0.08	0.46	0.08		1	-0.0297	0.48	0.08	0.44	1	0.2916	0.87	0.13
Colwell JM213B	pilosa	3	0.3091	0.09	0.04	0.84	0.04		3	0.268	0.16	0.05	0.79	1	0.2701	0.88	0.12
Cronquist 11001	nana	1	0.3162	0.87	0.03	0.07	0.03		1	0.275	0.88	0.03	0.08	1	0.3257	0.91	0.09
Cronquist 1214	longispica	4	0.0705	0.07	0.14	0.26	0.53		3	0.031	0.13	0.35	0.52	2	0.0983	0.35	0.65
Cronquist 1593	longispica	2	0.2093	0.03	0.58	0.08	0.31		2	0.2346	0.05	0.78	0.17	2	0.3108	0.11	0.89
Cronquist 2556	longispica	4	0.2019	0.13	0.16	0.15	0.56		2	0.0996	0.25	0.43	0.32	2	0.1129	0.37	0.63
Cronquist 7067	pilosa	4	0.1323	0.1	0.19	0.18	0.53		2	0.0764	0.18	0.43	0.38	2	0.1059	0.34	0.66
Cronquist 7817	pilosa	2	0.0928	0.01	0.55	0.04	0.4		2	0.3671	0.02	0.91	0.07	2	0.4344	0.04	0.96
Cronquist 8218	pilosa	2	0.2206	0.03	0.6	0.11	0.25		2	0.2021	0.06	0.71	0.23	2	0.3038	0.13	0.87
Davis 148 32	longispica	3	0.2294	0.05	0.06	0.73	0.16		3	0.3201	0.07	0.1	0.83	1	0.0661	0.57	0.43
Davis 216	longispica	4	0.2239	0.04	0.16	0.09	0.72		2	0.1957	0.1	0.62	0.28	2	0.2658	0.18	0.82
Davis 556	longispica	3	0.3071	0.05	0.06	0.84	0.05		3	0.3543	0.08	0.08	0.84	1	0.1396	0.69	0.31
Denton ALD 01	pilosa	4	-0.0373	0.02	0.41	0.05	0.52		2	0.3222	0.03	0.86	0.1	2	0.3634	0.07	0.93

Specimen	Taxon															
Denton 3886	pilosa	2	0.0584	0.26	0.37	0.13	0.25	1	0.0467	0.41	0.38	0.21	2	0.0481	0.46	0.54
Egger 1225	pilosa	2	0.2052	0.06	0.51	0.17	0.26	2	0.1429	0.12	0.57	0.31	2	0.2133	0.24	0.76
Egger 1232	pilosa	4	-0.0039	0.02	0.4	0.07	0.52	2	0.2765	0.03	0.84	0.14	2	0.3754	0.06	0.94
Egger 1495	pilosa	3	0.2079	0.08	0.15	0.65	0.12	3	0.2591	0.12	0.17	0.71	1	0.0537	0.56	0.44
Egger 754	longispica	2	0.2232	0.08	0.53	0.19	0.2	2	0.0959	0.13	0.53	0.33	2	0.156	0.28	0.72
Egger 755	longispica	4	0.1604	0.01	0.14	0.03	0.82	2	0.3113	0.02	0.86	0.12	2	0.4014	0.05	0.95
Egger 959a	pilosa	4	0.036	0.04	0.36	0.08	0.52	2	0.2919	0.07	0.76	0.17	2	0.3425	0.12	0.88
Egger 969a	pilosa	2	0.3562	0.02	0.8	0.05	0.13	2	0.2846	0.04	0.83	0.13	2	0.3405	0.1	0.9
Egger 969b	pilosa	4	0.039	0.03	0.34	0.08	0.54	2	0.2859	0.06	0.76	0.17	2	0.3401	0.12	0.88
Egger 980d	longispica	3	0.3231	0.07	0.09	0.77	0.07	3	0.3633	0.1	0.11	0.79	1	0.1299	0.65	0.35
Egger 984a	longispica	3	0.2032	0.05	0.11	0.62	0.22	3	0.2749	0.07	0.18	0.75	2	0.0041	0.43	0.57
Ertter 5728	pilosa	2	0.3603	0.02	0.83	0.06	0.1	2	0.2308	0.04	0.79	0.17	2	0.3106	0.1	0.9
Ertter 8808	longispica	2	0.2922	0.06	0.64	0.1	0.21	2	0.2334	0.11	0.68	0.21	2	0.2565	0.2	0.8
Ferris 11084	nana	1	0.0865	0.46	0.08	0.29	0.17	1	0.1077	0.56	0.12	0.32	1	0.2329	0.79	0.21
Fox 689	longispica	3	0.1382	0.24	0.09	0.45	0.22	3	0.0692	0.33	0.13	0.54	1	0.1869	0.75	0.25
Gentry 2138	nana	3	0.0754	0.39	0.06	0.48	0.07	1	-0.0222	0.49	0.06	0.45	1	0.3084	0.9	0.1
Gilman 2015 036	nana	1	0.2746	0.86	0.02	0.1	0.02	1	0.2776	0.88	0.02	0.1	1	0.3833	0.94	0.06
Gilman 2015 038	nana	1	0.1592	0.76	0.03	0.18	0.03	1	0.1635	0.8	0.03	0.17	1	0.3916	0.94	0.06
Gross 717	pilosa	2	0.058	0.04	0.39	0.36	0.22	3	0.0293	0.06	0.41	0.53	2	0.1614	0.22	0.78
Halse 1831	pilosa	3	0.2746	0.06	0.08	0.81	0.05	3	0.335	0.11	0.09	0.8	1	0.1488	0.72	0.28
Hitchcock 5613	nana	1	0.3708	0.96	0.01	0.03	0.01	1	0.3424	0.95	0.01	0.04	1	0.4357	0.97	0.03
Hitchcock s.n.	pilosa	2	0.3767	0.01	0.87	0.04	0.09	2	0.2675	0.02	0.87	0.11	2	0.3842	0.05	0.95
Holmgren 10760	nana	1	0.0662	0.62	0.04	0.28	0.06	1	0.1409	0.7	0.05	0.25	1	0.3483	0.92	0.08
Holmgren 1233	pilosa	4	0.0747	0.03	0.21	0.13	0.63	2	0.1344	0.06	0.63	0.3	2	0.2475	0.15	0.85
Holmgren 1447	nana	1	0.2712	0.83	0.02	0.1	0.05	1	0.2436	0.86	0.03	0.11	1	0.3457	0.92	0.08
Holmgren 1569	nana	1	0.4131	0.96	0.01	0.03	0.01	1	0.3744	0.95	0.01	0.04	1	0.4343	0.96	0.04
Holmgren 1652	nana	1	0.3178	0.91	0.01	0.06	0.01	1	0.2849	0.91	0.02	0.07	1	0.399	0.95	0.05
Holmgren 2220	nana	1	0.4105	0.97	0	0.02	0.01	1	0.3712	0.96	0.01	0.03	1	0.4436	0.96	0.04

Holmgren 2235	nana	1	0.0546	0.64	0.04	0.27	0.05	1	0.0577	0.7	0.04	0.26	1	0.3722	0.92	0.08
Holmgren 2812	nana	1	0.2059	0.53	0.16	0.08	0.24	1	0.1939	0.66	0.22	0.12	1	0.0084	0.58	0.42
Holmgren 2847	nana	1	0.281	0.86	0.02	0.1	0.02	1	0.2541	0.87	0.03	0.11	1	0.3897	0.93	0.07
Holmgren 2879	steenensis	4	0.1165	0.21	0.12	0.18	0.49	1	0.0819	0.38	0.27	0.34	1	0.0061	0.54	0.46
Holmgren 2913	nana	1	0.1967	0.59	0.19	0.1	0.13	1	0.2484	0.69	0.18	0.13	1	0.0841	0.65	0.35
Holmgren 2918	nana	4	0.051	0.13	0.3	0.12	0.44	2	0.1243	0.23	0.54	0.23	2	0.1619	0.31	0.69
Holmgren 2924	nana	1	0.1991	0.74	0.03	0.16	0.07	1	0.1762	0.78	0.04	0.17	1	0.3406	0.91	0.09
Holmgren 2926	pilosa	2	0.2941	0.1	0.62	0.11	0.18	2	0.1713	0.19	0.59	0.22	2	0.1747	0.28	0.72
Holmgren 2936	nana	1	0.0431	0.52	0.08	0.33	0.07	1	0.0761	0.6	0.08	0.32	1	0.2947	0.87	0.13
Holmgren 3619	pilosa	3	0.2052	0.05	0.07	0.7	0.18	3	0.2908	0.07	0.11	0.81	1	0.057	0.56	0.44
Holmgren 3625	nana	1	0.2537	0.74	0.07	0.12	0.06	1	0.2598	0.79	0.07	0.14	1	0.2651	0.84	0.16
Holmgren 4119	pilosa	3	0.1574	0.3	0.08	0.55	0.08	3	0.1724	0.38	0.09	0.53	1	0.2746	0.83	0.17
Holmgren 5347	longispica	3	0.3366	0.08	0.07	0.78	0.07	3	0.3275	0.13	0.09	0.79	1	0.18	0.73	0.27
Holmgren 8861	pilosa	2	0.0374	0.05	0.44	0.08	0.43	2	0.273	0.09	0.76	0.16	2	0.3033	0.14	0.86
Holmgren 9484	pilosa	2	0.3359	0.01	0.81	0.06	0.12	2	0.2224	0.03	0.81	0.16	2	0.3279	0.08	0.92
Holmgren 9618	pilosa	2	0.3916	0.01	0.84	0.04	0.1	2	0.281	0.04	0.84	0.12	2	0.3468	0.09	0.91
Ionta 00 14	pilosa	4	0.2696	0.03	0.12	0.08	0.77	2	0.1853	0.08	0.62	0.29	2	0.2783	0.16	0.84
Jacobs 2015 072	pilosa	4	-0.0664	0.18	0.29	0.14	0.39	2	0.0171	0.31	0.43	0.26	2	0.1007	0.4	0.6
Jacobs 2015 073	pilosa	1	0.1047	0.39	0.2	0.13	0.28	1	0.1521	0.54	0.26	0.2	1	-0.0014	0.57	0.43
Jacobs 2015 075	pilosa	4	-0.0341	0.06	0.3	0.18	0.46	2	0.1017	0.11	0.52	0.36	2	0.2015	0.24	0.76
Jacobs 2015 076	nana	3	0.1961	0.13	0.07	0.61	0.19	3	0.2094	0.19	0.11	0.71	1	0.1556	0.74	0.26
Jacobs 2015 084	pilosa	3	0.289	0.11	0.06	0.77	0.05	3	0.2582	0.18	0.07	0.75	1	0.2313	0.82	0.18
Jacobs 2015 093	nana	1	0.2379	0.82	0.02	0.12	0.03	1	0.2835	0.85	0.03	0.12	1	0.3933	0.94	0.06
Jacobs 2015 096	nana	1	0.2052	0.59	0.11	0.17	0.12	1	0.2232	0.68	0.12	0.2	1	0.1855	0.78	0.22
Jacobs 2015 102	nana	1	0.3197	0.76	0.07	0.1	0.07	1	0.3257	0.82	0.07	0.11	1	0.2259	0.84	0.16
Jacobs 2015 106	steenensis	4	0.0361	0.27	0.12	0.18	0.42	1	0.1209	0.44	0.25	0.31	1	0.0408	0.59	0.41
Johnson 850289	longispica	4	-0.0405	0.02	0.44	0.06	0.48	2	0.2925	0.03	0.85	0.13	2	0.3729	0.06	0.94

Jones 295	longispica	3	0.302	0.1	0.04	0.82	0.04	3	0.2616	0.18	0.06	0.77	1	0.2616	0.87	0.13
Kemper 143	longispica	3	0.0941	0.17	0.22	0.49	0.12	3	0.1512	0.24	0.2	0.57	1	0.099	0.65	0.35
Knoke 472	pilosa	2	0.3914	0.01	0.88	0.03	0.08	2	0.2863	0.04	0.85	0.11	2	0.3167	0.08	0.92
Knoke 577	pilosa	2	0.3971	0.01	0.9	0.03	0.07	2	0.2792	0.02	0.88	0.1	2	0.3721	0.05	0.95
Knoke 611	pilosa	3	0.3191	0.05	0.08	0.8	0.07	3	0.3704	0.08	0.1	0.82	1	0.1221	0.64	0.36
Kruckeburg 4167	longispica	2	0.3582	0.01	0.84	0.05	0.11	2	0.2626	0.02	0.85	0.13	2	0.385	0.06	0.94
Maguire 21128	nana	1	0.2961	0.82	0.05	0.09	0.04	1	0.2616	0.85	0.05	0.1	1	0.2873	0.88	0.12
Maguire 25800	nana	1	0.2244	0.83	0.02	0.12	0.02	1	0.2446	0.86	0.02	0.12	1	0.3855	0.95	0.05
Maguire 26487	pilosa	4	0.0736	0.03	0.27	0.18	0.53	2	0.0908	0.05	0.61	0.35	2	0.289	0.12	0.88
Moseley 591	longispica	3	0.1323	0.13	0.22	0.53	0.12	3	0.1996	0.19	0.2	0.61	1	0.073	0.6	0.4
Olmstead 717	pilosa	4	0.215	0.02	0.15	0.06	0.76	2	0.2378	0.06	0.72	0.22	2	0.3145	0.12	0.88
Olmstead 752	pilosa	3	0.2114	0.22	0.09	0.62	0.07	3	0.2378	0.3	0.1	0.6	1	0.2395	0.81	0.19
Olmstead 96 54	pilosa	2	0.3658	0.01	0.85	0.04	0.1	2	0.2623	0.02	0.85	0.12	2	0.3797	0.06	0.94
Ownbey 3412	longispica	4	0.1446	0.06	0.19	0.07	0.68	2	0.2455	0.14	0.67	0.19	2	0.2483	0.2	0.8
Packard 78 89	pilosa	2	0.36	0.01	0.79	0.06	0.13	2	0.2572	0.03	0.82	0.15	2	0.3845	0.07	0.93
Peck 18941	pilosa	4	0.1647	0.02	0.18	0.06	0.74	2	0.2604	0.06	0.77	0.18	2	0.3307	0.11	0.89
Peck 21406	pilosa	4	0.1475	0.01	0.19	0.05	0.75	2	0.28	0.03	0.82	0.15	2	0.4044	0.06	0.94
Pennell 22950	nana	1	0.3725	0.94	0.01	0.04	0.01	1	0.3828	0.94	0.01	0.05	1	0.4275	0.96	0.04
Reveal 2428	steenensis	4	0.1562	0.02	0.2	0.06	0.72	2	0.2757	0.04	0.77	0.19	2	0.3733	0.09	0.91
Reveal 2450	pilosa	2	0.3508	0.01	0.8	0.06	0.13	2	0.2404	0.03	0.81	0.16	2	0.3536	0.08	0.92
Rodman 681	pilosa	4	0.0099	0.02	0.39	0.08	0.52	2	0.2555	0.03	0.81	0.16	2	0.3717	0.07	0.93
Rogers 1006	pilosa	2	0.2232	0.04	0.57	0.12	0.26	2	0.193	0.07	0.68	0.25	2	0.302	0.15	0.85
Tiehm 10589	pilosa	3	0.2657	0.16	0.08	0.69	0.07	3	0.2938	0.23	0.1	0.67	1	0.2238	0.79	0.21
Tiehm 13189	pilosa	3	0.2522	0.06	0.12	0.75	0.07	3	0.308	0.1	0.13	0.77	1	0.0905	0.61	0.39
Tiehm 13885	pilosa	3	0.3042	0.12	0.1	0.67	0.12	3	0.2738	0.17	0.12	0.71	1	0.1626	0.7	0.3
Tiehm 8790	pilosa	2	0.1117	0.11	0.38	0.17	0.34	2	0.1229	0.17	0.53	0.3	2	0.1594	0.29	0.71
Tiehm 9598	pilosa	3	0.3235	0.03	0.06	0.88	0.04	3	0.3787	0.05	0.07	0.88	1	0.0968	0.62	0.38
Train 4258	nana	3	0.0331	0.42	0.05	0.48	0.05	1	-0.0409	0.54	0.05	0.41	1	0.3161	0.93	0.07

Vollmer 229	nana	1	0.0941	0.44	0.11	0.1	0.35	1	0.1954	0.63	0.21	0.16	1	0.0367	0.61	0.39
Whitehead 487	longispica	4	0.1833	0.02	0.13	0.04	0.81	2	0.2759	0.06	0.78	0.17	2	0.3193	0.1	0.9
Wilson s.n.	pilosa	3	0.1412	0.28	0.15	0.47	0.1	3	0.1662	0.35	0.16	0.49	1	0.1819	0.75	0.25
Wolf 5217	pilosa	3	0.0214	0.27	0.2	0.31	0.21	3	0.0083	0.37	0.25	0.39	1	0.0698	0.62	0.38
Zika 11219	steenensis	4	0.2203	0.04	0.17	0.11	0.68	2	0.1635	0.09	0.6	0.32	2	0.2798	0.18	0.82

CHAPTER 3: INCORPORATING ENVIRONMENTAL EVIDENCE TO DELIMIT SPECIES IN THE *CASTILLEJA AMBIGUA* SPECIES COMPLEX.

with David C. Tank

Abstract

Delimiting species boundaries is an important contribution to the biodiversity sciences, particularly conservation, where the status of *species* carries great weight. The last decade of delimitation work has relied heavily on molecular data, but more recently it has been widely advocated to apply multiple lines of evidence. Environmental data have historically been used to define species, but it has only recently been more widely included in species delimitation studies. Here we apply environmental data to the question of species boundaries in the *Castilleja ambigua* species complex (two taxonomic species). Given robustly estimated species ranges (using occurrence data from museum collections), we estimated niche models and extract climatic variables associated with focal taxa to use as an environmental line of evidence to corroborate molecular species boundaries. Here, disparate lines of evidence (molecular and environmental) are examined for congruent signals of delimitation.

Introduction

Status as a species carries with it important conservation implications (e.g. Myers *et al* 2000, Agapow *et al* 2004). Subsequently, species delimitation plays an important role in the biodiversity sciences, where the explicit quantification of biodiversity is necessary. Recently, there has been heavy reliance on molecular data to determine the boundaries between species (e.g., Fujita *et al* 2012, Pons *et al* 2006). However, it is clear that there are cases where molecular data alone are not sufficient for drawing species boundaries; for example, in incipient lineages that are in their earliest stages of diversification, and where any one line of evidence may provide a different signal of lineage boundaries than another. Moreover, these cases are often of particular interest and importance with respect to conservation implications. A common challenge associated with working in incipient systems is that newly diverged lineages often have restricted ranges and/or are relatively rare and known from very few populations. While they are commonly the target of management (Niemiller *et al* 2013), conservation decisions are frequently made with information from limited datasets. As such, analytical approaches that leverage publically available data with thorough analytical investigation are of great importance (Espíndola *et al* 2016).

As an example, a recent study (Jacobs *et al* 2018) applied multiple, independent molecular species delimitation approaches to a species complex in the taxonomically challenging plant genus

Castilleja (also known as 'the paintbrushes') using a limited molecular data set. After recovering incongruent delimitation schemes, Jacobs et al (2018) applied a *post-hoc* simulation-based approach to assess inferential error in the methods they applied. This study determined that in cases of incipient speciation, where the node(s) of interest are relatively shallow and where datasets are often limited in size, the signal of diversification can be difficult to detect. In this study, the molecular species delimitation approach implemented in BPP v.3.1 (Bayesian Phylogenetics and Phylogeography; Yang and Rannala 2014) was more sensitive to the signal of divergence than spedeSTEM, an alternative maximum likelihood-based approach (Ence and Carstens 2010). This sensitivity has recently been discussed in the literature (e.g., Carstens *et al* 2013, Barley *et al* 2017, Sukumaran and Knowles 2017) with a clear cautionary warning that population structure can be easily mistaken for species boundaries. In cases such as these, it is widely advocated to apply multiple lines of evidence to the question of species boundaries.

Ecological lines of evidence have long been used to infer species boundaries (Van Valen 1976, Andersson 1990). In the absence of experimental work to understand the physiological tolerances of species, the environmental niche is often used as a proxy to describe the biotic and abiotic characteristics of species (e.g. McCormack *et al* 2009, Morales *et al* 2016, McKelvy and Burbrink 2017). Recently, this has been accomplished through the estimation of species ranges and distributions and the creation of models that describe them (i.e., species delimitation models – SDMs; Raxworthy *et al* 2007, Rissler and Apodaca 2007, Dowell and Hekkala 2016, Morales *et al* 2016), thereby allowing one to predict or score novel regions as suitable or not for the species under question. The last decade has seen the widespread use of descriptive statistics and simulations to evaluate and compare environmental niches across different entities. More specifically, niche overlap has been a hallmark in the application of SDMs to the question of comparing species niches (Warren *et al* 2008).

There are limitations to the application of SDM outputs for quantifying niche overlap, however. Because the output of an SDM is the projection of the model into some geographic space (resulting in the suitability scores for each part of that space), the subsequent quantification of niche overlap is tightly linked to the extent of that geographic space and the resolution at which environments are heterogeneous. Measures of niche equivalency and similarity, which use niche overlap as a comparison statistic, are similarly linked to the extent of the geographic space being considered. Broennimann et al (2012) provided a statistical framework that aimed to overcome these limitations by calculating niche overlap in the context of a gridded or binned space, where each division of the space represents a unique set of environmental conditions found in the study area. The density of occurrence of an entity within that space, and the overall occupancy of the entity across

the space, is then used as the basis for quantifying niche overlap when compared with other entities. The benefit of an approach like this is that the measure of overlap is no longer biased by spatial resolution, geographic extent, or environmental heterogeneity (Broennimann *et al* 2012).

An additional contribution of Broennimann *et al* (2012), though (we think) underutilized, is the comparison of the quantification of the niche in climate space (through ordinations) versus geographic space (using SDM techniques *sensu* Warren *et al* 2008). In a simplified set of niche overlap simulations, Broennimann et al (2012) determine that ordination techniques typically outperform those of SDMs in accurately measuring niche overlap, but importantly lack, the often-desired capability of ranking and selecting which variables are most important for describing the environmental niche. So, depending on the needs of the study and the underlying structure of the environment, the choice of technique may impact niche quantification. For example, if one wants to identify what variables best discriminate a niche, and are therefore (theoretically) closely linked to the processes underlying distributions, then the SDM approach is useful. This is because SDMs apply a weighting scheme to prioritize variables that are good at identifying known occurrences. However, environmental complexity and collinear predictor variables can lead to spurious results in SDMs, where correlated variables (but not necessarily causative variables) are identified as important for discrimination, thus making predictions into new areas inconsistent with actual biological requirements of the entity (Broennimann *et al* 2012). Ordination techniques, on the other hand, aim to maximize variance in the datasets by finding orthogonal axes (thereby dealing with collinear variables) that best discriminate occurrences based on environmental conditions, rather than prioritizing the best predictors (Broennimann *et al* 2012).

Species delimitation studies have begun to incorporate ecological evidence in a number of ways. One can take a corroborative approach to delimitation where one examines congruence across lines of evidence to support species boundaries (e.g., Padial *et al* 2010, Schlick-Steiner *et al* 2010). Alternatively, one can fully integrate lines of evidence into a single analysis of species boundaries where, theoretically, each line of evidence contributes to the analysis of species boundaries (e.g. Gaussian clustering, Edwards and Knowles *2014*; modelling approaches, Guillot *et al* 2012, Solis-Lemus *et al* 2015). The former approach (corroborative evidence) has largely been applied in species delimitation studies using multiple types of data. For example, several studies have used climatic data as a corroborative line of evidence for species boundaries (e.g. Raxworthy *et al* 2007, Reeves *et al* 2011, Dagnino *et al* 2017, Gama *et al* 2017, Viera-barreto *et al* 2018). The later approach (fully integrated analyses) is, at present, limited to clustering techniques, though this has been met with some criticism (Meik *et al* 2015).

In this study, we extend the application of species delimitation of the *Castilleja ambigua* species complex to include ecological evidence. We apply these data to corroborate (or not) molecular evidence from a previous study (Jacobs *et al* 2018), in combination with a qualitative assessment of morphology in the complex. Due to the restricted ranges of two of the three members of this complex, in addition to the generally close proximity of ranges of these entities to one another, we apply both ordination and SDM techniques within the framework proposed by Broennimann *et al* (2012). A benefit of applying both approaches is the opportunity to characterize the geographic space inhabited by these putative lineages (through the projection of SDMs across specific geographic extents), as well as the climate space that they occupy (through ordination techniques based on environmental variables). Additionally, the comparison of climatic niche quantifications resulting from both techniques could prove valuable in assessing the reliability of niche quantification, as these different approaches could provide disparate signals, especially given the small, and often overlapping, ranges of putative lineages.

Methods

Study system

The *Castilleja ambigua* complex is composed of 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). The members of this complex occur along the western coast of North America, from southern California to British Columbia, and the islands of the Puget Sound where they occur in a variety of coastal habitats. Members of this complex are generally united by floral morphology (Egger *et al* 2012; Wetherwax *et al* 2016)—they share a pouchy lower lip, reduced beak, and relatively non-showy bracts common to most annual species of *Castilleja*. However, variability in the polymorphic *C. ambigua* has led to the description of three additional varieties that differ in geographic position (and presumably ecological preferences) and geographically consistent morphological variation (mostly in bract color, e.g. *C. ambigua* var. *humboldtiensis*) (Fairbarns & Egger 2007; Egger *et al* 2012).

The typic and most widespread of these varieties, *C. ambigua* var. *ambigua*, generally has white and yellow flowers and occurs on coastal bluffs and grasslands along the Pacific coast from southern California to British Columbia (Fig. 1). This typic variety can be divided into two main morpho groups that occur across the range: a fleshy morph with a single or few stems per plant that is found in marshy areas, and a less fleshy, often highly branched morph most often occurring in grasslands. Within each of these morpho groups there exists narrowly restricted populations of plants

that are consistently different in bract color that have been formally named as varieties. *Castilleja ambigua* var. *humboldtiensis* (D.D. Keck) J.M. Egger is a fleshy, less-branched variety that has primarily pink to rose-purple bracts. It occurs in salt marshes along the northern coast of California in Mendocino and Humboldt counties. The other, similarly 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. For the purposes of this study *C. ambigua* var. *ambigua, C. ambigua* var. *insalutata*, and *C. ambigua* var. *humboldtiensis* were modeled as a single species. Their ecological preferences and their geographic position (coastal, associated with grasslands and salt marshes; bounded to the north and the south by populations of *C. ambigua* var. *ambigua*) closely tie varieties *insalutata* and *humboldtiensis* to the typic form (var. *ambigua*), and previous phylogenetic work firmly place them within the *C. ambigua* var. *ambigua* lineage (Jacobs *et al* 2018). For these reasons, we do not consider these named varieties as putative lineages.

Recently, Egger et al (2012) described the variety *C. ambigua* var. *meadii* J.M. Egger & Ruygt. Vegetative morphology, geography, and ecological preferences readily distinguish *C. ambigua* var. *meadii* from its conspecifics; variety *meadii* is typically erect, with un-branched stems, and has leaves and bracts with narrow, linear lobes. It is restricted to the Atlas Peak Plateau district of Napa County, California, where it occurs in seasonally wet places associated with freshwater, ephemeral seeps. This taxon is known from only four extant populations (a fifth being recently documented as extirpated (Egger *et al.* 2012)). Recent molecular evidence (phylogenetic inference and molecular species delimitation) indicates that variety *meadii* seems to be an independently evolving lineage separate from other varieties, in contrast to varieties *insalutata* and *humboldtiensis*.

The final 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* has a short, compact, and single-stemmed habit with dull-brown vegetative coloration. The flowers of this species are yellow (in contrast to the bright, contrasting coloration of *C. ambigua*) and have a notable difference in position of the stigma at peak flowering time; *C. ambigua* exerts its stigma up and out of the corolla (a prominent feature in many outcrossing floral morphologies), while the stigma of *C. victoriae* remains inserted within the corolla, sitting low in the corolla tube. It has been

suggested that *C. victoriae* is self-compatible (capable of self-fertilization), though this has not been explicitly tested.

The examination of species boundaries in this complex is motivated by the ecological and morphological variation described above, in addition to the conservation and management implications associated with species status of the two range-restricted taxa *C. ambigua* var. *meadii* and *C. victoriae*. In light of the molecular evidence from previous work (Jacobs *et al* 2018), we focus on testing the ecological distinctiveness of the following three taxa: *Castilleja ambigua*, *C. ambigua* var. *meadii*, and *C. victoriae.*

Occurrence data—All known collections were assembled from regional herbaria (Stillinger Herbarium [ID], University of Washington Herbarium [WS]) and two online databases that collectively represent the known distributions of these three focal taxa (http://pnwherbaria.org, http://ucjeps.berkeley.edu). When present, GPS coordinates were taken directly from herbarium labels. However, in some cases, latitude and longitude were not provided by the collector, in which case, coordinates were estimated by hand using locality information and the GeoLocate web service (http://www.museum.tulane.edu/geolocate/default.html). Because the members of this complex are coastal, and coordinate estimates provided by software like GeoLocate are heavily impacted by the detail of the locality information provided, estimated latitude and longitude occasionally fell in uninhabitable areas (i.e., the ocean), or in city centroids (i.e. the result of poor detail in collection locality records). Therefore, final sets of coordinates for each putative lineage were loaded into Google Earth to visually confirm the position of the collection, and adjustments were made as necessary. Both *C. ambigua* var. *meadii* and *C. victoriae* are extremely range-restricted, known from only a few populations, and therefore poorly represented in herbaria and online databases. Therefore, we included historical records of now extirpated populations in our final dataset.

Pseudo-absence and background data points—Because all models used in our species distribution modeling appraoch require both presence and absence data, and because we did not have formally measured absence data, we estimated pseudo-absences based on our occurrence records. We set a radius of 50 km around each occurrence point, and randomly sampled points within that radius. Additionally, we estimated background data by randomly sampling points across the entire study region. A map of occurrence, pseudo-absences, and background datapoints used in this study is available in the supplemental material (Supplemental Fig. 1).

Climatic variables— Nineteen bioclimatic variables were downloaded from the WorldClim database (http://www.worldclim.org; Hijmans *et al* 2005) at approximately 1km^2 resolution (30 arc-second) for regions spanning the distribution of focal taxa. We assessed correlation of variables by calculating Pearson correlation coefficients for all pairwise combinations of variables given our occurrence dataset, and eliminated any whose correlation was above *r = 0.8*.

*Species distribution modeling—*We used the ensemble modeling approach available in the R package Biomod2 (Thuiller *et al* 2009, Thuiller *et al* 2016) to generate species distribution models (SDMs) for each of our focal taxa. A combination of nine algorithms were used for initial model building: generalized linear models (GLM), generalized boosted models (GBM), generalized additive models (GAM), classification tree analysis (CTA), artificial neural networks (ANN), surface range envelope (SRE), flexible discriminant analysis (FDA), multivariate adaptive regression splines (MARS), and random forests (RF). Because we had relatively few presence points in close proximity for two of our focal taxa (*C. ambigua* var. *meadii* and *C. victoriae*), we were unable to parse a separate dataset to evaluate model performance. Instead, we performed a three-fold cross validation of each model by randomly splitting our data into two subsets (a training and a testing set consisting of 70% and 30% of the original data, respectively). In the ensemble modeling step, the outputs of these nine models were evaluated by the relative operating characteristic (ROC), Cohen's kappa (KAPPA), and the true skill statistic (TSS), and only those meeting a threshold of 0.7 or greater were included in the final ensemble model projection (Araújo *et al* 2005). The SDM of each putative species was projected into the same area to calculate the probability of occurrence across the entire study area.

*Comparison of environmental niches—*Within the framework proposed by Broennimann et al (2012), for each putative species we quantified niches in both climate space and geographic space. The statistics calculated in each of these spaces is exactly the same; the difference between these two backdrops of comparison lies in the data used to calibrate the niche and calculate the occurrence densities. In climate space the input data are the raw climatic variables extracted from our BioClim layers; in geographic space, the data are the predicted probabilities from the SDMs. Comparisons in both climate space and geographic space begin by dividing the space into a user-provided number of 'grids' or 'bins' (i.e., the resolution of the space, *r*, here set to 100) that is bounded by the minimum and maximum values associated with the input data. The 'grid cells' or 'bins' within the space correspond to a unique set of environmental conditions present at one or more sites in the study area. The density of occurrence of an entity at each 'bin' is estimated using a kernel density function with standard Gaussian smoothing parameters, and describes how often each entity is found in 'bins'

corresponding to that particular set of environmental characteristics, scaled by the availability of that 'bin' throughout the range of the study area.

Given the calibration of the environmental niche and the calculation of the occurrence densities as described above, we measured niche overlap using the D metric (Schoener 1968; Warren et al 2008), which varies from 0 (no overlap) to 1 (complete overlap). We then tested niche equivalency by determining if the measure of niche overlap remains constant if we randomly shuffle and reallocate the occurrences of both entities. By comparing our measured overlap to a distribution of simulated overlap measures calculated on reshuffled occurrences, we can test the null hypothesis of equivalent niches. Niche similarity tests were also performed where the niche of one entity was held constant, while the niche of the second entity was randomized. Here again, the empirical measure of niche overlap was compared to a distribution of simulated overlap measures, to test the null hypothesis of the two compared niches being more similar than expected by chance.

For niche comparisons in climate space, we used the PCA-env approach (Broennimann *et al* 2012) that uses the first two axes of a principal components analysis of the raw climatic variables. Here we calibrated the niche (i.e., divided the environmental space into a grid) based on the climatic variables from the combined range of each pair of putative species analyzed. We limited our environmental variables to the same variables used to create the SDMs; however, we also performed analyses on the complete suite of environmental variables with no change in the results or interpretation (*not presented here*). For our niche comparisons, we performed 1000 simulations of niche equivalency and niche similarity.

Broennimann et al (2012) found that when calibrated on the ranges of both entities, SDMs consistently over-estimated niche overlap. Subsequently, for niche comparisons in geographic space, we used a vector of predicted probability of occurrences generated from an SDM of a single species only (i.e., the environmental space is calibrated on a single species). In this framework, the comparison of the overlap of the two entities is analyzed along a gradient of predictions resulting from the SDM of one of the species – i.e., the SDM of one species is 'projected' into the range of the other, and occurrence probabilities of the first entity's SDM are extracted from the range positions of the second entity. Each directional, pairwise comparison was performed. For example, using the *C. ambigua* SDM predictions, we calculate the overlap of *C. ambigua* to *C. ambigua* var. *meadii* and *C. ambigua* to *C. victoriae*; the same was performed using the SDM of *C. ambigua* var. *meadii* and *C. victoriae*. For each analysis, we performed 1000 simulations of niche equivalency and niche similarity.

Results

Taxon sampling, occurrence datasets, and climatic variables—Our final dataset consisted of 227 individual occurrence records (*C. ambigua* (n=185), *C ambigua* var. *meadii* (n=25), and *C. victoriae* (n=17)), representing the known ranges of our focal taxa (Fig. 1). Of the nineteen BioClim variables examined, only three met our threshold of correlation: Bio2 – mean diurnal range (mean of the monthly (max temp-min temp), Bio3 – isothermality (mean diurnal range/annual temperature range), and Bio15 – precipitation seasonality (coefficient of variation). Given that these species are all spring annuals, precipitation in the wettest month is likely an important factor in the distribution of these species we therefore also included Bio13 – precipitation in wettest month – in downstream analyses. The distribution of raw climatic variables grouped by focal taxon (Fig. 2) reveals overlap in temperature and precipitation among the ranges of these taxa. However, the bulk of the distribution of values often differs (for example, Bio2 in Fig. 2, top-left panel), suggesting that there are some differences that are consistent with the ranges of these entities.

Species distribution models—A visual assessment of the distribution models showed that our models predicted a high probability of occurrence in areas corresponding to known occurrences (dark red areas, Fig. 3). The contributions of each variable to models in the initial modeling steps show that Bio2 and bio3 primarily distinguish regions of high probability of occurrence for *C. ambigua*, bio2 and bio13 distinguish regions of high probability of occurrence of *C. ambigua* var. *meadii*, and bio2, bio13, and bio 15 distinguish regions of high probability of occurrence of *C. victoriae* (Table 1). All combinations of models were compared to generally assess the differences in the modeled distributions of each putative species. Mean, standard deviation, and correlation of probability of occurrence scores were calculated for each pairwise comparison of SDMs (Fig. 4). These comparisons indicated that many regions of high probability for each putative species do not correspond with high probability for either of the other two focal taxa (regions that are dark red in both mean and standard deviation plots). Additionally, cell-by-cell correlation plots revealed no correlation of probability scores in any of our comparisons (i.e., in areas of high probability for one taxon, there was low probability of occurrence for another).

Niche quantification and comparison—Figure 5 provides a visual interpretation of the quantified niches in climate space for each pairwise comparison of our three focal taxa. We see a great deal of overlap in the background environment in all three comparisons, reflecting the close proximity of the known range of each entity (discussed in detail below). The density of occurrence of each putative

lineage (the dark grey shading), however, occupies very different portions of climate space in each pairwise comparison. In all three PCA-env analyses, the first two axes accounted for ≥ 85% of the total inertia. We find similar patterns in the niches quantified in geographic space (Fig. 6). In these comparisons, niches calibrated on individual SDMs have a higher density of high probability occurrences relative to the compared entity. This is especially the case with *C. victoriae* and *C. ambigua* var. *meadii* (Fig. 6b, 6c) where the compared entity has a density of occurrence equal to zero for all possible scores of occurrence probabilities. We do find that that *C. victoriae* and *C. ambigua* overlap to some degree, though this occurs in regions of low probability of occurrence in the SDM calibrated on both taxa (Fig. 6a, 6b).

Niche overlap measures in both climate space and geographic space indicate little to no overlap in all pairwise comparisons of our focal taxa (Table 2). Additionally, we observed generally lower estimates of niche overlap in climate space, compared to that of geographic space. One exception to these general results is in the measure of overlap between *C. ambigua* and *C. victoriae*, which ranged from $D = 0$ (in climate space) to $D = 0.442$ (in geographic space calibrated on the *C. victoriae* SDM).

The hypothesis of niche equivalency is rejected in all taxon-by-taxon comparisons using both geographic and climate space (Table 2), indicating that no one niche can be considered interchangeable with another. Finally, we fail to reject the hypothesis of niche similarity in all comparisons, save those considering *C. victoriae* and *C. ambigua* in geographic space. This indicates that there is a good deal of similarity of niches, more so than expected by chance.

Discussion

Depicting the niche using climatic data in both environmental and geographic space
Niche quantification—Our measures of niche overlap are extremely low for all taxon comparisons in both environmental and geographic space; the only exception being our comparison of *Castilleja ambigua* and *C. victoriae*. In environmental space, we measure niche overlap at $D=0.005$, while in environmental space this measure ranges from $D=0.04$ to 0.442. We think this is largely an artifact of our use of the SDM probabilities in characterizing the niche in geographic space. For example, *C. ambigua* receives low to moderate values of probability of occurrence throughout the range of *C. victoriae*. As a result, they similarly occupy the same portions of the niche space—these portions have low probabilities for *C. ambigua*, but high probabilities for *C. victoriae*. This is also an example where the correction for prevalence of suitable habitat plays a big role in measurement of niche overlap. When we describe the niche in geographic space of the widespread (and therefore highly

represented) *C. ambiguga* and then project that model into the restricted range of *C. victoriae* where we know *C. ambigua* has low-moderate probabilities, we recover a larger degree overlap (D=0.215) than when we correct for the limited prevalence of *C.victoriae* suitable habitat in the comparison (D=0.041). Alternatively, when we describe the niche in geographic space using *C. victoriae* and project into the known range of *C. ambigua*, we measure very low niche overlap (D=0.075) until we correct for the prevalence of suitable habitat for *C. ambigua*, which increases the overlap measure (D=0.442). Had we filtered our SDMs by limiting the probabilities compared to only those greater than or equal to a probability of 0.7 (i.e., a high probability of occurrence), we would filter out much of the overlap areas and likely consistently measure a lower D between these two taxa (Supplemental Table 1, 2).

Niche comparisons—Niche equivalency simulations similarly fall at the smaller end of the distribution in all comparisons, both in environmental and geographic spaces (Table 2). Rejecting the null hypothesis of equivalency suggests that the niches inhabited by each taxon are not interchangeable. Given the disparity in niche variance between taxa (i.e., *C. ambigua* occupies a great deal more niche space than either of the other two taxa), this is, perhaps, not surprising. And this makes further sense, given additional qualitative ecological characteristics that were not explicitly included in these analyses. For example, both of the range restricted taxa, *C. ambigua* var. *meadii* and *C. victoriae*, are associated with ephemeral freshwater seeps and vernal pools, in contrast to the salt-marsh and/or coastal grassland associated *C. ambigua*.

Niche similarity measures are more variable and have contrasting outcomes (relative to measures of niche equivalency) in both environmental and geographic space. In environmental space, we observe our statistic at fairly high ends of the distribution. Here, we think this outcome is largely driven by the close proximity of ranges to one another and their corresponding, largely overlapping ranges. This analysis proceeds by holding the occurrence density of one taxon constant and randomly repositioning the density of the second taxon within its known boundaries. Given the large degree of overlap in the background 50% and 100% of sampling points (Fig. 5, solid lines), the measure of overlap should regularly be much higher. We confirmed this by reducing our background sampling contribution in environmental space and reanalyzing niche similarity (Supplemental Table 1, 2). In these analyses, we see little overlap in the 50% and 100% background samples and our observed statistic falls at a much lower end of the distribution. In geographic space, we think the results can be interpreted the same.

We think these results strongly support the environmental distinction of these taxa, despite the apparent inconsistencies in measures of similarity in environmental and geographic space. It is

clear, however, that the definition of range can influence the comparative measures we use here (especially niche similarity). We suggest the exploration of the impact that estimated ranges has on comparative measures is an important procedural step in quantifying or characterizing differences of niches between putative species using this approach. Additionally, either of these quantifications of niche (the environmental niche based on ordinations of raw climatic variables or the geographic niche based on probabilities from SDMs) would likely have been sufficient evidence to support the environmental distinction of these putative lineages; however, in our opinion, the congruence in these lines of evidence lends further weight to the distinction as it accommodates potential limitations of a single approach.

Taxonomic implications—Jacobs *et al* (2018) outlines the case for molecular distinction of these three focal taxa that begins with a phylogenetic inference at odds with taxonomy. Through the application of multiple independent molecular species delimitation approaches, and the subsequent assessment of inferential error with *post-hoc* simulations, Jacobs et al (2018) determined that molecular evidence supports the distinction of *C. ambigua* var. *meadii* as an independent lineage. In acknowledgement of the importance of multiple lines of evidence for robust species delimitation, the present study provides corroborative evidence beginning with a qualitative assessment of morphology. *Castilleja ambigua* and *C. ambigua* var. *meadii* are distinguished morphologically from *C. victoriae* by a reproductive morphology that suggests (though it has not been explicitly tested) differences in pollination syndrome—namely, the conventional outcrossing morphology that includes an exserted stigma at peak-reproductive maturity in *C. ambigua* and *C. ambigua* var. *meadii*, versus a common self-pollinating morphology of an inserted stigma held within the floral tube at the peak maturity found in *C. victoriae*. Vegetative morphology further distinguishes *C. ambigua* var. *meadii* from the rest of *C. ambigua*—likely as a result of its freshwater associations, variety *meadii* lacks the thick, fleshy stems and leaves of its (largely) salt-water associated conspecifics.

The primary contribution of this study to species delimitation in the *Castilleja ambigua* species complex is to provide a corroborative line of environmental evidence that these lineages are distinct from one another. The data we present here suggests different preferences for environmental and geographic spaces exhibited by these putative species. Geographically, the ranges of these taxa are very proximal—the widespread *C. ambigua* is flanked at the north by *C. victoriae* and to the southeast by *C. ambigua* var. *meadii*—but appear to be largely unsuitable for one-another in environmental space. We see this in the low correlation of probability of occurrences across SDMs of each taxon (Fig. 4), in addition to the mean and standard deviation of probabilities of occurrence

where a high probability in one taxon, in most cases, corresponds with a low probability of occurrence in another. Importantly, from the complementary nature of theses analyses in geographic and environmental spaces, we can conclude that this result is independent of the low prevalence of *C. ambigua* var. *meadii* and *C. victoriae* suitable habitat, as we recover the same results in our characterization of the niche of each entity in geographic space (Fig. 6), where we control for the frequency of suitable habitat. This signal is echoed in the characterization of environmental space in these three taxa (Fig. 5). Here, we observe again the proximity and overlap of these ranges (overlapping 50% and 100% background limits); however, the occupancy of regions of the niche are unique to each taxon (Fig. 5, non-overlapping areas with dark grey shading).

Given each of these methods, we have consistent evidence that the current range of any one taxon is not strongly suitable for that of any of the others considered here. Despite the proximity of both *C. victoriae* and *C. ambigua* var. *meadii* to the widespread *C. ambigua*, we find very low support for co-occurrence of any of these taxa. This is corroborated by an examination of the environmental niche, and the inference of very different core-niche occupancy by each taxon (Fig. 5). In so doing, we have generated a line of environmental evidence to corroborate species boundaries in the *C. ambigua* species complex. We consider the evidence provided here to support the recognition of three distinct lineages, and recommend the elevation of *C. ambigua* var. *meadii* to the status of species, which will be formally addressed in a subsequent publication.

Conservation implications—From a conservation standpoint, this study has the potential to affect conservation and management of these species. By applying these three approaches, we have more formally characterized and described the ranges of these taxa. This may have conservation and management implications, beyond those resulting in the elevation of *C. ambigua* var. *meadii* to species, through the identification of areas with high probability of occurrence that are not currently part of the known ranges of these taxa. For example, *C. ambigua* var. *meadii* has a moderate probability of occurrence in the Klamath Mountains of southwestern Oregon. Directed search efforts in these areas would be a good place to search for new, currently unknown populations.

From an evolutionary perspective, this study contributes to our knowledge of the early stages of the speciation process in this genus, specifically with respect to environmental characteristics of this species complex. While we have previously understood the range of *C. ambigua* to be much broader than that of the other two taxa, we can now add in this description a comment on the breadth of the niche occupied, relative to the other members of this complex. The extreme asymmetry in range size and niche breadth (as we have measured it here) could be in line with expectations associated with budding speciation (e.g. Grossenbacher *et al* 2014). However, one could not discount

the possibility that these range-restricted taxa could be remnant populations of a previously more widespread ancestor. The last two million years have seen extreme fluctuations in sea levels as a result of climatic fluctuations. As coastal species, these fluctuations would likely have impacted distributions. Demographic modeling of different scenarios (e.g. Ruffley *et al* 2018) could shed light on processes resulting in the current pattern of distribution and range size.

*Species delimitation with multiple lines of evidence—*It could be argued that a stable taxonomy is a primary goal of much species delimitation work, especially with the heavily reliance on the species status in conservation (Agapow *et al* 2004, Isaac *et al* 2004, Mace 2004, Morrison *et al* 2009). The articulation of the generalized lineage concept (de Queiroz 1998, 2005, 2007) and its subsequent adoption by many working on species delimitation has promoted the search for properties that identify diverging lineages (i.e. evolutionary significant units), rather than relying on a single criterion (e.g. reproductive isolation and the biological species concept, Mayr 1942). This is important because stability in taxonomic ranks, particularly at the species level and especially in incipient lineages, might (arguably) be highest when species are diagnosable based on two or more characters (Padial *et al* 2010, Carstens *et al* 2013).

When applying multiple lines of evidence to the question of species boundaries, one of two things can happen: either all the data point to the same conclusion, in which case congruent lines of evidence strengthen the support for the delimitation, or the data point to different delimitation schemes leaving the researcher to explore causes for incongruence. Ultimately, this can lead to subjectively prioritizing one scheme over another. Fully integrated analyses where multiple data types are combined into a single analysis (e.g. Guillot *et al* 2012, Zapata and Jiménez 2012, Edwards and Knowles 2014, Solis-Lemus *et al* 2014) provide an alternative that can bypass subjectivity in delimitations. Current methods and approaches that incorporate environmental data are limited in methodological scope (at present, only clustering methods are available). Meik et al (2015) argue, that because climatic variables are not intrinsic to the organism, they do not inherently reflect taxonomic signal. Furthermore, the interpretation of overlap between putative lineages based on independent lines of intrinsic and non-intrinsic data could be interpreted differently when combined and evaluated in the same statistical algorithm. Therefore, Meik *et al* (2015) suggest that climatic data should not be considered in clustering and ordination-based methods of integrated species delimitation. To our knowledge, there are no current methods that model the evolution of a species' niche, in the style of iBPP (Solis-Lemus *et al* 2014), which models morphological traits along the guide tree. Here, we have congruent lines of evidence; however, in the case of incongruent

delimitation schemes that utilize environmental data, the integration of multiple types of data will need to be carefully considered.
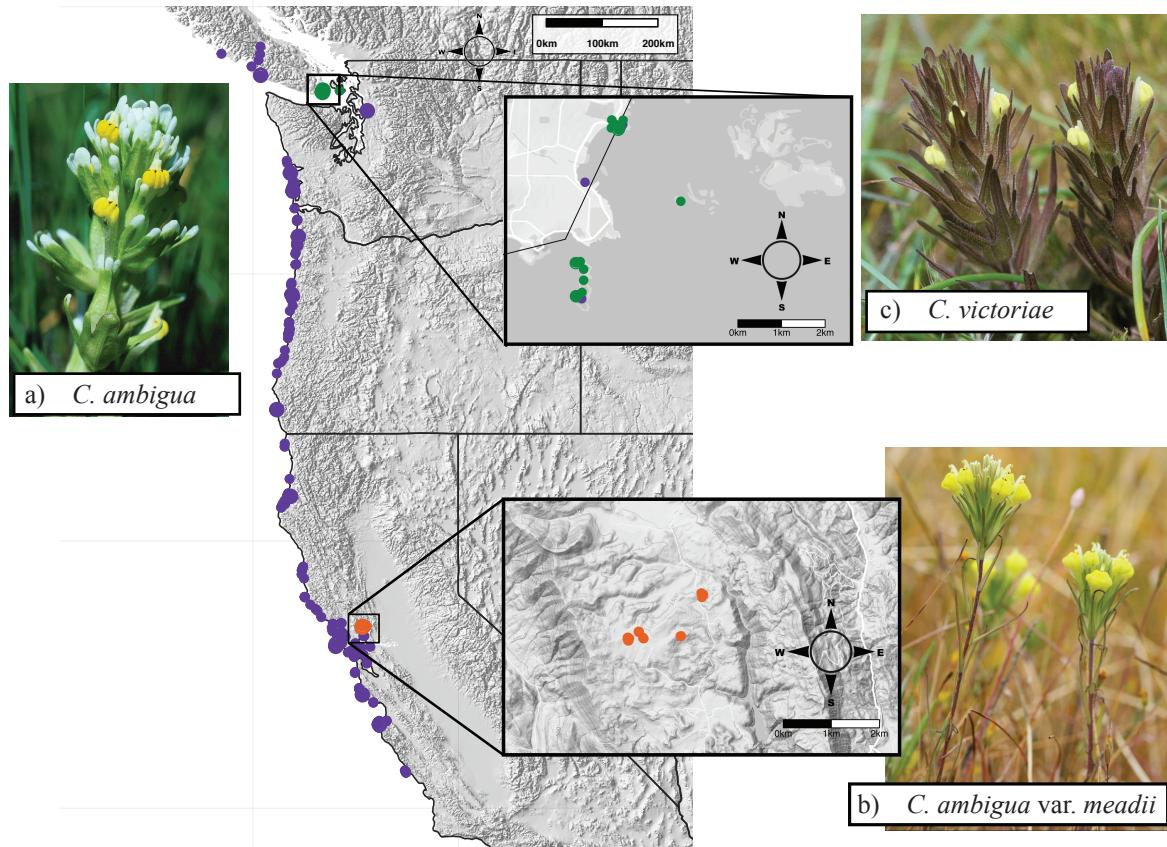
FIGURE 3.1. Distribution of members of the Castilleja ambigua species complex along the western coast of north America. (a) *Castilleja ambigua* (purple); (b) *Castilleja ambigua* var. *meadii* (orange); (c) *Castilleja victoriae* (green). Photographs by J.M. Egger.
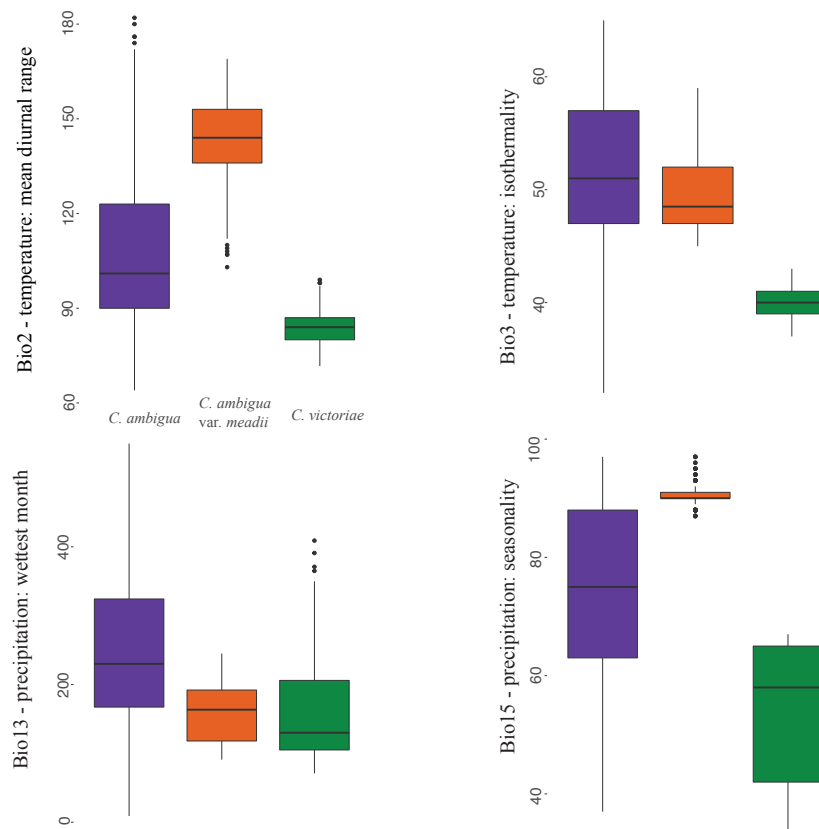
FIGURE 3.2. Distribution of raw climatic variables across the ranges of the three putative species considered here. Here, the range is composed of known occurrence records, in addition to pseudo-absences estimated within 50km of each known occurrence point. *Castilleja ambigua* (purple), *C. ambigua* var. *meadii* (orange), *C. victoriae* (green).
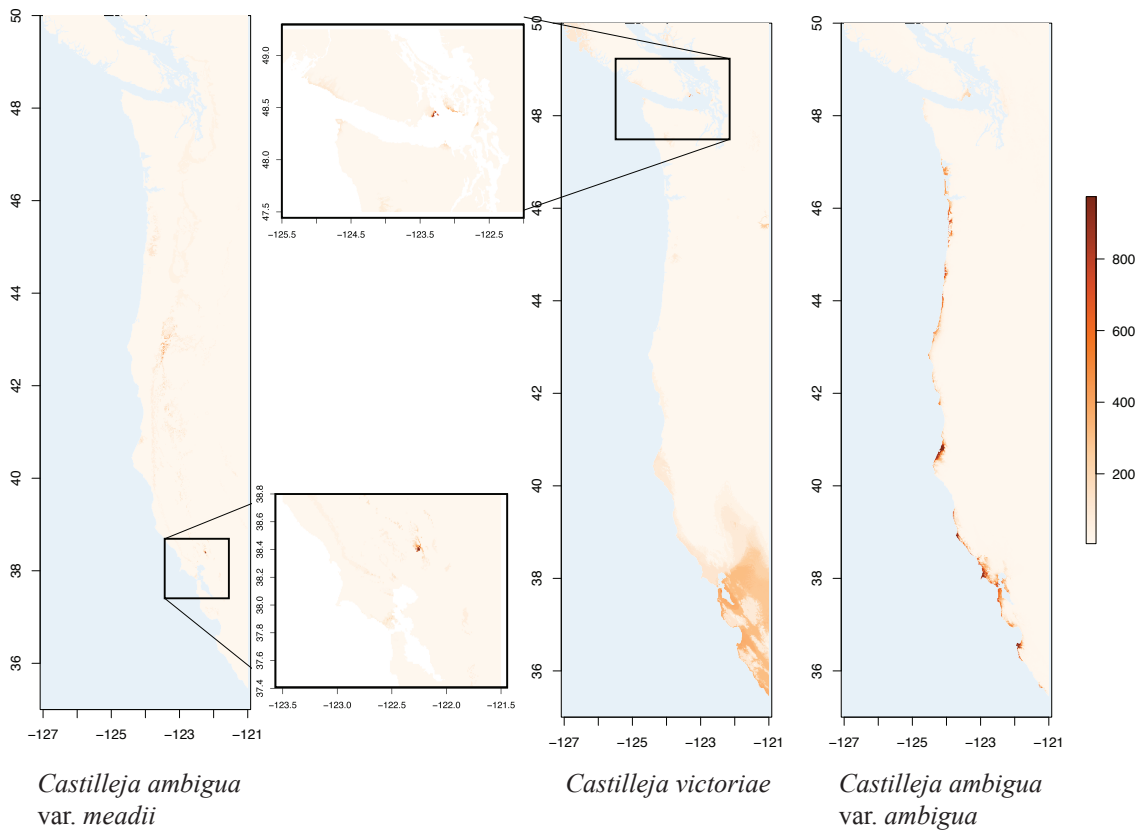
FIGURE 3.3. Species distribution models (SDMs) for each focal taxon. Shaded areas correspond to probability of occurrence where dark red colors correspond to high probabilities, and light orange colors correspond to areas of low probability. Latitude and longitude are given on the y and x axes, respectively. Left panel: *Castilleja ambigua* var. *meadii* and an inset panel zooming in on its known distribution; middle panel: *C. victoriae*, and an inset panel zooming in on its known distribution; right panel: *C. ambigua*.
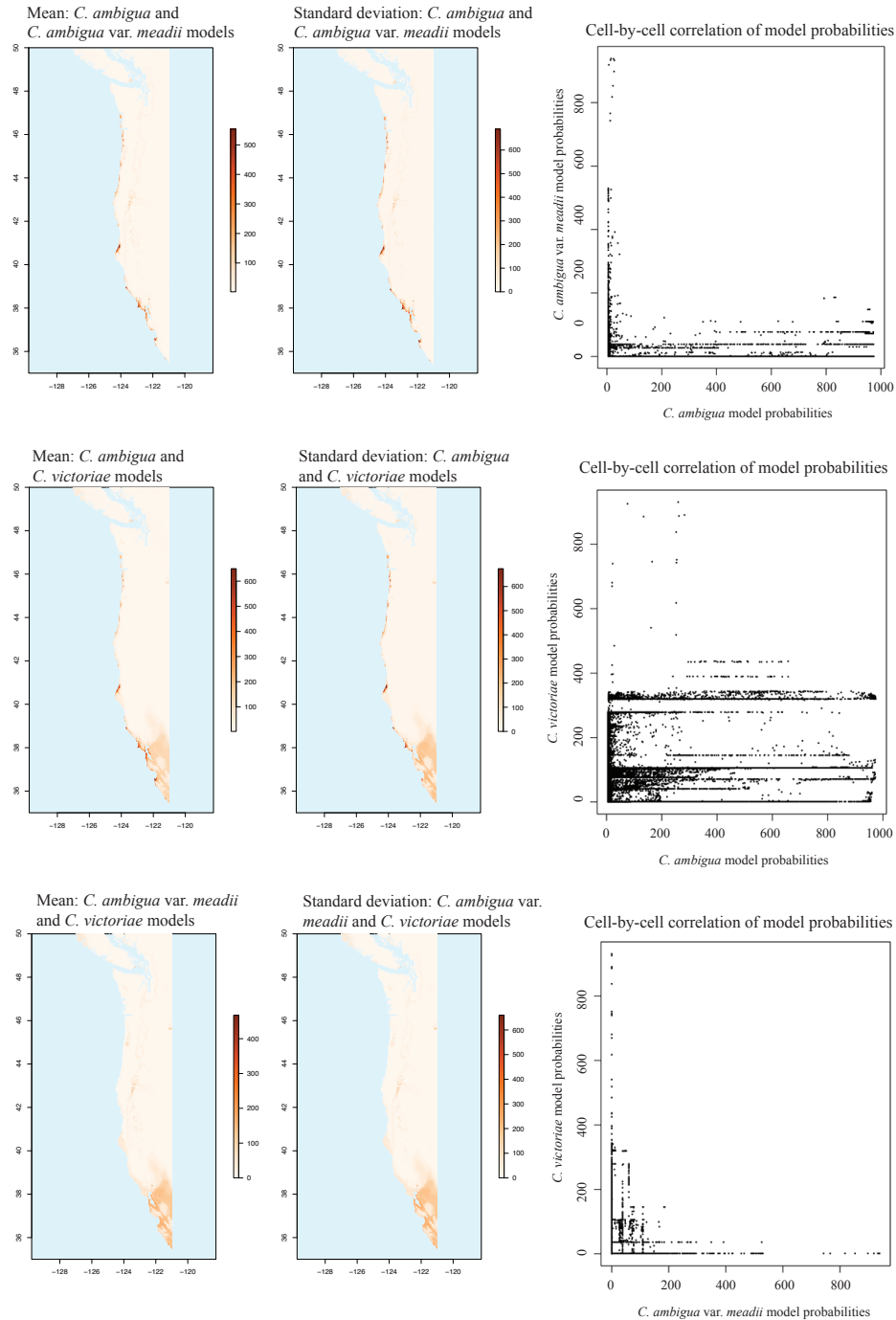
FIGURE 3.4. Mean, standard deviation, and model correlation of the three pairwise comparisons of SDMs for the *Castilleja ambigua* species complex. In each row the first two columns are the the mean (left) and standard deviation (middle) of SDM model probabilities. The third column (right) is a correlation plot of cell-by-cell probability scores. At top, comparisons of *C. ambigua* and *C. ambigua* var. *meadii*; in the middle, comparisons of *C. ambigua* and *C. victoriae*; on the bottom, comparisons of *C. ambigua* var. *meadii* and *C. victoriae*.

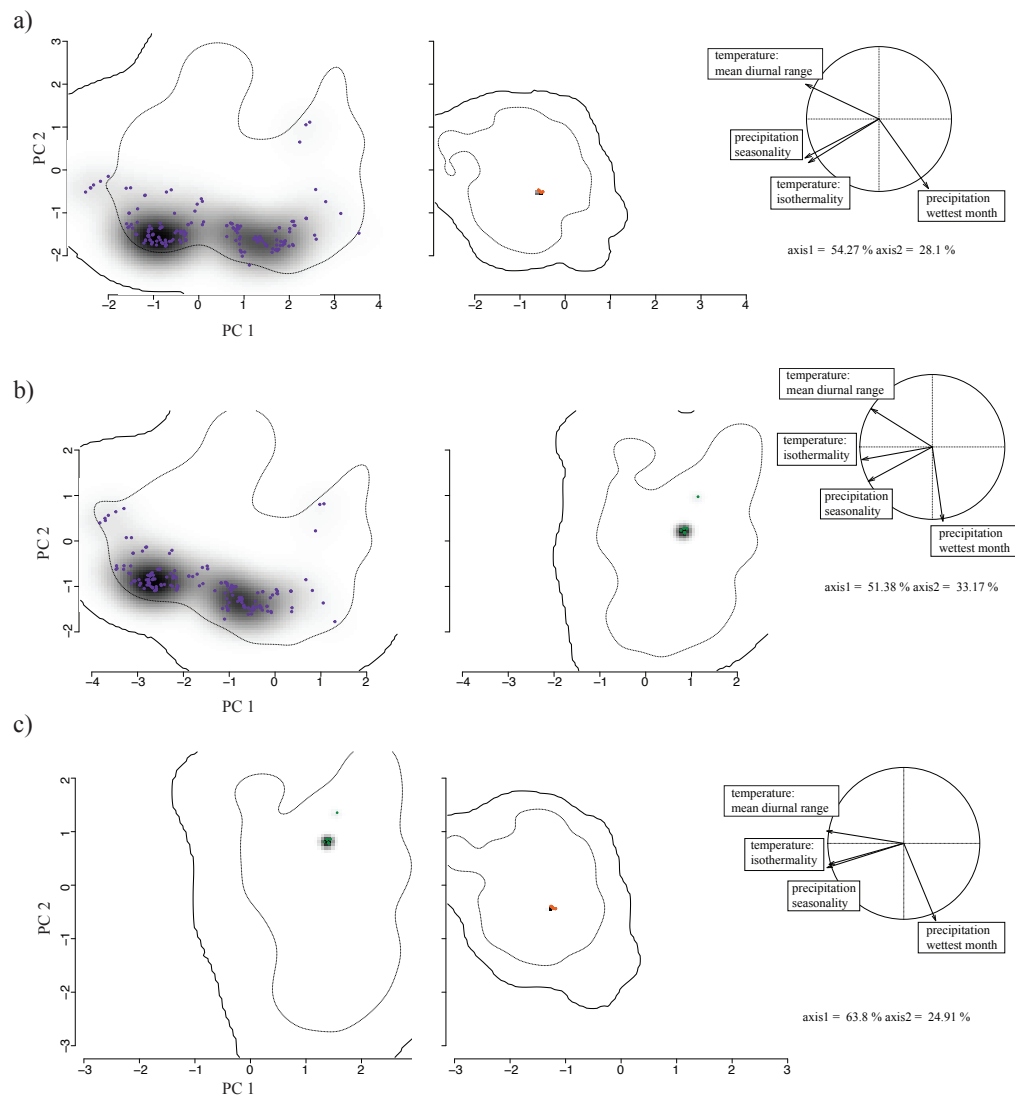FIGURE 3.5. The representation of the niche of each focal taxon in climate space calibrated on each alternative member of the *Castilleja ambigua* species complex. In each pair of panels, the niche of each member of a pair of putative species is plotted along the first two axes of a PCA-env in calibrated on the ranges of both members of the pair. The grey shading shows the density of occurrences of the species by cell. The inner and outer lines indicate 50% and 100% of the available (background) environment, respectively. Within each niche, points corresponding to known occurrences of each entity are plotted in purple (*C. ambigua*), orange (*C. ambigua* var. *meadii*), and green (*C. victoriae*). Next to these plots, we show the contribution of variables to the axes of the PCA and the percentage of inertia explained by the two axes. In panel a) (top) *Castilleja ambigua* (left) and *C. ambigua* var. *meadii* (right); panel b) (middle) *C. ambigua* (left) and *C. victoriae* (right); panel c) (bottom) *C. victoriae* (left) and *C. ambigua* var. *meadii* (right).

a)

C. ambigua SDM:
C. ambigua var. *meadii* (left)
and *C. victoriae* (right)

b)

C. victoriae SDM:
C. ambigua (left)
and *C. ambigua* var. *meadii* (right)

c)

C. ambigua var. meadii SDM:
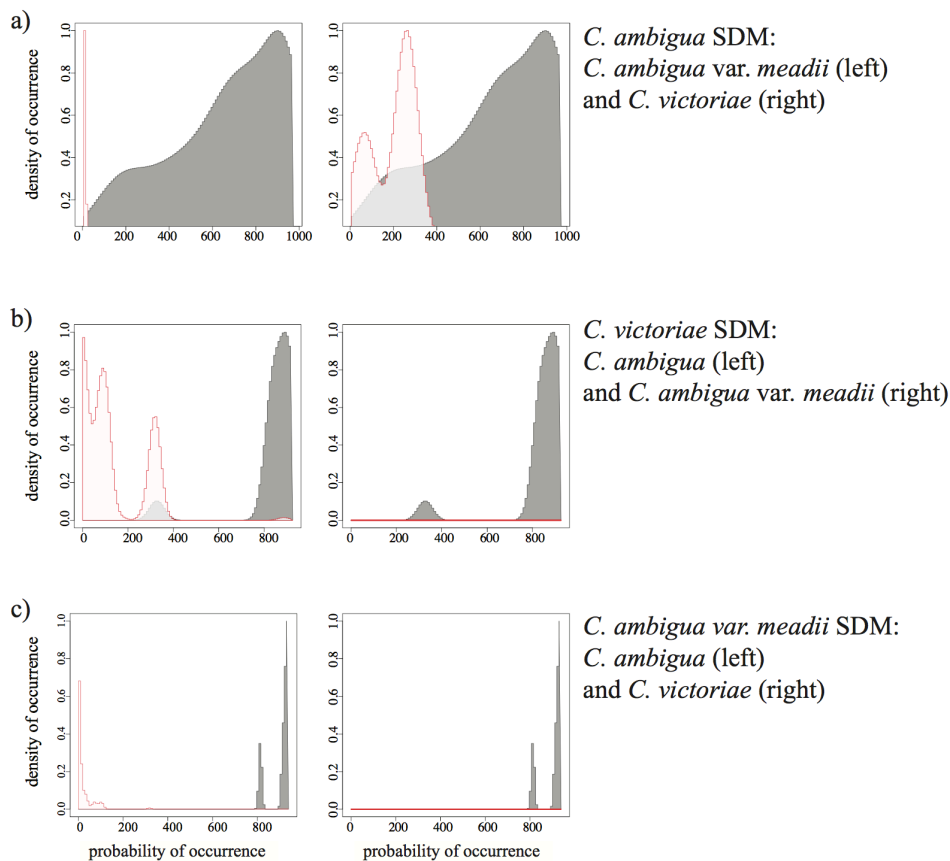*C. ambigua* (left)
and *C. victoriae* (right)

FIGURE 3.6. The representation of the niche of each focal taxon in geographic space calibrated on a single putative species of the *Castilleja ambigua* species complex. The x-axis indicates the distribution of probabilities of occurrence found within the study area; the y-axis indicates the density of occurrence of a given entity at a given probability of occurrence within the study area. In each panel, the densities of the entity whose SDM calibrated the niche is plotted in dark grey and the compared entity is plotted in lighter grey, bordered by red. In panel a) (top) we compare *C. ambigua* (dark grey) with *C. ambigua* var. *meadii* (left) and *C. victoriae* (right); in panel b) (middle) we compare *C. victoriae* (dark grey) with *C. ambigua* (left) and *C. ambigua* var. *meadii* (right); in panel c) (bottom), we compare *C. ambigua* var. *meadii* (dark grey) with *C. ambigua* (left) and *C. victoriae* (right).

TABLE 3.1. Contribution of each variable used in species distribution modeling. These values are averaged across the three-fold cross-validation runs for each of nine different algorithms used in the initial modeling step of our ensemble projection. Bio2 – mean diurnal range (mean of the monthly (max temp-min temp); Bio3 – isothermality (mean diurnal range/annual temperature range); and Bio15 – precipitation seasonality (coefficient of variation); Bio13 – precipitation in wettest month.

	C. ambigua		*C. ambigua* var. *meadii*		*C. victoriae*	
	mean	st.dev	mean	st.dev.	mean	st.dev.
bio2	0.64	0.37	0.75	0.04	0.92	0.10
bio3	0.79	0.02	0.58	0.04	0.23	0.24
bio13	0.23	0.01	0.77	0.04	0.82	0.14
bio15	0.19	0.03	0.21	0.02	0.43	0.32

TABLE 3.2. Quantification of niche overlap in gridded environmental space and subsequent estimates of niche equivalency and niche similarity, based on 1,000 simulations. At left, we report these statistics for e - space, which uses the first two axes of a PCA ordination of climatic variables. At right, we report statistics for g - space, which uses a single vector of predicted probabilities of occurrence derived from a single-species SDM. In each cell, the first value corresponds to the detected niche overlap corrected for climate prevalence (the density of occurrence / total density across entire climate space), and the second value corresponds to the detected overlap without the correction applied. Values with an asterisk are statistically significant.

e-space: calibrated on possible pair of entities

	C. ambigua and C. victoriae	C. ambigua and C. ambigua var. meadii	C. ambigua var. meadii and C. victoriae
Niche Overlap	0.005 ; 0.006	0.000 ; 0.001	0 ; 0
Niche equivalency	0.002 ; 0.002	0.002 ; 0.002	0.002 ; 0.002
Niche similarity			
entity 2 -> 1	0.65335 ; 0.7033	0.70529 ; 0.65335	2 ; 2
entity 1 -> 2	0.6014 ; 0.53746	0.33966 ; 0.31968	2 ; 2

g - space: calibrated on a single entity

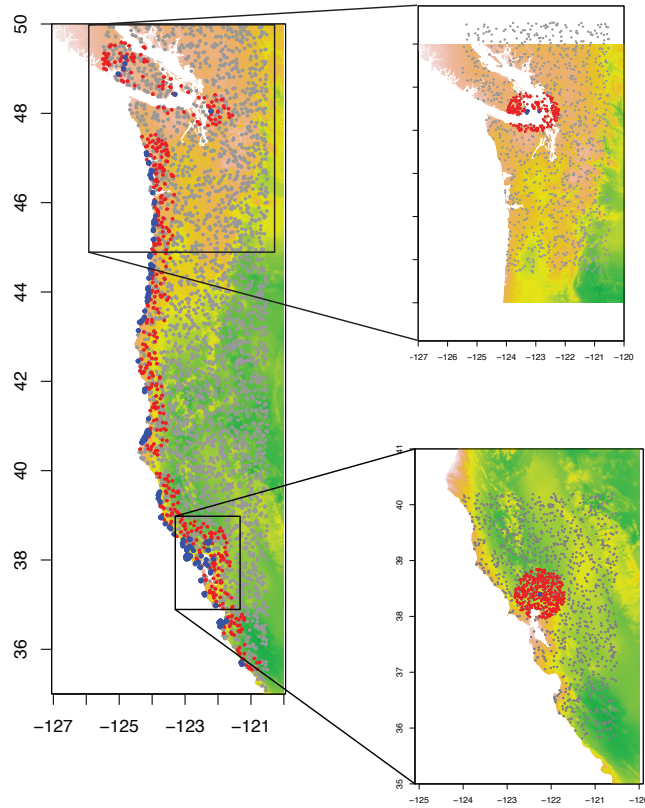|	C. ambigua		C. ambigua var. meadii		C. victoriae	
	C. ambigua var. meadii	C. victoriae	C. ambigua	C. victoriae	C. ambigua	C. ambigua var. meadii
Niche Overlap	0 ; 0.007	0.041 ; 0.215	0 ; 0	0 ; 0	0.442 ; 0.075	0 ; 0
Niche equivalency	0.002 ; 0.002	0.002 ; 0.002	0.002 ; 0.002	0.002 ; 0.002	0.00999 ; 0.00999	0.002 ; 0.002
Niche similarity						
entity 2 -> 1	0.42358 ; 0.42358	0.004 ; 0.004	1.73626 ; 1.73626	2 ; 2	0.002 ; 0.002	1.63437 ; 1.63437
entity 1 -> 2	0.1978 ; 0.1978	0.22777 ; 0.22777	1.3786 ; 0.13786	0.7952 ; 0.7952	0.00799 ; 0.00799	0.35165 ; 0.35165

REFERENCES

Agapow P.M., Bininda-Emonds, O.R.P, Crandall, K.A., Gittleman, J.L., Mace, G.M., Marshall, J.C., and A. Purvis. 2004. The impact of species concept on biodiversity studies. *The Quarterly Review of Biology*. 79: 161-179.

Andersson, L. 1990. The driving force: species concepts and ecology. *Taxon*. 39: 375-382.

Araújo M.B., Pearson R.G., Thuiller W., and M. Erhard. 2005. Validation of species-climate impact models under climate change. *Global Change Biology*. 11: 1504-1513.

Barley, A.J., Brown, J.M., and R.C. Thomson. 2017. Impact of model violations on the inference of species boundaries under the multispecies coalescent. *Systematic Biology*. 0: 1-17.

Broennimann, O., Fitzpatrick, M.C., Pearman, P.B., Petitpierre, B., Pellissier, L., Yoccoz, N.G., Thuiller, W., Fortin, M., Randin, C., Zimmermann, N.E., Graham. C.H., and A. Guisan. 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography*. 21: 481-497.

Carstens, B.C., Pelletier, T.A., Reid, N.M., and J.D. Satler. 2013. How to fail at species delimitation. *Molecular Ecology*. 22: 4369-4383.

Dagnino, D., Minuto, L. and G. Casazza. 2017. Divergence is not enough: the use of ecological niche models for the validation of taxon boundaries. *Plant Biology*. 19: 1003-1011.

de Queiroz, K. 1998. The general lineage concept of species, species criteria, and the process of speciation. In: *Endless Forms: Species and Speciation*. Eds. Howard S.J. and S.H. Berlocher), p. 57-78. Oxford University Press, New York.

de Queiroz, K. 2005. Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences USA*. 102: 6600-6607.

de Queiroz, K. 2007. Species concepts and species delimitation. *Systematic Biology*. 56: 879-886.

Dowell, S.A. and E.R. Hekkala. 2016. Divergent lineages and conserved niches: using ecological niche modeling to examine the evolutionary patterns of the Nile monitor (*Varanus nilotiucs*). *Evolutionary Ecology*. 30: 471-485.

Edwards, D.L. and L.L. Knowles. 2014. Species detection and individual assignment in species delimitation: can integrative data increase efficacy? *Proceedings of the Royal Society B*. 281: 1-7.

Egger, J.M., Ruygt, J.A. and D.C. Tank. 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., Ruffley, M., Smith, M.L., Carstens, B.C., Tank, D.C., and J. Sullivan. 2016. Identifying cryptic diversity with predictive phylogeography. *Proceedings of the Royal Society B*. 283: 1-9.

Fairbarns, M. and J.M. Egger. 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

Fujita, M.K, A.D Leaché, Burbrink, F.T., McGuire, J.A., and C. Moritz. 2012. Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology and Evolution*. 27: 480-488.

Gama, R., Aguirre-Gutiérrez, J., and M. Stech. 2017. Ecological niche comparison and molecular phylogeny segregate the invasive moss species *Campylopus introflexus* (Leucobryaceae, Bryophyta) from its closest relatives. *Ecology and Evolution*. 7: 8017-8031.

Grossenbacher, D.L, Veloz, S.D, and J.P. Sexton. 2014. Niche and range size patterns suggest that speciation begins in small, ecologically diverged populations in North American monkeyflowers (*Mimulus* spp.). *Evolution*. 68: 1270-1280.

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

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*. 25: 1965-1978.

Isaac, N.J.B., Mallet, J. and G.M. Mace. 2004. Taxonomic inflation: its influence on macroecology and conservation. *TRENDS in Ecology and Evolution*. 19: 464-469.

Jacobs, S.J., Kristofferson, C., Uribe-Convers, S., Latvis, M. and D.C. Tank. (*in press*). Incongruence in molecular species delimitation schemes: what to do when adding more data is difficult. *Molecular Ecology*. doi:10.5061/dryad.97d83g0.

Mace, G.M. 2004. The role of taxonomy in species conservation. *Philosophical Transactions of the Royal Society of London B*. 359: 711-719.

Mayr, E. 1942. Systematics and the origin of species, from the viewpoint of a zoologist. Harvard University Press.

McCormack, J.E., Zellmer, A.J., and L.L. Knowles. 2009. Does niche divergence accompany allopatric divergence in *Aphelocoma* jays as predicted under ecological speciation?: insights from tests with niche models. *Evolution*. 64: 1231-1244.

McKelvy, A.D. and F.T. Burbrink. 2017. Ecological divergence in the yellow-bellied kingsnake (*Lampropeltis calligaster*) at two North American biodiversity hotspots. *Molecular Phylogenetics and Evolution*. 106: 61-72.

Meik, J.M., Streicher, J.W., Lawing, A.M., Flores-Villela, O. and M.K. Fujita. 2015. Limitations of climatic data for inferring species boundaries: insights from speckled rattlesnakes. *PLOS One*. 10: 1-19.

Morales, A., Villalobos, F., Velazco, P.M., Simmons, N.B. and D. Piñero. 2016. Environmental niche drives genetic and morphometric structure in a widespread bat. *Journal of Biogeography*. 43: 1057-1068.

Morrison, W.R., Lohr., J.L., Duchen, P., Wilches, R., Trujillo, D., Mair, M. and S.S. Renner. 2009. The impact of taxonomic change on conservation: does it kill, can it save, or is it just irrelevant? *Biological Conservation*. 142: 3201-3206.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. *Nature*. 403: 853-858.

Niemiller, M.L., Near, T.J. and B.M. Fitzpatrik. 2011. Delimiting species using multilocus data: diagnosing cryptic diversity in the southern cavefish, *Typhlichthys subterraneus* (Teleostei: Amblyopsidae). *Evolution*. 66: 846-866.

Padial, J.M., Miralles, A., De la Riva, I. and M. Vences. 2010. The integrative future of taxonomy. *Frontiers in Zoology*. 7: 1-14.

Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D. and A.P. Vogler. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*. 55: 595-609.

Raxworthy, C.J., Ingram, C.M., Rabibisoa, N. and R.G. Pearson. 2007. Applications for ecological niche modeling for species delimitation: a review and empirical evaluation using day geckos (*Phelsuma*) from Madagascar. *Systematic Biology*. 56: 907-923.

Reeves, P.A. and C.M. Richards. 2011. Species delimitation under the general lineage concept: an empirical example using wild North American hops (Cannabaceae: *Humulus lupulus*). *Systematic Biology*. 60: 45-59.

Rissler, L.J. and J.J. Apodaca. 2007. Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). *Systematic Biology*. 56: 924-942.

Ruffley, M., Smith, M.L., Espíndola, A., Carstens, B.C., Sullivan, J. and D.C. Tank. 2018. Combining allele frequency and tree-based approaches improves phylogeographic inference from natural history collections. *Molecular Ecology*. doi:10.1111/mec.14491.

Schlick-Steiner, B.C., Steiner, F.M., Seifert, B., Stauffer, C., Christian, E. and R.H. Crozier. 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual Review of Entomology*. 55: 421-438.

Schoener, T.W. 1968. Anolis lizards of Bimini: resource partitioning in a complex fauna. *Ecology*. 61: 1051-1055.

Solís-Lemus, C., Knowles, L.L. and C. Ané. 2014. Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution*. 69: 492-507.

Sukumaran, J. and L.L. Knowles. 2017. Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Science*. 114: 1607-1612.

Thuiller, W. Lafourcade, B., Engler, R. and M.B. Araújo. 2009. BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography*. 32: 369-373.

Thuiller, W., Georges, D., Engler, R. and F. Breiner. 2016. biomod2: Ensemble platform for species distribution modeling. R-version 3.3-7 https://CRAN.R-project.org./package-biomod2.

Van Valen, L. 1976. Ecological species, multispecies, and oaks. *Taxon*. 25: 233-239.

Viera-Barreto, J.N., Pliscoff, P., Donato, M. and G. Sancho. 2018. Disentangling morphologically similar species of the Andean forest: integrating results from multivariate morphometric analyses, niche modelling and climatic space comparison in *Kaunia* (Eupatorieae: Asteraceae). *Botanical Journal of the Linnean Society*. 186: 259-272.

Warren, D.L., Glor, R.E. and M. Turelli. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution*. 62: 2868-2883.

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.

Yang Z, Rannala B (2014) Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution,* 31, 3125-3135.

Zapata F, Jiménez I (2012) Species delimitation: inferring gaps in morphology across geography. *Systematic Biology*, 61: 179–194.

SUPPLEMENTAL DATA

SUPPLEMENTAL FIGURE S3.1. Maps illustrating the geographic position of background points (grey), pseudo-absences (red), and occurrence (blue) for *Castilleja ambigua* (left panel), *C. ambigua* var. *meadii* (bottom-right panel), and *C. victoriae* (top-right panel). The range of each taxon is estimated by the occurrence records and the pseudo-absences.

SUPPLEMENTAL TABLE S3.1. Occurrence records used for species distribution modeling and niche quantification.

species	Database identifier	latitude	longitude
Castilleja ambigua	OSC-VP-91580	42.453144	-124.425206
Castilleja ambigua	OSC-VP-135174	43.345	-124.3292
Castilleja ambigua	HSC-VP-77506	43.135362	-124.418191
Castilleja ambigua	OSC-VP-54887	43.324414	-124.387176
Castilleja ambigua	HSC-VP-90798	43.438655	-124.23692
Castilleja ambigua	OSC-VP-54879	43.5707	-124.2231
Castilleja ambigua	OSC-VP-24470	43.7964	-124.1475
Castilleja ambigua	OSC-VP-6242	43.886681	-124.111857
Castilleja ambigua	OSC-VP-54880	43.929787	-124.117491
Castilleja ambigua	OSC-VP-44694	43.95232	-124.119067
Castilleja ambigua	OSC-VP-54876	43.971074	-124.099421
Castilleja ambigua	OSC-VP-105489	44.054967	-124.128794
Castilleja ambigua	OSC-VP-84305	44.0932	-124.115
Castilleja ambigua	ID-VP-129728	44.09375	-124.122
Castilleja ambigua	UBC-VP-161646	44.09375	-124.122
Castilleja ambigua	OSC-VP-194855	44.308012	-124.101281
Castilleja ambigua	OSC-VP-56388	44.409124	-124.031903
Castilleja ambigua	OSC-VP-142481	44.515101	-124.065332
Castilleja ambigua	OSC-VP-101265	44.611778	-124.034742
Castilleja ambigua	OSC-VP-142478	44.614398	-124.043527
Castilleja ambigua	OSC-VP-54882	44.8086	-124.0619
Castilleja ambigua	OSC-VP-194936	45.174356	-123.967976
Castilleja ambigua	OSC-VP-194957	45.1808	-123.944
Castilleja ambigua	OSC-VP-54888	45.4083	-123.9583
Castilleja ambigua	OSC-VP-78980	45.372799	-123.967406
Castilleja ambigua	OSC-VP-117673	45.5319	-123.95
Castilleja ambigua	OSC-VP-88560	45.7013	-123.896
Castilleja ambigua	OSC-VP-38085	45.988776	-123.931646
Castilleja ambigua	WTU-VP-134930	46.618499	-124.064295
Castilleja ambigua	WTU-VP-20050	46.54917	-124.02694
Castilleja ambigua	WS-VP-79358	46.642056	-124.05306
Castilleja ambigua	WTU-VP-134805	46.70667	-123.98056
Castilleja ambigua	WTU-VP-134803	46.725435	-124.028629
Castilleja ambigua	WS-VP-79356	46.89959	-124.099407
Castilleja ambigua	WS-VP-79361	46.882527	-124.101014

Castilleja ambigua	WTU-VP-134928	47.07111	-124.16472
Castilleja ambigua	WS-VP-79359	47.116815	-124.17554
Castilleja ambigua	WTU-VP-20048	48.044754	-122.211912
Castilleja ambigua	WTU-VP-134929	48.047059	-122.207159
Castilleja ambigua	UCSC1392	36.6085	-121.9588
Castilleja ambigua	UCD85171	36.632266	-121.74415
Castilleja ambigua	UCD84660	36.632807	-121.747575
Castilleja ambigua	UCD84651	36.640731	-121.751004
Castilleja ambigua	UCSC2497	36.9733	-122.0712
Castilleja ambigua	JEPS82395	36.983334	-122.066666
Castilleja ambigua	JEPS7866	37.227836	-122.410399
Castilleja ambigua	UC1537444	37.212304	-122.404044
Castilleja ambigua	UC406881	37.533699	-122.519124
Castilleja ambigua	GH365411	37.5371	-122.5192
Castilleja ambigua	JEPS9131	37.707499	-122.226184
Castilleja ambigua	JEPS9138	37.755061	-122.213197
Castilleja ambigua	CAS1034779	37.8291	-122.5341
Castilleja ambigua	UC27159	37.842408	-122.551708
Castilleja ambigua	UC27154	37.918724	-122.3855
Castilleja ambigua	JEPS3677	37.947991	-122.601674
Castilleja ambigua	JEPS34409	37.951183	-122.59675
Castilleja ambigua	UC1779741	38.0068	-122.3577
Castilleja ambigua	UC27158	38.022515	-122.141878
Castilleja ambigua	CHSC22508	38.027778	-122.960556
Castilleja ambigua	RSA120306	38.02659	-122.964646
Castilleja ambigua	RSA12228	38.0511	-122.9748
Castilleja ambigua	RSA189646	38.079923	-122.973277
Castilleja ambigua	JEPS77572	38.074852	-122.96424
Castilleja ambigua humboldtiensis	JEPS93142	38.0899	-122.8426
Castilleja ambigua humboldtiensis	JEPS93167	38.090295	-122.842877
Castilleja ambigua	JEPS66550	38.21154	-122.744995
Castilleja ambigua	CAS923156	38.2195	-122.9502
Castilleja ambigua humboldtiensis	JEPS93166	38.219852	-122.9499
Castilleja ambigua	JEPS110630	38.2211	-122.3087
Castilleja ambigua	JEPS110632	38.2211	-122.2904
Castilleja ambigua	JEPS86697	38.2561	-122.94151
Castilleja ambigua	JEPS20322	38.226249	-122.951332
Castilleja ambigua	UC726108	38.26784	-122.79573
Castilleja ambigua	POM179763	38.3103	-122.8532
Castilleja ambigua	GH365401	38.3127	-123.0647

Castilleja ambigua	RSA19277	38.3772	-123.0784
Castilleja ambigua	POM179718	38.3781	-123.0796
Castilleja ambigua	JEPS110631	38.4234	-122.2538
Castilleja ambigua	POM44312	38.4561	-123.0558
Castilleja ambigua	JEPS15564	38.4563	-122.867153
Castilleja ambigua	JEPS20321	38.462137	-122.866185
Castilleja ambigua	UCD122116	38.5767	-123.3365
Castilleja ambigua	JEPS9142	38.694756	-123.42903
Castilleja ambigua	UCD122118	38.7288	-123.4771
Castilleja ambigua	JEPS76862	38.952451	-123.737452
Castilleja ambigua	JEPS76202	38.954382	-123.739488
Castilleja ambigua	JEPS21445	38.95251	-123.736465
Castilleja ambigua humboldtiensis	HSC66431	39.3002	-123.7558
Castilleja ambigua	RSA80859	39.3767	-123.8157
Castilleja ambigua	JEPS17434	39.452555	-123.810365
Castilleja ambigua	UC1178862	39.457441	-123.806719
Castilleja ambigua	UCR131854	39.45333	-123.81
Castilleja ambigua	JEPS9143	39.457134	-123.80487
Castilleja ambigua	JEPS76198	39.507026	-123.783455
Castilleja ambigua	JEPS17465	39.52175	-123.775509
Castilleja ambigua humboldtiensis	OBI9388	40.627823	-124.312938
Castilleja ambigua humboldtiensis	HSC70448	40.6444	-124.3031
Castilleja ambigua humboldtiensis	CHSC68560	40.689722	-124.219167
Castilleja ambigua humboldtiensis	SJSU8462	40.7524	-124.2349
Castilleja ambigua humboldtiensis	HSC95751	40.772791	-124.195421
Castilleja ambigua humboldtiensis	UC1222799	40.807007	-124.144109
Castilleja ambigua humboldtiensis	HSC90264	40.8085	-124.1778
Castilleja ambigua humboldtiensis	HSC90430	40.8102	-124.176
Castilleja ambigua humboldtiensis	HSC82968	40.840593	-124.081631
Castilleja ambigua humboldtiensis	HSC38444	40.8321	-124.1678
Castilleja ambigua humboldtiensis	HSC59457	40.8356	-124.083
Castilleja ambigua humboldtiensis	HSC37568	40.8394	-124.1693
Castilleja ambigua humboldtiensis	JEPS76823	40.8429	-124.1708
Castilleja ambigua humboldtiensis	CHSC19743	40.854444	-124.085
Castilleja ambigua humboldtiensis	HSC95070	40.8561	-124.0988
Castilleja ambigua humboldtiensis	UC1224687	40.851469	-124.081018
Castilleja ambigua humboldtiensis	HSC78957	40.8762	-124.1353
Castilleja ambigua humboldtiensis	HSC20839	40.8898	-124.1421
Castilleja ambigua humboldtiensis	HSC91171	40.8919	-124.1438
Castilleja ambigua humboldtiensis	RSA18017	41.162539	-124.108019

Castilleja ambigua	UC278933	41.76528	-124.22972
Castilleja ambigua	UC1409070	41.824763	-124.225235
Castilleja ambigua	OSC-VP-170774	45.561699	-123.893289
Castilleja ambigua	OSC-VP-185809	45.584166	-123.947455
Castilleja ambigua humboldtiensis	UCR133050	38.1109	-122.891397
Castilleja ambigua	RSA491704	38.113674	-122.945825
Castilleja ambigua	JEPS66221	38.463342	-122.838954
Castilleja ambigua	SBBG34575	38.809129	-123.602057
Castilleja ambigua humboldtiensis	DS186867	40.806	-124.1432
Castilleja ambigua humboldtiensis	HSC32372	40.851	-124.0857
Castilleja ambigua	V-VP-9636	48.999535	-124.862104
Castilleja ambigua	V-VP-30974	49.112691	-124.825154
Castilleja ambigua	V-VP-42869	49.120193	-125.765118
Castilleja ambigua	V-VP-32275	49.253333	-124.816667
Castilleja ambigua	V-VP-62910	48.396815	-123.305572
Castilleja ambigua	V-VP-11812	48.424633	-123.304528
Castilleja ambigua	V-VP-2481	48.451266	-123.266019
Castilleja ambigua	V-VP-46011	48.882792	-125.033106
Castilleja ambigua humboldtiensis	UBC-VP-135665	40.627647	-124.312745
Castilleja ambigua	ID-VP-131829	43.362648	-124.30108
Castilleja ambigua	OSC-VP-6243	43.875031	-124.147976
Castilleja ambigua	HPSU-VP-7046	44.090273	-124.115452
Castilleja ambigua	WTU-VP-188084	44.574325	-123.969513
Castilleja ambigua	WTU-VP-188083	44.785577	-124.072843
Castilleja ambigua	WTU-VP-134809	44.83333	-124.06167
Castilleja ambigua	HPSU-VP-14113	45.52278	-123.88806
Castilleja ambigua	HPSU-VP-7151	46.012147	-123.911903
Castilleja ambigua	HPSU-VP-7045	46.015815	-123.927394
Castilleja ambigua	WTU-VP-193633	46.22538	-123.988629
Castilleja ambigua	WTU-VP-134810	46.882765	-124.103491
Castilleja ambigua	V-VP-83392	48.438431	-123.292811
Castilleja ambigua insalutata	PGM4557	36.590827	-121.964736
Castilleja ambigua	UC886452	35.666303	-121.274125
Castilleja ambigua insalutata	JEPS7430	35.708103	-121.30306
Castilleja ambigua insalutata	OBI32974	35.698606	-121.295152
Castilleja ambigua	RSA603985	36.508397	-121.940263
Castilleja ambigua insalutata	RSA603703	36.508407	-121.938689
Castilleja ambigua insalutata	PGM0376	36.595226	-121.961119
Castilleja ambigua insalutata	PGM0230	36.598681	-121.912281
Castilleja ambigua insalutata	PGM6961	36.578719	-121.862217

Castilleja ambigua insalutata	JEPS75300	36.609054	-121.954601
Castilleja ambigua insalutata	JEPS77459	36.609241	-121.954766
Castilleja ambigua insalutata	UC761831	36.609841	-121.956929
Castilleja ambigua insalutata	GH365421	36.636554	-121.930004
Castilleja ambigua	UC27161	36.62457	-121.916221
Castilleja ambigua insalutata	GH365415	36.625841	-121.916327
Castilleja ambigua insalutata	GH365420	36.6245	-121.915893
Castilleja ambigua insalutata	JEPS9132	36.624057	-121.915079
Castilleja ambigua	UCSC5957	37.235416	-122.415577
Castilleja ambigua	UCSC6159	37.248349	-122.417634
Castilleja ambigua	SBBG40478	37.245895	-122.417992
Castilleja ambigua	SD88447	37.55058	-122.512763
Castilleja ambigua	GH365391	37.871777	-122.307489
Castilleja ambigua	CAS1007814	37.915591	-122.690156
Castilleja ambigua	CAS1024374	38.090704	-122.966138
Castilleja ambigua	CAS928204	38.029501	-122.901404
Castilleja ambigua	CAS928406	38.029418	-122.921986
Castilleja ambigua	CAS928104	38.082765	-122.835705
Castilleja ambigua	CAS525948	38.027732	-122.963945
Castilleja ambigua	GH365404	38.087888	-122.50778
Castilleja ambigua	JEPS9136	38.393354	-122.34504
Castilleja ambigua	UC1334851	38.373254	-122.30536
Castilleja ambigua	GH365403	38.943685	-123.732275
Castilleja ambigua	OBI2912	38.95545	-123.741477
Castilleja ambigua	UCD33136	39.471414	-123.804591
Castilleja ambigua humboldtiensis	GH365414	40.69632	-124.275451
Castilleja ambigua humboldtiensis	HSC90779	40.811495	-124.160496
Castilleja ambigua insalutata	JEPS77474	35.667009	-121.276252
Castilleja ambigua insalutata	JEPS78256	35.667185	-121.271023
Castilleja ambigua insalutata	JEPS77472	35.676251	-121.285504
Castilleja ambigua	GH365407	36.538312	-121.927759
Castilleja ambigua insalutata	PGM3734	36.573715	-121.932479
Castilleja ambigua insalutata	JEPS78253	36.609723	-121.95559
Castilleja ambigua	UCSC5953	38.078395	-122.972616
Castilleja ambigua var. meadii	JEPS	38.39646	-122.25977
Castilleja ambigua var. meadii	JEPS	38.396646	-122.256487
Castilleja ambigua var. meadii	unvouchered	38.397773	-122.257585
Castilleja ambigua var. meadii	unvouchered	38.396905	-122.247623
Castilleja ambigua var. meadii	JEPS	38.404831	-122.242608
Castilleja ambigua var. meadii	JEPS	38.404008	-122.242443

Castilleja ambigua var. meadii	N/A	38.396575	-122.260132
Castilleja ambigua var. meadii	N/A	38.396437	-122.260027
Castilleja ambigua var. meadii	N/A	38.396111	-122.260029
Castilleja ambigua var. meadii	N/A	38.396351	-122.260255
Castilleja ambigua var. meadii	N/A	38.396496	-122.256336
Castilleja ambigua var. meadii	N/A	38.39642	-122.256463
Castilleja ambigua var. meadii	N/A	38.396519	-122.256571
Castilleja ambigua var. meadii	N/A	38.396441	-122.256306
Castilleja ambigua var. meadii	N/A	38.397732	-122.257452
Castilleja ambigua var. meadii	N/A	38.39771	-122.257642
Castilleja ambigua var. meadii	N/A	38.39765	-122.25742
Castilleja ambigua var. meadii	N/A	38.397588	-122.25755
Castilleja ambigua var. meadii	N/A	38.396945	-122.247411
Castilleja ambigua var. meadii	N/A	38.396869	-122.247502
Castilleja ambigua var. meadii	N/A	38.396851	-122.247438
Castilleja ambigua var. meadii	N/A	38.39682	-122.247483
Castilleja ambigua var. meadii	N/A	38.404337	-122.242133
Castilleja ambigua var. meadii	N/A	38.40458	-122.242076
Castilleja ambigua var. meadii	N/A	38.404561	-122.242318
Castilleja victoriae	UBC V29990	48.45106	-123.265982
Castilleja victoriae	UBC V29991	48.450404	-123.268268
Castilleja victoriae	V012317	48.450922	-123.265212
Castilleja victoriae	V013646	48.453465	-123.265211
Castilleja victoriae	V024401	48.405695	-123.306028
Castilleja victoriae	V052880	48.420103	-123.274812
Castilleja victoriae	V062471	48.437777	-123.293279
Castilleja victoriae	V085625	48.43947	-123.292719
Castilleja victoriae	V100273	48.438447	-123.295066
Castilleja victoriae	V133387	48.398337	-123.305477
Castilleja victoriae	V154160	48.40123	-123.304968
Castilleja victoriae	V154164	48.439516	-123.296226
Castilleja victoriae	V162119	48.43751	-123.295442
Castilleja victoriae	WTU 361293	48.427381	-122.889235
Castilleja victoriae	WTU 363026	48.403883	-123.304954
Castilleja victoriae	WTU 363323	48.450512	-123.266386
Castilleja victoriae	WTU 363324	48.437609	-123.296374

SUPPLEMENTAL TABLE S3.2. Quantification of niche overlap in gridded environmental space and subsequent estimates of niche equivalency and niche similarity based on 1,000 simulations using the full dataset (regular text), as well as using only occurrence data (i.e. no background data; bolded text). In each cell, the first value corresponds to the detected niche overlap corrected for climate prevalence (the density of occurrence / total density across entire climate space), and the second value corresponds to the detected overlap without the correction applied. Highlighted in grey with bolded text are measures associated with a reduced background dataset, where only occurrence records are used to quantify the niche. The full dataset values are provided (regular text, not highlighted) for comparison.

| | | e-space: calibrated on possible pair of entities | | |
		C. ambigua and *C. victoriae*	*C. ambigua* and *C. ambigua* var. *meadii*	*C. ambigua* var. *meadii* and *C. victoriae*
previous background	Niche Overlap	0.005 ; 0.006 **0.006 ; 0.001**	0.000 ; 0.001 **0.000 ; 0.001**	0 ; 0 **0 ; 0**
previous background	Niche equivalency	0.002 ; 0.002 **0.002 ; 0.002**	0.002 ; 0.002 **0.002 ; 0.002**	0.002 ; 0.002 **0.002 ; 0.002**
previous reduced background	Niche similarity entity *2 -> 1* entity *1 -> 2*	0.65335 ; 0.7033 0.6014 ; 0.53746 **0.002 ; 0.002** **0.73526 ; 0.73526**	0.70529 ; 0.65335 0.33966 ; 0.31968 **0.004 : 0.004** **0.04595 ; 0.04595**	2 ; 2 2 ; 2 **1.85215 ; 1.85215** **2 ; 2**

SUPPLEMENTAL TABLE S3.3. Quantification of niche overlap in binned geographic space and subsequent estimates of niche equivalency and niche similarity, based on 1,000 simulations using the full dataset (regular text), as well as using only occurrence data (i.e. no background data; bolded text). In each cell, the first value corresponds to the detected niche overlap corrected for climate prevalence (the density of occurrence / total density across entire climate space), and the second value corresponds to the detected overlap without the correction applied. Highlighted in grey with bolded text are measures associated with a reduced background dataset, where only occurrence records are used to quantify the niche. The full dataset values are provided (regular text, not highlighted) for comparison.

g - space: calibrated on a single entity

		C. ambigua		C. ambigua var. meadii		C. victoriae	
		C. ambigua var. meadii	C. victoriae	C. ambigua	C. victoriae	C. ambigua	C. ambigua var. meadii
previous	Niche Overlap	0 ; 0.007	0.041 ; 0.215	0 ; 0	0 ; 0	0.442 ; 0.075	0 ; 0
reduced background		**0.001 ; 0.021**	**0.001 ; 0.021**	**0 ; 0**	**0 ; 0**	**0.023 ; 0.048**	**0 ; 0.008**
previous	Niche equivalency	0.002 ; 0.002	0.002 ; 0.002	0.002 ; 0.002	0.002 ; 0.002	0.00999 ; 0.00999	0.002 ; 0.002
reduced background		**0.002 ; 0.002**	**0.002 ; 0.002**	**0.002 ; 0.002**	**0.002 ; 0.002**	**0.002 ; 0.002**	**0.002 ; 0.002**
previous	Niche similarity entity 2 -> 1	0.42358 ; 0.42358	0.004 ; 0.004	1.73626 ; 1.73626	2 ; 2	0.002 ; 0.002	1.63437 ; 1.63437
previous	entity 1 -> 2	0.1978 ; 0.1978	0.22777 ; 0.22777	1.3786 ; 0.13786	0.7952 ; 0.7952	0.00799 ; 0.00799	0.35165 ; 0.35165
reduced background		**2 ; 2**	**2 ; 2**	**2 ; 2**	**2 ; 2**	**0.0099 ; 0.0099**	**2 ; 2**
		0.40759 ; 0.40759	**0.41159 ; 0.41159**	**0.79321 ; 0.79321**	**0.80919 ; 0.80919**	**0.17982 ; 0.17982**	**0.17582 ; 0.17582**