American Journal of

Botany

INVITED SPECIAL ARTICLE
For the Special Issue: Using and Navigating the Plant Tree of Life

Next-generation lineage discovery: A case study of tuberous *Claytonia* L.

Thomas R. Stoughton[1,4], Ricardo Kriebel[2], Diana D. Jolles[1], and Robin L. O'Quinn[3]

PREMISE OF THE STUDY: Species formation is an intuitive endpoint of reproductive isolation, but circumscribing taxa that arise during speciation can be difficult because of gene flow, morphological continuity, hybridization or polyploidization, and low sequence variation among newly diverged lineages. Nonetheless, species complexes are ubiquitous, and their classification is essential for understanding how diversity influences ecosystem function.

METHODS: We used modern sequencing technology to identify lineages of perennial *Claytonia* L. and assessed correspondence between genetic lineages and morphological/ecological variation. Subsets of 18 taxa from 63 populations were used for (a) lineage discovery using network and coalescent analyses, (b) leaf shape analyses using elliptical Fourier analysis and ordination, and (c) ecological analyses (soil chemistry, climate) using ANOVA and ordination.

KEY RESULTS: Samples mainly aggregated into groups representing each of the previously recognized species in each of the genetic data sets. Compared to the double-digest restriction-site-associated DNA sequencing data set, genome skimming data provided more resolution and further opportunity to probe into patterns of nuclear and chloroplast genome diversity. Morphological and ecological associations are significantly different (albeit intergrading) among the taxa investigated. A new species, *Claytonia crawfordii,* is described based on morphological data presented here.

CONCLUSIONS: Genetic data presented in this study provide some of the first insights into phylogenetic relationships among recently diverged perennial *Claytonia* and are suggestive of past hybridization among caudicose and tuberous species. Given prior difficulties in understanding species boundaries among newly diverged plant lineages, this case study demonstrates the revolutionary breakthrough for systematics research that high throughput sequencing represents.

KEY WORDS   ddRAD; edaphic; elliptical Fourier analysis; genome skimming; Montiaceae; new species; next-generation sequencing; reticulation; statistical ordination; SVD Quartets.

Taxonomic circumscription at and below the species boundary can be as difficult technically (Hamilton and Reichard, 1992; Baum and Donoghue, 1995; Spaniel et al., 2012) as it is controversial (Hey, 2006; Pigliucci, 2003; Stace, 2013; Walsh, 2015; Zachos, 2015). Polyploidization or admixture resulting from hybridization and introgression can contribute to complex patterns of gene flow and phylogenetic diversification, morphological continuity, and low sequence variation among newly diverged lineages (Funk and Omland, 2003). Evidence of recent gene flow, however infrequent, violates the rule of reproductive isolation in the strict sense and

relative reproductive isolation (resulting from introgression) is difficult to quantify in a way that directly translates to taxonomic rank (Hamilton and Reichard, 1992; Cronin, 2006). Regardless of these challenges, groups of actively diverging or recently diverged lineages (i.e., species complexes) likely make up a considerable portion of the standing diversity in plant communities across the globe, undoubtedly have important ecological roles in ecosystems, and should be recognized at some taxonomic level for the purposes of conservation and management. Distinctions among species, metapopulations, and populations (Cronin, 2006) are essential for understanding whether sufficient isolation has occurred, and will be maintained, to warrant species rank. More and more workers (Baum, 2009; de Querioz, 2007) seem to favor the idea that speciation, rather than being a strictly punctuated event, is the gradual establishment of different types of isolation (e.g., phenological, ecological, gametic).

Species delimitation for recently diverged lineages based solely on morphological criteria lost favor among many taxonomists when PCR and Sanger sequencing in biological studies became more available and affordable. Before the advent of high-throughput sequencing technology, genetic data were sparse and therefore most often used simply to corroborate or contradict morphology-based phylogenetic hypotheses. As genetic data have become even cheaper and easier to obtain, molecular phylogenies are used as best approximations of the "true tree", and morphological data are increasingly used in the context of molecular-based hypotheses as post-hoc assessments of character evolution (e.g., adaptation). Morphology alone can be difficult to interpret and today may seem like scant support for species given the relative ease of obtaining molecular data and using modern computational phylogenetic methods to support or refute hypotheses. High-throughput next-generation sequencing (NGS) methods are also still relatively unapproachable for many researchers who identify as field biologists rather than molecular biologists.

Beyond the formidable task of learning a specialized lexicon, NGS methods are not a "one size fits all" for molecular systematics, especially in the case of species complexes. When considering a project, proximity to species boundaries and questions related to the biological integrity of these boundaries tend to dictate which approach is best applied in a given situation (Godden et al., 2012; Lemmon and Lemmon, 2013; Edwards et al., 2015; Zimmer and Wen, 2015; Andrews et al., 2016). A restriction enzyme-based approach such as genotyping by sequencing (Elshire et al., 2011), for example, is likely to be more informative for population genetics or phylogeography studies than alternatives like genome skimming (Cronn et al., 2012; Lemmon and Lemmon, 2013; Edwards et al., 2015; Zimmer and Wen, 2015; Andrews et al., 2016). On the other hand, whole-genome sequencing (e.g., Hillier et al., 2008 or Rubin et al., 2010) may reveal much more about the molecular evolution underlying a particular gene or suite of gene regions under selection than the fragmented genomic sequencing of restriction enzyme-based approaches (Edwards et al., 2015; Zimmer and Wen, 2015; Andrews et al., 2016). In situations where polyploidy or genome size differences may play a significant role in the evolution of a study system, or when large sample numbers ($n > 100$) must be squeezed onto a single NGS lane (i.e., mass-multiplexing of samples), reduced representation methods (e.g., double-digest restriction-site-associated DNA sequencing, as used in part for this study) may be among the few ways to ensure successful reduction of excess genome complexity so that a confident inference regarding relationships among many populations and/or species is possible (Andrews et al., 2016).

There is a trade-off with NGS between sampling breadth (measured both in number of loci and samples) and depth of coverage when it comes to sequencing: maximizing breadth and sacrificing depth (our metric for statistical confidence in a consensus sequence for any particular locus) would result in a less accurate phylogenetic estimate. As noted above, the list of sampling possibilities, alternative methods, and scenarios worth considering for starting a NGS project, including explanation of terms, have been discussed fairly thoroughly by other authors (Godden et al., 2012; Lemmon and Lemmon, 2013; Edwards et al., 2015; Zimmer and Wen, 2015; Andrews et al., 2016), so we focus instead on high-throughput sequencing methodologies as they are applied to species delimitation, a topic that has not yet been broadly addressed with regard to effectiveness and feasibility. Efforts by plant systematists to sample and sequence greater taxonomic breadth across the plant tree of life (Palmer et al., 2004; Matasci et al., 2014; Wickett et al., 2014) has yielded precious resources: genomic reference sequences for a variety of model and nonmodel plant species that can be used to explore microevolutionary patterns and processes. Delineating taxa at the specific and infraspecific level will be increasingly possible, allowing conservation botanists to better understand, identify, and manage taxa.

The genus *Claytonia* L. (Montiaceae, order Caryophyllales) is one of many plant lineages that underwent a recent radiation during or after the Miocene (crown age ca. 10–20 million years; Jeffers, 2015), and this diversification is likely associated with allopolyploidization (Miller and Chambers, 2006). Other examples of hybridization- and polyploidy-driven radiations in plants include the Hawaiian silverswords (*Argyroxiphium* D.C.; Barrier et al., 1999), North American *Silene* L. (Popp and Oxelman, 2007), *Penstemon* Schmidel (Wolfe et al., 2006), and the mustard family (Brassicaceae Burnett; Marhold and Lihová, 2006). Accurate estimates of standing biodiversity for these complex groups and others are needed because species are essential metrics in ecological studies that either use biodiversity as an experimental variable or aim to explain complex species interactions. In addition, many analytical methods for phylogenetic estimation require thoughtful sampling of taxa to estimate diversity parameters such as fixation indices with high accuracy. These phylogenetic analyses will be greatly improved by more precise taxonomic classifications, which will in turn enable accurate biodiversity assessments. Last, recognition of discrete lineages will advance our understanding of evolutionary processes and support sound conservation and decision-making by land managers. Extensive knowledge of morphological continuity in the context of both natural histories and genetic diversity appears to be the key factor limiting progress. For allopolyploid species complexes such as *Claytonia*, use of multiple lines of evidence, including high-throughput sequencing, geometric morphometrics, and natural history, promises to increase the accuracy and precision of lineage delimitation even when boundaries are somewhat obscured by phylogenetic reticulation.

Challenges associated with species delimitation are a hallmark of the genus *Claytonia*. Stoughton and Jolles (2013) discussed problems associated with the circumscription of *C. lanceolata* Pursh and ultimately recommended that its taxonomic classification be revisited. Subsequently, Stoughton et al. (2017) recognized eight taxa, one of which was treated previously in California (*C. lanceolata*) by Miller and Chambers (2006). New taxa were recognized on the basis of morphological, ecological, and DNA (Sanger) sequence data. Given what is known about lineage sorting and other causes of homoplasy in gene phylogeny estimates, consensus methods and hypothesis tests may perform better with increased and

random sampling of genetic data. Our acquisition of NGS data using double-digest restriction site-associated DNA sequencing (ddRAD; see Peterson et al., 2012) and genome skimming (SKIM; see Straub et al., 2012) allows us to more accurately circumscribe species groups in a phylogenetic context. More importantly, these new genetic data may also prove useful for illuminating sister relationships (i.e., resolving internal branches), which Stoughton et al. (2017) were previously unable to address due to limited sequence variation (i.e., 98 parsimony-informative characters out of 639 total base pairs of sequence) in a single nuclear region, nrITS.

The purpose of this study was to assess the utility of a method for reference-guided assembly of phylogenetic markers for simple lineage discovery using both ddRAD and low-coverage genome skimming (SKIM) high-throughput sequencing data. We determined whether (1) the degree of genetic differentiation and (2) patterns of variation in both cauline leaf morphology and ecological properties (soil chemical and climatic) reflect the patterns of lineage diversification inferred from the genetic data. We assessed these relatively new methods using *Claytonia* because it is a species-rich lineage that includes numerous polyploid species complexes (Miller and Chambers, 2006) and has very recently been reclassified based on multiple sources of evidence (Stoughton et al., 2017). In demonstrating these methods with a case study, including the description of a new species of *Claytonia*, we hope to contribute to an ongoing conversation about which genomic and morphological sampling approaches are best for studies concerning different kinds of biological problems (especially resolution of species complexes) and to encourage the adoption of powerful new methodologies.

MATERIALS AND METHODS

Taxon sampling

Our study required broad taxon sampling (Appendix S1; see Supplemental Data with this article) to assess the effectiveness of our methods at different phylogenetic depths. Therefore, 63 natural populations were sampled for morphological, ecological, and/or genetic study, representing 15 species (18 taxa) of tuberous and caudicose perennial *Claytonia* (our ingroup). Guided by results of a tribal-level phylogenetic study by O'Quinn and Hufford (2005), five outgroup samples were selected for the genetic analyses, namely, *Calandrinia menziesii* (Hook.) Torr. & A.Gray, *Lewisia rediviva* Pursh, *Montia parvifolia* (Moc. ex DC.) Greene, *Claytonia saxosa* Brandegee, and *Claytonia sarmentosa* C.A.Mey. Samples selected to represent our ingroup have unknown chromosome numbers, but previous research provided chromosome counts for some of these species ranging from $2n = 12$ to $2n = $ ca. 191 (for summary of references, see Miller and Chambers, 2006), indicating that they represent multiple polyploid complexes. Preliminary flow cytometry data suggest that many of the Californian taxa included in this study are polyploids (T. Stoughton, unpublished data), but no chromosome counts have been obtained for our sample set.

DNA isolation, NGS library preparation and sequencing

Total genomic DNA samples used for this study were isolated from 70 silica-dried leaf and/or stem samples obtained directly from 39 natural populations representing 17 ingroup taxa in addition to five outgroup samples (75 total). To estimate lineage diversity and phylogenetic relationships, multiple samples per population were included for six species [*C. lanceolata*, *C. obovata* Rydberg, *C. panamintensis* T.R.Stoughton, *C. peirsonii* (Munz & Johnston) T.R.Stoughton, *C. serpenticola* T.R.Stoughton, and *C. umbellata* S.Watson], and multiple populations per species were included for seven (*C. lanceolata*, *C. obovata*, *C. panamintensis*, *C. peirsonii*, *C. tuberosa* Pall. ex Willd., *C. umbellata*, and *C. virginica* L.) of the 17 ingroup taxa selected for genetic study. DNA was isolated from samples using minor modifications of the cetyltrimethylammonium bromide (CTAB) method of Doyle and Doyle (1987), mainly differing in the inclusion, exclusion, or reduction of the use of phenol, chloroform, and/or ammonium acetate for DNA purification, with two exceptions. First, two *Claytonia peirsonii* subsp. *peirsonii* DNAs were extracted using Power Plant Pro DNA Isolation Kits (MO BIO Laboratories, Carlsbad, CA, USA) following instructions from the manufacturer. Second, nine samples (indicated by asterisks [*] in Appendix S1) were isolated at Global Biologics L.L.C. (Columbia, MO, USA) using magnetic bead DNA isolation; the rest of the samples were isolated at Rancho Santa Ana Botanic Garden (Claremont, CA, USA) using the modified CTAB method.

Following total genomic DNA isolation, Global Biologics prepared double-digest restriction site-associated DNA sequencing (ddRAD) libraries, with standardization according to the procedure of Peterson et al. (2012) with minor modifications, using EcoRI and MspI enzymes. Uniquely barcoded ddRAD libraries were sequenced on an Illumina HiSeq 2500 (Illumina, San Diego, CA, USA). Two lanes of single-end (1×100) and one lane of paired-end (2×100) reads were obtained. FastQ files generated during sequencing were decompressed and de-multiplexed (i.e., sorted by molecular barcode) using pyRAD v. 2.1.7 (Eaton and Ree, 2013). Data from these three independent runs were combined for downstream ddRAD data assembly and phylogenetic marker selection. In addition to ddRAD libraries, standard genome skimming (SKIM) libraries were prepared in similar fashion to the procedure of Straub et al. (2012) and paired-end sequences (2×150 and 2×250) were obtained from each of two separate sequencing lanes using "Rapid Run Mode" on an Illumina HiSeq 2500. Information on sequencing results for individual samples is presented in Appendix S2.

Genomic data assembly

To provide an easy, scalable, and reproducible method for phylogenetic marker selection using ddRAD and SKIM data, we employed a reference-based assembly approach to locus identification using the recently published nuclear genome of *Beta vulgaris* L. (Dohm et al., 2014), another member of the order Caryophyllales Juss. ex Bercht. & J.Presl. Assemblies and alignments were conducted on a MacBook Pro computer with a 2.9 GHz Intel Core i7 processor and 8 gigabytes memory (Apple, Cupertino, CA, USA) using Geneious v. 9.0.2 (Kearse et al., 2012), whereas network and species tree estimates were generated on an iMac computer with a 4 GHz Intel Core i7 processor and 32 gigabytes memory (Apple). Reference-guided assemblies were conducted with the Geneious assembler to build contigs without further iteration using the default Medium-Low Sensitivity / Fast setting and utilizing the "Map Reads to Each Reference Sequence" workflow to iteratively assemble reads from each sample to each of the references (*Beta vulgaris* nuclear genome partitioned into nine references corresponding to nine chromosomes). Areas of the contigs with depth of coverage ≥10 sequence reads were retained for consensus sequence generation following removal of the reference sequence. Standard ambiguity codes were used at polymorphic

nucleotide positions, and each position was called using 75% majority rule. Consensus sequences for each sample, corresponding to each of the nine chromosomes of *Beta vulgaris*, were concatenated and these consensus sequences were retained for alignment and phylogenetic estimation. Alignments were generated using the MAUVE plugin for Geneious (Darling et al., 2004) and concatenated after the "strip alignment columns" tool was applied to standardize alignments at no more than 50% missing data. Additionally and for the SKIM data set only, whole-chloroplast alignments were constructed using the same methods as for the nuclear data but using an incomplete chloroplast reference obtained as bycatch from transcriptome data developed for *Claytonia virginica* by Michael Moore (Oberlin College, unpublished data). Features of the assemblies and alignments analyzed for this study, including information for individual chromosomal and plastome assemblies for each of the genetic data sets, are presented in Appendices S3–S5.

Species correspondence

The primary objective of this study was not to delimit species (discussed in great detail by Wiens [2007], and references therein), but rather to investigate whether our method for phylogenetic marker selection shows significant support for a pre-existing species-level classification. In essence, we estimated species correspondence sensu Stoughton et al. (2017) and Miller and Chambers (2006). Using the simple statistical criterion of bootstrap support extended from Felsenstein (1985) and discussed by White et al. (2014), statistical support for classification of these species was assessed at the 95% confidence level. Our ddRAD and SKIM nuclear data matrices were both analyzed using two different methods to assess correspondence for species (i.e., clustering of individual samples for a given taxon). First, we used SplitsTree v.4.13.1 (Huson and Bryant, 2006) to create a species network (excluding outgroups) with 1000 bootstrap replicates using the NeighborNet algorithm. Second, SVDQuartets analyses (Chifman and Kubatko, 2014) were implemented in PAUP* v. 4.0a147 (Swofford, 2002) to estimate species trees (including outgroups) under a multispecies coalescent model. Informed by the results of Leaché et al. (2015), SNPs were analyzed in the context of invariant and autapomorphic (i.e., parsimony uninformative) sites for all analyses. Ambiguities were treated as missing data in the alignments. One hundred thousand quartets were randomly sampled and these quartets were assembled automatically in PAUP* using the QFM algorithm (Reaz et al., 2014). Nonparametric bootstrapping with 1000 replicates was used to measure uncertainty in the estimation and these replicates were summarized in 70% majority-rule consensus trees. The networks resulting from our NeighborNet splits analyses, and the 70% majority-rule consensus trees from our multispecies coalescent analyses, were then visualized, and consistent sample clustering was assessed among both data sets and in comparison to the existing taxonomic classifications. Correspondence was also assessed between nuclear and chloroplast phylogenies estimated from the SKIM data using SVDQuartets and the same set of parameters.

Morphological differentiation

A single leaf was collected from each of 80 or more individual plants at each of 20 wild populations representing 13 taxa of tuberous *Claytonia*, with most populations sampled being geographically isolated from others (Appendix S1). Tuberous *Claytonia* typically have a single pair of opposite cauline leaves, which allowed us to select a single leaf from a homologous position on each individual for comparison across all individuals/taxa. Leaves were collected only from individuals that were flowering or past flowering to ensure that leaves were fully developed at the time of sampling. Following excision from the plants using surgical scissors (applied at the attachment point of leaf to stem), leaves collected from each individual were placed flat on newspaper and dried in a plant press using standard procedures for plant collecting described by Ross (1996). Leaves were flattened and dried because they often are naturally curved (e.g., when the petiole is erect but the blade is not), preventing accurate imaging of the leaf outlines. Dried leaves were spread onto a white 11 × 17-inch herbarium sheet along with a centimeter scale, and these were imaged using a HerbScan flatbed scanner (HerbScan Engineering, Chertsey, UK), with a few exceptions photographed from above with a hand-held camera while set against white paper. High resolution scans and photographs were converted to binary (black and white) images using either GIMP 2.8 (http://www.gimp.org) (Solomon, 2009) or Adobe Photoshop software (Adobe, San Jose, CA, USA), and these images were then used to create digital extractions of individual leaves, each representing a different individual plant from within a population. Extracted images were numbered sequentially by population. This procedure was repeated for each of the 20 populations included in the morphological portion of this study (Appendix S1), producing 2718 leaf images. A subset of these images (the first 100 per population, or 100 divided evenly among multiple populations when possible) were analyzed for 1288 images to maintain even sampling across taxa.

Analyses of morphological data were conducted in R version 3.1.2 (R Core Team, 2014) and can be summarized in a series of steps. First, the outlines of the 1288 leaves were imported into R, and coordinates describing each were extracted using Momocs (Bonhomme et al., 2014). Second, each leaf outline was centered, and four landmarks were placed around the leaf (one each at the leaf apex, leaf base, and in the middle of the leaf on each side) to ensure that no twisting of the outlines occurred during downstream analyses. Third, elliptical Fourier analysis (EFA) was conducted to calculate the harmonic coefficients, which were then used for principal component analysis (PCA) to summarize overall leaf shape variation within and among taxa. Shape variation was then visualized in bivariate morphospace. Axes of 50% confidence ellipses for each taxon were plotted onto the ordination plot to visualize the distribution of leaf shape variation among taxa using ggplot2 and Momocs (Bonhomme et al., 2014; Wickham, 2009). The PC scores were also used to test whether the mean and variance of leaf shape differed among taxa by conducting multivariate analysis of variance (MANOVA) in Momocs using taxon identity as a grouping factor. All taxa were compared in pairwise fashion simultaneously for statistical significance. In addition to the PCA and MANOVA, linear discriminant (DA) and discriminant prediction analyses were conducted using the MASS::lda (Venables and Ripley, 2002) and stats::predict (R Core Team, 2014) functions, respectively, to determine how effectively taxa could be identified using leaf shape alone. Scripts and leaf outlines used to conduct the EFA are deposited in Dryad (Stoughton et al., 2018).

Ecological differentiation

To explore the association between edaphic specificity and reproductive isolation, soils were collected at tuber depth for one individual at each of 30 natural populations (32 total samples), representing six species (Appendix S1). Sites for soil collection were

selected with the aim of representing the full range of slope and aspect diversity within populations and the array of substrate types that plants grow on throughout species' geographic distributions. The tuberous *Claytonia* species selected for this study often occur on colluvium associated with unstable talus slopes, making contamination by different soil horizons an undesirable possibility during soil collection, but we presume this artifact would have affected all samples more or less equally. With one exception, 4–8 populations were sampled per species. For two species, soils were collected in different years from the same geographic area (Appendix S1) to assess the consistency of our soil measurements.

Soil samples were analyzed by the Texas A & M University Soil, Water, and Forage Testing Laboratory (College Station, TX, USA). Thirteen soil properties were assayed (Appendix S6), including pH, nitrate levels (NO_3^-), electrical conductivity (Con), major plant available soil nutrients (P, K, Ca, Mg, and S), plant available soil micronutrients (Cu, Fe, Mn, and Zn), and sodium (Na). To expand our ecological assessment, two climatic variables were included to represent differences in temperature and moisture availability among natural populations (mean diurnal temperature range and annual precipitation). Climatic variables were obtained as 30-s raster layers from WorldClim 1.4 (Hijmans et al., 2005) and extracted for the 32 soil sample localities using the raster package (Hijmans, 2016) in R. Before statistical analyses, soil and climate variables were examined for correlation, multinormality, and outliers. Correlated variables were removed from the data set, and the remaining 14 variables (12 soil and 2 climate variables) were log-transformed. First, an ANOVA and Tukey's honestly significant difference (HSD) test were conducted for each soil and climatic variable to determine whether species exhibited significant differences in means. PCA was conducted using the stats::princomp function to determine whether suites of ecological variables can be used to distinguish among *Claytonia* species. Discriminant analyses (DAs) were conducted using the MASS::lda function (Venables and Ripley, 2002) to determine which variables, if any, could be used to most accurately predict where species occur. Discriminant analyses were conducted both with and without climatic variables included. The DA ordinations were visualized as bivariate plots to examine relationships among taxa that were maximized by DA and to expose differences supported by all or a subset of the soil chemical and climatic variables measured. Clustering of taxa in the ecological DA ordinations was visually compared with clusters inferred from the EFA ordination of leaf variation and the phylogenetic relationships inferred from multi-locus DNA sequences. We also used the stats::predict function to determine how effectively samples were classified using DA prediction similarly as with the morphology data. The number of correctly predicted samples was divided by the total number of samples for that taxon and a χ^2 test was used to determine whether the DA diagnosed the taxonomic identity of our samples with greater accuracy based upon ecological characteristics in comparison with random assignment. Scripts and soils data used to conduct the ecological analyses are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.m99v4 (Stoughton et al., 2018).

RESULTS
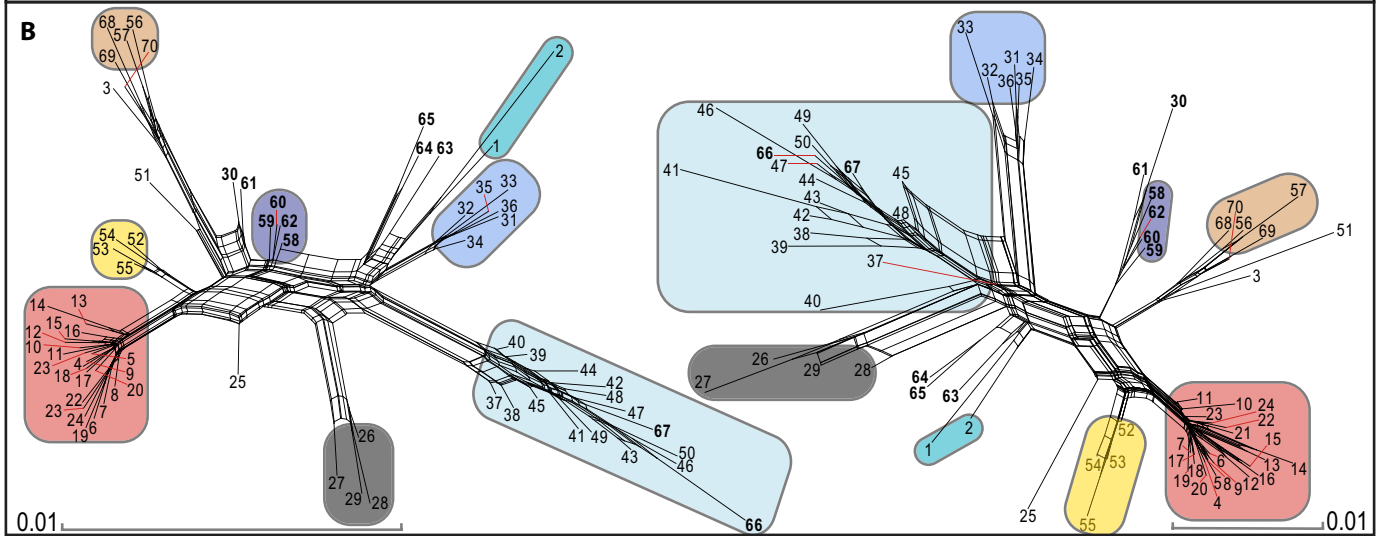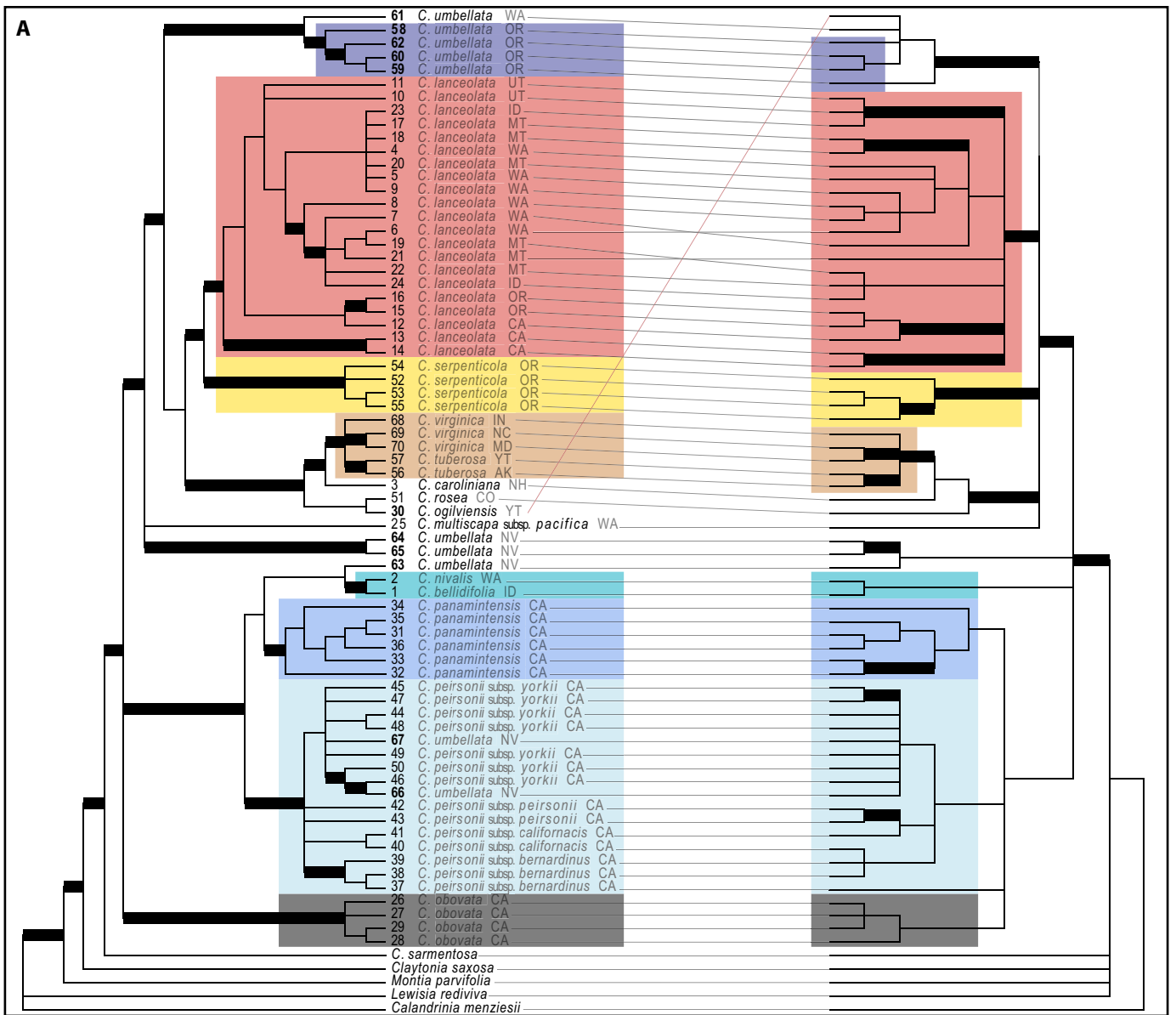
Genomic data assembly

Summary statistics from de-multiplexing sequencing libraries are reported in Appendix S2. A total of 295,808,182 sequence reads were produced across the three ddRAD sequencing lanes, about half of which were paired-end, and 532,597,344 paired-end sequence reads were produced across the two SKIM sequencing lanes. A skewed distribution of sequence reads was obtained from the ddRAD data with the most deeply covered sample (*Montia parvifolia*; outgroup) containing 31,848,680 reads and the most shallowly covered sample (*Lewisia rediviva*; outgroup) comprising 25,626 reads. The distribution of SKIM reads was considerably less skewed (Appendix S2). On average, we obtained 3,944,109 (± 4,316,404 standard deviation) reads/sample for the ddRAD data set, and 7,101,298 (± 1,881,101 standard deviation) reads/sample for the SKIM data set.

Reads from each sample were assembled to the references for all nine chromosomes of *Beta vulgaris*, resulting in a wide range of concatenated assembly lengths (10,995–1,420,829 bp and 1,340,592–3,181,685 for ddRAD and SKIM data sets, respectively) for different samples using the ≥10 read coverage threshold (Appendices S3, S4). SKIM plastome assemblies were less variable in length (Appendix S4). Nuclear genome assemblies were aligned across all samples by *Beta* chromosome number, and these alignments were concatenated resulting in a total concatenated alignment 15,035,340 bp in length for the ddRAD data set, and 31,425,399 bp in length for the SKIM data set, before standardization at ≤50% permissible missing data (Appendix S5). The final ddRAD alignment analyzed for this study was 73,574 bp in length and contained 10,621 parsimony informative characters (PICs, Appendix S5). The final SKIM alignments were 889,625 bp in length with 111,328 PICs and 121,275 bp in length with 3815 PICs for nuclear and chloroplast alignments, respectively.

Species correspondence

Multi-species coalescent trees and splits networks estimated from the ddRAD and SKIM nuclear data sets are largely congruent (Fig. 1), grouping samples in a pattern consistent with the species circumscriptions of Stoughton et al. (2017) and Miller and Chambers (2006), with the exception of *Claytonia umbellata*. In general, the ddRAD data set provided moderate support for a monophyletic *Claytonia* (82 BS), while the SKIM nuclear data set showed strong support for monophyly of *Claytonia* (100 BS) and *Montia parvifolia+Claytonia* (100 BS). SKIM nuclear data provided further support for caudicose and tuberous perennial *Claytonia* being monophyletic (100 BS) and sister to rhizomatous perennials (*C. sarmentosa*, 99 BS), with all of the perennials together being sister to the annual species (*C. saxosa*, 100 BS). These results mirror findings of a broad study of the Montieae Dumort. by O'Quinn and Hufford (2005). *Claytonia obovata*, *C. panamintensis*, *C. serpenticola*, *C. lanceolata*, and *C. peirsonii* are each strongly supported as distinct species (100 BS) in the SKIM analysis, whereas only *C. serpenticola* and *C. lanceolata* are strongly supported (100 BS) by the ddRAD data set. *Claytonia panamintensis* (79 BS), *C. obovata* (76 BS), and *C. peirsonii* (BS 74) were moderately supported by the ddRAD data except that one sample of *C. peirsonii* subsp. *bernardinus* T.R.Stoughton was not resolved. In the RAD data set, *C. virginica* was not resolved as monophyletic but is strongly supported as being allied with *C. tuberosa* (100 BS) and nested in a clade including *C. caroliniana* Michx. and *C. rosea* Rydberg (100 BS), while the SKIM data set resolved the species in this clade together with *C. ogilviensis* McNeill (95 BS). *Claytonia bellidifolia* Rydberg and *C. nivalis* English, the two caudicose species in the sample set and members of the *C. megarhiza* (A.Gray) Parry ex S.Watson complex

A

61	*C. umbellata*	WA
58	*C. umbellata*	OR
62	*C. umbellata*	OR
60	*C. umbellata*	OR
59	*C. umbellata*	OR
11	*C. lanceolata*	UT
10	*C. lanceolata*	UT
23	*C. lanceolata*	ID
17	*C. lanceolata*	MT
18	*C. lanceolata*	MT
4	*C. lanceolata*	WA
20	*C. lanceolata*	MT
5	*C. lanceolata*	WA
9	*C. lanceolata*	WA
8	*C. lanceolata*	WA
7	*C. lanceolata*	WA
6	*C. lanceolata*	WA
19	*C. lanceolata*	MT
21	*C. lanceolata*	MT
22	*C. lanceolata*	MT
24	*C. lanceolata*	ID
16	*C. lanceolata*	OR
15	*C. lanceolata*	OR
12	*C. lanceolata*	CA
13	*C. lanceolata*	CA
14	*C. lanceolata*	CA
54	*C. serpenticola*	OR
52	*C. serpenticola*	OR
53	*C. serpenticola*	OR
55	*C. serpenticola*	OR
68	*C. virginica*	IN
69	*C. virginica*	NC
70	*C. virginica*	MD
57	*C. tuberosa*	YT
56	*C. tuberosa*	AK
3	*C. caroliniana*	NH
51	*C. rosea*	CO
30	*C. ogilviensis*	YT
25	*C. multiscapa* subsp. *pacifica*	WA
64	*C. umbellata*	NV
65	*C. umbellata*	NV
63	*C. umbellata*	NV
2	*C. nivalis*	WA
1	*C. bellidifolia*	ID
34	*C. panamintensis*	CA
35	*C. panamintensis*	CA
31	*C. panamintensis*	CA
36	*C. panamintensis*	CA
33	*C. panamintensis*	CA
32	*C. panamintensis*	CA
45	*C. peirsonii* subsp. *yorkii*	CA
47	*C. peirsonii* subsp. *yorkii*	CA
44	*C. peirsonii* subsp. *yorkii*	CA
48	*C. peirsonii* subsp. *yorkii*	CA
67	*C. umbellata*	NV
49	*C. peirsonii* subsp. *yorkii*	CA
50	*C. peirsonii* subsp. *yorkii*	CA
46	*C. peirsonii* subsp. *yorkii*	CA
66	*C. umbellata*	NV
42	*C. peirsonii* subsp. *peirsonii*	CA
43	*C. peirsonii* subsp. *peirsonii*	CA
41	*C. peirsonii* subsp. *californacis*	CA
40	*C. peirsonii* subsp. *californacis*	CA
39	*C. peirsonii* subsp. *bernardinus*	CA
38	*C. peirsonii* subsp. *bernardinus*	CA
37	*C. peirsonii* subsp. *bernardinus*	CA
26	*C. obovata*	CA
27	*C. obovata*	CA
29	*C. obovata*	CA
28	*C. obovata*	CA
	C. sarmentosa	
	Claytonia saxosa	
	Montia parvifolia	
	Lewisia rediviva	
	Calandrinia menziesii	

B

0.01

0.01

are resolved by the ddRAD data as sister with moderate support (87 BS), but their relationship with respect to the rest of the tuberous species sampled is not resolved; the SKIM data set showed strong support for the caudicose species as sister (98 BS), and this pair was nested (95 BS) in a clade of tuberous species including *C. peirsonii*, *C. panamintensis*, and some of the *C. umbellata* samples. *Claytonia umbellata* is associated with a great deal of conflict in both splits networks (Fig. 1B) and was found to be paraphyletic (Fig. 1A) with samples being resolved in multiple clades (including two nested within the *C. peirsonii* clade).

The multi-species coalescent tree estimated using our SKIM plastome data set showed a topology and patterns of bootstrap support fairly similar to that of the SKIM nuclear data (Fig. 2). *Montia parvifolia* was resolved with strong support (100 BS) as sister to all of the *Claytonia* samples, and caudicose and tuberous species together were resolved as monophyletic with strong support (99 BS), the difference being that *C. saxosa* and *C. sarmentosa* were moderately supported as a sister pair (92 BS) with respect to these rather than as a grade. The signal for species (i.e., monophyly of samples from the same population/taxon) is high (100 BS) for all species with multiple samples, with the exceptions of *C. virginica* and *C. umbellata*, the latter of which is found to be widely paraphyletic as with the SKIM nuclear data. All four subspecific taxa included within *C. peirsonii* are also strongly supported (100 BS), albeit two *C. umbellata* samples were resolved in this clade with *C. peirsonii* subsp. *yorkii* T.R.Stoughton. The SKIM plastome and nuclear multi-species coalescent trees differ generally with respect to placement of the caudicose species (*C. bellidifolia* and *C. nivalis*), *C. obovata*, and *C. panamintensis*, and the relationship between the majority of the *C. umbellata* samples (including *C. ogilviensis*) and *C. lanceolata*, *C. serpenticola*, and the *C. virginica* species complex (including *C. tuberosa*).

Morphological differentiation

The first two principal components of the harmonic coefficients derived from the elliptic Fourier analysis explained 83.0% of the variation in the leaf shape (Fig. 3). The first component (PC1) explained 69.8% of the variation and described leaves that change from linear on one extreme to ovate on the other (Fig. 3). The second PC explains 13.2% and includes variation from lanceolate to oblanceolate leaves, but also explains differences in handedness associated with inflorescence orientation and leaf asymmetry (Fig. 3). The morphospace of the PC scores showed clustering of samples assignable to each taxon (13 taxa, $n = 1288$; Fig. 3), including a fairly distinct cluster belonging to a new taxon described in the taxonomic treatment below, but considerable overlap among 50% confidence intervals for the centroids exists in some cases, and individuals of the collective taxa form a continuum across PC1 and PC2. Pairwise MANOVA showed a significant multivariate taxon effect for the PC scores ($P < 0.001$, Appendix S7) for all pairwise taxon comparisons, strongly suggesting that leaf shapes differ among taxa. Additionally, taxonomic predictions from the DA were fairly accurate (65.8%, 848 of 1288 correctly predicted; Appendix S8) but indicate that classification accuracy based strictly on leaf shape varies among taxa. For

instance, *C. lanceolata*, *C. obovata*, *C. umbellata*, and *C. virginica* (in addition to the new taxon) were classified with greater than 70% accuracy, while the three subspecies of *C. peirsonii* (all but *C. peirsonii* subsp. *yorkii*) and *C. rosea* were classified correctly less than 50% of the time. Additional characters, like allopatric distributions and differences in ecology, betalain pigmentation, and inflorescence architecture, facilitate straight-forward diagnosis of the taxa that were not identified confidently by the DA.

Ecological differentiation

ANOVA and Tukey's HSD test of soil properties and climatic variables indicate that temperature and precipitation exhibit statistically significant differences in means for some, but not all, pairs of taxa. Soil property means were not significantly different (i.e., <0.05) for any pairs of taxa, but exhibited some noteworthy trends. With respect to mean diurnal temperature (bio2), *C. umbellata* differs significantly from *C. lanceolata*, *C. obovata*, *C. peirsonii*, and *C. serpenticola* ($F_{5, 26} = 9.101$, $P < 0.001$). Annual mean precipitation (bio12) differed significantly among eight pairs of taxa ($F_{5, 26} = 9.743$, $P < 0.001$). Tukey's HSD further indicates that temperature and precipitation associated with *C. panamintensis*, *C. umbellata*, and *C. peirsonii* habitats differed significantly from those of *C. obovata* (P ≤ 0.05) and *C. serpenticola* ($P ≤ 0.03$). The former two species habitats also differed significantly from that of *C. lanceolata* ($P ≤ 0.05$). The PCA indicates that ecological characteristics in combination (soils only and soils + climate) are not sufficient to distinguish among species habitats. When discriminant analyses were conducted to determine which characteristics could best distinguish among species' habitats, we found that the first linear discriminant (LD1, 60% of trace) was influenced most by mean diurnal temperature, soil conductivity, and sulfur, whereas LD2 (16% of trace) was most influenced by mean annual precipitation and pH. Habitat characteristics can be used to predict species occupancy with relatively high accuracy (Table 1; χ^2 $(5, n = 25) = 118.79$, $P < 0.001$). We conducted a second DA to determine whether climatic variables were overwhelming adaptive signal of soil properties, removing temperature and precipitation from the data set. Omission of climate resulted in weaker separation (LD1 = 40%, LD2 = 26%, LD3 = 20%) and predictive accuracy (χ^2 [5, $n = 25$] = 89.926, $P < 0.001$) of species ecologies, but increased the predictive accuracy for *C. peirsonii* (Table 1), a species with multiple subspecies that exhibits wider ranges of both temperature and precipitation conditions compared with all other species. When climate is removed, *C. umbellata*, *C. serpenticola*, and *C. obovata* have lower prediction accuracy. In both predictive DA analyses, several samples are misclassified as residing in *C. lanceolata*, *C. personii*, or *C. umbellata* habitats, all of which experience the highest ranges in annual precipitation.

DISCUSSION

Species complexes can arise in many different ways, including incomplete cladogenesis or genomic duplication, both being

---

FIGURE 1. Phylogenetic trees and splits networks resulting from analysis of nuclear data derived from genome skimming (SKIM) and double-digest restriction-site-associated DNA sequencing (ddRAD). Numbered labels correspond to a list of vouchered specimens contained within Appendix S1, with *Claytonia umbellata* samples (and *C. ogilviensis*) bolded for ease of reference. Colored boxes correspond with core clades comprised of samples clustering together in most analyses. (A) Multispecies coalescent trees (70% consensus) estimated from SKIM (left) and (right) ddRAD data. Thickened branches correspond to bootstrap support greater than 95%. (B) Neighbor-net splits networks estimated from SKIM (left) and (right) ddRAD data.
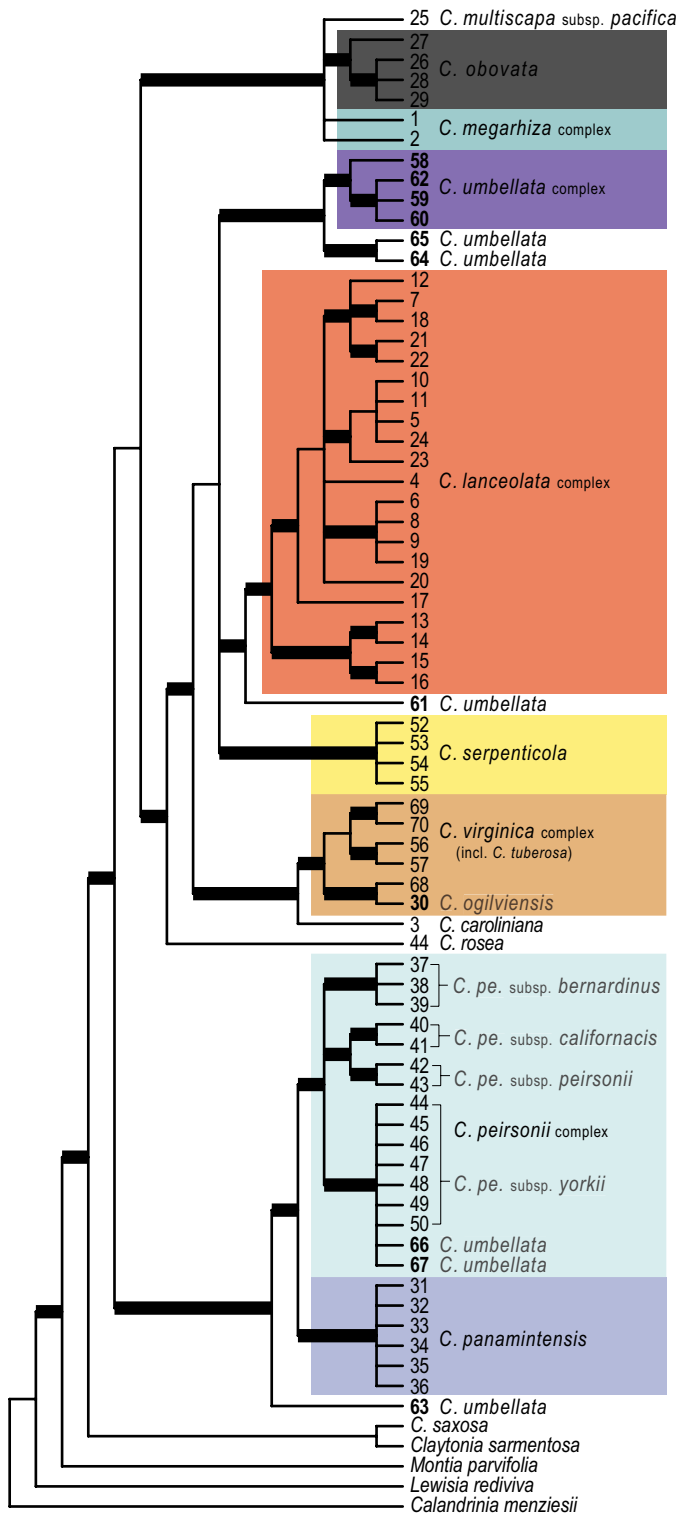
FIGURE 2. Multispecies coalescent tree (70% consensus) derived from genome skimming plastome analysis. Numbering and labelling is identical to that described for Fig. 1.

human population growth), it is more important than ever before to bolster our basic understanding of biodiversity and evolutionary relationships to better understand patterns of adaptation in isolated lineages. Yet diagnosis of distinct taxa (evolutionarily significant lineages) is difficult in many plant species complexes, and this is true for the *C. lanceolata* species complex recently recircumscribed by Stoughton et al. (2017). Five species (eight taxa) were recognized by Stoughton et al. (2017), where only one had been recognized previously by Miller and Chambers (2006). Molecular data were presented by Stoughton et al. (2017), but they provided limited resolution for the group of interest and sampling of *Claytonia* species and populations was sparse.

By contrast, the genetic data presented here provide significant support for formal recognition of all eight taxa recently circumscribed by Stoughton et al. (2017). The SKIM data generally provided more phylogenetic resolution than the ddRAD data for our sample set. It should be noted, however, that sequence data were considerably less evenly distributed across samples in the ddRAD data set, suggesting that missing data related to quality of sequencing library preparation may have disproportionately affected ddRAD phylogenetic estimations. On the other hand, the wide paraphyly of *C. umbellata*, instability of the placement of *C. ogilviensis*, and close association of the caudicose species (*C. bellidifolia* and *C. nivalis*) with tuberous *Claytonia* sampled in this study are suggestive of past hybridization among species that are otherwise treated in separate taxonomic sections by some researchers (i.e., sect. *Caudicosae* (A.Gray) Poelln. and sect. *Claytonia*). In other words, *C. umbellata* may represent a species with multiple origins involving past gene flow among tuberous and caudicose lineages. Although other processes may also explain the observed pattern, our conclusion that hybridization and/or introgression has occurred is supported by the conflict associated with *C. umbellata* in the splits networks (Fig. 1), and morphological intermediacy of subterranean storage organs exhibited by "problem" taxa; *C. umbellata* and *C. ogilviensis* have taproots extending below their tubers, rather than rounded tubers with fibrous roots (as in other tuberous species, members of sect. *Claytonia*) or well-developed caudices (as in *C. bellidifolia* and *C. nivalis,* members of sect. *Caudicosae*). In addition, tuberous species are united morphologically by having one cotyledon (presumably two that are fused) whereas caudicose, rhizomatous, and annual species have retained the symplesiomorphic condition of two free cotyledons. Incongruent resolution and paraphyly of *C. umbellata* samples between SKIM plastome and nuclear phylogenies are also suggestive of chloroplast capture. However, samples of *C. umbellata* (#66 and #67 in Figs. 1–2) nested in the *C. peirsonii* subsp. *yorkii* clade likely represent contamination of genomic isolates considering that one sample from the same population (#65) is resolved with samples of *C. umbellata* from other populations in all analyses. An alternative explanation for the placement of #66 and #67 is that they represent a cryptic taxon, or introgressed hybrids, but the identity and ancestry of samples from this population near the type locality of *C. umbellata* must be explored in future work. *Claytonia peirsonii* and *C. umbellata* are allopatric, although their respective ranges do come close to overlapping in the Sierra Nevada where *C. peirsonii* subsp. *yorkii* occurs. Fortunately, our nuclear phylogenetic estimates are robust to the exclusion of the "problem" taxa (*C. umbellata* and *C. ogilviensis*), indicating that our estimates are representative of true relationships among the sampled taxa (data not shown).

Regarding the morphological analyses, we found that most of the variation in the morphological data set (~83%) was associated
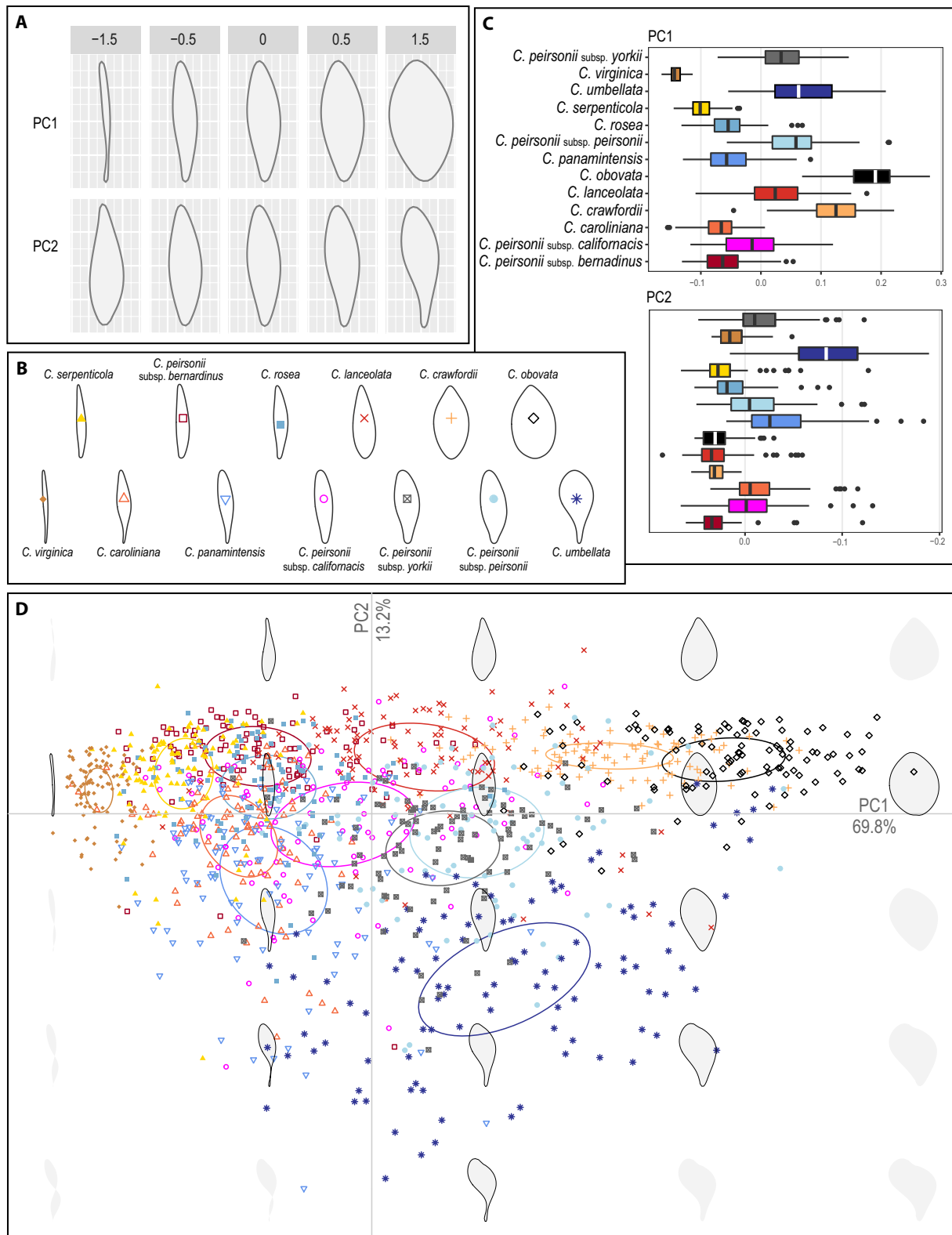
intimately linked to diversification in many organisms, especially angiosperms (Soltis et al., 2009). Considering the current challenges to global ecosystem health (i.e., climate change, deforestation,

FIGURE 3. Summary of the elliptic Fourier analysis of leaf shape. (A) Changes in leaf shape associated with each of the first two principal components. Each row corresponds to variation explained in each PC, and the columns correspond to the mean shape (center column) and plus or minus 1.5 standard deviations from the mean on each side. (B) Mean shape for each of the 13 taxa (colored) included in the leaf shape analysis. (C) Boxplots derived from PCA ordination of elliptical Fourier descriptors showing means and ranges of variation in overall leaf shape assignable to each taxon (colored) in the morphological analysis for PC1 and PC2. (D) The same PCA ordination showing 50% confidence intervals around clusters for each taxon (colored).

TABLE 1. Ecological prediction accuracy based on discriminant function analysis (DA) of 14 variables (12 soils and 2 climate) for six species. For each species, the number of samples (*n*) used to develop the model and the predicted assignment under that model (columns 3–8) are given for DA conducted with (outside parentheses) and without (inside parentheses) climate variables, i.e., temperature and precipitation. The total number of samples assigned to each species' ecology (last row) indicates how ecology prediction differs from actual number of samples included for each species (*n*).

Species	*n*	*lanc.*	*obov.*	*pana.*	*peir.*	*serp.*	*umb.*
C. lanceolata	6	5 (5)	—	—	1 (1)	—	—
C. obovata	5	1 (2)	4 (3)	—	—	—	—
C. panamintensis	1	—	—	1 (1)	—	—	—
C. peirsonii	8	1 (1)	—	—	6 (7)	—	1 (0)
C. serpenticola	4	1 (1)	—	—	—	3 (2)	0 (1)
C. umbellata	8	—	—	—	0 (2)	0 (1)	8 (5)
Total	32	8 (9)	4 (3)	1 (1)	7 (10)	3 (3)	9 (6)

with statistically significant differences among taxa (based upon MANOVA) corresponding to changes in leaf width and whether the widest point of the leaf was above or below the middle. This result is consistent with a morphometric study of *Viburnum* L. leaves by Schmerler et al. (2012) using EFA, in which leaf width was found to represent a large proportion of the variation. Handedness, an asymmetrical artifact arising during the collecting process, contributes negligibly to variation in the morphospace, presumably affects all species equally, and is correlated with degree of petiolar differentiation along PC2 (Fig. 3A). This trait is not biologically meaningful and ideally its contribution would be reduced by, for example, sampling consistently with respect to development of cauline leaves relative to terminal inflorescences. An attempt was made to mirror some extracted leaf images that were incorrectly oriented, but reflecting them to make a symmetrical shape may also prove useful in reducing the unwanted contribution of leaf handedness to the total variation in the morphological data set. Differential drying of highly succulent *Claytonia* leaves appears to contribute negligibly to the major PCs, although this artifact may be reduced in the future by using plexiglass to flatten (and photograph) fresh leaves. Based on our understanding of phylogenetic relationships inferred from multi-locus DNA sequences, patterns of morphological variation identified in our morphometric study of these taxa may best be explained by past hybridization events, although retention of ancestral polymorphism or incomplete lineage sorting may also be invoked as explanations. For instance, the position of members of the *C. peirsonii* complex in the morphospace (Fig. 3) is consistent with the genetic intermediacy of this complex to *C. panamintensis* and *C. obovata* from California (Fig. 1B). In all, species (and subspecies) circumscriptions are reflected fairly well in the clustering of samples in the PCA ordination in some cases (e.g., *C. virginica*), but a great deal of overlap (e.g., *C. peirsonii* subsp. *peirsonii* and *C. peirsonii* subsp. *yorkii*) and highly variable leaf shapes (e.g., *C. peirsonii* subsp. *californacis* T.R.Stoughton) exemplify most taxa. Fortunately, leaf shape is not the only morphological character used to discriminate these mostly allopatric taxa. We echo Stoughton et al. (2017) in suggesting that numerous individuals should be observed while making collections in the field so as to document the full range of variation in populations of *Claytonia*.

It appears, based on our analyses, that closely related species occupy divergent habitats, except in cases where allopatric speciation may have occurred. *Claytonia serpenticola* and *C. lanceolata*, for instance, have been inferred to be sister species (Figs. 1, 2), yet

they occupy peripatric distributions and distinctly different habitats: *C. lanceolata* habitat was never misclassified as *C. serpenticola* habitat (Table 1). As for the predictive analysis, our χ^2 test was significant ($p \leq 0.001$), and prediction accuracy was relatively high ($\geq 75\%$ for all individual species), indicating that soil chemistry may reveal a great deal of information about how niches differ. Coincidentally, most of the inaccurate predictions resulted in misidentification of one or another species as *C. lanceolata*, many of which were treated as that species before the recent study of *C. lanceolata* sensu lato by Stoughton et al. (2017). Historic treatment of these taxa as *C. lanceolata* may be related to the fact that, in addition to overlapping leaf shape variation demonstrated in the current study (Fig. 3D), the concept of *C. lanceolata* under the treatment by Stoughton et al. (2017) is that of a generalist species occupying a wide range of substrates and approaching occupied habitats of at least a few of these other taxa across its broad geographic distribution.

CONCLUSIONS

There are many methods for species delimitation, but it is preferable to be more conservative in species delimitation (i.e., to lump or take no taxonomic action) than to delimit entities that do not actually represent evolutionary lineages, such as those that (unbeknownst to the taxonomist) result from phenotypic plasticity (Carstens et al., 2013). Here, we provide strong molecular (Figs. 1, 2), morphological (Fig. 3), and ecological (Table 1) justification for a recent taxonomic treatment of tuberous *Claytonia* by Stoughton et al. (2017) in which many of the taxa included in the present study were circumscribed. Beyond that, we present strong morphological justification for a new species treated in the taxonomic treatment below and validate morphological features of leaf shape that distinguish this and other species of both eastern and western North American tuberous *Claytonia* that are largely allopatric and supported by genetic data. We utilize an extremely simple, reproducible method of reference-based phylogenetic marker selection using both ddRAD and SKIM data that proves to be useful to delimit species boundaries, even in a taxonomically difficult complex of several polyploid species (Miller and Chambers, 2006; Stoughton and Jolles, 2013). Coupling genetic analyses with statistical ordination provides additional insight regarding the degree to which morphological and ecological variation reflect phylogenetic differentiation. Furthermore, genetic data discussed here produced a fairly robust phylogenetic hypothesis of species relationships suggestive of past hybridization among caudicose and tuberous *Claytonia*.

In this new age of modern molecular methods, the ability of analytical and interpretive methods development to keep up with the sheer volume and resolution of genetic data being generated will be a challenge, as will be our ability to link speciation patterns with morphology, ecology, and geography. For the genus *Claytonia*, we adopt a unified concept of species (sensu de Quieroz, 2005) that relies on multiple secondary species criteria linked with natural history (e.g., morphology, ecology, geography). When genetic data are lacking or unavailable, morphological and ecological characteristics (i.e., substrate type, landform, vegetation assemblage) may be utilized to support the hypothesis that a particular species should be recognized. Species hypotheses should be tested, when possible, using molecular methods and the criterion of reciprocal monophyly. As more and more

genomic resources become available for a wide variety of organisms as a result of on-going research (e.g., efforts like the 1000 Plants (1KP) project [Matasci et al. 2014]), the requisite genomic reference sequence from a reasonably close relative will become considerably less difficult to obtain, making integrated analyses of molecular and other data types more and more feasible.

TAXONOMIC TREATMENT

Claytonia crawfordii T.R.Stoughton sp. nov.—TYPE: USA California: Tuolumne Co., vicinity of Bald Mountain between Long Barn and Strawberry, Sierra Nevada, 2 April 2016, *T. R. Stoughton 2172 with Diana Jolles* (holotype: PSH; isotypes: CAS, JEPS, RSA, NY, US).

Diagnosis

Most similar in general proportions to *Claytonia obovata* but principally differing by the combination of having a relatively elongate peduncle (which is short or absent in *C. obovata*), narrower basal and cauline (Fig. 3) leaves, and raised secondary veins present on the adaxial surfaces of the leaves (Fig. 4) that are reminiscent of the *C. peirsonii* species complex.

Additional specimens examined

USA California: Tuolumne Co., 1 mile down (USFS road) 4n15 from top of Crandall Peak, 31 March 1990, *Peggy Carkeet s.n.* (OSC); between Long Barn and Pinecrest (on NE side of Bald Mt.), 24 March 1978, *L. R. Heckard 4741 with R. Bacigalupi and G. Stebbins* (JEPS); (Near USFS road 6N96, Stanislaus NF), 12 May 2016, *J. Poore s.n.* (SPIF).

Distribution and habitat

Claytonia crawfordii is a Sierra Nevada endemic known only from Tuolumne County where it occurs in highly exposed, relatively treeless areas among volcanic rocks mostly on north-facing slopes just below ridgeline.

Discussion

This species was first discovered in 2015 at JEPS wherein the specimen (*L. R. Heckard 4741*) was identified as *C. obovata* based on leaf shape. It was noted in the annotation that certain other morphological features (such as the evidently elongate peduncles) did not fit the description of *C. obovata*, a species that would be well out of range in the Sierra Nevada near Long Barn, California. This specimen was mentioned by Stoughton et al. (2017) as belonging
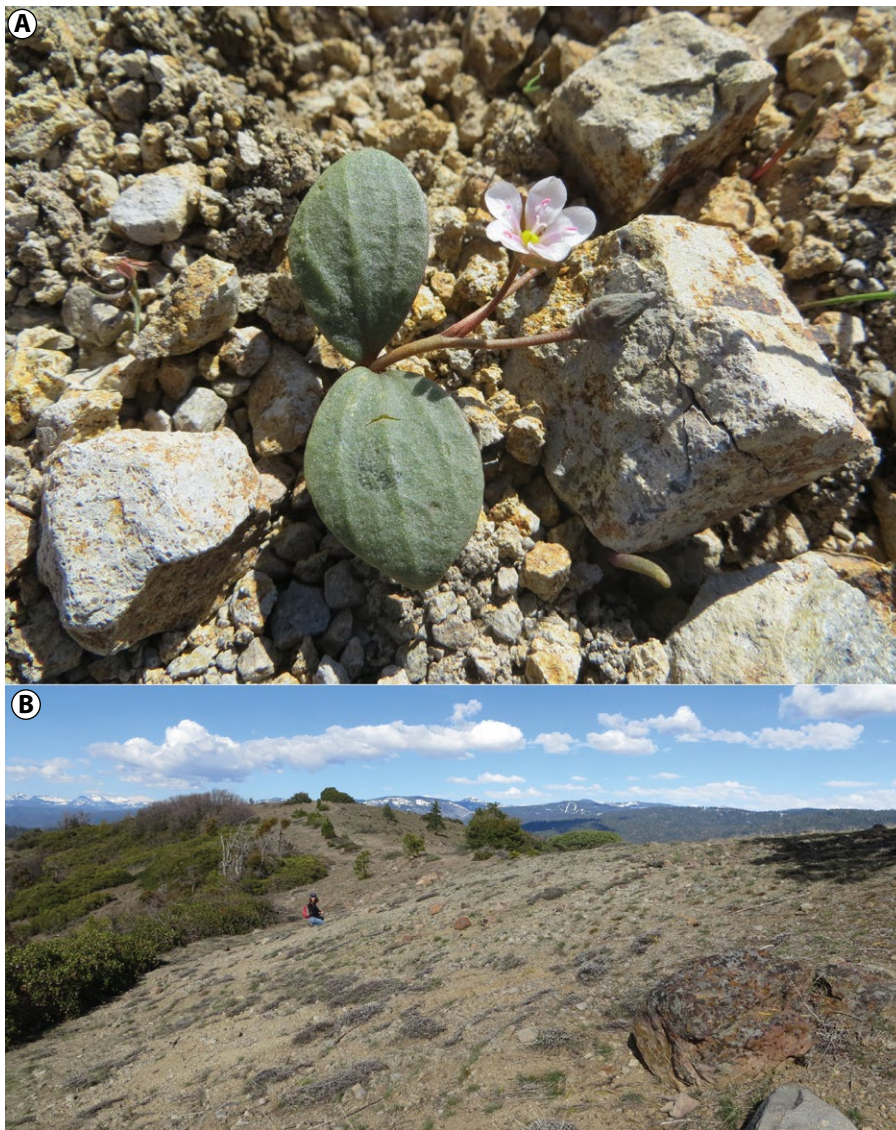
FIGURE 4. *Claytonia crawfordii.* (A) Aerial stem morphology. (B) Representative habitat at the type locality near Long Barn, California.

to a putatively new taxon closely allied with either *C. obovata* or *C. peirsonii*, based on its leaf shape and venation, respectively, but that more research was needed. Given the unique ecologies and wide distances between these two species and *C. crawfordii* (which combines some morphological features of both), we maintain that it is most reasonable to treat this easily diagnosable taxon at the species level. The morphological data on leaf shape presented in this study provide additional justification for its recognition as such. It may well prove that *C. crawfordii* is closely related to *C. obovata* or some member of the *C. peirsonii* complex, but genetic data are lacking; additional data (and answers concerning relationships) are forthcoming.

Etymology

This species is named for Larry and Suzanne Crawford, strangers who welcomed the first author, T. R. Stoughton, into their home and rekindled his passion for sharing botanical knowledge with others.

ACKNOWLEDGEMENTS

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

LITERATURE CITED

Andrews, K. R., J. M. Good, M. R. Miller, G. Luikart, and P. A. Hohenlohe. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* 17: 81–92.

Barrier, M., B. G. Baldwin, R. H. Robichaux, and M. D. Purugganan. 1999. Interspecific hybrid ancestry of a plant adaptive radiation: allopolyploidy of the Hawaiian silversword alliance (Asteraceae) inferred from floral homeotic gene duplications. *Molecular Biology and Evolution* 16: 1105–1113.

Baum, D. A. 2009. Species as ranked taxa. *Systematic Biology* 58: 74–86.

Baum, D. A., and M. J. Donoghue. 1995. Choosing among alternative "phylogenetic" species concepts. *Systematic Botany* 20: 560–573.

Bonhomme, V., S. Picq, C. Gaucherel, and J. Claude. 2014. Momocs: outline analysis using R. *Journal of Statistical Software* 56: 1–24.

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

Chifman, J., and L. Kubatko. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30: 3317–3324.

Cronin, M. A. 2006. A proposal to eliminate redundant terminology for intra-species groups. *Wildlife Society Bulletin* 34: 237–241.

Cronn, R., B. J. Knaus, A. Liston, P. J. Maughan, M. Parks, J. V. Syring, and J. Udall. 2012. Targeted enrichment strategies for next-generation plant biology. *American Journal of Botany* 99: 291–311.

Darling, A. C. E., B. Mau, F. R. Blattner, and N. T. Perna. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* 14: 1394–1403.

de Queiroz, K. 2005. A unified concept of species and its consequences for the future of taxonomy. *Proceedings of the California Academy of Sciences* 56: 196–215.

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

Dohm, J. C., A. E. Minoche, D. Holtgräwe, S. Capella-Gutiérrez, F. Zakrzewski, H. Tafer, O. Rupp, et al. 2014. The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature* 505: 546–549.

Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

Eaton, D. A., and R. H. Ree. 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology* 62: 689–706.

Edwards, S. V., A. J. Shultz, and S. C. Campbell-Staton. 2015. Next-generation sequencing and the expanding domain of phylogeography. *Folia Zoologica* 64(3): 187–206.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One* 6: e19379.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783–791.

Funk, D. J., and K. E. Omland. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics* 34: 397–423.

Godden, G. T., I. E. Jordon-Thaden, S. Chamala, A. A. Crowl, N. Garcia, C. C. Germain-Aubrey, J. M. Heany, et al. 2012. Making next-generation sequencing work for you: approaches and practical considerations for marker development and phylogenetics. *Plant Ecology and Diversity* 5: 427–450.

Hamilton, C. W., and S. H. Reichard. 1992. Current practice in the use of subspecies, variety, and forma in the classification of wild plants. *Taxon* 41: 485–498.

Hey, J. 2006. On the failure of modern species concepts. *Trends in Ecology & Evolution* 21: 447–450.

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

Hijmans, R.J. 2016. raster: geographic data analysis and modeling. R package version 2.5-8. Available at https://CRAN.R-project.org/package=raster.

Hillier, L. W., G. T. Marth, A. R. Quinlan, D. Dooling, G. Fewell, D. Barnett, P. Fox, et al. 2008. Whole-genome sequencing and variant discovery in *C. elegans*. *Nature Methods* 5: 183–188.

Huson, D. H., and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23: 254–267.

Jeffers, S. 2015. Spring beauty (*Claytonia*, Montiaceae) in Beringia: new evidence on species delineation from morphometrics and phylogenetic analysis of plastid and nuclear DNA sequence data. M.S. thesis, University of Alaska, Fairbanks, AK, USA.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, et al. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649.

Leaché, A. D., B. L. Banbury, J. Felsenstein, A. N. M. De Oca, and A. Stamatakis. 2015. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology* 64: 1032–1047.

Lemmon, E. M., and A. R. Lemmon. 2013. High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 44: 99–121.

Marhold, K., and J. Lihová. 2006. Polyploidy, hybridization and reticulate evolution: lessons from the Brassicaceae. *Plant Systematics and Evolution* 259(2–4): 143–174.

Matasci, N., L. H. Hung, Z. Yan, E. J. Carpenter, N. J. Wickett, S. Mirarab, N. Nguyen, et al. 2014. Data access for the 1,000 Plants (1KP) project. *GigaScience* 3: 17.

Miller, J. M., and K. L. Chambers. 2006. Systematics of *Claytonia* (Portulacaceae). *Systematic Botany Monographs* 78: 1–236.

O'Quinn, R., and L. Hufford. 2005. Molecular systematics of Montieae (Portulacaceae): Implications for taxonomy, biogeography and ecology. *Systematic Botany* 30: 314–331.

Palmer, J. D., D. E. Soltis, and M. W. Chase. 2004. The plant tree of life: an overview and some points of view. *American Journal of Botany* 91: 1437–1445.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. S. Hoekstra. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS One* 7: e37135.

Pigliucci, M. 2003. Species as family resemblance concepts: The (dis-) solution of the species problem? *BioEssays* 25: 596–602.

Popp, M., and B. Oxelman. 2007. Origin and evolution of North American polyploid *Silene* (Caryophyllaceae). *American Journal of Botany* 94: 330–349.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at https://www.R-project.org/.

Reaz, R., M. S. Bayzid, and M. S. Rahman. 2014. Accurate phylogenetic tree reconstruction from quartets: a heuristic approach. *PloS One* 9: e104008.

Ross, T. S. 1996. Herbarium specimens as documents: purposes and general collecting techniques. *Crossosoma* 22: 3–39.

Rubin, C. J., M. C. Zody, J. Eriksson, J. R. Meadows, E. Sherwood, M. T. Webster, L. Jiang, et al. 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464: 587–591.

Schmerler, S. B., W. L. Clement, J. M. Beaulieu, D. S. Chatelet, L. Sack, M. J. Donoghue, and E. J. Edwards. 2012. Evolution of leaf form correlates with tropical–temperate transitions in *Viburnum* (Adoxaceae). *Proceedings of the Royal Society, B, Biological Sciences* 279: 3905–3913.

Solomon, R. W. 2009. Free and open source software for the manipulation of digital images. *American Journal of Roentgenology* 192: W330–W334.

Soltis, D. E., V. A. Albert, J. Leebens-Mack, C. D. Bell, A. H. Paterson, C. Zheng, D. Sankoff, et al. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* 96: 336–348.

Spaniel, S., J. Zozomová-Lihová, N. G. Passalacqua, and K. Marhold. 2012. Infraspecific classification of *Alyssum diffusum* (Brassicaceae) in Italy. *Willdenowia* 42(1): 37–56.

Stace, C. A. 2013. The study of infraspecific variation. *Commentaries in Plant Science* 2: 15.

Stoughton, T. R., and D. D. Jolles. 2013. Discovery of *Claytonia lanceolata* var. *peirsonii* in the San Bernardino Mountains perpetuates a history of taxonomic uncertainty. *Aliso* 31: 35–42.

Stoughton, T. R., D. D. Jolles, and R. L. O'Quinn. 2017. The western spring beauties, *Claytonia lanceolata* (Montiaceae): a review and revised taxonomy for California. *Systematic Botany* 42(2): 283–300.

Stoughton, T. R., R. Kriebel, D. D. Jolles, and R. L. O'Quinn. 2018. Data from: Next Generation Lineage Discovery: a case study of tuberous *Claytonia* L. *Dryad Digital Repository*. https://doi.org/10.5061/dryad.m99v4.

Straub, S. C., M. Parks, K. Weitemier, M. Fishbein, R. C. Cronn, and A. Liston. 2012. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *American Journal of Botany* 99: 349–364.

Swofford, D. L. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer, Sunderland, MA, USA.

Venables, W. N., and B. D. Ripley. 2002. Modern applied statistics with S. 4th edn. Springer, NY, NY, USA.

Walsh, N. G. 2015. Elevation of rank for *Leucochrysum albicans* var. *tricolor* (Asteraceae: Gnaphalieae). *Muelleria* 34: 11–13.

White, B. P., E. M. Pilgrim, L. M. Boykin, E. D. Stein, and R. D. Mazor. 2014. Comparison of four species-delimitation methods applied to a DNA barcode data set of insect larvae for use in routine bioassessment. *Freshwater Science* 33(1): 338–348.

Wickett, N. J., S. Mirarab, N. Nguyen, T. Warnow, E. Carpenter, N. Matasci, S. Ayyampalayam, et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences, USA* 111(45): E4859–E4868.

Wickham, H. 2009. ggplot2: elegant graphics for data analysis. Springer, NY, NY, USA.

Wiens, J. J. 2007. Species delimitation: new approaches for discovering diversity. *Systematic Biology* 56: 875–878.

Wolfe, A. D., C. P. Randle, S. L. Datwyler, J. J. Morawetz, N. Arguedas, and J. Diaz. 2006. Phylogeny, taxonomic affinities, and biogeography of *Penstemon* (Plantaginaceae) based on ITS and cpDNA sequence data. *American Journal of Botany* 93: 1699–1713.

Zachos, F. E. 2015. Taxonomic inflation, the phylogenetic species concept and lineages in the tree of life–a cautionary comment on species splitting. *Journal of Zoological Systematics and Evolutionary Research* 53(2): 180–184.

Zimmer, E. A., and J. Wen. 2015. Using nuclear gene data for plant phylogenetics: Progress and prospects II. Next-gen approaches. *Journal of Systematics and Evolution* 53: 371–379.