

# UC Santa Cruz

## UC Santa Cruz Previously Published Works

### Title

Phylogenomic analysis does not support a classic but controversial hypothesis of progenitor-derivative origins for the serpentine endemic *Clarkia franciscana*

### Permalink

<https://escholarship.org/uc/item/0r9630xt>

### Journal

Evolution, 76(6)

### ISSN

0014-3820

### Authors

Sianta, Shelley A  
Kay, Kathleen M

### Publication Date

2022-06-01

### DOI

10.1111/evo.14484

Peer reviewed

# Phylogenomic analysis does not support a classic but controversial hypothesis of progenitor-derivative origins for the serpentine endemic *Clarkia franciscana*

Shelley A. Sianta<sup>1,2,3</sup> and Kathleen M. Kay<sup>1</sup>

<sup>1</sup>Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, California 95060

<sup>2</sup>Current Address: Department of Plant and Microbial Biology, University of Minnesota, St. Paul, Minnesota 55108

<sup>3</sup>E-mail: ssianta@umn.edu

Received July 4, 2021

Accepted March 4, 2022

Budding speciation involves isolation of marginal populations at the periphery of a species range and is thought to be a prominent mode of speciation in organisms with low dispersal and/or strong local adaptation among populations. Budding speciation is typically evidenced by abutting, asymmetric ranges of ecologically divergent sister species and low genetic diversity in putative budded species. Yet these indirect patterns may be unreliable, instead caused by postspeciation processes such as range or demographic shifts. Nested phylogenetic relationships provide the most conclusive evidence of budding speciation. A putative case of budding speciation in the serpentine endemic *Clarkia franciscana* and two closely related widespread congeners was studied by Harlan Lewis, Peter Raven, Leslie Gottlieb, and others over a 20-year period, yet the origin of *C. franciscana* remains controversial. Here, we reinvestigate this system with phylogenomic analyses to determine whether *C. franciscana* is a recently derived budded species, phylogenetically nested within one of the other two putative progenitor species. In contrast to the hypothesized pattern of relatedness among the three *Clarkia* species, we find no evidence for recent budding speciation. Instead, the data suggest the three species diverged simultaneously. We urge caution in using contemporary range patterns to infer geographic modes of speciation.

**KEY WORDS:** budding speciation, *Clarkia*, incomplete lineage sorting, phylogenomics, rapid divergence, serpentine.

“There are many rich descriptions of [plant] species and how they are reproductively isolated by various mechanisms, but there is little specific evidence about the course of their divergence. Thus, it remains critical to examine particular cases of speciation, and to find out whether the general models of the processes are consistent with the facts.”

Leslie Gottlieb (2004)

Biologists have verbalized many models of speciation, but it is difficult to definitively pinpoint the model by which a particular species evolved, given postspeciation changes. Comparative analyses that regress geographic range characteristics of sister taxa (e.g., range overlap and range asymmetry) are often used to identify patterns of speciation. Although allopatric speciation via vicariance is supported in five mammal clades

(Fitzpatrick and Turelli 2006), studies across clades spanning the tree of life have found evidence of budding speciation (Barraclough and Vogler 2000; Malay and Paulay 2010; Claremont et al. 2012; Anacker and Strauss 2014; Grossenbacher et al. 2014). Budding speciation occurs when marginal populations become reproductively isolated from the remainder of the species and encompasses multiple named models of speciation, such as peripatric speciation (Mayr 1954), quantum speciation (Grant 1981), and catastrophic speciation (Lewis 1962). In organisms that experience high levels of local adaptation and population structure, such as plants, budding speciation is hypothesized to be a particularly common mode of speciation (Kisel and Barraclough 2010), and it has been suggested to play a major role in the

diversification of the species-rich California Floristic Province (Crawford 2010; Anacker and Strauss 2014; Grossenbacher et al. 2014). Budding speciation results in a species pair that is often termed “progenitor-derivative” because one geographically widespread member retains the phenotypic and ecological niche of the ancestral species, whereas the other exhibits extensive change (in phenotype, ecology, and/or chromosomal patterning) and occupies a smaller, abutting range. However, patterns of highly asymmetrical and abutting geographic ranges between ecologically divergent taxa can result from postspeciation processes—for example, allopatric speciation followed by range expansion and/or contraction in one of the sister species (Losos and Glor 2003)—and may be misleading for identifying cases of budding speciation.

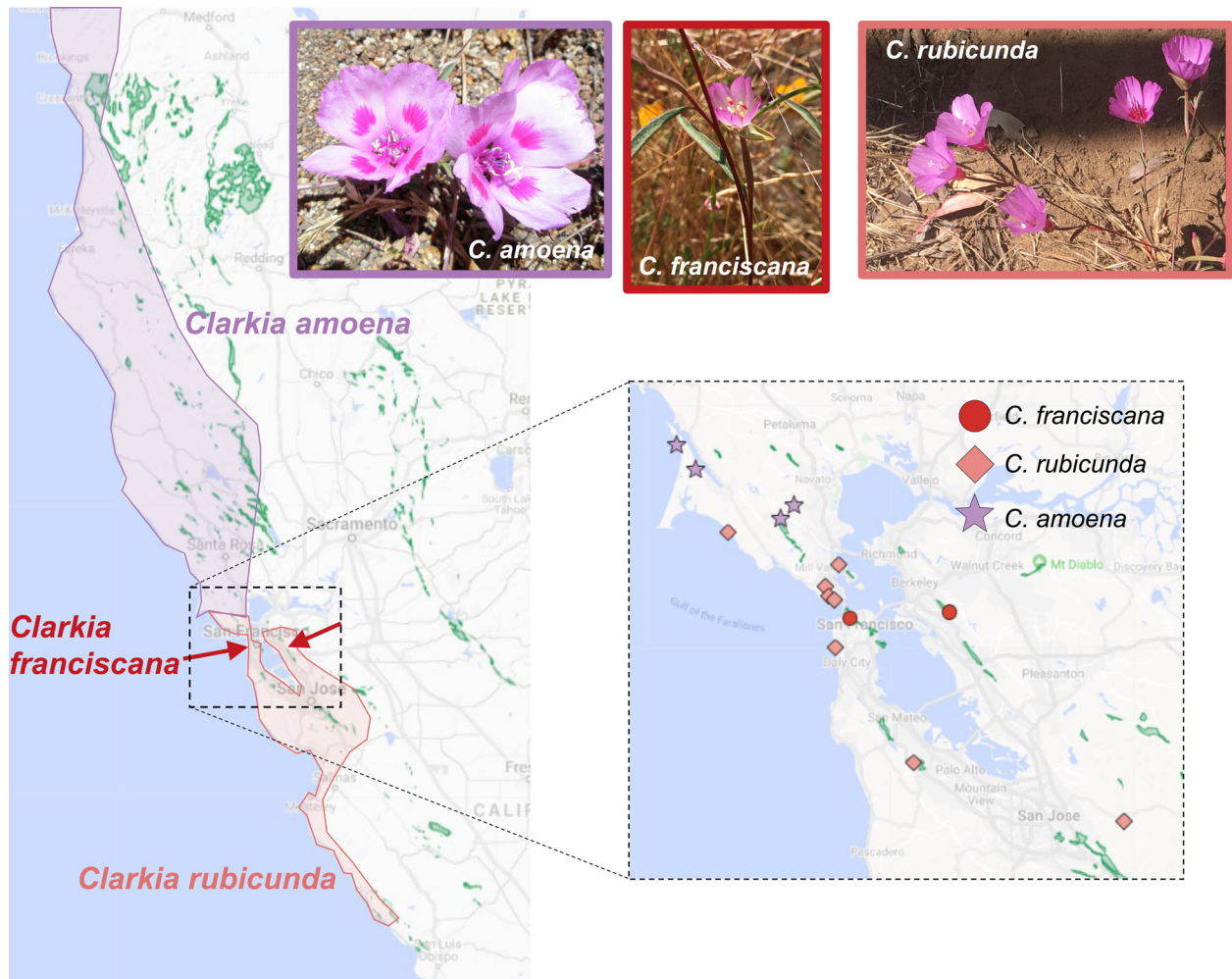
Phylogenetic evidence is the most conclusive way to identify progenitor-derivative species pairs (Crawford 2010), yet the phylogenetic resolution necessary is often elusive. Instead of showing reciprocal monophyly, derivative species are expected to be monophyletic and nested within a paraphyletic progenitor species (Rieseberg and Brouillet 1994), with the derivative species most closely related to the peripheral populations from which it evolved. For example, in one of the best empirical demonstrations of budding speciation, the narrowly distributed serpentine endemic *Layia discoidea* (Asteraceae) was found to be phylogenetically nested within the widespread *L. glandulosa* (Baldwin 2005). However, the topology of one gene tree may not represent the true species tree, and lineage sorting and intraspecific gene flow with recombination are expected to gradually erase the paraphyly of the progenitor species over time, eventually resulting in two reciprocally monophyletic species when considering the dominant species tree from limited phylogenetic data (Rieseberg and Brouillet 1994).

Phylogenomic analysis with population-level sampling may help uncover budding speciation due to the increased signal provided by sampling hundreds of genes, allowing for more accurate inference of the dominant species tree, and the explicit modeling of gene tree variation due to incomplete lineage sorting (ILS), introgression, and/or ancestral population structure (Degnan and Rosenberg 2009; García et al. 2017; Carlsen et al. 2018; Morales-Briones et al. 2018, 2021). If budding speciation occurred relatively recently, we expect the dominant species tree to show a topology consistent with budding, that is, the derivative species nested within the progenitor species. However, if budding speciation was ancient, we expect most gene trees to show a topology consistent with reciprocal monophyly (i.e., sister species relationships) because of lineage sorting and intraspecific gene flow with recombination. In this case, gene trees discordant with the reciprocal monophyly topology can still provide evidence for budding speciation, because we would expect a higher proportion of the discordant gene trees to be consistent with the hypothesized

progenitor-derivative pattern of nestedness than other topologies. If speciation did not occur through budding speciation, then we expect no bias in discordant gene trees for one topology over another.

Throughout the 1950s–1970s, *Evolution* published a suite of influential papers about the prominence of rapid and recent progenitor-derivative budding speciation in the western North American genus *Clarkia* (Onagraceae) by Harlan Lewis, Peter Raven, Leslie Gottlieb, and others (Lewis 1953, 1962; Lewis and Roberts 1956; Lewis and Raven 1958; Bartholomew et al. 1973; Gottlieb 1973, 1974b) that have collectively been cited hundreds of times (ISI Web of Science). Lewis (1962) proposed that speciation in *Clarkia* occurred through rapid isolation of peripheral populations through “catastrophic selection” that involved abrupt adaptation to harsher environments accompanied by barriers to gene flow, unlike Mayr (1954) and Grant’s (1981) peripheral speciation models that invoked a strong role of genetic drift. Lewis predicted the derivative species would be in ecologically marginal and recent habitats, have a smaller and abutting range to the progenitor species, and be morphologically similar but reproductively isolated from the progenitor species. Reproductive isolation was thought to be quickly achieved through rapid chromosomal repatterning associated with catastrophic selection events, rendering hybrids sterile (Lewis 1962).

A classic case of putative budding speciation is found in *Clarkia franciscana* (H. Lewis and P. H. Raven), a restricted serpentine endemic, and two morphologically similar species *C. rubicunda* ([Lindl.] H. Lewis and M. Lewis) and *C. amoena* ([Lehm.] A. Nelson and J. F. Macbr.) (Lewis and Raven 1958; Bartholomew et al. 1973; Gottlieb 1973, 1974a). The three species vary in their range sizes, but all ranges overlap in the San Francisco Bay area (Fig. 1). The most widespread and ecologically diverse species, *Clarkia amoena*, was hypothesized to be the progenitor species of *C. rubicunda*, through a scenario in which populations at the more arid southern range edge of *C. amoena* gave rise to locally adapted individuals that survived catastrophic selection, resulting in a derivative species with different chromosomal patterning. *Clarkia rubicunda* then went through a similar process, giving rise to the highly selfing *C. franciscana* that colonized harsh serpentine soil habitats. Lewis and Raven (1958) proposed the evolution of *C. franciscana* happened both rapidly and recently since the last glacial maximum. Multiple studies documented variation in the chromosomal patterns of the three species and found that the chromosomal rearrangements among them rendered interspecific hybrids sterile, making introgression unlikely to explain patterns of relatedness among the species (Lewis and Raven 1958; Snow 1963, 1964; Bartholomew et al. 1973). To test the hypothesis of budding speciation, Gottlieb (1973) used isozymes to determine whether *C. franciscana* contained a subset of alleles present in *C. rubicunda*, a prediction expected if the



**Figure 1.** Geographic ranges (polygons) and sampling localities (inset) of *Clarkia franciscana*, *C. rubicunda*, and *C. amoena*. Green layers indicate serpentine patches. Photo credits from left to right: *Clarkia amoena* by KMK, *C. franciscana* by SAS, and *C. rubicunda* by SAS.

former recently evolved as a derivative species. Surprisingly, Gottlieb found that *C. franciscana* harbored unique alleles at six of the eight isozyme systems tested, suggesting that *C. franciscana* was older than hypothesized by Lewis and Raven (1958). However, his results regarding progenitor-derivative speciation were inconclusive because genealogical relationships could not be established among isozyme alleles.

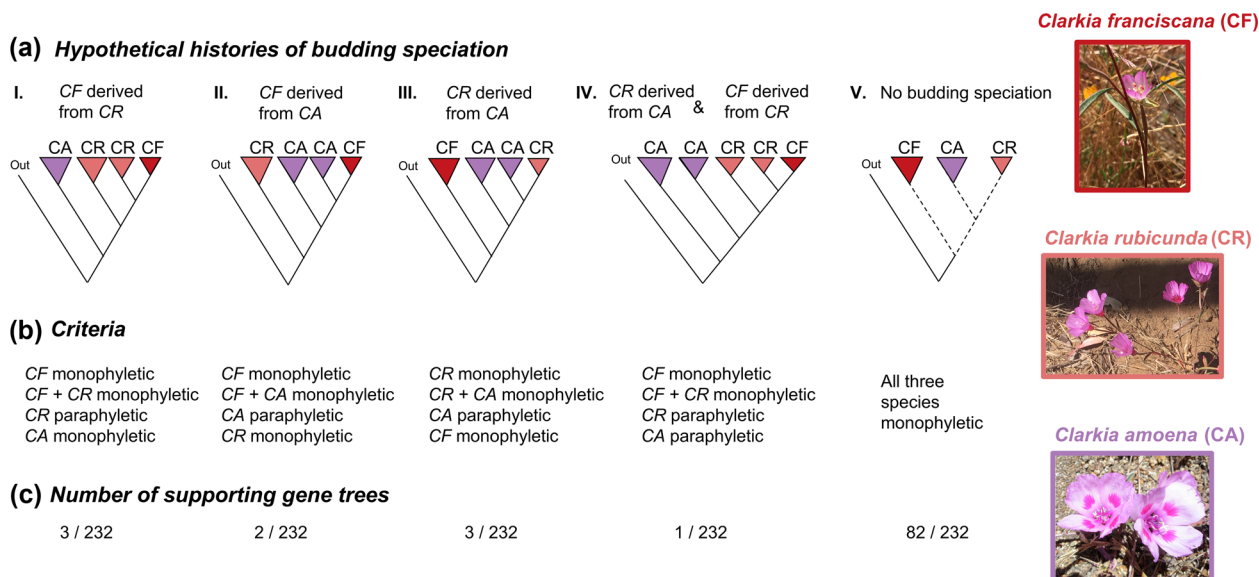
Here, we reevaluate the hypothesized story of budding speciation in the *C. franciscana*-*C. rubicunda*-*C. amoena* triad using phylogenomic analyses. We sample multiple populations of *C. franciscana*, *C. rubicunda*, and *C. amoena* with targeted sequencing of low-copy genes to infer gene trees and species trees. We explicitly ask whether there is evidence consistent with budding speciation. In particular, we assess the evidence for reciprocal monophyly versus nestedness in the species trees and gene trees consistent with various hypothesized budding speciation events. First, we ask whether *C. franciscana* is a derivative species of *C. rubicunda* (Fig. 2a, scenario I) or *C. amoena* (Fig. 2a, sce-

nario II). Second, we ask whether *C. rubicunda* is a derivative species of *C. amoena* (Fig. 2a, scenario III). Third, we ask whether there are multiple events of budding speciation in this group, as proposed by Lewis and Raven (1958), where *C. franciscana* is a derivative species of *C. rubicunda*, which in turn is derivative species of *C. amoena* (Fig. 2a, scenario IV). Lastly, we ask whether there is primary support for no budding speciation, that is, monophyly of each species, in this group (Fig. 2a, scenario V).

## Methods

### TAXON SAMPLING

We focused our taxon sampling on areas of the ranges of *C. amoena*, *C. rubicunda*, and *C. franciscana* at or near the abutting range boundaries in the San Francisco Bay area of California. We sampled a total of 23 individuals from 14 populations across the 3 species (Fig. 1; Table S1). *Clarkia franciscana* is a



**Figure 2.** There is a low level of gene tree support for budding speciation in the *Clarkia* triad. (a) Hypothetical scenarios of budding speciation (I–IV), or not (V), among the three species. Triangles at tips represent multiple individuals from a given species. Dashed branches in scenario V indicate ambiguity as to the relationship among the three species. (b) We used monophyly criteria to characterize a gene tree as supporting each of the five scenarios. (c) The number of RAxML gene trees that fit the criteria for each scenario. To minimize the effect of gene tree error, we used gene trees that have branches with less than 33% bootstrap support collapsed.

California state- and federally listed endangered species that occurs in only two locations, each on chemically harsh serpentine soils. *Clarkia rubicunda* is a relatively widespread serpentine tolerator, with populations both on and off serpentine. As one of our aims was to understand the ecological transitions associated with progenitor-derivative speciation, we sampled a mix of serpentine and nonserpentine populations of *C. rubicunda* to test which ecotype gave rise to the serpentine endemic *C. franciscana*. We used one *C. arcuata* individual as an outgroup because it was the closest diploid relative of the species triad based on a preliminary phylogeny of *Clarkia* built with the same dataset used here, and is supported as sister to *C. franciscana* in a partially sampled phylogeny of the genus (Gottlieb and Ford 1996). Some of the tissue we used was collected in the field, whereas other tissue was collected from growing field-collected seeds in the greenhouse.

**TARGETED SEQUENCING**

We used a targeted-genome enrichment approach using a bait set designed from transcriptomes of two *Oenothera* (Onagraceae) species, *O. serrulata* and *O. berlandieri* (Cooper et al. 2021; Patsis et al. 2021). Briefly, transcriptomes were assembled and then mapped to a set of 956 single- or low-copy nuclear loci shared among *Arabidopsis*, *Populus*, *Vitis*, and *Oryza* (Duarte et al. 2010). Of the 956 loci, 322 loci in the *Oenothera* transcripts were randomly selected for bait design (Cooper et al. 2021). Baits were 120 nucleotides in length and were designed to have a 60-nucleotide overlap (2× tiling), for a total of 19,994

baits. The bait set was manufactured by MYcroarray (now Arbor Biosciences, Ann Arbor, MI, USA). This sequencing method is advantageous for phylogenomic studies for three reasons: it results in alignable contigs (genic regions) across all taxa that can be used to build per-locus gene trees, contigs will include parts of more variable “splash-zone” sequences (i.e., intron and flanking gene sequences) that may be more informative at lower phylogenetic levels, and it is suitable for organisms without a reference genome.

We extracted DNA with a modified CTAB extraction protocol, incubating leaf samples in the CTAB solution at 55°C overnight (Doyle and Doyle 1987). Libraries for the samples used in this study were prepared in a larger run with 72 other *Clarkia* and *Camissonia* (Onagraceae) samples. We sonicated 200 ng of genomic DNA per sample, targeting 550 bp fragment sizes. We prepared sequencing libraries with the Illumina TruSeq Nano HT DNA Library Preparation Kit (San Diego, CA, USA) following the manufacturer’s protocol at half reagent volumes following the second addition of AMPure beads (Beckman Coulter, Beverly, MA, USA). We ligated Illumina i5 and i7 barcode indices to all libraries. We hybridized libraries to baits following the MYcroarray protocol. We pooled 12–17 samples in one hybridizing reaction, inputting 100 ng of each library into the hybridization pool. We pooled samples roughly by taxonomic association (e.g., samples within species were pooled together, or closely related species were pooled together). Hybridization was performed at 65°C for 18 h. We reamplified enriched libraries with 14–18

PCR cycles and performed a final PCR cleanup step with the Qiagen QiaQuick PCR cleanup (Qiagen, Hilden, Germany). We checked molarity and ensured the fragment lengths were appropriate for sequencing using a Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). We combined all hybridization pools into one run at equimolar ratios (4 nM) with a 1% molar ratio of PhiX Control (Illumina) on an Illumina MiSeq (600 cycles, version 3 chemistry). We recovered a total of 7,031,356 paired-end 300-bp reads for our 24 samples (26,974,129 paired-end reads for the whole run of 96 samples) and an average of 292,973 paired-end reads per sample used in this study.

### BIOINFORMATIC PROCESSING OF SEQUENCES

We used the bcl2fastq version 2.18.0.12 Illumina Conversion Software to demultiplex reads and convert the raw basecall files to fastq files. We used Trimmomatic (Bolger et al. 2014) to remove Illumina adapters and filter reads for quality. We removed bases at the leading and trailing ends that were under a phred33 quality score of 10, and trimmed sequences once a sliding window of 4 bases averaged below a quality score of 20. We removed reads that were less than 20 bases and reads that did not have a mated pair. After quality control, there were a total of 6,276,835 paired-end reads, with an average of 261,534 paired-end reads per sample.

We then used HybPiper (Johnson et al. 2016) to assemble reads into contigs and sort them into gene directories using *O. serrulata* and *O. berlandieri* as reference sequences. An average of 75% of all trimmed and filtered reads per sample were sorted into a gene directory. HybPiper assembled contigs de novo for each gene separately using SPAdes (Bankevich et al. 2012). The program Exonerate (Slater and Birney 2005) was used to align translated contigs to the translated target sequence for each gene. If multiple contigs overlapped by at least 20 bp, they were merged into a supercontig. If no contigs overlapped, the longest contig was retained. If there were multiple, long contigs that spanned the target sequence length, HybPiper flagged the gene directory with a paralog warning. An average of 23 of the 322 loci per sample flagged paralog warnings. We use the supercontigs containing both exon and intron sequences in downstream analyses in an effort to include more variable, noncoding regions (Weitemier et al. 2014). Supercontigs had an average coverage depth of  $98\times$  (with a standard deviation of  $24\times$ ). An average of 302 genes per sample mapped with contiguous sequences. An average of 252 genes and 158 genes per sample were at least 50% and 75% of the reference sequence length, respectively.

We removed one sample that had low sequencing and low enrichment efficiency. We further filtered our set of genes to include only those that were present in at least 22 of the remaining 23 samples and then subsequently removed any loci that flagged a paralog warning in any of the samples, resulting in a remaining

dataset of 232 loci. Gene sequences were aligned with MAFFT version 7.130b (Katoh and Standley 2013) under the `-auto` setting. We used TrimAl version 1.4.rev.15 (Capella-Gutiérrez et al. 2009) to trim columns with their `-automated1` heuristic method.

### PHYLOGENOMIC ANALYSES

We inferred species trees in two ways: with a concatenated supermatrix and a coalescent-based summary method (Mirarab et al. 2014). Concatenation has the advantage of adding gene matrices that individually have low phylogenetic signal to increase the power to resolve relationships. However, concatenation assumes that all sites have evolved according to a single evolutionary tree, an assumption that is violated with recombination among genes and admixture (Degnan and Rosenberg 2009). Given that supermatrices implicitly are composed of hundreds of genes, concatenation methods can lead to highly supported but wrong species trees (Edwards et al. 2007). In contrast, coalescent-based methods explicitly model gene discordance that is expected due to ILS (Liu et al. 2009). The accuracy of concatenation versus coalescent-based methods is dependent on the level of ILS in the samples, with the latter being more accurate in high ILS situations (Kubatko and Degnan 2007; Roch and Warnow 2015). Given that we sampled three species hypothesized to have evolved recently, ILS should play a large role in discordance among gene trees. However, species tree methods that model discordance due to ILS are sensitive to gene tree estimation error. Because many of our samples are population-level samples, we expect a relatively high level of gene tree estimation error with the loci we used. Thus, we build and contrast a concatenation-based species trees and a multispecies coalescent species tree.

We built a species tree from a concatenated supermatrix of all of our genes in RAxML-HPC version 8.2.0 (Stamatakis 2014). We concatenated all of our aligned gene sequences into a supermatrix, and created a partition file that characterized the boundary of each gene sequence, which allows different models of evolution to be fit for each of the genes. We used the GTRGAMMA model and the rapid bootstrap analysis (100 bootstraps).

We then used the program ASTRAL version 4.10.2 (Mirarab et al. 2014; Mirarab and Warnow 2015; Sayyari and Mirarab 2016) to build a species tree that incorporates ILS. We constructed gene trees for ASTRAL input with RAxML, using *Clarkia arcuata* as an outgroup, the GTRGAMMA model, and the rapid bootstrap analysis (100 bootstraps per gene tree). As gene tree estimation error can introduce bias into branch length estimates (Sayyari and Mirarab 2016), we collapsed branches with less than 33% support in each RAxML gene tree using DendroPy and SumTrees (version 4.5.2; Sukumaran and Holder 2010). We used the ASTRAL algorithm that computes local posterior probability support values for every branch based on gene tree quartet frequencies (Sayyari and Mirarab 2016). Branch

length units are in coalescent units, which are the ratio of the number of generations to the effective population size (Degnan and Rosenberg 2009). Shorter internal branch lengths could reflect fewer generations that have passed since divergence or a higher effective population size, both of which being scenarios where discordance among gene trees is more likely.

We were primarily interested in two aspects of the species tree topology: whether each species was monophyletic, and the relationships among the three species. Although both of our species trees had high support for the monophyly of each species (see *Results*), they differed in the relationship among the three species. We used the single-site log likelihood (SLL) method developed by Walker et al. (2018) to determine if outlier loci were causing the discordance in among-species relationships between the two species trees. The SLL method calculates per-site log likelihoods for the two species trees at each site in the supermatrix (in RAxML with the “-f g” command). The differences in log likelihoods between the two species trees across all sites are plotted to assess outlier loci—that is, loci that strongly support one species tree topology over the other. We identified 11 outlier loci as those falling outside of two standard deviations away from the mean difference in log likelihoods (see *Results*). The most extreme outliers were four genes that strongly supported the RAxML species tree. Because RAxML assumes all sites evolve according to a single evolutionary tree, as opposed to modeling variation in tree topology in ASTRAL, we reran the concatenated RAxML supermatrix analysis without the 11 outlier loci.

### DISCORDANCE AMONG GENE TREES

We explored gene tree discordance in two ways. First, we used the program PhyParts (Smith et al. 2015) to quantify the level of gene tree discordance in topology along each branch of the species tree. We used gene trees with branches under 33% bootstrap support collapsed, rooted them with the *C. arcuata* outgroup, and generated a rooted ASTRAL tree to use in the analysis. The output from PhyParts was visualized on the ASTRAL topology with `phypartspiecharts.py` (available at [github.com/mossmatters/phyloscripts](https://github.com/mossmatters/phyloscripts)). The PhyParts analysis outputs the numbers of gene trees that are concordant with the species tree topology at each branch, discordant with the species tree, or are uninformative (i.e., the gene tree has support values lower than 33% at that branch).

Second, we explicitly assessed the number of gene trees that show support for hypothesized progenitor-derivative relationships (i.e., a monophyletic derivative species nested within a paraphyletic progenitor species). If budding speciation occurred relatively recently, the majority of informative gene trees should support the pattern of nestedness consistent with a budding speciation scenario. If budding speciation was ancient, such that

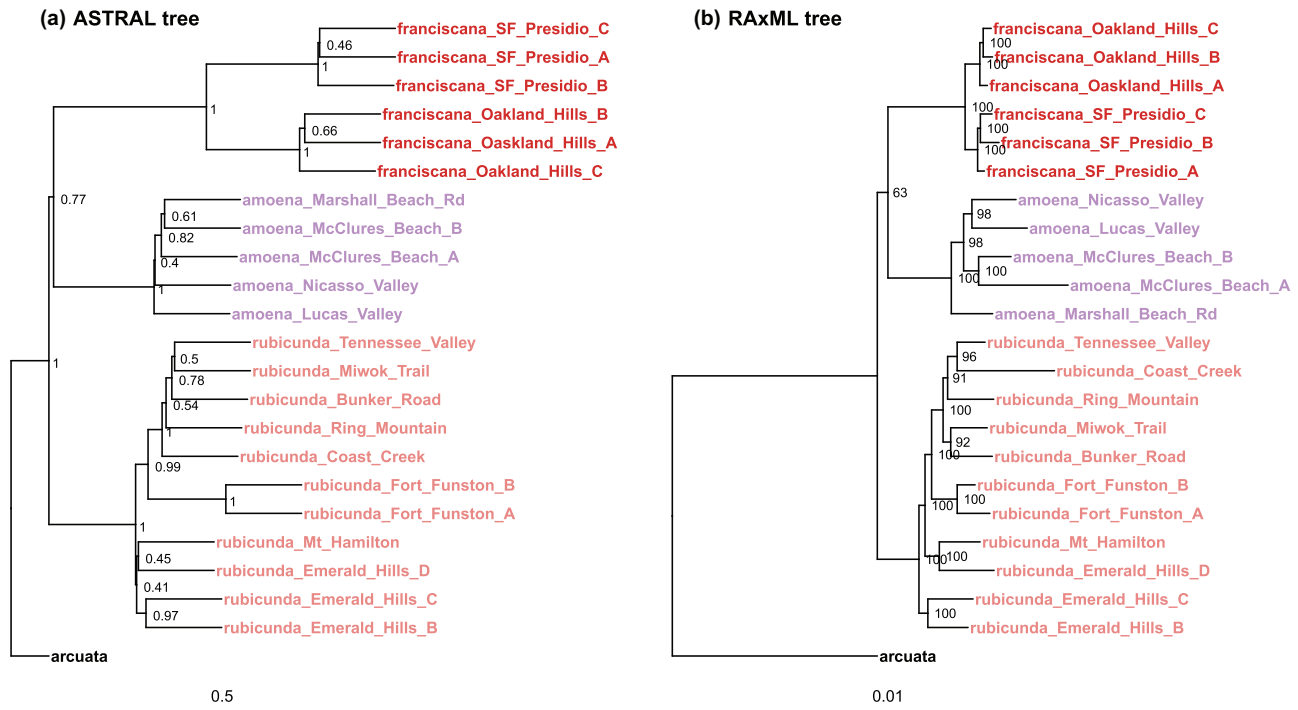
within-species gene flow erased the predominant pattern of nestedness, we predict that the majority of gene trees would support reciprocal monophyly and that there will be a sharp imbalance in the minority discordant gene trees, with the most frequent discordant gene tree supporting the ancient budding speciation scenario. We grouped gene trees given different patterns of nestedness and monophyly (Fig. 2) with code developed by Carlsen et al. (2018; `monophyly.R`; available from <https://github.com/tomas-fer/scripts/>) to qualitatively test these predictions. In a progenitor-derivative species pair, the derivative should be monophyletic, the derivative + progenitor should be monophyletic, and the progenitor should be paraphyletic. For every gene tree, we assessed whether the following groups were monophyletic: *C. franciscana*, *C. rubicunda*, *C. amoena*, *C. franciscana* + *C. rubicunda*, *C. franciscana* + *C. amoena*, and *C. rubicunda* + *C. amoena*. If *C. franciscana* was derived from within *C. rubicunda*, we would expect the *C. franciscana* and *C. franciscana* + *C. rubicunda* clades to be monophyletic, but *C. rubicunda* not to be monophyletic. We counted the number of gene trees that met these criteria. We used analogous criteria to quantify the number of gene trees that support *C. franciscana* being nested within *C. amoena*, *C. rubicunda* being nested within *C. amoena*, a double nested scenario where *C. rubicunda* is derived from *C. amoena* and *C. franciscana* is derived from *C. rubicunda*, and reciprocal monophyly of all three species (no budding speciation).

## Results

### PHYLOGENOMIC ANALYSES

The original concatenated supermatrix was 523,781 sites in length. The first concatenated RAxML tree we inferred had 100% bootstrap support for each species being monophyletic (Fig. S1). Nearly all of the among-population relationships within each species had 100% bootstrap support, which is likely an artifact of the overinflated bootstrap support that is commonly seen in concatenation analyses (Gadagkar et al. 2005). The relationships among the three species in the maximum likelihood tree showed *C. franciscana* and *C. rubicunda* as a clade, with *C. amoena* as its sister with 89% bootstrap support.

The ASTRAL species tree also showed strong local posterior probability support for the monophyly of each species (Fig. 3a). The within-species support values were generally lower than in the RAxML tree. However, each of the two populations of *C. franciscana* were still highly supported as monophyletic groups (local posterior probability = 1). Branch lengths, in coalescent units, separating the two *C. franciscana* populations (1.48, 1.23) were comparable in length to those leading to each species. The ASTRAL tree resolved *C. franciscana* as sister to *C. amoena* and that clade sister to *C. rubicunda*, with a local posterior probability



**Figure 3.** Species trees inferred from (a) ASTRAL, a coalescent-based summary method, and (b) a concatenated supermatrix without outlier loci in RAXML recover the same among-species relationships. (a) Node support values are local posterior probabilities. Branch lengths are measured in coalescent units, and terminal branches are scaled to 1. (b) Node support values are bootstrap supports. Branch lengths are measured in substitutions per site.

of 0.77. The branch length supporting that relationship, however, was very short (0.06).

We analyzed the discrepancy between our two species trees using the single site log likelihood test (Walker et al. 2018). We compared the summed log likelihoods for each gene and identified 11 outlier genes, the most extreme of which (AT4G29490\_12728, AT5G02250\_8194, AT5G03905\_25421, and AT5G50930\_27107) showed much greater support for the initial RAXML species tree than the ASTRAL species tree (Fig. S2). Outlier loci may be caused by misalignments (Walker et al. 2018), although we did not find evidence of large-scale misalignment within these genes, or unrecognized paralogs (Brown and Thomson 2017). The RAXML gene trees for these genes did not show a consistent topology, instead showing polyphyletic relationships among the species (Figs. S3, S4).

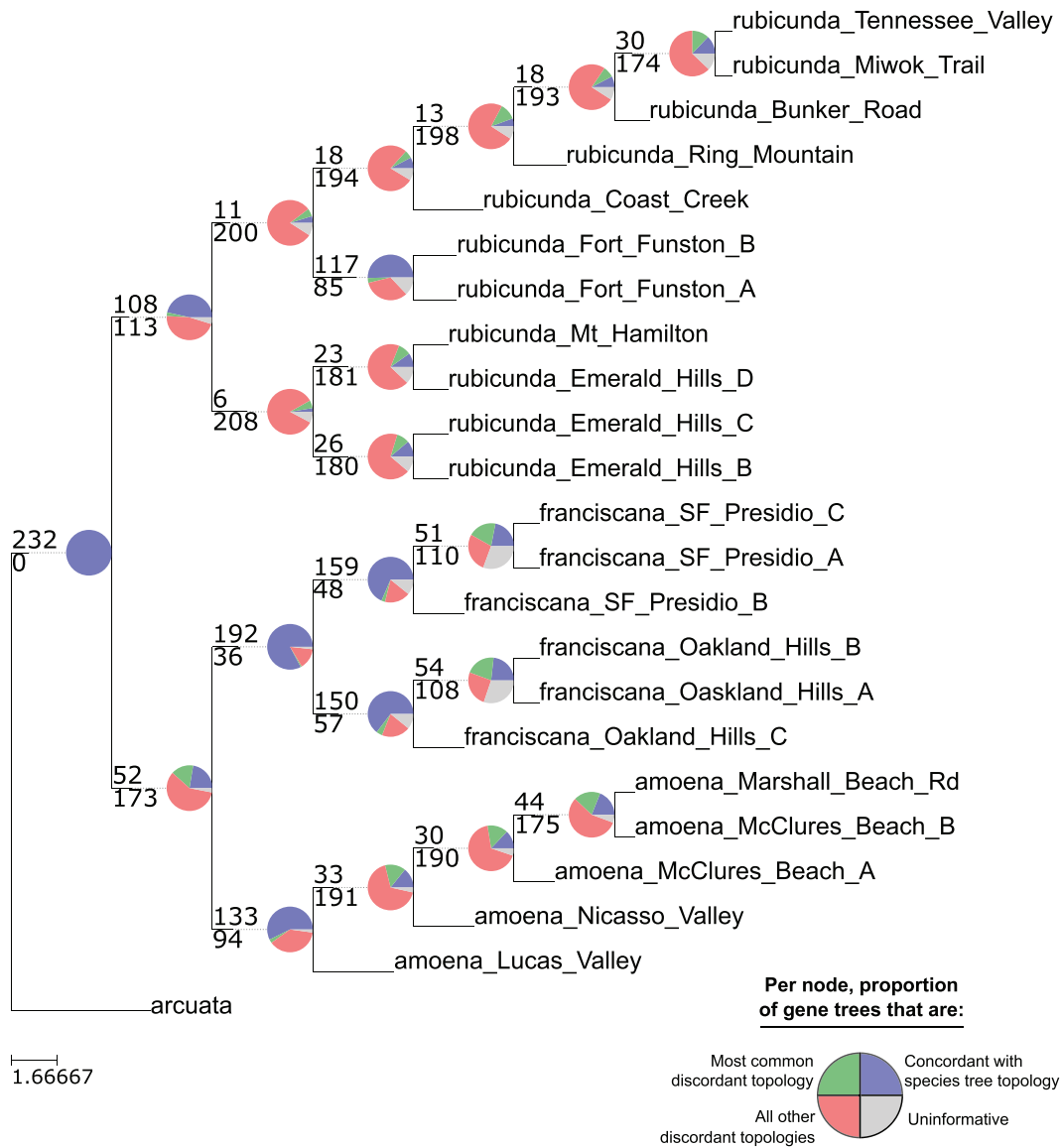
We created a new supermatrix without the 11 outlier loci that had 481,032 sites and reran the RAXML analysis. The inferred phylogenetic tree resolved each species as monophyletic (100% bootstrap support for each species, Fig. 3b). The new RAXML phylogeny supported the same relationship among the three species as the ASTRAL phylogeny—*C. franciscana* sister to *C. amoena*, and *C. rubicunda* sister to the former clade. Similar to the ASTRAL tree, the *C. franciscana*-*C. amoena* clade had low support (63% bootstrap support).

## GENE TREE DISCORDANCE

Our Phyparts analysis of gene tree discordance with the ASTRAL species tree showed that, although the branch supporting *C. franciscana* as sister to *C. amoena* has moderate support (0.77 local posterior probability), there was significant discordance at that node (Fig. 4). Of the 232 gene trees, only seven were uninformative, suggesting that the discordance was not due to a lack of information. Only 52 of the 232 gene trees supported *C. franciscana* and *C. amoena* in a clade sister to *C. rubicunda* (blue slice of pie chart, Fig. 4). There were 37 gene trees that supported *C. franciscana* as sister to a clade composed of *C. rubicunda* and *C. amoena* and 37 gene trees that supported *C. franciscana* and *C. rubicunda* in a clade sister to *C. amoena*. The remainder of informative discordant trees ( $n = 99$ ) supported polyphyletic relationships.

We explicitly counted the number of gene trees that would support hypothesized progenitor-derivative relationships. Specifically, we counted the number of trees in 5 scenarios: (1) *C. franciscana* is nested within *C. rubicunda*, with *C. amoena* as the basal sister, (2) *C. franciscana* is nested within *C. amoena*, with *C. rubicunda* as the basal sister, (3) *C. rubicunda* nested within *C. amoena*, with *C. franciscana* as the basal sister, (4) *C. franciscana* nested within *C. rubicunda*, which is turn is nested within *C. amoena*, and lastly (5) reciprocal monophyly of all three species,





**Figure 4.** Gene tree discordance visualized on the ASTRAL species tree. The pie charts at each node represent the number of gene trees that fall into one of three categories: concordant with the species tree (blue), discordant with the species tree (green for the most common alternative, and red for all other alternatives), and uninformative (gray, i.e., gene trees with less than 33% BS at that node). The number on top of each branch is the number of gene trees concordant with the species tree topology at that node (blue slice). The number on the bottom of each branch is the number of informative discordant topologies at that node (green + red slices).

agnostic of their relationship to one another. There were three gene trees that supported *C. franciscana* as being derived from within *C. rubicunda*, two gene trees that supported *C. franciscana* as being derived from within *C. amoena*, three gene trees that supported *C. rubicunda* as being derived from within *C. amoena*, and one gene tree that supported *C. franciscana* derived from *C. rubicunda*, which in turn is derived from *C. amoena* (Fig. 2). Eighty-two of the gene trees supported each species being reciprocally monophyletic. The remaining 141 gene trees were either

uninformative (i.e., polytomies among the species) or showed polyphyletic topologies.

### Discussion

Budding speciation that results in progenitor-derivative species pairs is thought to be a common phenomenon, especially in plants (Rieseberg and Brouillet 1994; Crawford 2010; Anacker and Strauss 2014; Grossenbacher et al. 2014). Unfortunately, it is

difficult to positively identify progenitor-derivative species pairs because many of the lines of evidence used (geographic range overlap and asymmetry, mating system transitions, ecological shifts, and phylogenetic relationships at a given locus) can change postspeciation. Here, we revisited a hypothesized case of recent budding speciation among three species—wherein the serpentine endemic *C. franciscana* was putatively derived from *C. rubicunda*, which in turn was putatively derived from *C. amoena* (Lewis and Raven 1958). Prior work on this group of species drew evidence from range distributions, habitat affinities, chromosomal rearrangements, morphology, mating system, and electrophoretic isozyme similarity, and yet the mode of speciation remained controversial. We took a phylogenomic approach, analyzing the history of 232 genes, to test Lewis and Raven's (1958) hypothesis.

### PHYLOGENOMIC DATA SUGGESTS RAPID, ALTHOUGH NOT RECENT, DIVERGENCE AMONG THE THREE CLARKIA SPECIES

Our phylogenomic analyses do not support the hypotheses that *C. franciscana* was recently derived from *C. rubicunda*, and that *C. rubicunda* was recently derived from *C. amoena*. Instead of finding patterns of phylogenetic nestedness consistent with budding speciation, we find that monophyly of each of the three species is the predominant and most supported topology, both by the species trees and the majority of the non-polyphyletic gene trees. If budding speciation happened far in the past, we expect to recover a signal of the progenitor-derivative relationships within the gene trees that are discordant with the species tree topology. However, we found only two out of 232 (0.9%) gene trees that placed *C. franciscana* as a monophyletic clade nested within *C. rubicunda*, and three out of 232 (1.1%) gene trees that show an alternative pattern of nestedness—with *C. franciscana* nested within *C. amoena*. Four (1.7%) gene trees supported *C. rubicunda* being nested within *C. amoena*. That these discordant tree topologies are so low in number and that no pattern of nestedness is qualitatively greater than the others suggests that their discordance with the species tree is the result of ILS instead of a remaining signal of budding speciation.

Our analysis does support the claim by Lewis and Raven (1958) that speciation in this triad happened rapidly. Rapid, almost simultaneous, speciation of the three species is indicated by the low support values of the branch supporting *C. franciscana* and *C. amoena* as sister in both the ASTRAL (0.87 local posterior probability) and the RAxML (25% bootstrap support) trees. Given the sheer number of sites used in the RAxML analysis and the tendency for concatenation analyses to overestimate bootstrap support (Edwards et al. 2007), it is likely that the RAxML analysis shows a true hard polytomy among the three species. Hard polytomies, indicative of rapid divergence and near simultane-

ous speciation, have historically been hard to distinguish from soft polytomies, which are due to the lack of phylogenetically informative characters (Purvis and Garland 1993). For example, a recent phylogenomic analysis of the Zingiberales found similar levels of gene tree discordance and low support values, leading the authors to conclude that this historically hard-to-resolve tropical group truly radiated rapidly (Carlsen et al. 2018). Lewis and Raven (1958) hypothesized that catastrophic selection, which fixes chromosomal deviants, drove speciation in this group. Indeed, given that the chromosomal rearrangements among species (primarily a suite of translocations) causes F1 hybrid sterility (Lewis and Raven 1958), it is likely that fixation of unique chromosomal rearrangements in *C. rubicunda* and *C. franciscana* was a mechanism contributing to rapid speciation in this triad.

Is it possible that budding speciation occurred very long ago and that there has been so much gene flow among progenitor populations to erase the signal of budding speciation? The amount of gene flow needed to erase the signal of budding speciation should be dependent on the degree of divergence among the progenitor populations at the time of budding speciation. If there is substantial population structure in the progenitor species at the time of budding speciation, more subsequent gene flow would be needed to erase the budding speciation signal than if there was little population structure. Previous morphological and cytological surveys of the three *Clarkia* species found substantial intraspecific variation among populations of the presumed progenitors, *C. rubicunda* and *C. amoena*, supporting their taxonomic division into two and five subspecies, respectively (Lewis and Lewis 1955; Lewis and Raven 1958). Our species trees also show clades supporting distinct groups of populations, particularly in *C. rubicunda*, for which our samples span more of the current species range. For example, in both the ASTRAL and RAxML species trees (Fig. 3) there is strong support for a clade consisting of *C. rubicunda* populations occurring in Marin County, north of the Golden Gate Bridge, with populations south of the bridge forming basal clades. The amount of intraspecific variation in these two species is inconsistent with the idea that a large amount of gene flow among populations within progenitor species has completely erased the signal of budding speciation.

Lewis and Raven (1958) hypothesized that speciation in this triad was recent, occurring more recently than the last glacial maximum (approximately 21,500 years ago). We use the coalescent branch lengths from the ASTRAL species tree and back-of-the-envelope math to speculate what effective population sizes per species would be needed for divergence to have happened more recently than the last glacial maximum. Coalescent branch lengths are in units of the number of generations divided by the number of effective haploid genomes in a population (Degnan and Rosenberg 2009). We calculated the branch lengths from each individual to the node with the most recent common

ancestor of the three species and averaged the total branch lengths within each species. We set the number of generations from the most recent common ancestor as 21,500, as these species are all annuals. The average effective population sizes for each species are 1553, 3021, and 3353 for *C. franciscana*, *C. amoena*, and *C. rubicunda*, respectively. This approximation assumes that effective population sizes are constant over time, which is likely not the case for the primarily self-fertilizing species, *C. franciscana*. Nevertheless, these approximated effective population sizes are extremely low and refute the hypothesis that this radiation occurred more recently than the last glacial maximum. Even the endangered *C. franciscana* is locally abundant, with patches comprising thousands of individuals (Gottlieb 1973; U.S. Fish and Wildlife Service 2010).

Gottlieb (1974a) was right to take a critical view of progenitor-derivative speciation in this group, and our results are consistent with his isozyme work but far more conclusive. *Clarkia franciscana* contained several unique isozyme alleles, and *C. rubicunda* and *C. amoena* were also distinct at a number of loci. This pattern of allelic variation could have resulted from the sorting of ancestral polymorphism into the three species at those isozyme loci, or from genic evolution within the isolated species following budding speciation. Our work shows that *C. franciscana* alleles are not generally derivative of *C. rubicunda* alleles, nor are *C. rubicunda* alleles derivative of *C. amoena*. Given the high discordance along the branch supporting the relationship of *C. franciscana* and *C. amoena*, it seems most likely that sorting of ancestral polymorphism led to the distinct number of isozyme loci in the three species seen by Gottlieb (1974a).

Ecological divergence was one of the leading lines of evidence to suggest that *C. franciscana* evolved from a *C. rubicunda*-like ancestor. *Clarkia franciscana* is currently strictly endemic to naturally toxic serpentine soils, whereas *C. rubicunda* has populations occurring on and off of serpentine. Because of its nested geographic range within *C. rubicunda* and its endemism to serpentine habitats, *C. franciscana* was thought to be a classic case of neoendemism (Stebbins and Major 1965). Neoendemics are recently evolved taxa specialized to habitat islands and are hypothesized to evolve from a small group of initial founders, thus representing a product of budding speciation (Stebbins and Major 1965; Kay et al. 2011). In contrast, paleoendemics are taxa that were once widespread species but became restricted to a narrow ecological niche (Stebbins and Major 1965). Our phylogenomic evidence is not consistent with the hypothesis that *C. franciscana* is a serpentine neoendemic because we do not find evidence of budding speciation. Additionally, the two *C. franciscana* populations on either side of the San Francisco Bay formed well-supported, divergent clades in our analyses, characterized by low levels of discordance, a result consistent with comparisons of isozyme loci between the populations (Gottlieb and Edwards

1992). It seems more likely that *C. franciscana* was once more widespread throughout the San Francisco Bay area and underwent a process of biotype depletion, wherein nonserpentine populations went extinct as the climate and competitive environment changed (Raven and Axelrod 1978; Anacker and Harrison 2011). The self-fertilization mating system and natural fluctuations in population size (Gottlieb 1973) could have facilitated rapid sorting of ancestral alleles in these two populations.

Future work in this group could employ population genomic studies, sampling both populations and genomes more densely, to continue to unravel the history of this group. For example, the timing of the species splits could be estimated with demographic analyses, and we could better understand fine-scale population structure and population changes that have occurred since speciation. It would be particularly fascinating to relate species and population histories to the topographic history of the recently formed California coast ranges.

#### WHAT CAN PHYLOGENOMIC DATA TELL US ABOUT BUDDING SPECIATION?

Budding speciation invokes dispersal to a new habitat, establishment of a new population, local adaptation, and the build-up of reproductive isolation. Given the intimate connection between local adaptation and reproductive isolation (Dobzhansky 1937; Sobel et al. 2010), it is unsurprising that budding speciation could occur, particularly in organisms with low dispersal capabilities. That many sister species pairs in topographically and ecologically diverse areas show some of the geographic and ecological patterns consistent with budding speciation suggests that it may be prevalent (Anacker and Strauss 2014; Grossenbacher et al. 2014). Although our results reject the hypothesis of budding speciation in the *C. franciscana*-*C. rubicunda*-*C. amoena* triad, our results cannot be extrapolated to all hypothesized cases of budding speciation. Rather, our results motivate the critical eye with which we should view speciation histories more generally, particularly given contemporary geographic patterns. For example, a recent phylogenomic study overturned the speciation history of the textbook example of pollinator and habitat divergence between largely parapatric and presumed sister species *Mimulus lewisii* and *Mimulus cardinalis* by showing that these two species respectively belong to clades with other species of similar pollination syndrome instead of forming a monophyletic pair (Nelson et al. 2021).

New evidence falsifying previously hypothesized speciation histories does not demote a system nor the work that has been done on it. It simply changes the questions asked or the framework in which results are interpreted. For example, work on *M. lewisii* and *M. cardinalis* (Schemske and Bradshaw 1999; Ramsey et al. 2003; Angert et al. 2008) can be viewed through the lens of species maintenance during secondary contact. Likewise,

our study does not discount the importance of ecological, morphological, and chromosomal divergence in speciation. Shifts to new soil substrates, mating systems, and chromosomal patterning have likely all played a role in reproductive isolation among the three *Clarkia* species, just not in the context of budding speciation. Moreover, the imbalance in anagenesis of ecological and reproductive traits across species, especially in *C. franciscana*, is particularly interesting. Rather than resulting from recent “catastrophic” speciation, our data show a relatively old species in decline. Its adaptation to harsh serpentine soils may have come at the cost of competitive ability off serpentine, as has been shown in other serpentine endemics (Sianta and Kay 2019), and its restriction to isolated serpentine patches may have promoted the shift to a highly self-fertilizing mating system, potentially compounding its long-term decline.

The ability to detect budding speciation with phylogenomic data will be maximized when the derivative species evolves from a progenitor with substantial population structure, because the low levels of gene flow among populations of the progenitor would preserve a predominant phylogenetic signal of budding speciation. For example, in *Layia*, one of the best-known examples of budding speciation, the progenitor species *L. glandulosa* shows strong population structure and ecological variation in soil affinities (Baldwin 2005). The budded species *L. discoidea* is phylogenetically nested within *L. glandulosa* and most closely related to geographically nearby *L. glandulosa* populations with similar soil affinities. In contrast, it would be difficult to detect a phylogenetic signal of budding speciation when the derivative species originated from within a species with little population structure, because the high levels of gene flow among populations of the progenitor should swamp out the signal of budding speciation and lead to reciprocal monophyly.

Introgression among diverged taxa can also produce patterns of gene tree discordance with the species tree and may be difficult to distinguish from a phylogenetic signal of budding speciation, depending on the timing of the speciation event and the extent of introgression. Unless introgression was very rampant and not countered by selection, we would not expect introgression to produce a dominant topology consistent with budding speciation. However, when the dominant topology shows reciprocal monophyly but an imbalanced subset of discordant gene trees shows nested relationships, the interpretation is more difficult. It could indicate that budding speciation occurred long enough ago that the signal has eroded, or it could indicate postspeciation introgression between one species and a subset of its sister species. An imbalance of discordant gene trees underlies the well-known test for introgression with the Patterson’s *D*-statistic (Durand et al. 2011). In fact, ancestral population structure is known to produce false positives for Patterson’s *D* when a subset of one species has been more isolated from the sister species, for example, as a re-

sult of a partial barrier to gene flow in the ancestral geographic range (Slatkin and Pollack 2008; Durand et al. 2011). This type of ancestral population structure in the progenitor is essentially the same as that invoked in the concept of budding speciation, along with more anagenesis in the derivative species. Thus, budding speciation and introgression may be indistinguishable with tests of “treeness” alone (i.e., the suite of tests using Patterson’s *D*) when the dominant topology shows reciprocal monophyly. Nevertheless, tests of introgression that incorporate branch length variation and/or distributions in coalescent timing may be able to distinguish the two if introgression is more recent than the budding speciation event (Hibbins and Hahn 2021). Lastly, the signal of budding speciation should be species wide for the derivative species (i.e., samples from all of the populations of the derivative species should be monophyletic and nested within the progenitor), whereas we would expect introgression to show a more local signal (i.e., only populations of the “derivative” species that experience introgression would show a signal of nestedness).

In our case, chromosomal rearrangements and the resulting strong postzygotic isolation among the *Clarkia* species make contemporary introgression unlikely, and we find strong support for reciprocal monophyly in both the species trees and gene trees. Therefore, we did not further investigate introgression as a possible cause of gene tree discordance. Future studies could use simulation to better understand which, if any, current tests of introgression could be used to distinguish ancient budding speciation from introgression under a variety of speciation parameters.

## Conclusions

Comparative analyses that use geographic range features of closely related species give insight into patterns of speciation modes that may be operating across clades in the tree of life. However, our results pertaining to Lewis and Raven’s (1958) classic story are another cautionary tale of using current species distributions as evidence of when and how speciation happened (Losos and Glor 2003). Although species distributions and additional lines of circumstantial evidence such as shifts in mating system, specialization to ecologically marginal habitats, and unique chromosomal arrangements (Crawford 2010) have been indicators of known instances of budding speciation (e.g., Lewis and Roberts 1956; Gottlieb 1974b, 2004; Baldwin 2005), they should not be taken as conclusive. The *C. franciscana*, *C. rubicunda*, and *C. amoena* species all showed multiple patterns consistent with budding speciation, and yet our phylogenomic analyses indicate that rapid budding speciation did not happen in this group. The high occurrence of gene tree discordance in our study system reinforces the importance of sampling a large number of

genes to understand evolutionary relationships, particularly when speciation occurred rapidly. Population-level phylogenomic analyses are becoming more accessible with the standardization of markers and protocols (e.g., Baker et al. 2021) and are our best method for positively identifying progenitor-derivative species pairs. In this way, we can better determine whether, as Leslie Gottlieb (2004) said, the “general models of the processes are consistent with the facts.”

## AUTHOR CONTRIBUTIONS

SAS and KMK designed the study and contributed to plant collections. SAS collected, processed, and analyzed phylogenomic data. SAS wrote the original manuscript and both authors contributed to the final version.

## ACKNOWLEDGMENTS

This work was funded partly by a National Science Foundation Dimensions in Biodiversity grant (DEB 1342873) to K. Skogen (Chicago Botanic Garden). We thank the team at the Chicago Botanic Garden (N. Wickett, K. Skogen, J. Fant, R. Overson, M. Johnson, and L. Bechen) for their guidance and advice, lab space, and resources in preparing samples for sequencing and in initial processing of sequence data. We thank L. Naumovich and Golden Hour Restoration Institute for help coordinating collection of *Clarkia franciscana* at Redwood Regional Park in Oakland, CA and E. Pimentel at the Presidio Trust for help coordinating collection of *C. franciscana* at the Presidio, San Francisco, CA. All collections were conducted with appropriate permits for *C. franciscana* (California Department of Fish and Wildlife permit No. 2081(a)-16-011-RP, National Park Service permit PRSF-2017-SCI-0001, United States Fish & Wildlife Service permit TE17017C-0) and *C. amoena* and *C. rubicunda* (National Park Service permits PORE-2007-SCI-0026 and GOGA-2007-SCI-0011). SAS was supported by a National Science Foundation Graduate Research Fellowship (NSF-DGE-1339067) and KMK was supported by a National Parks Ecological Research Postdoctoral Fellowship while collecting samples. We thank A. Keuter at the UCSC herbarium for help preparing specimens.

## DATA ARCHIVING

Bait sequences are available at <https://github.com/wickettlab/HybSeqFiles>.

Raw sequence reads are in the NCBI Sequence Read Archive under BioProject ID: PRJNA800094 and BioSample IDs: SAMN25232224–25232247.

Code and data (alignments, gene trees, and species trees) are deposited in Dryad (<https://doi.org/10.7291/D1F38Q>).

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## LITERATURE CITED

- Anacker, B.L. & Harrison, S.P. (2011) Climate and the evolution of serpentine endemism in California. *Evol. Ecol.*, 26, 1011–1023.
- Anacker, B.L. & Strauss, S.Y. (2014) The geography and ecology of plant speciation: range overlap and niche divergence in sister species. *Proc. R. Soc. B Biol. Sci.*, 280, 20132980.
- Angert, A.L., Bradshaw, H.D.Jr. & Schemske, D.W. (2008) Using experimental evolution to investigate geographic range limits in Monkeyflowers. *Evolution*, 62, 2660–2675.
- Baker, W.J., Dodsworth, S., Forest, F., Graham, S.W., Johnson, M.G., McDonnell, A., Pokorny, L., Tate, J.A., Wicke, S. & Wickett, N.J. (2021) Exploring Angiosperms353: an open, community toolkit for collaborative phylogenomic research on flowering plants. *Am. J. Bot.*, 108, 1059–1065.
- Baldwin, B.G. (2005) Origin of the serpentine-endemic herb *Layia discoidea* from the widespread *L. glandulosa* (Compositae). *Evolution*, 59, 2473–2479.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.*, 19, 455–477.
- Barracough, T.G. & Vogler, A.P. (2000) Detecting the geographical pattern of speciation from species-level phylogenies. *Am. Nat.*, 155, 419–434.
- Bartholomew, B., Eaton, L.C. & Raven, P.H. (1973) *Clarkia rubicunda*: a model of plant evolution in semiarid regions. *Evolution*, 27, 505–517.
- Bolger, A.M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114–2120.
- Brown, J.M. & Thomson, R.C. (2017) Bayes factors unmask highly variable information content, bias and extreme influence in phylogenomic analyses. *Syst. Biol.*, 66, 517–530.
- Capella-Gutiérrez, S., Silla-Martínez, J.M. & Gabaldón, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25, 1972–1973.
- Carlsen, M.M., Fér, T., Schmickl, R., Leong-Škorničková, J., Newman, M. & Kress, W.J. (2018) Resolving the rapid plant radiation of early diverging lineages in the tropical Zingiberales: pushing the limits of genomic data. *Mol. Phylogenet. Evol.*, 128, 55–68.
- Claremont, M., Reid, D.G. & Williams, S.T. (2012) Speciation and dietary specialization in *Drupa*, a genus of predatory marine snails (Gastropoda: Muricidae). *Zool. Scr.*, 41, 137–149.
- Cooper, B.J., Moore, M.J., Douglas, N.A., Wagner, W.L., Johnson, M.G., Overson, R.P., McDonnell, A.J., Levin, R.A., Raguso, R.A., Flores-Olivera, H., et al (2021) Target enrichment and extensive population sampling help untangle the recent, rapid radiation of *Oenothera* sect. *Calylophus*. *bioRxiv*, <https://doi.org/10.1101/2021.02.20.432097>.
- Crawford, D.J. (2010) Progenitor-derivative species pairs and plant speciation. *Taxon*, 59, 1413–1423.
- Degnan, J.H. & Rosenberg, N.A. (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.*, 24, 332–340.
- Dobzhansky, T. (1937) Genetics and the origin of species. Columbia Univ. Press, New York.
- Doyle, J. & Doyle, J. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.*, 19, 11–15.
- Duarte, J.M., Wall, P.K., Edger, P.P., Landherr, L.L., Ma, H., Pires, J.C., Leebens-Mack, J. & DePamphilis, C.W. (2010) Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evol. Biol.*, 10, 61.
- Durand, E.Y., Patterson, N., Reich, D. & Slatkin, M. (2011) Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.*, 28, 2239–2252.
- Edwards, S.V., Liu, L. & Pearl, D.K. (2007) High-resolution species trees without concatenation. *Proc. Natl. Acad. Sci. USA*, 104, 5936–5941.
- Fitzpatrick, B.M. & Turelli, M. (2006) The geography of mammalian speciation: mixed signals from phylogenies and range maps. *Evolution*, 60, 601–615.
- Gadagkar, S.R., Rosenberg, M.S. & Kumar, S. (2005) Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *J. Exp. Zool.*, 304B, 64–74.

- García, N., Folk, R.A., Meerow, A.W., Chamala, S., Gitzendanner, M.A., deOliveira, R.S., Soltis, D.E. & Soltis, P.S. (2017) Deep reticulation and incomplete lineage sorting obscure the diploid phylogeny of rain-lilies and allies (Amaryllidaceae tribe Hippeastreae). *Mol. Phylogenet. Evol.*, 111, 231–247.
- Gottlieb, L.D. (1973) Enzyme differentiation and phylogeny in *Clarkia franciscana*, *C. rubicunda* and *C. amoena*. *Evolution*, 27, 205–214.
- Gottlieb, L.D. (1974a) Gene duplication and fixed heterozygosity for alcohol dehydrogenase in the diploid plant *Clarkia franciscana*. *Proc. Natl. Acad. Sci. USA*, 71, 1816–1818.
- Gottlieb, L.D. (1974b) Genetic confirmation of the origin of *Clarkia linguata*. *Evolution*, 28, 244–250.
- Gottlieb, L.D. (2004) Rethinking classic examples of recent speciation in plants. *New Phytol.*, 161, 71–82.
- Gottlieb, L.D. & Edwards, S.W. (1992) An electrophoretic test of the genetic independence of a newly discovered population of *Clarkia franciscana*. *Madroño*, 39, 1–7.
- Gottlieb, L.D. & Ford, V.S. (1996) Phylogenetic relationships among the sections of *Clarkia* (Onagraceae) inferred from the nucleotide sequences of PgiC. *Syst. Bot.*, 21, 45–62.
- Grant, V. (1981) Plant speciation. Columbia Univ. Press, New York.
- Grossenbacher, D.L., Veloz, S.D. & Sexton, J.P. (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.
- Hibbins, M.S. & Hahn, M.W. (2021) Phylogenomic approaches to detecting and characterizing introgression. *EcoEvoRxiv*, <https://doi.org/10.32942/osf.io/uahd8>.
- Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J.C. & Wickett, N.J. (2016) HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl. Plant Sci.*, 4:1600016.
- Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, 30, 772–780.
- Kay, K.M., Ward, K.L., Watt, L.R. & Schemske, D.W. (2011) Plant speciation. Pp. 71–97 in S. P.Harrison and N.Rajakaruna, eds. *Serpentine: the evolution and ecology of a model system*. Univ. of California Press, Berkeley, CA.
- Kisel, Y. & Barraclough, T.G. (2010) Speciation has a spatial scale that depends on levels of gene flow. *Am. Nat.*, 175, 316–334.
- Kubatko, L.S. & Degnan, J.H. (2007) Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst. Biol.*, 56, 17–24.
- Lewis, H. (1953) The mechanism of evolution in the genus *Clarkia*. *Evolution*, 7, 1–20.
- Lewis, H. (1962) Catastrophic selection as a factor in speciation. *Evolution*, 16, 257–271.
- Lewis, H. & Lewis, M. (1955) The genus *Clarkia*. *Univ. Calif. Publ. Bot.*, 20, 241–392.
- Lewis, H. & Roberts, M.R. (1956) The origin of *Clarkia linguata*. *Evolution*, 10, 126–138.
- Lewis, H. & Raven, P.H. (1958) Rapid evolution in *Clarkia*. *Evolution*, 12, 319–336.
- Liu, L., Yu, L., Kubatko, L., Pearl, D.K. & Edwards, S.V. (2009) Coalescent methods for estimating phylogenetic trees. *Mol. Phylogenet. Evol.*, 53, 320–328.
- Losos, J.B. & Glor, R.E. (2003) Phylogenetic comparative methods and the geography of speciation. *Trends Ecol. Evol.*, 18, 220–227.
- Malay, M.C.M.D. & Paulay, G. (2010) Peripatric speciation drives diversification and distributional pattern of reef hermit crabs (Decapoda: Diogenidae: Calcinus). *Evolution*, 64, 634–662.
- Mayr, E. (1954) Change of genetic environment and evolution. Pp. 157–180 in J.Huxley, A. C.Hardy & E. B.Ford, eds. *Evolution as a process*. Collier Books, New York.
- Mirarab, S. & Warnow, T. (2015) ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*, 31, i44–i52.
- Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S. & Warnow, T. (2014) ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*, 30, 541–548.
- Morales-Briones, D.F., Liston, A. & Tank, D.C. (2018) Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). *New Phytol.*, 218, 1668–1684.
- Morales-Briones, D.F., Kadereit, G., Tefarikis, D.T., Moore, M.J., Smith, S.A., Brockington, S.F., Timoneda, A., Yim, W.C., Cushman, J.C. & Yang, Y. (2021) Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in Amaranthaceae s.l. *Syst. Biol.*, 70, 219–235.
- Nelson, T.C., Stathos, A.M., Vanderpool, D.D., Finseth, F.R., Yuan, Y.W. & Fishman, L. (2021) Ancient and recent introgression shape the evolutionary history of pollinator adaptation and speciation in a model monkeyflower radiation (*Mimulus* section *Erythranthe*). *PLoS Genet.*, 17, 1–26.
- Patsis, A., Overson, R.P., Skogen, K.A., Wickett, N.J., Johnson, M.G., Wagner, W.L., Raguso, R.A., Fant, J.B. & Levin, R.A. (2021) Elucidating the evolutionary history of *Oenothera* sect. *Pachylophus* (Onagraceae): a phylogenomic approach. *Syst. Bot.*, 46, 799–811.
- Purvis, A. & Garland, T. (1993) Polytomies in comparative analyses of continuous characters. *Syst. Biol.*, 42, 569–575.
- Ramsey, J., Bradshaw, H.D. & Schemske, D.W. (2003) Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution*, 57, 1520–34.
- Raven, P.H. & Axelrod, D.I. (1978) Origins and relationships of the California flora. 72nd ed. Univ. of California Press, Berkeley, CA.
- Rieseberg, L.H. & Brouillet, L. (1994) Are many plant species paraphyletic? *Taxon*, 43, 21–32.
- Roch, S. & Warnow, T. (2015) On the robustness to gene tree estimation error (or lack thereof) of coalescent-based species tree methods. *Syst. Biol.*, 64, 663–676.
- Sayyari, E. & Mirarab, S. (2016) Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.*, 33, 1654–1668.
- Schemske, D.W. & Bradshaw, H.D. (1999) Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). *Proc. Natl. Acad. Sci. USA*, 96, 11910–5.
- Sianta, S.A. & Kay, K.M. (2019) Adaptation and divergence in edaphic specialists and generalists: serpentine soil endemics in the California flora occur in barer serpentine habitats with lower soil calcium levels than serpentine tolerators. *Am. J. Bot.*, 109, 1–14.
- Slater, G.S.C. & Birney, E. (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, 6, 1–11.
- Slatkin, M. & Pollack, J.L. (2008) Subdivision in an ancestral species creates asymmetry in gene trees. *Mol. Biol. Evol.*, 25, 2241–2246.
- Smith, S.A., Moore, M.J., Brown, J.W. & Yang, Y. (2015) Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evol. Biol.*, 15, 150.
- Snow, R. (1963) Cytogenetic studies in *Clarkia*, section *Primigenia*. I. A cytological survey of *Clarkia amoena*. *Am. J. Bot.*, 50, 337–348.
- Snow, R. (1964) Cytogenetic studies in *Clarkia*, section *Primigenia*. III. Cytogenetics of monosomics in *Clarkia amoena*. *Genetica*, 35, 205–235.
- Sobel, J.M., Chen, G.F., Watt, L.R., Schemske, D.W. & Rausher, M. (2010) The biology of speciation. *Evolution*, 64, 295–315.

- Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312–1313.
- Stebbins, G.L. & Major, J. (1965) Endemism and speciation in the California flora. *Ecol. Monogr.*, 35, 1–35.
- Sukumaran, J. & Holder, M.T. (2010) DendroPy: a Python library for phylogenetic computing. *Bioinformatics*, 26, 1569–1571.
- U.S. Fish and Wildlife Service (2010) *Clarkia franciscana* (Presidio clarkia) 5-year review: summary and evaluation. U.S. Fish & Wildlife Service, Washington, D.C.
- Walker, J.F., Brown, J.W. & Smith, S.A. (2018) Analyzing contentious relationships and outlier genes in phylogenomics. *Syst. Biol.*, 67, 916–924.
- Weitemier, K., Straub, S.C.K., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A. & Liston, A. (2014) Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. *Appl. Plant Sci.*, 2:1400042.

Associate Editor: D. Filatov  
Handling Editor: A. McAdam

### *Supporting Information*

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

SUPPORTING INFORMATION