doi:10.1111/j.1558-5646.2011.01521.x

TEMPO AND MODE OF MATING SYSTEM EVOLUTION BETWEEN INCIPIENT *CLARKIA* SPECIES

James B. Pettengill[1] and David A. Moeller[1,2]

[1]*Department of Plant Biology, University of Minnesota, Saint Paul, Minnesota 55108*

[2]*E-mail: moeller@umn.edu*

Received March 10, 2011
Accepted October 31, 2011

Mating systems are among the most labile characteristics of flowering plants, with transitions frequently occurring among populations or in association with speciation. The frequency of mating system shifts has made it difficult to reconstruct historical evolutionary dynamics unless transitions have been very recent. Here, we examine molecular and phenotypic variation to determine the polarity, timescale, and causes of a transition between outcrossing and self-fertilization in sister subspecies of *Clarkia xantiana*. Phylogenetic analyses and coalescent-based estimates of the time to most recent common ancestor indicated that outcrossing is ancestral to selfing and that there has been a single origin of selfing. Estimates of divergence time between outcrossing and selfing subspecies were 10,000 (95% CI [credible interval]: 3169–66,889) and 65,000 years ago (95% CI: 33,035–151,448) based on two different methods, suggesting a recent and rapid evolutionary transition. Population genetic data indicated that the transition to selfing was associated with a 80% reduction in molecular diversity, which is much greater than the 50% reduction expected under a shift from obligate outcrossing to obligate self-fertilization alone. Our data also suggest that this severe loss of diversity was caused by colonization bottlenecks. Together with previous studies, evidence for reproductive assurance in *C. xantiana* now connects variation in plant–pollinator interactions in the field to phenotypic and molecular evolution.

KEY WORDS: **Coalescence, divergence time, incomplete lineage sorting, phylogeography, reproductive assurance, self-fertilization, species delimitation.**

Transitions between outcrossing and self-fertilization have occurred repeatedly throughout the evolutionary history of angiosperms (Stebbins 1974; Holsinger 2000). Selfing has arisen in some cases through the loss of a self-incompatibility system (reviewed in Igic et al. 2008) but frequently occurs when self-compatible populations transition between outcrossing and autonomous selfing as primary forms of reproduction (Richards 1999). Although phylogenetic studies have revealed important insights into the macroevolution of mating systems (e.g., Schoen et al. 1997; Goodwillie 1999; Barrett 2002; Igic et al. 2006; Foxe et al. 2009), details of the historical dynamics of transitions have been challenging to reconstruct because mating systems are highly labile. The evolutionary history of conspecific populations and recently diverged taxa that differ in mating system has been of particular interest for understanding the polarity of transitions, the timing of divergence, and the mechanisms responsible for mating system shifts (e.g., Foxe et al. 2010; Ness et al. 2010).

The direction of evolutionary change has typically been assumed to occur from outcrossing to selfing. However, it has been debated as to whether mating system transitions are unidirectional and whether selfing is an evolutionary dead end (Stebbins 1957; Takebayashi and Morrell 2001; Escobar et al. 2010). In higher level phylogenies, it is difficult to infer whether and how often a character state is gained or lost, particularly when that character evolves rapidly and when diversification and extinction rates of lineages are associated with character states (Goldberg and Igic 2008). In the past, genomic and computational resources limited the feasibility of studies investigating the historical dynamics of

mating system transitions at the population level or among closely related species. However, given the increasing ease of obtaining sequence data and the development of divergence population genetic approaches based on the coalescent (Hey and Nielsen 2004; Drummond and Rambaut 2007), it is now feasible to examine the recent evolutionary history of shifts in reproductive strategies.

Two general models have been proposed to explain the evolution of self-fertilization (mating system models reviewed in Uyenoyama et al. 1993; Goodwillie et al. 2005). One model suggests that selfing can evolve due to an inherent 50% transmission advantage of selfers over obligate outcrossers (Fisher 1941; Jain 1976). This automatic transmission model rests on the condition that selfing individuals can contribute genes to the next generation not only by selfing but also by siring offspring on other individuals, resulting in a fitness advantage over obligately outcrossing individuals. A second model of mating system evolution suggests that selfing evolves as a form of reproductive assurance in environments where outcrossing is unlikely to occur because pollinators or mates are unreliable (Stebbins 1957; Lloyd 1979, 1992; Cheptou 2004; Morgan and Wilson 2005). This model has received increasing support from field studies showing that selfing elevates seed production (Herlihy and Eckert 2002; Elle and Carney 2003; Kalisz et al. 2004; Busch 2005) and that the selfing phenotype is favored by natural selection when limited mates and effective pollinators cause pollen limitation (Moeller and Geber 2005; Moeller 2006). It is important to note that these two models are not necessarily mutually exclusive; however, they provide substantially different predictions about the process of mating system evolution.

Although field studies provide insights into the adaptive significance of alternative mating strategies, they are less successful in identifying the original causes of mating system transitions. It has been proposed that the two different causes of the evolution of selfing leave different signatures at the molecular level and that population genetic analyses can provide information regarding the history and mode of mating system divergence (Schoen et al. 1996). Under both models, selection on mating system modifier loci (and hitchhiking effects on linked loci) should cause losses of molecular variation. In addition, high rates of selfing should cause as much as a 50% reduction in neutral diversity across the genome (Charlesworth et al. 1993; Nordborg 2000); for transitions from mixed mating (partial selfers) to higher rates of selfing, the expected loss of neutral variation should be less than 50%. The key difference between models is that an advantage of selfing under reproductive assurance is the potential for populations to colonize new sites void of pollinators or mates (Baker 1955, 1967; Pannell and Barrett 1998). Frequent population bottlenecks associated with colonization (i.e., founder effects) under the reproductive assurance model are expected to result in strong genome-wide reductions in neutral diversity, greater than the 50% reduction in molecular diversity that is expected for a shift from obligate outcrossing to obligate self-fertilization alone (Schoen et al. 1996). Population bottlenecks are also expected to result in high variance in the frequency distribution of polymorphisms among loci due to random genetic drift (Charlesworth et al. 2003; Wright and Gaut 2005).

Under the automatic transmission model, by contrast, the advantage of selfing is dependent on the availability of mates (i.e., relatively large population sizes) and pollen vectors for outcrossing (Jain 1976). Frequent population bottlenecks are not predicted to accompany the evolution of selfing via automatic selection and therefore populations should maintain higher levels of neutral genetic variation than under reproductive assurance (Schoen et al. 1996). One problem for differentiating between models is that the transition from outcrossing to selfing must have occurred recently. When mating system transitions have occurred in the distant past, it is difficult to distinguish whether extant patterns of genetic diversity were affected by mating system divergence versus other demographic and evolutionary phenomena that may have occurred after the speciation process (i.e., complete reproductive isolation and genealogical exclusivity). Therefore, examining the signature of mating system evolution using molecular population genetic data is most powerful when taxa have only recently begun to diverge (Schoen et al. 1996).

In this study, we examine the evolutionary history of mating system evolution in *Clarkia xantiana* A. Gray (Onagraceae), which consists of a predominantly selfing subspecies, *parviflora*, and a mixed mating but primarily outcrossing subspecies, *xantiana*. The geographic distribution, pollination ecology, and floral biology of this species are well documented (Eckhart and Geber 1999; Runions and Geber 2000; Moeller 2006; Eckhart et al. 2011), which has allowed us to comprehensively sample populations and floral variation across the species' range. Both taxa are self-compatible and preliminary studies indicate that they are cross-compatible to some extent, producing viable and fertile F1 hybrids in controlled environments (D.A. Moeller, unpubl. data). The two subspecies can be distinguished primarily by their flowers and flowering time (Eckhart and Geber 1999). Where the subspecies are sympatric and co-occur, they are distinct and hybrids are not commonly observed.

Compared to other investigations of mating system evolution using molecular population genetics, this system is unusual in that a series of field experiments have supported the reproductive assurance model. Specifically, selfing provides reproductive assurance and the selfing phenotype is favored by natural selection in environments where few mates and effective pollinators limit opportunities for outcrossing (Fausto et al. 2001; Geber and Eckhart 2005; Moeller and Geber 2005). The pattern of mating system divergence among populations is also correlated with the

distribution and abundance of specialized pollinators (Moeller 2006). In this article, we examine the evolutionary history of outcrossing and selfing taxa of *C. xantiana* using DNA sequence data (eight nuclear and three chloroplast loci) and microsatellite variation (four loci) collected from 154 individuals representing 31 geographic populations. We used phylogenetic- and coalescent-based methods to determine the number of transitions between mating strategies, the ancestral mating system, and the time of divergence. We then used taxon-wide and population-specific patterns of sequence polymorphism to test theoretical predictions for patterns of molecular variation under the reproductive assurance and automatic transmission models of mating system evolution.

Methods

GEOGRAPHY AND NATURAL HISTORY

Clarkia xantiana A. Gray ssp. *xantiana* and *C. xantiana* A. Gray ssp. *parviflora* (Eastw.) Harlan Lewis & P. H. Raven are winter annuals that are endemic to the southern Sierra Nevada foothills of Kern and Tulare Counties, California. Their distributions extend south through the Tehachapi Mountains to the Transverse Ranges (Liebre and San Gabriel Mountains), where populations are less common (Fig. 1). The subspecies are parapatric with a narrow zone of sympatry (~5 km) at *xantiana*'s eastern and *parviflora*'s western range margin (Eckhart and Geber 1999; Eckhart et al. 2010). In the zone of sympatry, subspecies often occur within meters to tens of meters of one another in the same site. Their distributions occupy different sections of a dominantly west-to-east environmental gradient where the eastern section, occupied by *parviflora*, receives less and more variable precipitation compared to the western section, occupied by *xantiana* (Eckhart et al. 2010, 2011).

In *xantiana*, outcrossing is effected mainly by solitary bees, some of which are specialized on the genus *Clarkia* (Moeller 2005; Eckhart et al. 2006; Moeller 2006). Outcrossing is also promoted by physical separation of the anthers and receptive stigma (herkogamy) and the temporal precedence of anther dehiscence over stigma receptivity (protandry). Floral manipulation experiments in the field have demonstrated that autonomous selfing does not contribute significantly to seed production in *xantiana* (Moeller et al. 2012). However, geitonogamous selfing (between flowers on the same individual) and biparental inbreeding are likely common due to multiflowered inflorescences and limited dispersal. In *parviflora* populations, visits by pollinators are uncommon and the specialist solitary bee pollinators of *Clarkia* are not present (Moeller 2006). Herkogamy and protandry are strongly reduced and autonomous selfing occurs readily just prior to, coincident with, or shortly after flower opening (Eckhart and Geber 1999; Runions and Geber 2000; Moeller 2006). Rates of autonomous self-fertilization in pollinator-free greenhouse environments typically exceed 90% for *parviflora* and less than 10% for *xantiana* (Moeller 2006). Within-population studies indicate outcrossing rates range from 0.62—0.73 (95% CI [credible interval]: 0.58–0.76) for six *xantiana* populations and 0.12 and 0.13 (95% CI: 0.10–0.15) for two *parviflora* populations (Moeller et al. 2012).

POPULATION SAMPLING AND MEASUREMENT OF FLORAL TRAITS

We sampled 31 populations (15 *xantiana*, 16 *parviflora*) spanning the species' distribution except for a section of the Tehachapi Mountains that is privately owned and inaccessible (Fig. 1, Table S1). Populations of the two subspecies that share the same number as a label (e.g., 5x and 5p) co-occur in the narrow zone of sympatry (Fig. 1, Table S1). We planted one seed from each of five haphazardly chosen field-collected maternal families from each population (four individuals from 68p). Plants were initially grown in environmental chambers, where leaves were collected for DNA extraction, and subsequently grown in a greenhouse, where floral traits were measured. Therefore, molecular data and floral phenotypes were ascertained from the same individuals. Floral traits were measured on the first two flowers that opened on each plant; we used the average of these two measurements in our analyses (see also Moeller 2006). We measured petal length, petal width, herkogamy, and dichogamy. Herkogamy was measured as the distance from the receptive stigma to the nearest anther and was measured just as the stigma became receptive. Dichogamy was measured as the amount of time between long-anther dehiscence and stigma receptivity, which was determined by repeatedly observing flowers throughout floral development. We estimated individual autofertility (autonomous selfing rates) by determining fruit set across all flowers on each plant (mean of 35.5 flowers/plant). Plants were spaced widely and not disturbed during flowering to prevent contamination.

DNA SEQUENCING AND MICROSATELLITE GENOTYPING

DNA was extracted using Qiagen DNeasy plant mini kits (Valencia, CA). We used PCR to amplify eight nuclear genomic regions (nDNA) and three chloroplast regions (cpDNA) from the 154 individuals representing the 31 populations. The PCR primers used to amplify single-copy nDNA loci were designed from EST sequences isolated from *C. breweri* flower buds (see Moeller et al. 2011). The three cpDNA (*psbA-trnH*, Kress and Erickson 2007; *trnT-trnL*; and *trnL-trnF*, Taberlet et al. 1991) regions were chosen because they are among the most variable regions of the chloroplast (Shaw et al. 2005).

PCR products were sequenced directly except for *g3pdh* and a subset of other loci. For these exceptions, PCR products were cloned into pGem-T Easy vectors (Promega Corp., Madison, WI) and one to five cloned products were sequenced.

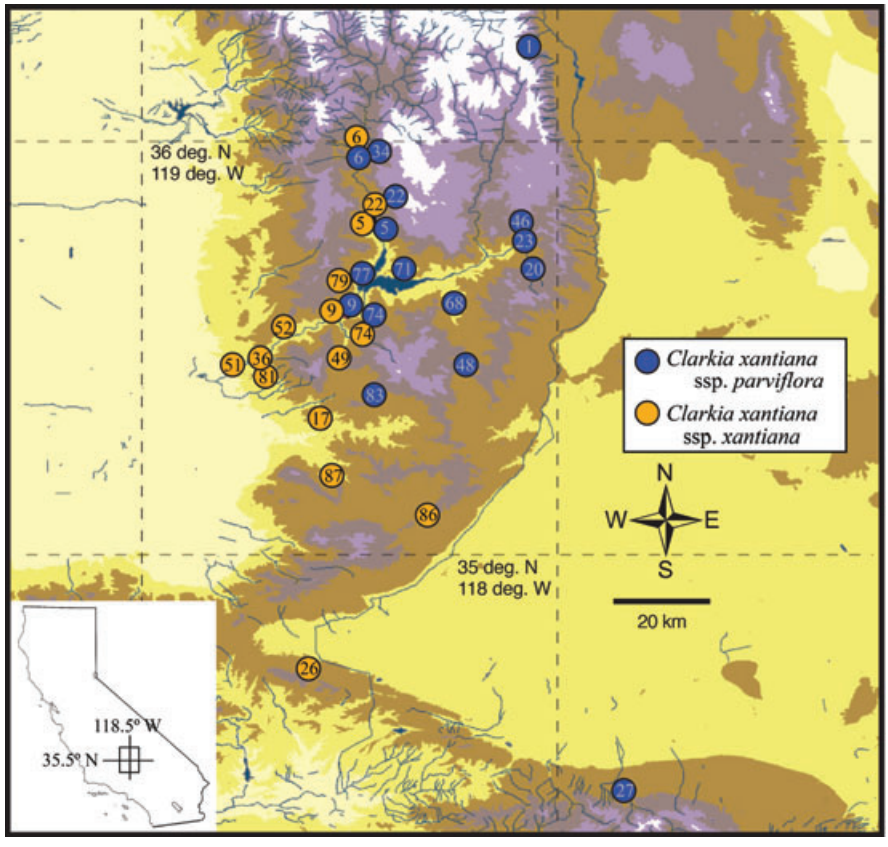


Figure 1. **Map showing the location of the 16 *C. xantiana* ssp. *parviflora* and 15 *C. xantiana* ssp. *xantiana* populations sampled in this study. The numbers that label each population correspond to those in Table S1. Color shading shows elevations in 500-m intervals from deepest canyons (pale yellow, 0–500 m) to highest peaks (white, >2500 m).**

Sanger sequencing was conducted using the ABI BigDye version 3.1 (Applied Biosystems, Foster City, CA) chemistry in conjunction with an ABI 3730xl DNA sequencer. All chromatograms were manually inspected using SEQUENCHER version 4.8 (Gene Codes Corp., Ann Arbor, MI) to determine polymorphic sites, which were coded according to IUPAC ambiguity codes, and assemble consensus sequences from bidirectional sequence reads. We used the program PHASE version 2.1 (Stephens et al. 2001; Stephens and Donnelly 2003) to infer the haplotypes treating cloned fragments as known haplotypes. To avoid biasing results due to nucleotide misincorporation into cloned products, rare polymorphisms detected from cloned loci were confirmed by directly sequencing PCR products. Alignments were inferred using the program MUSCLE (Edgar 2004) with the default settings and were minimally edited to reduce erroneous homology statements. The total length of reliable sequence was used for the phylogenetic analyses (see Table 1 for GenBank accession numbers).

Genomic DNA from the same 154 individuals was used to genotype four dinucleotide repeat microsatellite loci. Primer sequences and amplification conditions can be found in Moeller et al. (2011). PCR was conducted separately for each locus with the 6-FAM and NED dyes; amplified products were combined for fragment separation on an ABI 3730xl using LIZ-500 as a size standard. A subset of individuals was re-run to confirm alleles. All fragment sizes were determined by directly examining each electropherogram.

PHYLOGENETIC INFERENCE AND EVOLUTIONARY DISTINCTIVENESS

We estimated the phylogenetic relationships among individuals using both the maximum-likelihood approach of GARLI version 0.951 (Genetic Algorithm for Rapid Likelihood Inference; Zwickl 2006) and the Bayesian method implemented in MRBAYES version 3.1 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003). To evaluate the range of topologies and associated likelihood scores across independent runs, we conducted 1000 analyses of the original nonbootstrapped dataset with GARLI using the default parameter settings. Statistical support for topological relationships based on the maximum-likelihood method was assessed through 1000 bootstrap replicates. All analyses with GARLI were performed using the computing resources associated with The Lattice Project (Bazinet and Cummings 2008). We used the default settings within MRBAYES except for the following: GTR + Γ + I nucleotide substitution model, a heating

Table 1. Summary statistics for 11 sequenced loci: number of sequences obtained (*N*), length of fragment used in the phylogenetic analyses (L_p), length of fragment for phased coalescent methods (L_c), segregating sites (*S*), haplotype diversity (H_d), population structure (F_{st}), genealogical sorting index (gsi_T), average posterior probability of phased haplotypes (*Phase*), nucleotide substitution model (*NST*), and GenBank accession numbers.

	N							*xantiana*	*parviflora*	*xantiana*	*parviflora*			
Locus	All	*xantiana*	*parviflora*	L_p	L_c	*S*	H_d	F_{st}	F_{st}	gsi_T	gsi_T	Phase	*NST*	Genbank accessions
nDNA														
a23	79	19	60	473	473	25	0.88	0.42	0.7	0.05*	0.30*	0.86	HKY + I	JQ009828–JQ009844, JQ010748–JQ010836
d13	100	30	70	839	334	21	0.43	0.43	0.76	0.06*	0.25*	0.98	HKY + G + I	JQ009731–JQ009827
d5	143	65	78	392	391	37	0.94	0.15	0.59	0.11*	0.16*	0.89	GTR + G + I	JQ009584–JQ009730
f9	152	73	74	543	542	80	0.88	0.33	0.44	0.06*	0.40*	0.93	HKY + G	JQ010595–JQ010747
g3pdh[1]	138	61	72	504	503	100	0.91	0.38	0.54	0.04	0.45*	0.9	GTR + G + I	JQ009845–JQ009982
i11	154	75	79	424	424	48	0.84	0.14	0.46	0.21*	0.23*	0.83	HKY + G	JQ010441–JQ010594
ipi2	145	70	75	914	913	25	0.69	0.15	0.83	0.13*	0.01	0.9	GTR + I	JQ010291–JQ010440
k22[1]	153	75	78	649	648	83	0.95	0.19	0.74	0.19*	0.34*	0.9	GTR + G + I	JQ010137–JQ010290
cpDNA														
psbA– trnH	145	66	79	228	67	12	0.04	0.48	0.73	0.01	0.2*	–	K81uf + I	JQ009448–JQ009583
trnT– trnL	120	53	67	122	41	1	0.45	–	–	0.13	0.13	–	K81uf	[3]
trnL– trnF	154	75	79	416	415	4	0.34	–	–	0.14	0.12	–	HKY	JQ009983–JQ010136
Overall[2]	1483	662	811	5504	4751	436	0.67	0.3	0.64	0.58	0.75	0.9	–	–

[1]Violated the assumptions of a molecular clock.

[2]Sums are provided for all metrics except for gsi_T values, which are estimates based on 100 bootstrap replicates of the concatenated dataset, and the phase posterior probabilities, which represent the mean.

[3]Sequences were <200 bp and could not be deposited in GenBank. They are available upon request from the corresponding author.

*Significant at $P < 0.05$

parameter of 0.03, and 0.01 as the mean of the prior on branch length. Bayesian Markov chain Monte Carlo (MCMC) posterior probabilities were estimated after removing 25% of the 4.1×10^6 generations as burnin. To provide polarity among nucleotide states and infer ancestral relationships, we used six congeneric outgroup taxa (*C. concinna*, *C. davyi*, *C. arcuata*, two accessions of *C. amoena* ssp. *huntiana*, and *C. amoena* ssp. *caurina*).

Given that the phylogenetic analyses cannot reconstruct reticulate relationships that result from recombination or hybridization, we also estimated relationships using the phylogenetic network method NeighborNet within SplitsTree4 (Huson and Bryant 2006). Outgroups were excluded from the analyses and the network was constructed based on Kimura-2-parameter distances (Kimura 1980). Support for edges were estimated from 1000 bootstrap replicates.

To quantify the evolutionary distinctiveness of the two taxa based on the degree of genealogical exclusivity, we calculated the ensemble genealogical sorting index (gsi_T; Cummings et al. 2008) using 100 bootstrap replicates from the analyses with Garli. The gsi_T is a normalized index under which 0 represents a random arrangement of conspecifics on the tree and 1 represents monophyly. Significance of the gsi_T index was assessed through 5000 permutations. We also estimated the degree of incomplete lineage sorting among the loci by calculating the gsi_T index from 100 bootstrap replicates of Garli analyses run for each gene separately.

We used the program InStruct (Gao et al. 2007) to further examine the distinctiveness of the subspecies and investigate patterns of introgression. InStruct groups individuals into subpopulations and accounts for inbreeding by relaxing the assumption of Hardy-Weinberg equilibrium within subpopulations (Gao et al. 2007). Using both the phased haplotype and microsatellite data, we ran two independent MCMC chains of 2.0×10^6 generations with 1.0×10^6 generations as burnin under a model allowing for admixture. The number of populations (i.e., k) was fixed at two. Convergence between the chains was assessed based on the Gelman–Rubin statistic and we present the results from the chain with the lowest Deviance Information Criterion.

TMRCA AND DIVERGENCE TIME

We used the *Beast method within Beast version 1.6.1 (Drummond and Rambaut 2007) to estimate the time to most recent common ancestor (TMRCA) of each taxon and the time of divergence between the two taxa (i.e., the time after which there has not been significant gene flow; Heled and Drummond 2010). We defined two taxon sets corresponding to the two subspecies but did not enforce monophyly. Outgroups were not included in the analysis as the multispecies coalescent can be used to infer root position (Heled and Drummond 2010). To avoid biased results due to recombination (Strasburg and Rieseberg 2009), we used the largest nonrecombining segment within each nDNA locus as identified by the program *IMgc* (Woerner et al. 2007). Loci were partitioned separately and assigned the best-fitting nucleotide substitution model based on the results from ModelTest (Posada and Crandall 1998) and using Akaike's Information Criterion (Akaike 1974). Loci that violated the assumption of a molecular clock, based on analyses run using Paup* (Swofford 2003), were assigned a relaxed lognormal molecular clock. Trees were unlinked, except for the nonrecombining cpDNA, such that parameter estimates are based on integrating across loci. Using a Yule model of speciation, analyses were run for 1.0×10^8 generations removing the initial 1.0×10^7 as burnin. We present the mode of time since divergence and the average across all loci of the mode of TMRCA for each taxon. Sufficient sampling of parameter space throughout the MCMC sampler was evaluated using Tracer version 1.5 (Rambaut and Drummond 2009) and replicate runs were conducted to ensure accuracy of results.

We also estimated time of divergence using the isolation-with-migration model implemented in IMa2 (Hey and Nielsen 2007). Unlike *Beast, IMa2 allows for gene flow during taxon divergence and does not assume a Yule model. We ran a full model treating each taxon as a separate group. As with the *Beast analyses, only the nDNA loci were treated as independent and the largest nonrecombining block was used. The DNA sequences were assigned the HKY nucleotide substitution model (Hasegawa et al. 1985) and the appropriate inheritance scalar (i.e., 1 for autosomal nDNA and 0.5 for cpDNA loci). Analyses with IMa2 also included the microsatellite loci, which were assigned the stepwise mutation model and an inheritance scalar of 1. We ran 20 chains under a geometric heating scheme with appropriate priors for the migration, effective population sizes, and divergence time parameters that were identified through preliminary exploratory analyses. Runs consisted of 3.0×10^5 steps as burnin followed by 3.0×10^6 steps. Marginal histograms were evaluated and compared to ensure consistent parameter estimates.

To convert parameter estimates from *Beast and IMa2 to demographic timescales (e.g., years instead of mutations) we used the following mutation rates: 1.5×10^{-8} (Koch et al. 2001) and 0.9×10^{-9} (Berry et al. 2004) substitutions per site per year for the nDNA and cpDNA loci, respectively, and 0.00024 mutations per locus per generation for microsatellites (Thuillet et al. 2002). To convert the microsatellite mutation rate from a per generation to per year timescale, we used a generation time of 1.77 years based on data from a demographic study of 20 *xantiana* populations, which included study of seed germination and dormancy (Eckhart et al. 2011). Generation time was inferred from matrix population models using the method of Cochran and Ellner (1992).

DEMOGRAPHIC HISTORY AND POPULATION STRUCTURE

For nDNA, we identified synonymous and nonsynonymous sites by aligning each haplotype to a reference coding sequence from the EST library or one identified using BLAST (Altschul et al. 1990). Using the program SITES (Hey and Wakeley 1997), we calculated the average number of segregating nucleotides per site, θ_w (Watterson 1975), the degree of linkage disequilibrium, r^2, the population recombination rate, ρ, and Tajima's D (Tajima 1989). If populations of the selfing taxon have been affected commonly by population bottlenecks, as predicted under the reproductive assurance hypothesis, then we expect elevated variance in Tajima's D relative to neutral equilibrium expectations (Wright and Gaut 2005). Because geographic sampling strategy can affect inferences about the site frequency spectrum (Ptak and Przeworski 2002; De and Durrett 2007), we calculted θ_w and Tajima's D for each population separately and at the subspecies level. We were unable to estimate r^2 and ρ for many local population samples due to small sample sizes and low levels of polymorphism. Therefore, we only present results based on estimates calculated at the subspecies level. We used DNASP version 5.0 (Librado and Rozas 2009) to estimate the degree of population structure, F_{st}, within each subspecies (Hudson et al. 1992).

We used a two-tailed t-test to test for significant differences between means of each statistic. We also tested whether the variance among estimates of Tajima's D in each subspecies was elevated relative to expectations under mutation-drift equilibrium using the program HKA (Hey and Nielsen 2004). We excluded cpDNA from these analyses because they contained very low levels of polymorphism.

For microsatellite loci, we used the program GENODIVE version 2.0b21(Meirmans and Van Tienderen 2004) to calculate the number of alleles, allele size ranges within each subspecies, observed heterozygosity, H_o, and expected heterozygosity, H_e, within populations for each locus, inbreeding coefficients, F_{is}; Weir and Cockerham 1984, within populations for each locus, and population structure among populations within subspecies, F_{st}; Weir and Cockerham 1984. The program FSTAT version 2.9.3.2 (Goudet 1995) was used to estimate rarefied allelic richness. We also conducted two-tailed t-tests to determine whether allelic richness, F_{is} and F_{st} differed significantly between the two subspecies.

Results

PHYLOGENETIC INFERENCE AND EVOLUTIONARY DISTINCTIVENESS

Phylogenetic analyses were based on 1483 sequences and 5504 bp of which 425 were variable. Because of the few hybrid individuals in our sample and recent divergence, the two subspecies did not harbor any fixed differences between them; however, *xantiana* and *parviflora* averaged 30 and 5 unique polymorphisms per locus, respectively. The average number of polymorphisms shared between the subspecies per locus was six. Subspecies were not reciprocally monophyletic on the best topology estimated under a maximum-likelihood or Bayesian method, but there were only two samples of *xantiana* and one sample of *parviflora* that caused the taxa to be polyphyletic (Fig. 2). One clade with a high posterior probability contains all *parviflora* samples and only a few apparently admixed *xantiana* individuals (see below), which suggests that *parviflora* is a distinct group and that *xantiana* paraphyletic.

The SPLITSTREE network showed two distinct groups corresponding to each subspecies, with considerable genetic distance between them, and a few individuals from sympatric populations found among heterospecifics. Similar to the phylogenetic analyses, we found that statistical support was greatest within *parviflora* where individuals from the same population often form a strongly supported group (Figs. 2 and S1).

The results from INSTRUCT illustrated a sharp genetic difference between the two taxa. At $k = 2$, individuals were clustered into groups according to subspecies designation, which contrasts with the phylogenetic results that placed several *xantiana* individuals within a clade that included all *parviflora* samples. INSTRUCT results also showed little indication of widespread admixture or hybridization. What admixture is present is found mainly within *xantiana* samples from sympatric populations (populations 5x, 6x, and 22x; Figs. 1 and 2).

The genealogical sorting index (gsi_T) suggested a high degree of genealogical exclusivity for each taxon; gsi_T values were 0.75 ($P < 0.01$) and 0.58 ($P < 0.01$) for *parviflora* and *xantiana*, respectively (Table 1). The gsi_T values for each taxon based on individual gene trees were substantially less than observed for the concatenated dataset. Values for *xantiana* were also lower than *parviflora* for nearly all loci, which is likely the result of *xantiana* being paraphyletic (i.e., the node uniting all *xantiana* samples includes all *parviflora* samples and, thus, drives the gsi_T values toward 0).

TMRCA AND DIVERGENCE TIME

The estimates of TMRCA for each of the subspecies corroborated the phylogenetic results in indicating that *xantiana* is the ancestral taxon (Fig. 3). The mode of the posterior distribution of coalescence time to the ancestor uniting all *xantiana* alleles was 170,852 years ago (95% CI: 49,877–920,373) and of all *parviflora* alleles was 95,071 years ago (95% CI: 32,112–588,610).

The time since divergence (i.e., the time at which the subspecies began to diverge and evolve along separate evolutionary trajectories) between the two subspecies from *BEAST was 9420 years (95% CI: 3169–66,889; Fig. 4). The estimate of the time since divergence from IMa2 was 66,152 years ago (95% CI:

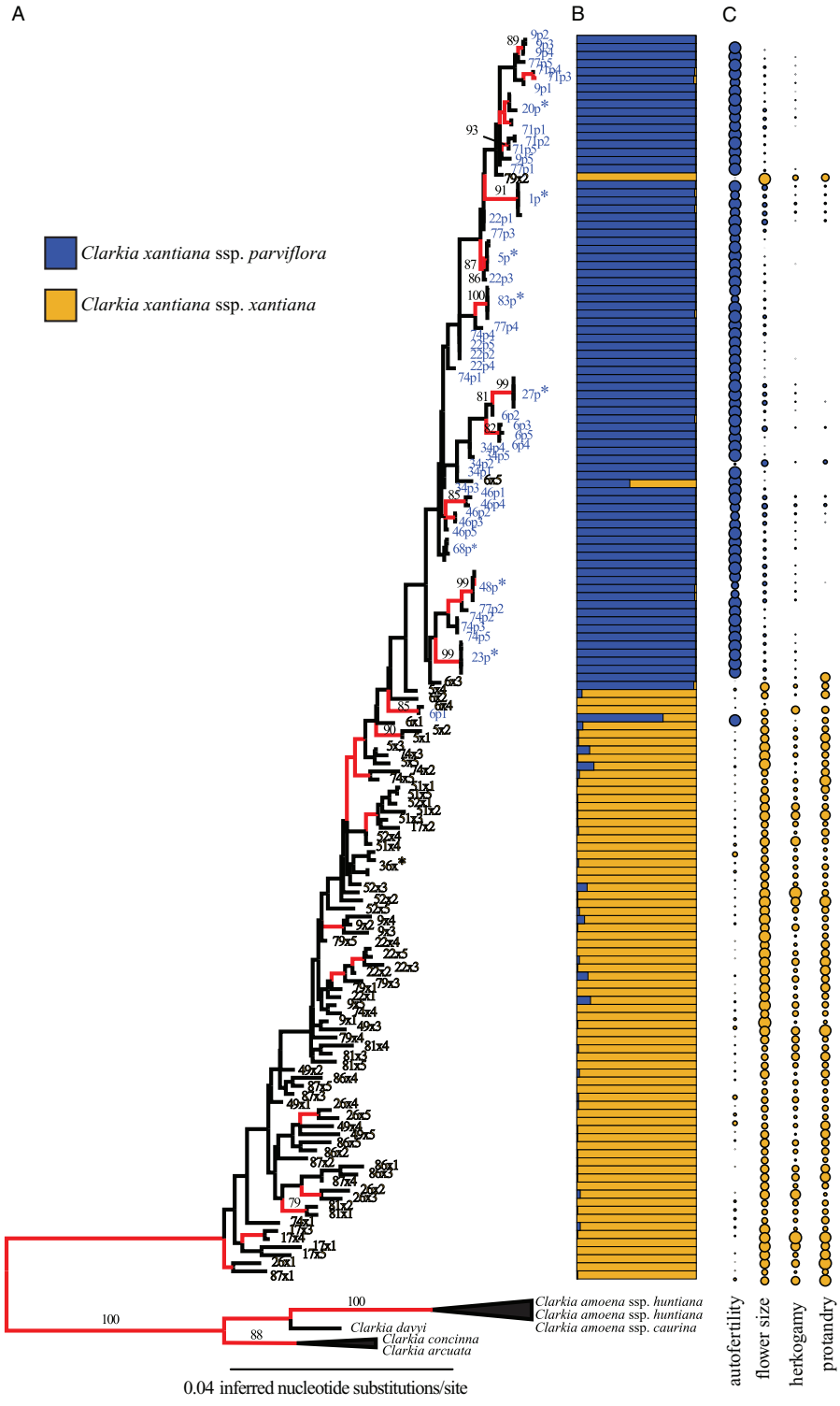


Figure 2. (A) The phylogenetic tree with the highest maximum-likelihood from the 1000 replicate analyses of the nonbootstrapped dataset with Garli. Numbers associated with branches represent the percent of 1000 bootstrap replicates supporting that clade; only bootstrap values >80 are shown. Branches in red are those with greater than 0.80 posterior probability from the Bayesian analysis. Asterisks at tips denote clades consisting of all samples from the same population. (B) InStruct results showing the fractional assignment of individuals to clusters at $k = 2$. Individuals are shown in the same order as they appear on the phylogeny and color designates the proportion of an individual's multilocus genotype that belonged to one of the two clusters. (C) Observed autofertility and standardized values of flower size, herkogamy, and protandry for each individual.

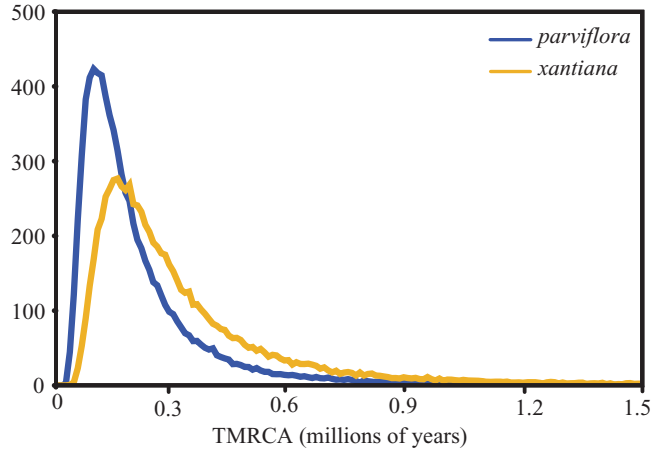


Figure 3. Density plot of the average across all loci of TMRCA for each subspecies estimated with *BEAST.

33,035–151,448; Fig. 4). Estimates of N_e and population migration rates, $2N_em$, from IMa2 can be found in Figure S2.

DEMOGRAPHIC HISTORY AND POLYMORPHISM

Estimates of population genetic parameters based on all sites and a dataset with only noncoding sites were strongly correlated (θ_w, $r = 0.85$; Tajima's D, $r = 0.95$) and produced similar results. Therefore, we only present results based on analyses of all sites. Nucleotide polymorphism within populations (θ_w) was approximately fivefold greater in *xantiana* than *parviflora* ($t = -11.80$, df $= 237$, $P < 0.001$; Fig. 5; Table S2). A similar pattern was present when estimates of θ_w were calculated from taxon-wide samples, where individuals within a subspecies were pooled (Fig. 5; Table S3). Coalescent-based estimates of N_e from IMa2 were similarly fourfold greater in *xantiana* than *parviflora* (Figure S2). Genetic variation was strongly partitioned among populations in *parviflora*, with significantly greater differentiation among populations of *parviflora* ($\overline{F}_{st} = 0.64$) than *xantiana* ($\overline{F}_{st} = 0.3$; $t = 5.28$, df $= 16$, $P < 0.01$; Fig. 5).

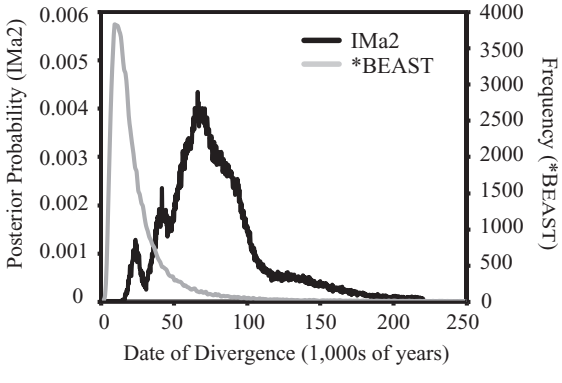


Figure 4. Distributions of the divergence time between the two subspecies based on two different methods, IMa2 and *BEAST

Tajima's D did not differ significantly between the two subspecies regardless of whether it was calculated on taxon-wide or population-specific samples ($t = 1.50$, df $= 14$, $P = 0.156$ and $t = -0.99$, df $= 163$, $P = 0.323$, respectively; Fig. 5; Tables S2 and S3). The mean estimate of D from population-specific samples was close to zero for both subspecies; however, pooling individuals within subspecies caused the mean estimate of D to become negative (Fig. 5). The variance among population-specific estimates of D was significantly greater than equilibrium expectations for *parviflora* ($\hat{\sigma}^2_{obs} = 2.22$, $\hat{\sigma}^2_{simulated} = 0.83$; $P < 0.05$) but not *xantiana* ($\hat{\sigma}^2_{obs} = 0.26$, $\hat{\sigma}^2_{simulated} = 0.84$) suggesting that population bottlenecks have affected *parviflora* populations.

We also observed strong phylogenetic structure among *parviflora* populations, which also suggests that population bottlenecks have occurred. Individuals from eight of the 16 *parviflora* populations clustered together in the same clade (1p, 5p, 20p, 23p, 27p, 48p, 68p, and 83p), often with appreciable bootstrap support and high posterior probabilities (Fig. 2). By contrast, there was only one clade that included all individuals from the same population of *xantiana* (36x).

Linkage disequilibrium (r^2) was significantly greater in *parviflora* ($\bar{r}^2 = 0.14$) than *xantiana* ($\bar{r}^2 = 0.07$; $t = 2.51$, df $= 14$, $P = 0.03$; Fig. 5). A similar difference was found using the population recombination rate, ρ, which was correlated with r^2 ($r = 0.75$).

Patterns of microsatellite variation were similar to those from sequence data. Allelic richness and observed heterozygosity were greater in *xantiana* (71 alleles; $H_o = 0.40$) than *parviflora* (37 alleles; $H_o = 0.06$; Table 2); however, rarefied allelic richness was not significantly different ($t = -1.3583$, df $= 6$, $P = 0.22$). As expected, we also found significantly higher inbreeding coefficients in *parviflora* ($\overline{F}_{is} = 0.80$) than *xantiana* populations ($\overline{F}_{is} = 0.38$; $t = 5.17$, df $= 81$, $P < 0.01$; Table 2) and greater population structure in *parviflora* ($\overline{F}_{st} = 0.24$) than *xantiana* ($\overline{F}_{itst} = 0.15$; Table 2); however, this difference was not significant ($t = 0.80$, df $= 6$, $P = 0.45$).

Discussion

Our results provide evidence that the predominantly outcrossing *xantiana* is the ancestral taxon from which the predominantly selfing *parviflora* evolved. Although there are multiple lines of evidence that support the distinctiveness of the two taxa (i.e., phylogenetic, genotypic clustering, and morphological), we also found that divergence began recently and that there has been an insufficient amount of time for the sorting of ancestral polymorphism. Given that the two taxa have not completed the speciation process (i.e., acquired complete reproductive isolation and genealogical exclusivity), inferences about the mode of mating system

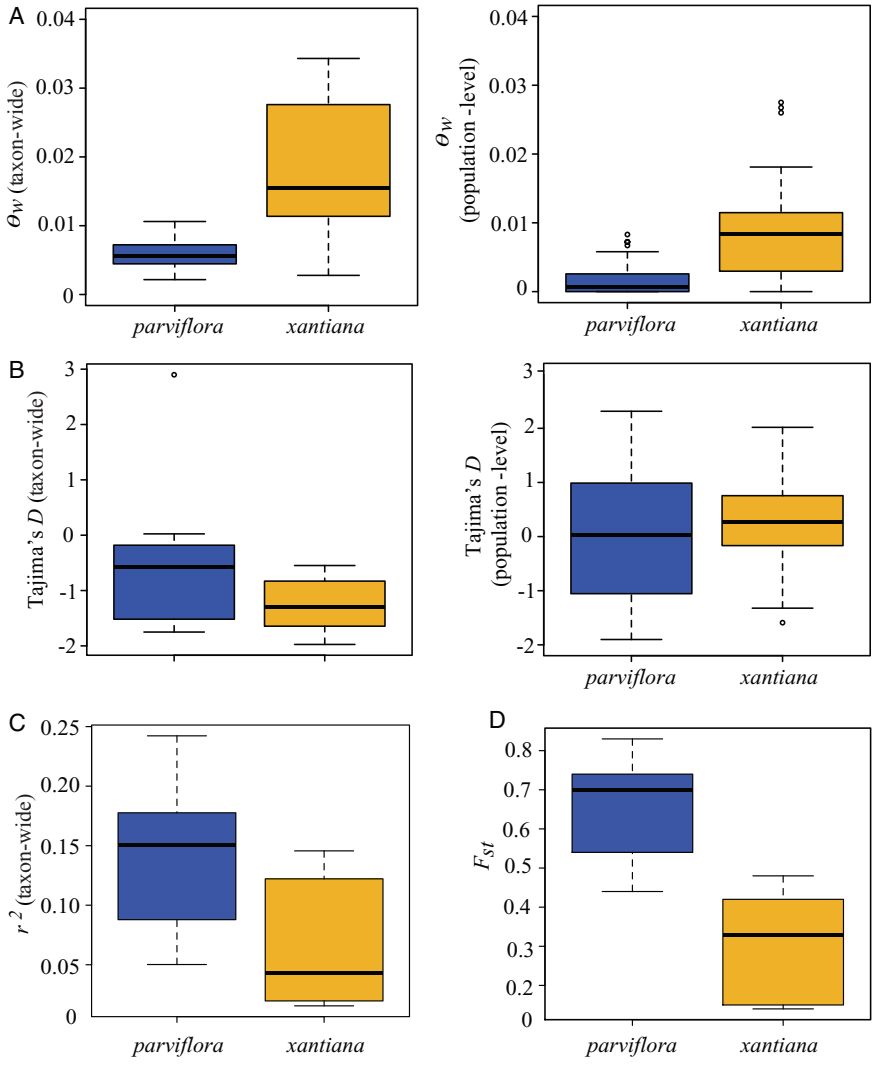


Figure 5. Comparison of the two subspecies in (A) θ_w, (B) Tajima's *D*, (C) linkage disequilibrium (r^2), and (D) population structure (F_{st}). Boxes show the interquartile range, bars illustrate the median, and the whiskers extend out to 1.5 times the interquartile range.

evolution are not likely confounded by events that occur postspeciation. Molecular population genetic analyses indicated substantially reduced molecular diversity and strong genetic bottlenecks in *parviflora*, which are consistent with theoretical predictions for the reproductive assurance model. This is the first system that we are aware of where evidence for reproductive assurance connects plant–pollinator interactions in the field to patterns of natural selection on phenotypes to patterns of sequence variation.

EVOLUTIONARY HISTORY AND INCIPIENT SPECIATION

Studies of mating system transitions are often framed around the assumption that selfing is derived from outcrossing. In some cases, phylogenetic studies have shown that selfing has arisen multiple times from outcrossing progenitors (Schoen et al. 1997; Goodwillie 1999) but in many other cases, it remains unclear as to the polarity of mating system shifts. In this study, we found support that *xantiana* is the ancestral progenitor from which *parviflora* arose based on two different types of analyses. Specifically, an outgroup rooted phylogeny placed *xantiana* basal to *parviflora*. In addition, coalescent analyses showed that the TMRCA was approximately 100,000 years earlier for *xantiana* than *parviflora*.

The two methods we used to estimate time of divergence suggested that *xantiana* and *parviflora* began to diverge approximately 66,152 (IMa2; 95% CI: 33,035–151,448) or 9420 (*BEAST; 95% CI: 3169–66,889) years ago during the upper Pleistocene. The discrepancy between the two methods can be explained by considering the different assumptions they make about the divergence process. The estimate of divergence time from *BEAST represents the point at which significant gene flow ceases to occur between taxa (Heled and Drummond 2010). The estimate of divergence time from IMa2 explicitly allows for migration during the divergence process. Given that the INSTRUCT analyses suggest that there has been recent but limited hybridization between the two

Table 2. Summary statistics for four microsatellite loci: percent successful amplification, allelic ranges, number of alleles per locus (A), allelic richness (A_R), observed heterozygosity (H_o), expected heterozygosity (H_e), inbreeding coefficients (F_{IS}), and population structure (F_{ST}).

		Allelic range (bp)		A	A_R		H_o		H_e		F_{IS}		F_{ST}	
Locus	% Amplification	*xantiana*	*parviflora*		*xantiana*	*parviflora*	*xantiana*	*parviflora*	*xantiana*	*parviflora*	*xantiana*	*parviflora*	*xantiana*	*parviflora*
cx3	0.99	244–268	249–255	13	12.9	3	0.44	0	0.68	0.12	0.32	1	0.16	0.1
cx7	0.99	113–131	107–125	10	8.9	4.8	0.33	0.04	0.44	0.14	0.2	0.74	0.17	0
cx9	0.97	154–220	194–203	30	29	6	0.32	0	0.83	0.33	0.61	1	0.12	0.34
cx11	0.96	88–152	122–164	37	20	22.6	0.5	0.21	0.75	0.4	0.32	0.48	0.13	0.52
Overall	0.98	n/a	n/a	90	17.7	9.1	0.4	0.06	0.67	0.25	0.36	0.81	0.15	0.24

taxa, our samples likely violate the assumption of no recent gene flow in *BEAST. Recent introgression following secondary contact also violates the assumption of constant gene flow throughout the divergence process in IMa2, which may result in inflated estimates of time of divergence (Becquet and Przeworski 2009). As a result of these violations, the two methods will produce estimates of time since divergence that are biased in opposite directions (i.e., *BEAST will produce shallower estimates and IMa2 will produce deeper estimates).

Our results also suggest that the limited introgression between subspecies is asymmetric, primarily from *parviflora* to *xantiana*. Specifically, the results from INSTRUCT (Fig. 2) show multiple instances where individuals with *xantiana*-like phenotypes have admixed genomes, but we did not observe this for individuals with *parviflora*-like phenotypes. We also found that asymmetric migration rates from IMa2 were greater in the direction of *parviflora* to *xantiana* than the opposite direction (Fig. S2). Asymmetric introgression is also supported by how samples from sympatric sites (i.e., areas where introgression is most likely to occur) are grouped in the phylogenetic analyses. *Xantiana* samples from some sympatric sites (5x and 6x) are embedded within the strongly supported clade that includes all *parviflora* samples (Fig. 2). By contrast, *parviflora* samples from the same sympatric populations (5p and four individuals from 6p) comprise strongly supported clades that are nested among other *parviflora* individuals (Fig. 2). The same pattern of asymmetric introgression has also been observed between sister selfing and outcrossing *Mimulus* species (Sweigart and Willis 2003). One possibility is that hybrids are formed when outcrossers sire seeds on selfers and the resulting F1 hybrids are more likely to backcross to the outcrossing parent. In *C. xantiana*, we have observed that pollinators prefer and frequently visit both parental *xantiana* and F1 hybrids, but discriminate against the smaller flowered *parviflora* (D.A. Moeller, unpubl. data). This pattern of pollinator behavior increases the likelihood for backcrossing to *xantiana* rather than *parviflora*.

SPECIES DELIMITATION AND INCOMPLETE LINEAGE SORTING

The TMRCA predated the time of divergence between *xantiana* and *parviflora* by approximately 100,000 years. Although TMRCA is expected to predate the time of divergence, this very broad difference further emphasizes how recently they began to diverge. Such shallow species tree divergence also makes it particularly difficult to evaluate the evolutionary distinctiveness of the two taxa under the criterion of genealogical exclusivity (Baum and Shaw 1995). The substantial amount of DNA polymorphism, principally within *xantiana*, suggests that the lack of reciprocal monophyly is due to retention of shared ancestral polymorphism rather than a lack of nucleotide variation. Based on

simulation studies, the time necessary to achieve genealogical exclusivity for a majority of the genome is 4–7 N_e (Hudson and Coyne 2002). Because of the high levels of polymorphism in *xantiana*, the time necessary for *xantiana* and *parviflora* to achieve genealogical exclusivity is on the order of hundreds of thousands of years.

Despite the expected time necessary for *xantiana* and *parviflora* to each achieve complete genealogical exclusivity, there are multiple lines of evidence to suggest that the two taxa are incipient species and represent independent evolutionary lineages. For example, gsi_T values for each taxon based on a concatenated matrix are significant and close to one. Additionally, the lack of reciprocal monophyly is not caused by individuals from throughout each taxon's range but rather is restricted to only a few samples from sympatric populations. Overall, the distinctiveness of the two taxa is supported by a combination of morphological differences (Figs. 2 and S3) and the formation of distinct genotypic clusters with no evidence for widespread admixture (Fig. 2; Mallett 1995). The general paraphyletic relationship of *xantiana* with respect to a nearly monophyletic *parviflora* is also expected of taxa at an intermediate state on the continuum of divergence (i.e., polyphyly to paraphyly to monophlyly; Rosenberg 2003).

TEMPO OF MATING SYSTEM EVOLUTION

Our estimate of the time since divergence suggests that the rate of evolution to a selfing phenotype has been quite rapid. The phylogenetic placement of *xantiana* as ancestral to *parviflora* suggests that the selfing phenotype of *parviflora* likely evolved from a phenotype characterized by larger flowers and higher levels of protandry and herkogamy (Fig. 2; see also ancestral character state reconstructions in Fig. S4). Consequently, the distinct morphological differences between the two subspecies (Fig. 2) are primarily the result of phenotypic changes over the past 65,000 years along the lineage leading to *parviflora*. Such a rapid rate of phenotypic evolution contrasts with the divergence of the selfing *Arabidopsis thaliana* from the outcrossing *A. lyrata*, which may have begun one million years ago (Tang et al. 2007). Other instances of rapid evolution of selfing (i.e., within the last 100,000 years) include the divergence of *Capsella rubella* from its self-incompatible progenitor, *C. grandiflora* (Foxe et al. 2009; Guo et al. 2009), and of selfing populations of *Leavenworthia alabamica* from self-incompatible populations (Busch et al. 2011). Additional studies estimating the temporal dynamics of divergence between closely related taxa with different reproductive strategies will help to clarify the rate at which such phenotypic changes occur.

DEMOGRAPHIC HISTORY AND EVIDENCE FOR REPRODUCTIVE ASSURANCE

The transition from outcrossing to selfing, alone, has dramatic effects on the genome, including reductions of up to 50% of genetic diversity (reviewed in Charlesworth and Wright 2001; Glemin et al. 2006). For transitions from partial selfers, such as *xantiana*, to more predominant selfers, such as *parviflora*, the reductions in genetic variation are expected to be less than 50%. Theoretical models have predicted that the reproductive assurance and automatic transmission modes of mating system evolution leave different signatures at the molecular level (Schoen et al. 1996). Under the former, the reduction in genetic diversity should be much greater than 50% owing to colonization bottlenecks (i.e., founder effects), whereas under the latter, extreme losses of diversity are not expected. We found that nucleotide diversity within *parviflora* was one fifth of that in *xantiana* and that the variance of Tajima's *D* in *parviflora* was significantly elevated relative to neutral equilibrium expectations. Both results suggest strong bottleneck events following the divergence of *parviflora* from *xantiana* (Charlesworth et al. 2003) and provide support for the reproductive assurance hypothesis.

Founder effects and population bottlenecks can also explain the greater degree of population structure (F_{st}) within *parviflora* relative to *xantiana* (Tables 2 and 3; Fig. 5). This result is consistent with previous studies showing greater population structure in selfers than outcrossers (Hamrick and Godt 1996; Nybom 2004). Population structure may also explain the discrepancy between estimates of Tajima's *D* based on taxon-wide versus population-specific samples. Taxon-wide samples resulted in somewhat negative values of *D* for both taxa (an excess of rare polymorphisms) whereas population-specific samples resulted in mean values of *D* near zero. This effect of geographic sampling strategy has been commonly found in recent molecular population genetic studies (Pool and Aquadro 2006; Arunyawat et al. 2007; Moeller et al. 2007) and simulations suggest that pooling across local populations can skew the site frequency spectrum (Ptak and Przeworski 2002; De and Durrett 2007) particularly in nonequilibrium populations (Städler et al. 2009).

FIELD EXPERIMENTS AND EVIDENCE FOR REPRODUCTIVE ASSURANCE

The reproductive assurance model predicts that selfing in plants evolves under chronic pollen limitation of reproduction, which can arise when pollinators or mates are unreliable (Lloyd 1992; Schoen et al. 1996; Morgan and Wilson 2005). However, processes that are independent of mating system evolution (e.g., genetic bottlenecks associated with the speciation process) could also result in the genetic signature similar to that predicted for the reproductive assurance model. For example, automatic selection could have driven the evolution of selfing, with population bottlenecks occurring subsequently for reasons unrelated to the mating system transition. Field studies, which provide direct insight into the adaptive significance of selfing versus outcrossing, serve as an independent functional assessment of mating system variation

and can be used to validate the results of molecular population genetic studies.

In *C. xantiana*, the results of field experiments provide additional support that the reproductive assurance model best explains the shift in mating system between *xantiana* and *parviflora*. Outcross pollination in *xantiana* is effected by solitary bees, particularly those specialized on the genus *Clarkia* (Moeller 2005; Eckhart et al. 2006), whereas in *parviflora* populations, pollinator visits are uncommon and specialist pollinators have not been found (Fausto et al. 2001; Moeller 2006). Moreover, patterns of population differentiation in mating system mirror geographic patterns of pollinator abundance and community composition (Moeller 2006). Reciprocal transplant experiments between the geographic ranges of *xantiana* and *parviflora* have also shown that *xantiana* suffers strong pollen limitation in *parviflora*'s range but neither subspecies was strongly pollen limited in *xantiana*'s range (Geber and Eckhart 2005). Emasculation experiments have verified that selfing elevates female fertility in field populations of *parviflora*, suggesting that selfing provides reproductive assurance (Moeller 2006). Finally, manipulative field experiments have shown that under pollinator and mate limitation, natural selection favors the selfing phenotype (reduced herkogamy and protandry), whereas selection maintains the outcrossing phenotype when pollinators and mates are reliable (Moeller and Geber 2005). In sum, patterns of DNA sequence variation are consistent with field studies in suggesting that selfing arose as a mechanism of reproductive assurance in environments where outcrossing is unlikely.

ACKNOWLEDGMENTS

We are indebted to M. Geber and V. Eckhart for sharing their extensive knowledge of this system, P. Tiffin for assisting with the initiation of this project, and B. Morris for estimating generation time from demographic data. We also thank E. Beckman, R. Bier, R. Carter, J. Iverson, C. Kelly, and J. Reese for help collecting and processing the molecular data. Computational resources and assistance were provided by the Minnesota Supercomputing Institute (MSI) and A. Bazinet and M. P. Cummings. J. Kelly and two anonymous reviewers provided helpful comments that improved the manuscript. Funding for this project was provided by the National Science Foundation (DEB-1025004 to D. A. Moeller), the University of Minnesota, and the University of Georgia.

LITERATURE CITED

Akaike, H. 1974. New look at statistical model identification. IEEE T. Automat. Contr. 19:716–723.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Arunyawat, U., W. Stephan, and T. Städler. 2007. Using multilocus sequence data to assess population structure, natural selection, and linkage disequilibrium in wild tomatoes. Mol. Biol. Evol. 24:2310–2322.

Baker, H. G. 1955. Self–compatibility and establishment after long–distance dispersal. Evolution 9:349–350.

———. 1967. Support for Baker's law—as a rule. Evolution 21:853–856.

Barrett, S. C. H. 2002. The evolution of plant sexual diversity. Nat. Rev. Genet. 3:274–284.

Baum, D. A., and K. L. Shaw. 1995. Genealogical perspectives on the species problem. Pp. 289–303 *in* P. C. Hock and A. G. Stevenson, eds. Experimental and molecular approaches to plant biosystematics. Missouri Botanical Garden, St. Louis.

Bazinet, A. L., and M. P. Cummings. 2008. The Lattice Project: a grid research and production environment combining multiple grid computing models *in* M. H. W. Weber, ed. Distributed & grid computing—science made transparent for everyone. Principles, applications and supporting communities. Tectum Publishing House, Marburg.

Becquet, C., and M. Przeworski. 2009. Learning about modes of speciation by computational approaches. Evolution 63:2547–2562.

Berry, P. E., W. J. Hahn, K. J. Systma, J. C. Hall, and A. R. Mast. 2004. Phylogenetic relationships and biogeography of *Fuchsia* (Onagraceae) based on noncoding nuclear and chloroplast DNA data. Am. J. Bot. 91:601–614.

Busch, J. W. 2005. The evolution of self–compatibility in geographically peripheral populations of *Leavenworthia alabamica* (Brassicaceae). Am. J. Bot. 92:1503–1512.

Busch, J. W., S. Joly, and D. J. Schoen. 2011. Demographic signatures accompanying the evolution of selfing in *Leavenworthia alabamica*. Mol. Biol. Evol. 28:1717–1729.

Charlesworth, B., D. Charlesworth, and N. H. Barton. 2003. The effects of genetic and geographic structure on neutral variation. Annu. Rev. Ecol. Syst. 34:99–125.

Charlesworth, D., M. T. Morgan, and B. Charlesworth. 1993. Mutation accumulation in finite outbreeding and inbreeding populations. Genet. Res. 61:39–56.

Charlesworth, D., and S. I. Wright. 2001. Breeding systems and genome evolution. Curr. Opin. Genet. Dev. 11:685–690.

Cheptou, P. O. 2004. Allee effect and self-fertilization in hermaphrodites: reproductive assurance in demographically stable populations. Evolution 58:2613–2621.

Cochran, M. E., and S. Ellner. 1992. Simple methods for calculating age-based life history parameters for stage-structured populations. Ecol. Monogr. 62:345–364.

Cummings, M. P., M. C. Neel, and K. L. Shaw. 2008. A genealogical approach to quantifying lineage divergence. Evolution 62:2411–2422.

De, A., and R. Durrett. 2007. Stepping-stone spatial structure causes slow decay of linkage disequilibrium and shifts the site frequency spectrum. Genetics 176:969–981.

Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214.

Eckhart, V. M., and M. A. Geber. 1999. Character variation and geographic distribution of *Clarkia xantiana* (Onagraceae): flowers and phenology distinguish two subspecies. Madroño 46:117–125.

Eckhart, V. M., M. A. Geber, W. F. Morris, E. S. Fabio, P. Tiffin, and D. A. Moeller. 2011. The geography of demography: long-term demographic studies and species distribution models reveal a species border limited by adaptation. Am. Nat. 178:526–543.

Eckhart, V. M., N. S. Rushing, G. M. Hart, and J. D. Hansen. 2006. Frequency–dependent pollinator foraging in polymorphic *Clarkia xantiana* ssp. *xantiana* populations: implications for flower colour evolution and pollinator interactions. Oikos 112:412–421.

Eckhart, V. M., I. Singh, A. M. Louthan, A. J. Keledjian, A. Chu, D. A. Moeller, and M. A. Geber. 2010. Plant–soil water relations and species border of *Clarkia xantiana* ssp. *xantiana* (Onagraceae). Int. J. Plant Sci. 171:749–760.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nuc. Acids Res. 32:1792–1797.

Elle, E., and R. Carney. 2003. Reproductive assurance varies with flower size in *Collinsia parviflora* (Scrophulariaceae). Am. J. Bot. 90:888–896.

Escobar, J. S., A. Cenci, J. Bolognini, A. Haudry, S. Laurent, J. David, and S. Glemin. 2010. An integrative test of the dead–end hypothesis of selfing evolution in Triticeae (Poaceae). Evolution 64:2855–2872.

Fausto, J. A., V. M. Eckhart, and M. A. Geber. 2001. Reproductive assurance and the evolutionary ecology of self–pollination in *Clarkia xantiana* (Onagraceae). Am. J. Bot. 88:1794–1800.

Fisher, R. A. 1941. Average excess and average effect of an allelic substitution. Ann. Eugenics 11:53–63.

Foxe, J. P., T. Slotte, E. A. Stahl, B. Neuffer, H. Hurka, and S. I. Wright. 2009. Recent speciation associated with the evolution of selfing in *Capsella*. Proc. Nat. Acad. Sci. USA 106:5241–5245.

Foxe, J. P., M. Stift, A. Tedder, A. Haudry, S. I. Wright, and B. K. Mable. 2010. Reconstructing origins of loss of self-incompatibility and selfing in North American *Arabidopsis lyrata*: a population genetic context. Evolution 64:495–510.

Gao, H., S. Williamson, and C. D. Bustamante. 2007. A Markov Chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. Genetics 176:1635–1651.

Geber, M. A., and V. M. Eckhart. 2005. Experimental studies of adaptation in *Clarkia xantiana*. II. Fitness variation across a subspecies border. Evolution 59:521–531.

Glemin, S., E. Bazin, and D. Charlesworth. 2006. Impact of mating systems on patterns of sequence polymorphism in flowering plants. Proc. R. Soc. Lond. B. 273:3011–3019.

Goldberg, E. E., and B. Igic. 2008. On phylogenetic tests of irreversible evolution. Evolution 62:2727–2741.

Goodwillie, C. 1999. Multiple origins of self-compatibility in *Linanthus* section *Leptosiphon* (Polemoniaceae): phylogenetic evidence from internal–transcribed–spacer sequence data. Evolution 53:1387–1395.

Goodwillie, C., S. Kalisz, and C. G. Eckert. 2005. The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence. Annu. Rev. Ecol. Evol. Syst. 36:47–79.

Goudet, J. 1995. FSTAT v1.2: a computer program to calculate *F*- statistics. J. Hered. 86:485–486.

Guo, Y. L., J. S. Bechsgaard, T. Slotte, B. Neuffer, M. Lascoux, D. Weigel, and M. H. Schierup. 2009. Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-incompatibility and an extreme bottleneck. Proc. Nat. Acad. Sci. USA 106:5246–5251.

Hamrick, J. L, and M. J. W. Godt. 1996. Effects of life history traits on genetic diversity in plant species. Philos. Trans. R. Soc. Lond. B. 351:1291–1298.

Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.

Heled, J., and A. J. Drummond. 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27:570–580.

Herlihy, C. R., and C. G. Eckert. 2002. Genetic cost of reproductive assurance in a self-fertilizing plant. Nature 416:320–323.

Hey, J., and R. Nielsen. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167:747–760.

———. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proc. Nat. Acad. Sci. USA 104:2785–2790.

Hey, J., and J. Wakeley. 1997. A coalescent estimator of the population recombination rate. Genetics 145:833–846.

Holsinger, K. E. 2000. Reproductive systems and evolution in vascular plants. Proc. Nat. Acad. Sci. USA 97:7037–7042.

Hudson, R. R., and J. A. Coyne. 2002. Mathematical consequences of the genealogical species concept. Evolution 56:1557–1565.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.

Hudson, R. R., M. Slatkin and W. P. Maddison. 1992. Estimation of levels of gene flow from DNA sequence data. Genetics 132:583–589.

Huson, D. H., and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23:254–267.

Igic, B., L. Bohs, and J. R. Kohn. 2006. Ancient polymorphism reveals unidirectional breeding system transitions. Proc. Nat. Acad. Sci. USA 103:1359–1363.

Igic, B., R. Lande, and J. R. Kohn. 2008. Loss of self-incompatibility and its evolutionary consequences. Int. J. Plant Sci. 169:93–104.

Jain, S. K. 1976. Evolution of inbreeding in plants. Annu. Rev. Ecol. Syst. 7:469–495.

Kalisz, S., D. W. Vogler, and K. M. Hanley. 2004. Context-dependent autonomous self-fertilization yields reproductive assurance and mixed mating. Nature 430:884–887.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16:111–120.

Koch, M., B. Haubold, and T. Mitchell–Olds. 2001. Molecular systematics of the Brassicaceae: evidence from coding plastidic *matK* and nuclear *Chs* sequences. Am. J. Bot. 88:534–544.

Kress, J. W., and D. L. Erickson. 2007. A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH–psbA* spacer region. PLoS ONE 2:e560.

Librado, P., and Rozas, J. 2009. DNASP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452.

Lloyd, D. G. 1979. Some reproductive factors affecting the selection of self-fertilization in plants. Am. Nat. 113:67–79.

———. 1992. Self- and cross-fertilization in plants. II. The selection of self-fertilization. Int. J. Plant Sci. 153:370–380.

Mallett, J. 1995. A species definition for the Modern Synthesis. Trends Ecol. Evol. 10:294–299.

Meirmans, P. G., and P. H. Van Tienderen. 2004. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Mol. Ecol. Notes 4:792–794.

Moeller, D. A. 2005. Pollinator community structure and sources of spatial variation in plant–pollinator interactions in *Clarkia xantiana* ssp. *xantiana*. Oecologia 142:28–37.

———. 2006. Geographic structure of pollinator communities, reproductive assurance, and the evolution of self-pollination. Ecology 87:1510–1522.

Moeller, D. A., and M. A. Geber. 2005. Ecological context of the evolution of self-pollination in *Clarkia xantiana*: population size, plant communities, and reproductive assurance. Evolution 59:786–799.

Moeller, D. A., M. I. Tenaillon, and P. Tiffin. 2007. Population structure and its effects on patterns of nucleotide polymorphism in teosinte (*Zea mays* ssp. *parviglumis*). Genetics 176:1799–1809.

Moeller, D. A., M. A. Geber, and P. Tiffin. 2011. Population genetics and the evolution of geographic range limits in an annual plant. Am. Nat. 178:S44–S61.

Moeller, D. A., M. A. Geber, V. M. Eckhart, and P. Tiffin. 2012. Reduced pollinator service and elevated pollen limitation at the geographic range limit of an annual plant. Ecology: *In press*.

Morgan, M. T., and W. G. Wilson. 2005. Self-fertilization and the escape from pollen limitation in variable pollination environments. Evolution 59:1143–1148.

Ness, R. W., S. I. Wright, and S. C. H. Barrett. 2010. Mating-system variation, demographic history and patterns of nucleotide diversity in the tristylous plant *Eichhornia paniculata*. Genetics 184:381–392.

Nordborg, M. 2000. Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. Genetics 154:923–929.

Nybom, H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol. Ecol. 13:1143–1155.

Pannell, J. R., and S. C. H. Barrett. 1998. Baker's law revisited: reproductive assurance in a metapopulation. Evolution 52:657–668.

Pool, J. E., and C. F. Aquadro. 2006. History and structure of sub-Saharan populations of *Drosophila melanogaster*. Genetics 174:915–929.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818.

Ptak, S. E., and M. Przeworski. 2002. Evidence for population growth in humans is confounded by fine-scale population structure. Trends. Genet. 18:559–563.

Rambaut, A., and A. J. Drummond. 2009. TRACER. Available at http://tree.bio.ed.ac.uk/software/tracer. University of Edinburgh, Edinburgh, U.K.

Richards, A. 1999. Plant breeding systems. Chapman and Hall, London, U.K.

Ronquist, F., and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Rosenberg, N. A. 2003. The shapes of neutral gene genealogies in two species: probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. Evolution 57:1465–1477.

Runions, C. J., and M. A. Geber. 2000. Evolution of the self-pollinating flower in *Clarkia xantiana* (Onagraceae). I. Size and development of floral organs. Am. J. Bot. 67:1439–1450.

Schoen, D. J., M. O. Johnston, A. L'Heureux, and J. V. Marsolais. 1997. Evolutionary history of the mating system in *Amsinckia* (Boraginaceae). Evolution 51:1090–1099.

Schoen, D. J., M. T. Morgan, and T. Bataillon. 1996. How does self-pollination evolve? Inferences from floral ecology and molecular genetic variation. Philos. Trans. R. Soc. Lond. B. 351:1281–1290.

Shaw, J., E. B. Lickey, J. T. Beck, S. B. Farmer, W. S. Liu, J. Miller, K. C. Siripun, C. T. Winder, E. E. Schilling, and R. L. Small. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am. J. Bot. 92:142–166.

Stadler, T., B. Haubold, C. Merino, W. Stephan, and P. Pfaffelhuber. 2009. The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. Genetics 182:205–216.

Stebbins, G. L. 1957. Self-fertilization and population variability in the higher plants. Am. Nat. 91:337–354.

———. 1974. Flowering plants: evolution above the species level. Harvard Univ. Press, Cambridge, MA.

Stephens, M., and P. Donnelly. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am. J. Hum. Genet. 73:1162–1169.

Stephens, M., N. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68:978–989.

Strasburg, J. L., and L. H. Rieseberg. 2009. How robust are "Isolation with Migration" analyses to violations of the IM model? A simulation study. Mol. Biol. Evol. 27:297– 310.

Sweigart, A. L., and J. H. Willis. 2003. Patterns of nucleotide diversity are affected by mating system and asymmetric introgression in two species of *Mimulus*. Evolution 57:2490–2506.

Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0. Sinauer Associates, Sunderland.

Taberlet, P., L. Gielly, G. Pautou, and J. Bouvet. 1991. Universal primers for amplification of three noncoding regions of chloroplast DNA. Plant Mol. Biol. 17:1105–1109.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

Takebayashi, N., and P. L. Morrell. 2001. Is self-fertilization an evolutionary dead end? Revisiting an old hypothesis with genetic theories and a macroevolutionary approach. Am. J. Bot. 88:1143–1150.

Tang, C., C. Toomajian, S. Sherman-Broyles, V. Plagnol, Y. L. Guo, T. T. Hu, R. M. Clark, J. B. Nasrallah, D. Weigel, and M. Nordborg. 2007. The evolution of selfing in *Arabidopsis thaliana*. Science 317:1070–1072.

Thuillet, A. C., D. Bru, J. David, P. Roumet, S. Santoni, P. Sourdille, and T. Bataillon. 2002. Direct estimation of mutation rate for 10 microsatellite loci in durum wheat, *Triticum turgidum* (L.) Thell. ssp *durum* desf. Mol. Biol. Evol. 19:122–125.

Uyenoyama, M. K., K. E. Holsinger, and D. M. Waller. 1993. Ecological and genetic factors directing the evolution of self-fertilization. Oxford Surv. Evol. Biol. 9:327–381.

Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. 7:256–275.

Weir, B. S., and C. C. Cockerham. 1984. Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358–1370.

Wright, S. I., and B. S. Gaut. 2005. Molecular population genetics and the search for adaptive evolution in plants. Mol. Biol. Evol. 22:506–519.

Woerner, A. E., M. P. Cox, and M. F. Hammer. 2007. Recombination-filtered genomic datasets by information maximization. Bioinformatics 23:1851–1853.

Zwickl, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. diss., The University of Texas at Austin.

Associate Editor: J. Kelly

Supporting Information

The following supporting information is available for this article:

Figure S1. NeighborNet network from SPLITSTREE4 showing reticulate relationships among lineages.
Figure S2. Marginal posterior probability for (A) current *Ne* and ancestral *Ne*, and (B) asymmetric population migration rates (2*Nem*) estimated with IMa2.
Figure S3. Boxplots illustrating the observed differences in (A) herkogamy, (B) protandry, (C) flower size, and (D) autofertility *xantiana* and *parviflora*.
Figure S4. Ancestral character reconstructions based on the best topology from GARLI (Fig. 2) for (A) protandry, (B) flower size, (C) autofertility, and (D) herkogamy.
Table S1. Population codes (Site #), location, number of samples (*N*), elevation, and geographic coordinates of individuals included in this study.
Table S2. Population-level estimates of θw and Tajima's *D*.
Table S3. Taxon-level estimates of θw and Tajima's *D*.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

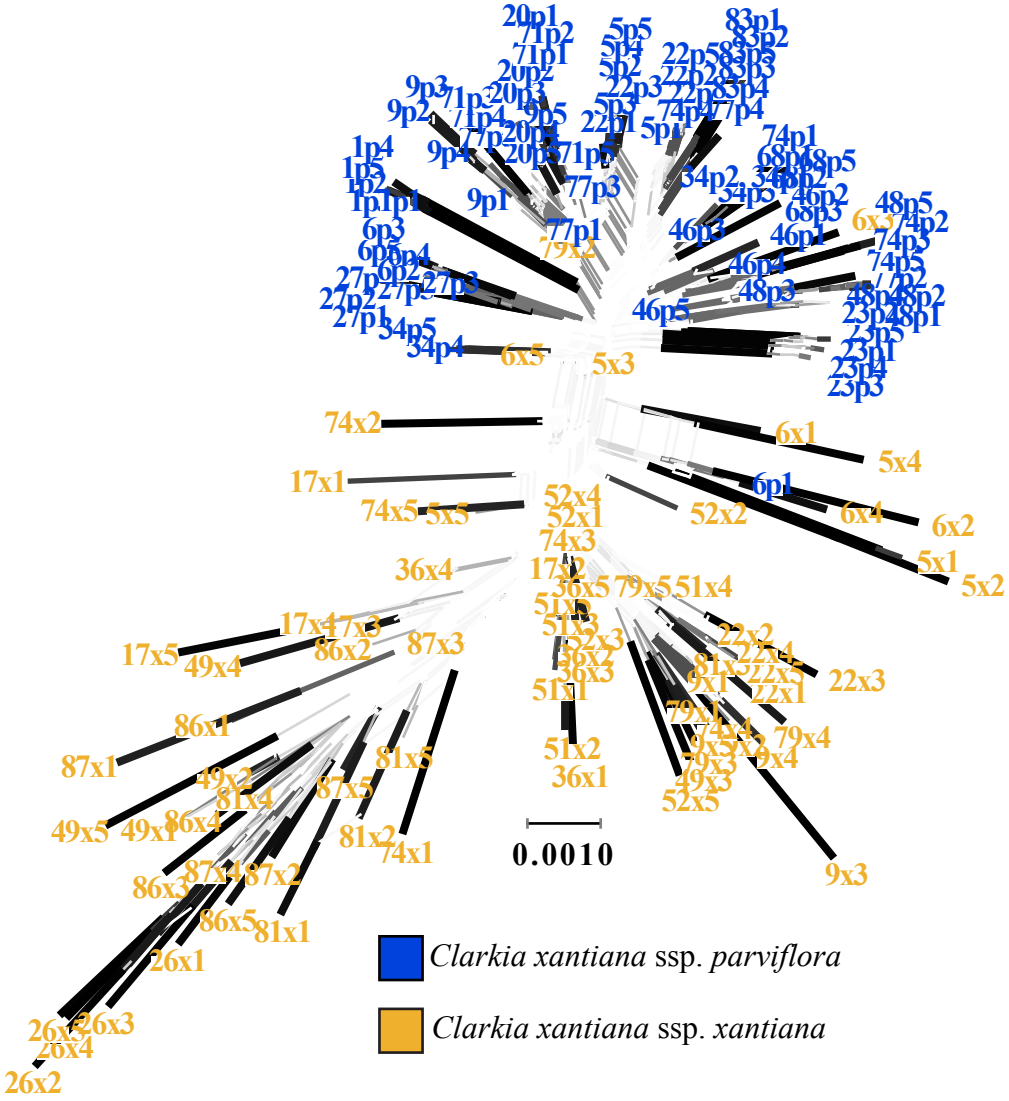
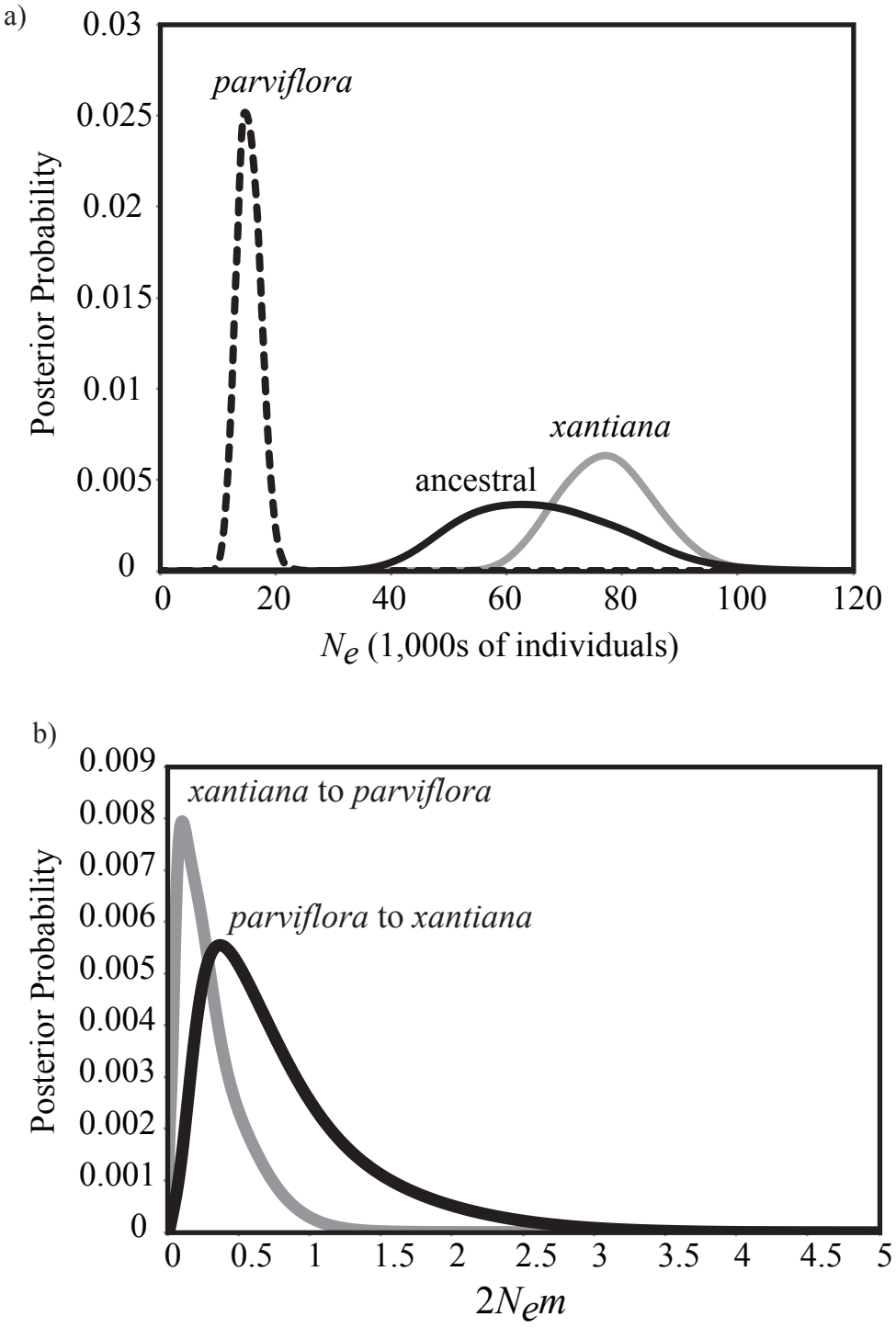


Figure S1. NeighborNet network from SPLITSTREE4 showing reticulate relationships among lineages. The widest and darkest lines represent edges that were present within > 90% of 1000 bootstrap replicates; the narrowest and lightest lines indicate < 50% bootstrap support.



Supplemental Figure S2. Marginal posterior probability for a) current N_e and ancestral N_e, and b) asymmetric population migration rates ($2N_em$) estimated with IMa2.

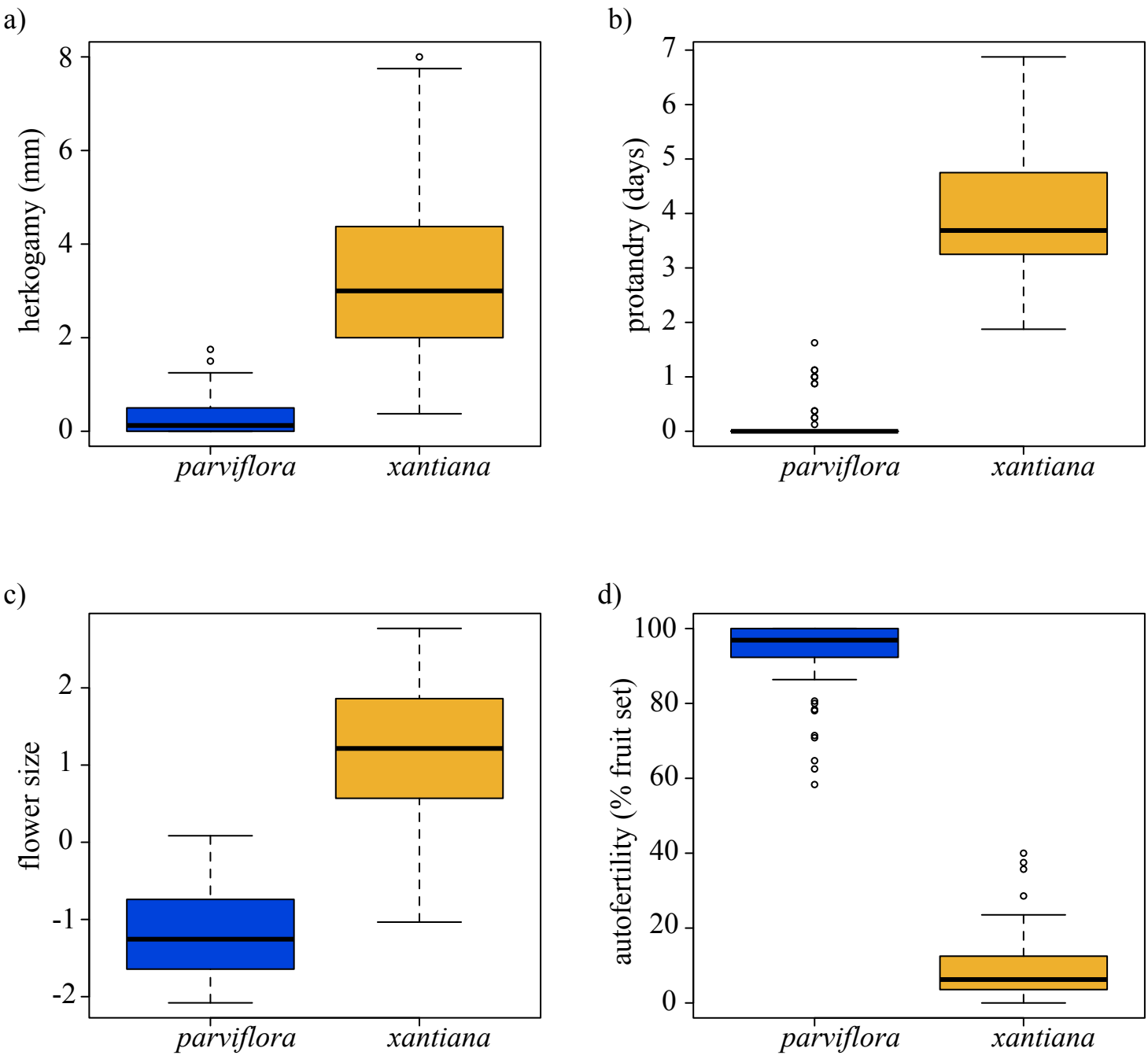


Figure S3. Boxplots illustrating the observed differences in a) herkogamy, b) protandry, c) flower size, and d) autofertility *xantiana* and *parviflora.* Flower size is based on principal component scores; see text for additional information.

Ancestral state reconstructions

Method

Ancestral state reconstructions for autofertility (percent fruit set in a pollinator-free environment) and standardized values of three floral traits: 1) herkogamy, 2) dichogamy and, 3) flower size (the first principal component of petal length and width) were conducted using the best topology from the GARLI phylogenetic analysis. Reconstructions were inferred using the maximum-likelihood method (Schluter et al. 1997) of the function ACE within the R package APE (Paradis et al. 2004). Individuals for which measurements were not available (10 samples and all outgroup samples) were pruned from the tree before running the analyses. See Figure S1 for the observed differences between taxa.

Results

Variation in autofertility and floral characters (flower size, herkogamy, and protandry) was correlated with taxonomic designation and further illustrates the distinctiveness of the taxa (Fig. 2). Low autofertility and floral characters associated with outcrossing (large flowers, high herkogamy and protandry) were the ancestral states inferred on the best phylogeny from GARLI (Fig. S2). Because there appears to be a single origin of the selfing phenotype, reconstructed character states for nodes uniting *parviflora* individuals are all indicative of a selfing phenotype.

Literature Cited

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

Schluter D., Price T. Mooers A. O. and Ludwig D. 1997. Likelihood of ancestral states in adaptive radiation. Evolution 51: 1699-1711.

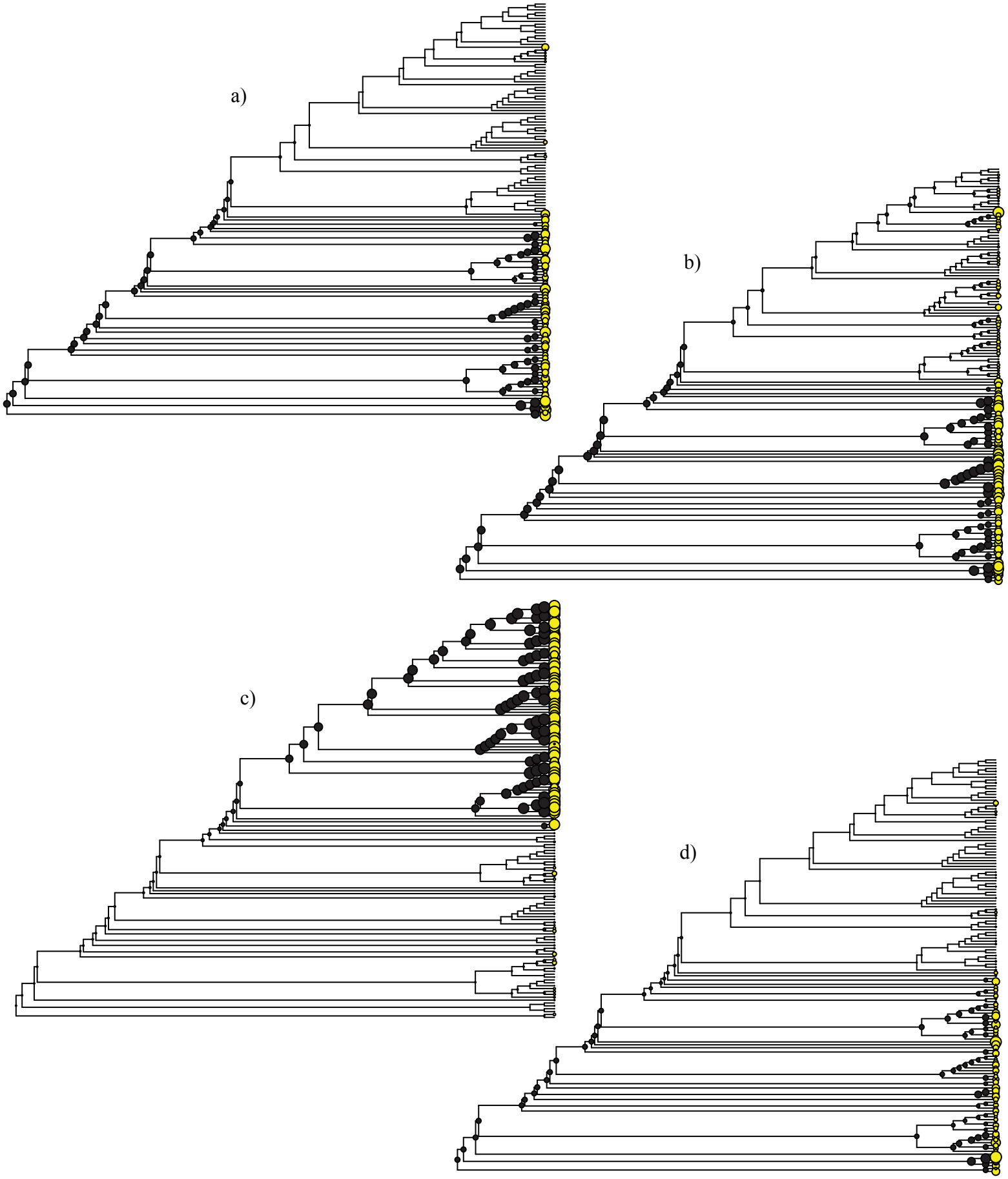


Figure S4. Ancestral character reconstructions based on the best topology from GARLI (Fig. 2) for a) protandry, b) flower size, c) autofertility and, d) herkogamy. Yellow circles represent standradized observed values and black circles are inferred ancestral states. For reconstructions, polytomies were randomly resolved and the Grafen method (Grafen 1989) of branch transformation with a power of 0.8 was used.

(Grafen, A. 1989. The phylogenetic regression. Philos. T. Roy. Soc. B. 326:119–157.)

Table S1. Population codes (Site #), location, number of samples (*N*), elevation, and geographic coordinates of individuals included in this study.

Subspecies	Site #	*N*	Elevation (m)	NAD27 northing (m)	NAD27 easting (m)
Clarkia xantiana ssp. *parviflora*	1p	5	1750	4010422	404057
	5p*	5	900	3964219	368479
	6p*	5	1150	3981714	366072
	9p*	5	800	3984492	366600
	20p	5	1400	3950976	403925
	22p*	5	950	3966924	369076
	23p	5	1200	3958344	402278
	27p	5	1190	3810885	422764
	34p	5	1520	3982402	369247
	46p	5	1600	3963973	400626
	48p	5	1450	3925682	388719
	68p	4	900	3941814	386723
	71p	5	900	3950678	375435
	74p*	5	1400	3937720	366519
	77p*	5	1000	3949048	366071
	83p	5	1200	3918062	368769
Total	16	79			
Clarkia xantiana ssp. *xantiana*	5x*	5	900	3964219	368479
	6x*	5	1150	3984492	366600
	9x*	5	800	3938627	362163
	17x	5	950	3911691	356262
	22x*	5	950	3966924	369076
	26x	5	1030	3844927	353557
	36x	5	1271	3925688	349283
	49x	5	1200	3927612	361200
	51x	5	350	3925456	338745
	52x	5	500	3927801	344267
	74x*	5	1400	3937720	366519
	79x*	5	1200	3948316	363149
	81x	5	850	3922433	343001
	86x	5	1300	3885762	381001
	87x	5	950	3898329	358115
Total	15	75			

* Sympatric and co-occurring populations

Table S2. Population-level estimates of θ_w and Tajima's *D*.

Locus	Population	Subspecies	θ_w	*D*
A23	1p	*parviflora*	0	NA
A23	20p	*parviflora*	0.00231	0.592
A23	22p	*parviflora*	0.0015	0.019
A23	23p	*parviflora*	0	NA
A23	27p	*parviflora*	0	NA
A23	34p	*parviflora*	0	NA
A23	46p	*parviflora*	0.00278	-0.447
A23	48p	*parviflora*	0	NA
A23	5p	*parviflora*	0.00082	-1.055
A23	6p	*parviflora*	0.00082	1.444
A23	71p	*parviflora*	0.00424	NA
A23	74p	*parviflora*	0.00374	0.124
A23	77p	*parviflora*	0.003	1.652
A23	83p	*parviflora*	0.003	0.143
A23	9p	*parviflora*	0.00245	0.458
A23	22x	*xantiana*	0.00082	0.334
A23	36x	*xantiana*	0.00116	1.633
A23	49x	*xantiana*	0.00186	1.032
A23	5x	*xantiana*	0.00371	0.974
A23	74x	*xantiana*	0.00847	NA
A23	79x	*xantiana*	0.00424	NA
A23	81x	*xantiana*	0.00371	0.149
A23	87x	*xantiana*	0.0104	1.874
D13	1p	*parviflora*	0	NA
D13	20p	*parviflora*	0	NA
D13	22p	*parviflora*	0	NA
D13	23p	*parviflora*	0	NA
D13	27p	*parviflora*	0	NA
D13	34p	*parviflora*	0.00106	-1.112
D13	46p	*parviflora*	0	NA
D13	48p	*parviflora*	0	NA
D13	5p	*parviflora*	0	NA
D13	68p	*parviflora*	0	NA
D13	6p	*parviflora*	0.00231	0.414
D13	71p	*parviflora*	0	NA
D13	74p	*parviflora*	0	NA
D13	77p	*parviflora*	0	NA
D13	83p	*parviflora*	0	NA
D13	9p	*parviflora*	0	NA
D13	17x	*xantiana*	0.01049	1.284
D13	22x	*xantiana*	0.00393	0.338
D13	26x	*xantiana*	0.01143	1.323

Locus	Population	Subspecies	θ_w	D
D13	36x	*xantiana*	0.00299	NA
D13	49x	*xantiana*	0.0118	0.088
D13	51x	*xantiana*	0.00163	-0.612
D13	52x	*xantiana*	0	NA
D13	5x	*xantiana*	0.00231	0.242
D13	6x	*xantiana*	0	NA
D13	74x	*xantiana*	0.00163	-0.612
D13	79x	*xantiana*	0	NA
D13	81x	*xantiana*	0.0118	1.196
D13	87x	*xantiana*	0.00262	-1.132
D13	9x	*xantiana*	0	NA
D5	1p	*parviflora*	0.00451	-1.136
D5	20p	*parviflora*	0.00098	-1.055
D5	22p	*parviflora*	0.0018	0.019
D5	23p	*parviflora*	0	NA
D5	27p	*parviflora*	0.00451	1.889
D5	34p	*parviflora*	0.0018	1.033
D5	46p	*parviflora*	0.00721	-0.822
D5	48p	*parviflora*	0	NA
D5	5p	*parviflora*	0	NA
D5	68p	*parviflora*	0	NA
D5	6p	*parviflora*	0.00361	-1.667
D5	71p	*parviflora*	0.00451	2.091
D5	74p	*parviflora*	0.00361	0.988
D5	77p	*parviflora*	0.0018	1.844
D5	83p	*parviflora*	0.00271	-1.562
D5	9p	*parviflora*	0.00271	-0.507
D5	17x	*xantiana*	0.01574	-0.559
D5	22x	*xantiana*	0.00984	-0.55
D5	26x	*xantiana*	0.00902	0.75
D5	36x	*xantiana*	0.00902	-0.894
D5	49x	*xantiana*	0.00541	0.761
D5	51x	*xantiana*	0.00992	0.282
D5	52x	*xantiana*	0.01082	-0.075
D5	5x	*xantiana*	0.00894	0.611
D5	6x	*xantiana*	0.00974	-0.389
D5	74x	*xantiana*	0.00451	-0.329
D5	79x	*xantiana*	0.01082	-1.322
D5	81x	*xantiana*	0.00721	0.025
D5	86x	*xantiana*	0.00902	0.667
D5	87x	*xantiana*	0.00984	-0.137
D5	9x	*xantiana*	0.00295	0.839
F9	1p	*parviflora*	0	NA
F9	20p	*parviflora*	0.00134	1.641

Locus	Population	Subspecies	θ_w	D
F9	22p	*parviflora*	0	NA
F9	23p	*parviflora*	0.00067	0.82
F9	27p	*parviflora*	0	NA
F9	34p	*parviflora*	0	NA
F9	46p	*parviflora*	0.00134	0.019
F9	48p	*parviflora*	0.00067	1.303
F9	5p	*parviflora*	0.00067	0.015
F9	68p	*parviflora*	0	NA
F9	6p	*parviflora*	0.00671	0.026
F9	71p	*parviflora*	0	NA
F9	74p	*parviflora*	0.00067	-1.112
F9	77p	*parviflora*	0.00067	-1.112
F9	83p	*parviflora*	0	NA
F9	9p	*parviflora*	0.00067	-1.112
F9	17x	*xantiana*	0.00667	0.862
F9	22x	*xantiana*	0.00664	1.029
F9	26x	*xantiana*	0.00396	-0.019
F9	36x	*xantiana*	0.00659	-1.59
F9	49x	*xantiana*	0.01439	-0.699
F9	51x	*xantiana*	0.00791	0.026
F9	52x	*xantiana*	0.01187	-0.315
F9	5x	*xantiana*	0.00536	-0.449
F9	6x	*xantiana*	0.00604	1.188
F9	74x	*xantiana*	0.00669	-0.977
F9	79x	*xantiana*	0.01131	-0.403
F9	81x	*xantiana*	0.00923	0.274
F9	86x	*xantiana*	0.01059	-0.482
F9	87x	*xantiana*	0.00989	0.278
F9	9x	*xantiana*	0.0095	-0.194
g3pdh	1p	*parviflora*	0.0029	-0.822
g3pdh	20p	*parviflora*	0.00147	-0.691
g3pdh	22p	*parviflora*	0	NA
g3pdh	23p	*parviflora*	0.00072	0.015
g3pdh	27p	*parviflora*	0	NA
g3pdh	34p	*parviflora*	0.0058	2.195
g3pdh	46p	*parviflora*	0.0029	2.134
g3pdh	48p	*parviflora*	0	NA
g3pdh	5p	*parviflora*	0	NA
g3pdh	68p	*parviflora*	0	NA
g3pdh	6p	*parviflora*	0.0058	0.025
g3pdh	71p	*parviflora*	0	NA
g3pdh	74p	*parviflora*	0.0029	0.022
g3pdh	77p	*parviflora*	0.00079	0.334
g3pdh	83p	*parviflora*	0.00145	0.019

Locus	Population	Subspecies	θ_w	D
g3pdh	9p	*parviflora*	0.00072	0.015
g3pdh	17x	*xantiana*	0	NA
g3pdh	22x	*xantiana*	0.0087	1.612
g3pdh	26x	*xantiana*	0.01202	-0.042
g3pdh	36x	*xantiana*	0.01343	1.108
g3pdh	49x	*xantiana*	0.0275	0.99
g3pdh	51x	*xantiana*	0.00715	0.611
g3pdh	52x	*xantiana*	0.01441	0.876
g3pdh	5x	*xantiana*	0.0126	0.002
g3pdh	6x	*xantiana*	0.00712	0.545
g3pdh	74x	*xantiana*	0.026	1.039
g3pdh	79x	*xantiana*	0.01524	-1.031
g3pdh	81x	*xantiana*	0.02674	0.705
g3pdh	86x	*xantiana*	0.01425	0.682
g3pdh	87x	*xantiana*	0.01153	1.046
g3pdh	9x	*xantiana*	0.00726	0.054
i11	1p	*parviflora*	0	NA
i11	20p	*parviflora*	0.00083	-1.112
i11	22p	*parviflora*	0.00083	-1.112
i11	23p	*parviflora*	0	NA
i11	27p	*parviflora*	0	NA
i11	34p	*parviflora*	0	NA
i11	46p	*parviflora*	0.00083	1.464
i11	48p	*parviflora*	0	NA
i11	5p	*parviflora*	0.00417	-0.582
i11	68p	*parviflora*	0.00364	-0.222
i11	6p	*parviflora*	0.00333	-1.245
i11	71p	*parviflora*	0.00083	1.303
i11	74p	*parviflora*	0.00417	-1.741
i11	77p	*parviflora*	0.00083	0.82
i11	83p	*parviflora*	0.00167	-1.401
i11	9p	*parviflora*	0.00333	-0.762
i11	17x	*xantiana*	0.01	-0.139
i11	22x	*xantiana*	0.00333	0.385
i11	26x	*xantiana*	0.00917	0.743
i11	36x	*xantiana*	0.00667	0.636
i11	49x	*xantiana*	0.0075	0.148
i11	51x	*xantiana*	0.00667	0.025
i11	52x	*xantiana*	0.00834	-0.448
i11	5x	*xantiana*	0.005	1.975
i11	6x	*xantiana*	0.00667	0.263
i11	74x	*xantiana*	0.0075	-0.494
i11	79x	*xantiana*	0.01	-0.21
i11	81x	*xantiana*	0.005	1.325

Locus	Population	Subspecies	θ_w	D
i11	86x	*xantiana*	0.00917	0.947
i11	87x	*xantiana*	0.01	0.192
i11	9x	*xantiana*	0.01167	-0.055
ipi2	1p	*parviflora*	0.00077	0.019
ipi2	20p	*parviflora*	0.0031	2.195
ipi2	22p	*parviflora*	0	NA
ipi2	23p	*parviflora*	0	NA
ipi2	27p	*parviflora*	0	NA
ipi2	34p	*parviflora*	0.00039	1.303
ipi2	46p	*parviflora*	0.00042	0.334
ipi2	48p	*parviflora*	0	NA
ipi2	5p	*parviflora*	0	NA
ipi2	68p	*parviflora*	0	NA
ipi2	6p	*parviflora*	0.00194	0.023
ipi2	71p	*parviflora*	0.00039	0.015
ipi2	74p	*parviflora*	0.00271	0.025
ipi2	77p	*parviflora*	0	NA
ipi2	83p	*parviflora*	0	NA
ipi2	9p	*parviflora*	0.00048	0.851
ipi2	17x	*xantiana*	0	NA
ipi2	22x	*xantiana*	0.00039	0.015
ipi2	26x	*xantiana*	0	NA
ipi2	36x	*xantiana*	0	NA
ipi2	49x	*xantiana*	0	NA
ipi2	51x	*xantiana*	0	NA
ipi2	52x	*xantiana*	0	NA
ipi2	5x	*xantiana*	0.00042	0.334
ipi2	6x	*xantiana*	0.00116	0.021
ipi2	74x	*xantiana*	0.00116	0.7
ipi2	79x	*xantiana*	0.00253	0.518
ipi2	81x	*xantiana*	0	NA
ipi2	86x	*xantiana*	0	NA
ipi2	87x	*xantiana*	0	NA
ipi2	9x	*xantiana*	0.00232	0.371
K22	1p	*parviflora*	0	NA
K22	20p	*parviflora*	0	NA
K22	22p	*parviflora*	0.00721	2.315
K22	23p	*parviflora*	0	NA
K22	27p	*parviflora*	0	NA
K22	34p	*parviflora*	0.00443	0.025
K22	46p	*parviflora*	0.00724	2.315
K22	48p	*parviflora*	0	NA
K22	5p	*parviflora*	0.00056	-1.112
K22	68p	*parviflora*	0	NA

Locus	Population	Subspecies	θ_w	D
K22	6p	*parviflora*	0.0083	0.143
K22	71p	*parviflora*	0.00502	-1.901
K22	74p	*parviflora*	0.00055	-1.112
K22	77p	*parviflora*	0.00722	2.315
K22	83p	*parviflora*	0	NA
K22	9p	*parviflora*	0.00059	-1.112
K22	17x	*xantiana*	0.01111	-0.549
K22	22x	*xantiana*	0.01051	-0.763
K22	26x	*xantiana*	0.00841	0.587
K22	36x	*xantiana*	0.01275	-0.102
K22	49x	*xantiana*	0.01811	-0.284
K22	51x	*xantiana*	0.01326	1.334
K22	52x	*xantiana*	0.0162	-0.741
K22	5x	*xantiana*	0.01279	0.74
K22	6x	*xantiana*	0.01333	0.625
K22	74x	*xantiana*	0.01445	0.109
K22	79x	*xantiana*	0.01776	0.047
K22	81x	*xantiana*	0.01388	-0.092
K22	86x	*xantiana*	0.01505	-0.083
K22	87x	*xantiana*	0.01171	1.143
K22	9x	*xantiana*	0.01476	0.443

Table S3. Taxon-level estimates of θ_w and Tajima's *D*.

Locus	θ_w *xantiana*	θ_w *parviflora*	*D* *parviflora*	*D* *xantiana*
a23	0.011	0.006	-0.419	-0.927
d13	0.012	0.002	-1.347	-0.548
d5	0.014	0.008	-0.728	-0.873
f9	0.025	0.005	-1.75	-1.56
g3pdh‡	0.034	0.011	-0.133	-0.441
i11	0.017	0.006	-1.689	-1.598
ipi2	0.003	0.004	0.617	-1.974
k22‡	0.031	0.006	2.899	-1.488