PNAS RESEARCH ARTICLE | EVOLUTION

The extent of introgression between incipient *Clarkia* species is determined by temporal environmental variation and mating system

Shelley A. Sianta[a,1], David A. Moeller[a], and Yaniv Brandvain[a]

Edited by Douglas Schemske, Michigan State University, East Lansing, MI; received September 18, 2023; accepted February 7, 2024

Introgression is pervasive across the tree of life but varies across taxa, geography, and genomic regions. However, the factors modulating this variation and how they may be affected by global change are not well understood. Here, we used 200 genomes and a 15-y site-specific environmental dataset to investigate the effects of environmental variation and mating system divergence on the magnitude of introgression between a recently diverged outcrosser-selfer pair of annual plants in the genus *Clarkia*. These sister taxa diverged very recently and subsequently came into secondary sympatry where they form replicated contact zones. Consistent with observations of other outcrosser-selfer pairs, we found that introgression was asymmetric between taxa, with substantially more introgression from the selfer to the outcrosser. This asymmetry was caused by a bias in the direction of initial F1 hybrid formation and subsequent backcrossing. We also found extensive variation in the outcrosser's admixture proportion among contact zones, which was predicted nearly entirely by interannual variance in spring precipitation. Greater fluctuations in spring precipitation resulted in higher admixture proportions, likely mediated by the effects of spring precipitation on the expression of traits that determine premating reproductive isolation. Climate-driven hybridization dynamics may be particularly affected by global change, potentially reshaping species boundaries and adaptation to novel environments.

introgression | speciation | environmental variation | temporal variation | mating system

Secondary contact provides the ultimate test of whether incipient species have diverged sufficiently to remain distinct. Population genomic analyses have revealed that introgression is quite common for taxa in secondary contact (1–4) and that introgression varies among contact zones (5–7) and parts of genomes (1). However, the ecological factors and phenotypic traits that modulate the direction and magnitude of introgression, and its effects on species boundaries, remain unresolved.

Variation in introgression among independent contact zones may be caused by multiple mechanisms, e.g., differences in relative abundance (6), the local genetic architecture of reproductive isolation, or environmental context (7). The environment can mediate both the expression of traits influencing hybridization (8, 9) and the fitness consequences of hybridization and introgression (10), making reproductive isolation context dependent. While studies of geographic clines reveal how introgression varies across a species' range (11–13), we have limited empirical examples that link specific environmental factors to the extent of introgression in replicated contact zones.

The extent of introgression can also differ between two hybridizing species. In animals, there is evidence that ancestry from species with more attractive males introgresses into sister taxa with less attractive males than vice versa (14, 15). Such asymmetric introgression is also common in plants. Causes of asymmetric introgression in plants have been tied to differences in ploidy (16), relative abundance (17–19), and mating system (i.e., the degree of self-fertilization in a population) (20–27). The evolution of self-fertilization (selfing) from cross-fertilization (outcrossing) is one of the most common evolutionary transitions in plants (28) and is often associated with divergence at shallow evolutionary scales (29). Multiple studies have demonstrated both asymmetric reproductive isolation between outcrossing and selfing taxa (30–32) and asymmetric introgression, with greater introgression from selfing to outcrossing lineages than the reverse (20–27).

The field of speciation genomics is moving beyond the simple yes/no question of "do these taxa exchange genes" to the loftier goals of understanding the factors that facilitate or prevent introgression. Addressing this question is key to understanding what promotes or limits the stable maintenance of new species. The extent of introgression appears to vary both across species pairs and across populations of a species pair, but quantitative estimates of this variation are required to test meaningful hypotheses. Here we used weather

Significance

Recent genomic analyses show that introgression, or genetic exchange between species, is quite common. Yet, the emergence of stable species requires some limit on the extent of introgression. To understand the factors that limit introgression, we analyzed whole-genome sequences from six independent secondary contact zones of two recently diverged sister taxa. We found that the magnitude of introgression varied substantially across contact zones and that this variation was strongly predicted by temporal variance in the environment, specifically by interannual variance in precipitation. Our results have implications for how global climate change may affect hybridization dynamics and highlight the key role that environmental factors play in the speciation process.

Author affiliations: [a]Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN 55108

Preprint server: BioRxiv, https://doi.org/10.1101/2023.8.30.555593, CC BY-NC 4.0 International Copyright License.

Author contributions: D.A.M. and Y.B. designed research; S.A.S. and D.A.M performed research; S.A.S. and Y.B. analyzed data; and S.A.S., D.A.M., and Y.B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: ssianta@umn.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2316008121/-/DCSupplemental.

Published March 11, 2024.

station data to identify a major environmental driver of introgression across six independent regions of secondary contact between two recently diverged taxa.

We examined well-studied sister taxa, currently classified as subspecies of *Clarkia xantiana* (33), which diverged recently (25), differ in mating system (34, 35), and form replicated contact zones across a narrow region of secondary sympatry (25, 36). These two taxa are endemic to the foothills of the southern Sierra Nevada of California, U.S.A., with subspecies *xantiana* (hereafter, the "outcrosser" or *xantiana*) occurring in the more mesic oak-pine savanna habitats in the western foothills of the Sierra Nevada and subspecies *parviflora* (hereafter, the "selfer" or *parviflora*) occurring in more xeric scrub habitats in the eastern foothills of the Sierra Nevada (34). Their ranges overlap in a narrow north–south band, approximately 5 × 30 km, in the Kern River drainage, which covers a complex topographic landscape. Previous work showed that reproductive isolation is >98% for both taxa (31). However, preliminary evidence indicates that introgression is asymmetric, with higher gene flow from the selfer to the outcrosser (25).

We used 200 whole-genome sequences from 29 populations across the range of both taxa to first confirm prior studies of divergence time (25), divergence history (36), and population structure (36, 37). These historical insights provide key context for the design of introgression analyses and the interpretation of admixture results. We then combine sequence data with 15 y of site-specific weather data to examine the contribution of environmental variation and mating system divergence to the substantial quantitative variation we found in introgression across six independent contact zones occurring over <30 km. Specifically, we tested whether mean and variance in precipitation and temperature, which are important drivers of demographic parameters and vary strongly across the region of sympatry (38), predict levels of introgression across contact zones. We also ask whether mating system divergence drives asymmetry in introgression, as seen in other systems. We find that two factors, mating system and interannual variance in spring precipitation, together explain 93% of the variation in introgression across the contact zones. Mating system causes asymmetry in the direction of hybridization and interannual variance in spring precipitation modulates the frequency of hybridization. These results have implications for the role of climate change in the maintenance of reproductive barriers between ecologically distinct populations and species.

Results

Genome Divergence and Population Structure. We assembled and annotated a reference genome of an inbred line of the selfer, *parviflora*, using PacBio sequencing (50×) and Bionano optimal maps for scaffolding. The reference genome, the first for the genus *Clarkia*, was assembled into 59 scaffolds totaling 695 Mb, with 45 k annotated genes. We then performed whole-genome resequencing (WGS) on 200 individuals from across the range of each taxon (*SI Appendix*, Table S1 and Fig. 1 *A* and *B*). We subset our filtered dataset to coding region sequences and generated two high-quality matrices to use in analyses: all fourfold degenerate synonymous sites (variant and invariant, 1,751,603 bp) and fourfold degenerate synonymous biallelic single nucleotide polymorphisms (SNPs; 353,538 bp).

A principal components analysis (PCA) showed that the two taxa are genetically distinct (Fig. 1*C*). Consistent with expectations from their difference in mating system, the selfer had one-quarter the diversity within populations (π), half of the absolute divergence among populations (D_{xy}), and four times the relative differentiation among populations (F_{ST}) relative to the outcrosser, *xantiana* (Table 1 and *SI Appendix*, Fig. S1).

Population structure within each taxon largely mirrored geography. Within the outcrosser, southern populations (THPO, OC, LCW, DLE) were differentiated strongly from northern populations; the latter group comprised populations from the region of secondary sympatry and adjacent allopatric populations (Fig. 1*D* and *SI Appendix*, Fig. S2 *B–D*). Within the zone of secondary sympatry, southern outcrosser populations (SM, S7, GRW) were closely related to nearby allopatric outcrosser populations (BR, KYE; Fig. 1*D* and *SI Appendix*, Fig. S2*D*). By contrast, northern sympatric outcrosser populations (S22, SAW, GC) form a distinct cluster that is divergent from the remaining outcrosser populations (Fig. 1*D* and *SI Appendix*, Fig. S2*D*). Within the selfer, populations from the region of sympatry and adjacent allopatric sites form a cluster that is strongly divergent from allopatric populations farther from the region of sympatry (Fig. 1*E* and *SI Appendix*, Fig. S2 *E* and *F*).

Allopatric Origin of Selfer with Gene Flow in Secondary Sympatry. Prior analyses based on a smaller dataset showed that the selfer diverged from the outcrosser in allopatry ca. 65,000 y bp and later came into secondary contact (25, 36). Consistent with those results, we found that allopatric selfer populations were more similar to the outcrosser on PC1 than sympatric selfer populations (Fig. 1*C*). Low genetic diversity within and low pairwise D_{xy} between sympatric selfer populations (Fig. 1*E*) suggest that serial bottlenecks accompanied the selfer's westward range expansion.

Patterns of population structure also supported a model of "budding speciation", wherein the selfer is recently derived from one lineage within the outcrosser. This divergence history causes variation in the degree of divergence between the selfer and different outcrosser populations. Populations of the outcrosser differed dramatically in their divergence from the selfer (median D_{xy} ranges from 0.86 to 1.36%), while divergence from the outcrosser was consistent across populations of the selfer (median D_{xy} ranges from 1.11 to 1.13%, Fig. 2*A*). Of all allopatric outcrosser populations in our dataset, LCW was the least diverged from the selfer (D_{xy} of 1.11%) and can serve as a proxy for the ancestral progenitor population from which the selfer arose. Using LCW as an outgroup, we followed Hudson et al. (39) to estimate a divergence time of ca. 56,000 y bp.

Patterns of interspecific divergence (D_{xy}) are consistent with introgression into the outcrosser but not into the selfer. Interspecific divergence from the selfer is lower for sympatric (mean D_{xy} = 1.05%) than for allopatric (mean D_{xy} = 1.17%) outcrosser populations (Fig. 2*A*). This result is robust to the removal of the two southernmost allopatric outcrosser populations (THPO and OC), which are substantially divergent from all other samples (*SI Appendix*, Fig. S2 *B–D*). However, the extent of reduced interspecific D_{xy} varied across contact zones. There was little introgression in the southern region of sympatry—D_{xy} between the selfer and southern sympatric outcrosser populations (GRW, S7, SM) was comparable to D_{xy} between the selfer and nearby allopatric outcrosser populations (Fig. 2*A*). By contrast, D_{xy} between the selfer and northern sympatric outcrosser populations decreased by between 8.4% and 20% relative to average allopatric interspecific D_{xy}, suggesting extensive introgression in the northern region of sympatry. By contrast, sympatric and allopatric selfer populations had similar D_{xy} from the outcrosser (Fig. 2*A*); this lack of variation in interspecific D_{xy} among selfer populations reflects its recent divergence from a taxon with higher diversity (Table 1 and *SI Appendix, SI Text 2*).

Introgression Is Asymmetric and Independent across Contact Zones. We first designed ABBA-BABA tests to formally test for introgression into each taxon at each contact zone. We found

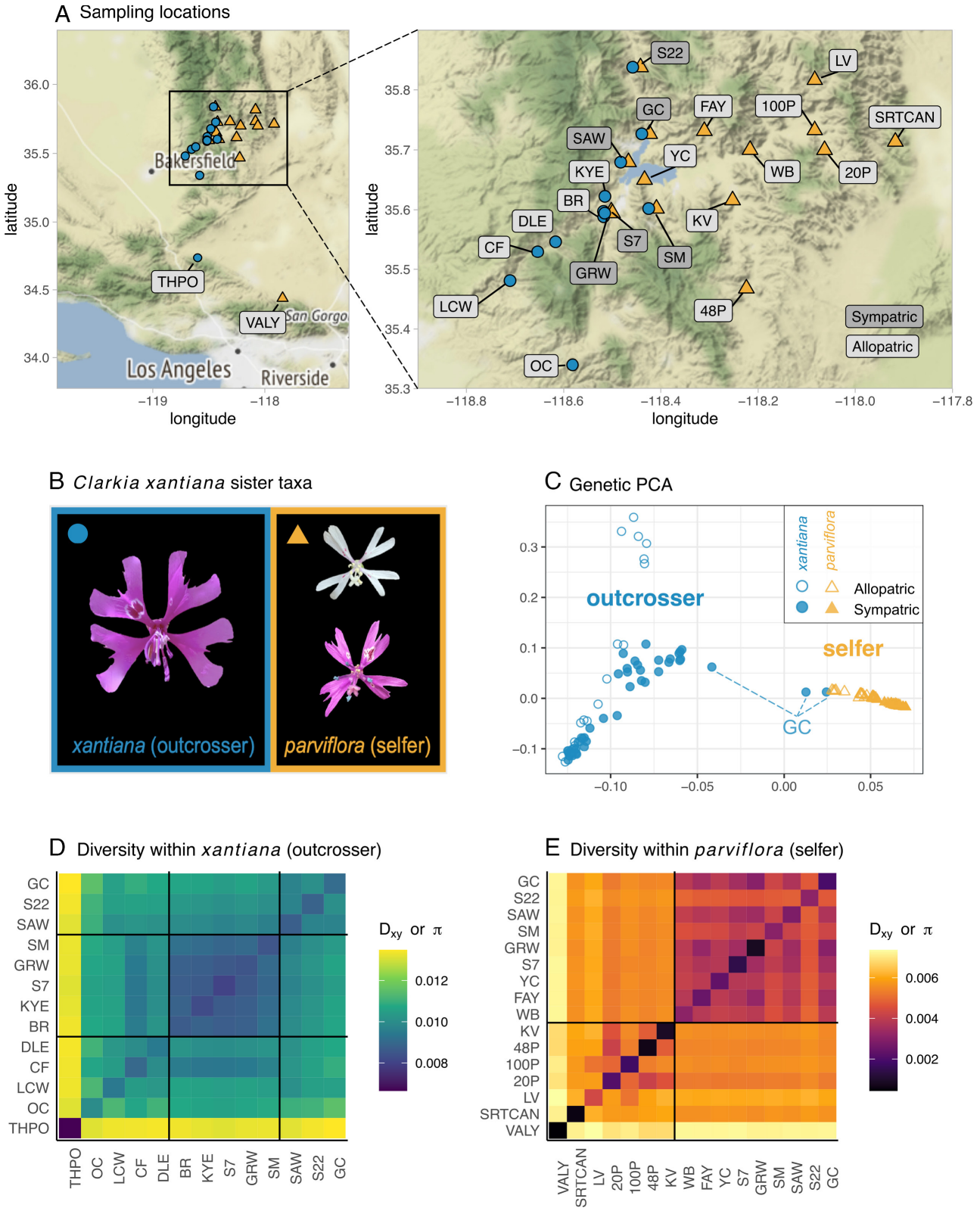

Fig. 1. Sister taxa of *C. xantiana* are genetically distinct and structured primarily by geography. (*A*) Sampling locations of populations. Sympatric contact zones where both taxa co-occur are labeled in dark gray, allopatric populations are labeled in light gray. (*B*) Flowers of each taxon. The selfer has small flowers, varies in flower color, and has less spatial and temporal separation between mature male and female reproductive organs. (*C*) Genetic PCA between taxa based on fourfold degenerate SNPs from coding sequences. *X*-axis is PC1 and *y*-axis is PC2. Each point represents an individual, with symbol fill reflecting whether that individual is from sympatry or allopatry. We highlight three individuals of *xantiana* (the outcrosser) from the GC contact zone that cluster near *parviflora* (the selfer). (*D* and *E*) Genetic diversity within and among populations of each taxon. Diagonals represent π, diversity within a population. Off-diagonals represent D_{xy}, divergence between populations. Solid black lines demarcate qualitative groupings of populations based on pairwise D_{xy}. Photo credits in *B*: Taz Mueller.

Table 1. Diversity and divergence within and between taxa are consistent with mating system shifts

	π	D_{xy}	F_{ST}
parviflora (selfer)	0.20% (0.06%)	0.53% (0.02%)	0.41 (0.03)
xantiana (outcrosser)	0.87% (0.05%)	1.06% (0.03%)	0.10 (0.02)
parviflora X *xantiana*	NA	1.12% (0.01%)	0.36 (0.01)

Average (2 * SE) neutral diversity within populations (π), absolute pairwise divergence among populations (D_{xy}), and relative pairwise divergence among populations (F_{ST}).

evidence for introgression into the outcrosser in all contact zones (significant D-statistic) except for the most southern site, GRW (Fig. 2*B* and *SI Appendix*, Table S2). By contrast, we found no evidence for introgression into the selfer (Fig. 2*B*). We then designed ABBA-BABA tests to evaluate the extent to which introgression is independent across contact zones by asking whether there is an excess of discordant trees that group the outcrosser and selfer from the same contact zone in a clade (*SI Appendix*, Fig. S3 *A–C*). We found evidence for independent introgression occurring in at least the three sympatric sites with substantial introgression (GC, S22 and S7; *SI Appendix*, Fig. S3*D*). The remaining sites do not show any signs of local introgression, suggesting that these contact zones either share ancestral introgressed ancestry (i.e., genomic regions that have been introgressed from another taxon) or that we lacked power to detect independence because introgression has been rare.

Phylogenetic analysis of SNPs in the chloroplast resolved a well-supported clade of nearly all outcrosser individuals (95% bootstrap support) that is either sister to (51% bootstrap support) or nested within (49% bootstrap support) a clade of primarily selfer individuals (Fig. 2*C* and *SI Appendix*, Figs. S4 and S5). Within the selfer clade, there were ten outcrosser individuals from three contact zones (*SI Appendix*, Fig. S5). Within the outcrosser clade, there were no selfer individuals. This asymmetric chloroplast capture demonstrates that hybrids were sired by the outcrosser on selfer maternal plants and that hybrids backcrossed to the outcrosser through receipt of outcrosser pollen.

Extensive Variation in the Magnitude of Introgression among Contact Zones. We next quantified individual-level admixture proportions (i.e., percent of the genome that is introgressed) using a Hidden Markov Model (HMM) (40). We found evidence of introgression into each taxon at all contact zones (Fig. 3 and *SI Appendix*, Table S3) that is largely asymmetric, with the outcrosser having higher admixture proportions than the selfer. The asymmetry is particularly stark at contact zones with higher overall admixture proportions (S22, GC, SAW, and SM) and is not observed in populations with low admixture proportions (S7, GRW).

Unlike ABBA-BABA, the HMM identified outcrosser ancestry in the selfer. Admixture proportions in the selfer were generally low: four populations had average admixture proportions under 1.8% and the remaining two populations had average admixture proportions of 3.8% (SM) and 6.9% (S22). Notably, both S22 and SM had the highest sequence diversity (π) among and greatest divergence from the other sympatric selfer populations (D_{xy}, Fig. 1*E* and *SI Appendix*, Fig. S6), consistent with a history of introgression. We also estimated admixture proportions with the f_d statistic (41) and detected no introgression into the selfer; however, this is likely a consequence of the low diversity within the selfer (*Discussion* and *SI Appendix, SI Texts 2 and 4*). Given the

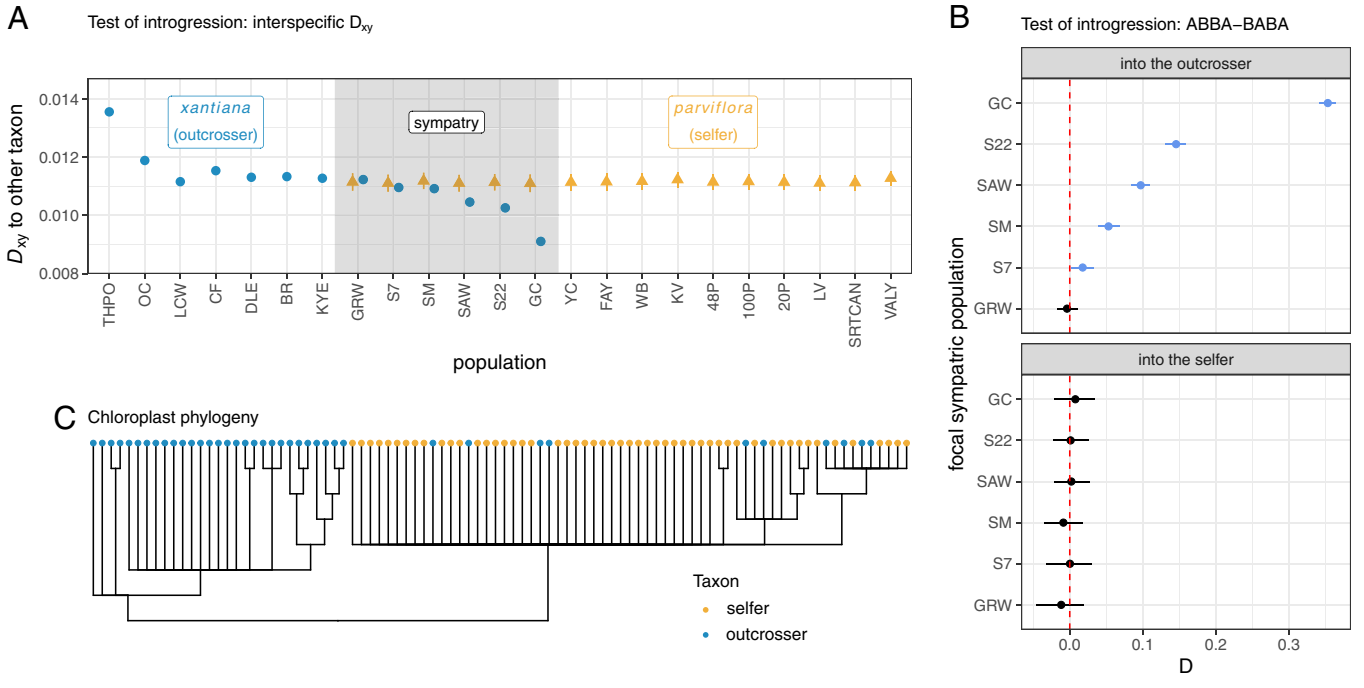

Fig. 2. Tests for the presence of introgression find evidence for introgression into the outcrosser. (*A*) Interspecific D_{xy} between an outcrosser (blue) or selfer (orange) population at a given site (*x*-axis) and the other taxon. Points represent mean pairwise interspecific D_{xy} between the focal population and all other populations of the other taxon; error bars represent one SE. Sympatric sites are highlighted by the gray rectangle. (*B*) ABBA-BABA tests of introgression into the outcrosser (*Top* facet) or selfer (*Bottom* facet). Points are the observed D-statistic and error bars represent bootstrapped 95% CI. Tests of introgression into the outcrosser take the form of (P1= allopatric outcrosser, P2 = focal sympatric outcrosser (*y*-axis), P3 = allopatric *parviflora*, and O = *C. unguiculata*). Tests of introgression into the selfer take the form of (P1= allopatric selfer, P2 = focal sympatric selfer (*y*-axis), P3 = allopatric outcrosser, and O = *C. unguiculata*). (*C*) Dendrogram of chloroplast phylogeny, with tip points colored by taxon, shows the 10 outcrosser individuals that fall within the selfer clade. Seventy percent of individuals within each population are subset out for space; full phylogeny is shown in *SI Appendix*, Fig. S4.

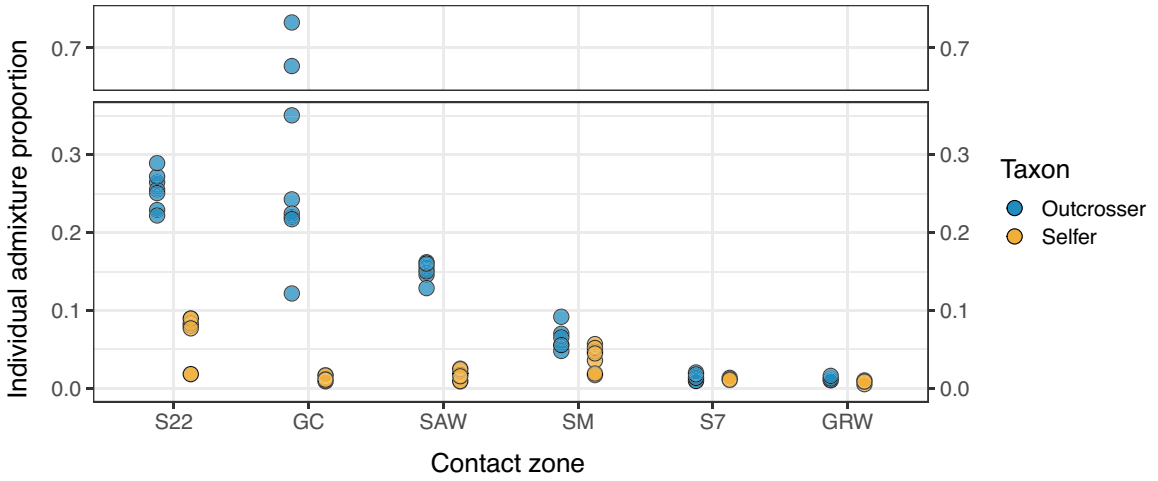

Fig. 3. The extent of introgression is largely asymmetric between taxa and varies across contact zone sites. HMM estimates of individual-level admixture proportions. The HMM estimates a probability of ancestry at each site in the genome and individual-level admixture proportions are calculated as proportion heterospecific ancestry, weighted by probability.

discrepancy among tests, we validated that the HMM truly detected admixed tracts instead of unsorted ancestral polymorphism in the two contact zones with the highest admixture proportions by comparing divergence (D_{xy}) at selfer genomic regions called admixed or unadmixed by the HMM with an allopatric outcrosser population (see *SI Appendix, SI Text 3* for full details). Our validation tests found significant evidence supporting introgression from the outcrosser to the selfer at S22 ($P = 0.029$) and marginally significant evidence at site SM ($P = 0.075$).

The magnitude of introgression into the outcrosser was highly variable among contact zones whereas the selfer exhibited little variation. There were high levels of selfer ancestry in outcrosser populations in the northern contact zones (mean admixture proportions of 35%, 25%, and 15% in GC, S22, and SAW, respectively), and less introgression into southern contact zones (admixture proportions of 6.5%, 1.4%, and 1.3% in SM, S7, and GRW, respectively). The HMM identified early-generation hybrids with exceptionally high admixture proportion at site GC, consistent with observations in our PCA ("GC" labels on Fig. 1*C*). We also estimated admixture proportions with the f_d statistic and with a D_{xy}-based calculation; both methods produced admixture proportions highly correlated with the those from the HMM (*SI Appendix, SI Text 4* and Fig. S7).

Taken together, the HMM results suggest that introgression was largely asymmetric, with more introgression from the selfer to the outcrosser than the reverse. The HMM also confirmed that introgression dynamics have been quite variable across contact zones, particularly for introgression into the outcrosser.

Greater Introgression Occurs in More Variable Environments. Given that introgression into the outcrosser was highly variable among contact zones, we tested whether admixture proportion in the outcrosser was related to a set of eight environmental variables collected over 15 y. We included mean and variance of temperature and precipitation in the winter (November to January) and spring (February to May), because natural plants primarily germinate in winter and primarily grow and flower in spring. Our approach to model selection—leave-one-out (LOO) regression—identified variance in spring precipitation as the strongest predictor and the only variable to outperform an intercept-only model in the LOO test. Its performance also exceeded expectations from the null (Permutation-based $P = 0.003$; Fig. 4*A*). A simple linear regression similarly identified a strong relationship between admixture proportion and variance in spring precipitation ($R^2 = 0.93$; $t = 7.02$, $df = 4$, $P = 0.002$; Fig. 4*B*), with the standardized regression coefficient of 0.96 close to its maximum value of 1.0.

Lastly, we asked how much of the total variance in admixture proportions across taxa and contact zones was explained by both mating system and variance in spring precipitation. These two factors and their interaction explained 93% of the variation in admixture proportion ($R^2 = 0.93$, $F_{3,8} = 36.5$, $P < 0.001$).

Discussion

Disentangling the factors that modulate the extent to which incipient species remain isolated upon secondary contact is central to understanding the generation of biodiversity. Here, we show that secondary contact between two recently diverged taxa has resulted in extensive variation in the magnitude of introgression, both in the direction between taxa and across six independent contact zones that span a narrow 5 × 30 km region. Mating system divergence caused asymmetry in introgression, with considerably more introgression from the selfer to the outcrosser than the reverse. Admixture proportion in the outcrosser, which varied strongly across contact zones, was largely explained by temporal variance in the environment. Environmental fluctuations may result in episodes where reproductive isolation is weakened, thereby increasing the opportunity for hybridization and gene flow.

While variation in hybridization dynamics and introgression across contact zones has been documented in other systems (42–45), rarely have studies identified factors modulating that variation [but see Mandeville et al. (5) and Lewanski et al. (6)]. In *C. xantiana*, introgression into the outcrosser varies clinally across geography, with higher admixture proportions in northern (range: 15 to 35%) than southern contact zones (range: 1 to 6%). Such spatial variation in introgression across sites likely reflects spatial variation in factors that modulate the opportunity for hybridization [e.g., Zhang et al. (8)] and/or spatially variable selection on introgressed ancestry (10, 46). We found that one environmental variable—interannual variance in spring precipitation—explained nearly all variation in the outcrosser's admixture proportion across contact zones.

Variation in spring precipitation is associated with variation in flowering time among closely related species (47–50). Indeed, flowering time divergence between the *Clarkia* taxa in this study is likely driven by adaptive divergence to the onset of late-spring

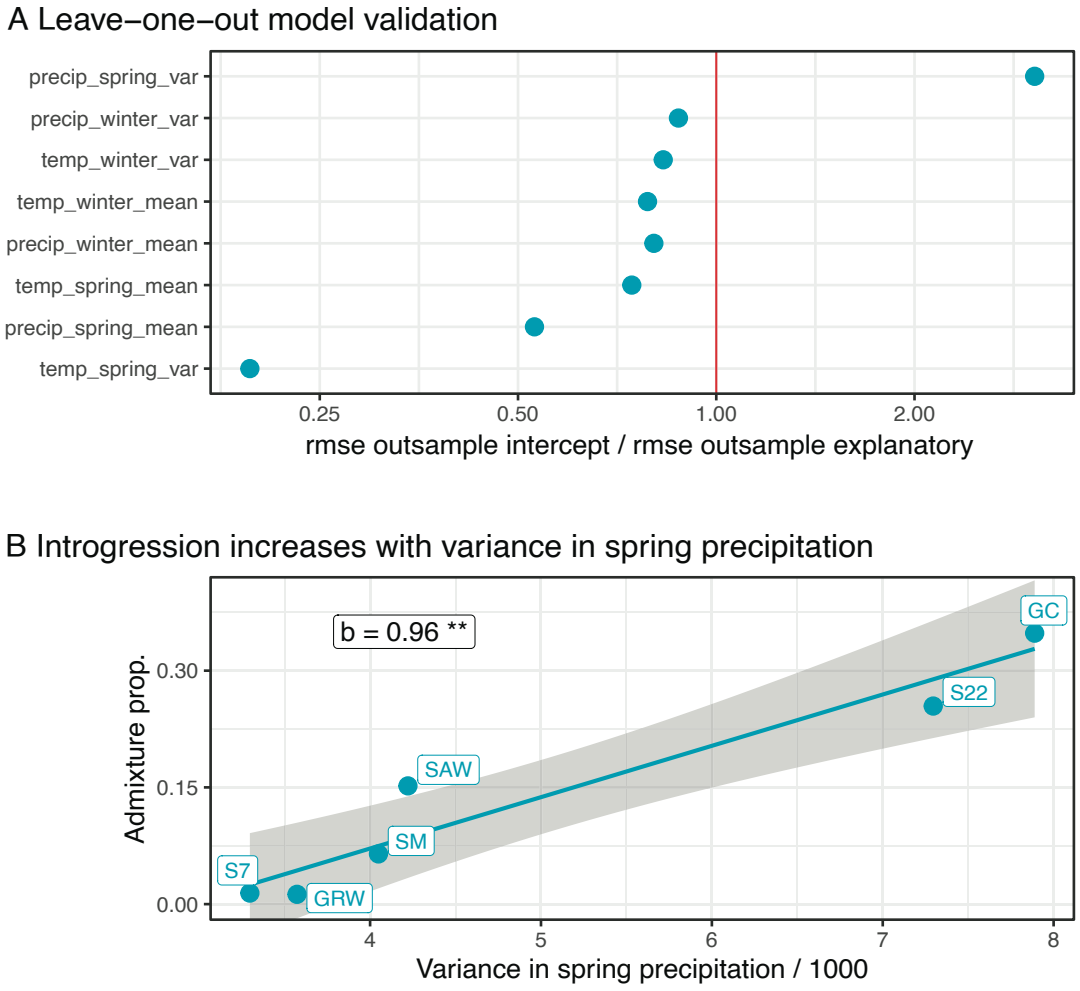

Fig. 4. Variance in spring precipitation predicts the admixture proportion across six contact zones. (*A*) We present the root mean square error (rmse) for outsample prediction from LOO validation for single-variable regressions for each of eight environmental variables. To aid in interpretation, we present this as the ratio of outsample rmse for a model with just the intercept to one with an intercept and slope predicted by the variable of interest (on the *y* axis). Values exceeding one indicate that the variable of interest predicts admixture proportion in the outsample better than does a simple model with just an intercept. (*B*) The linear relationship between admixture proportion in the outcrosser and the variance in spring precipitation. The standardized regression coefficient, *b*, is presented. ***P* = 0.002.

drought: The selfer is a drought avoider and flowers earlier than the outcrosser, which is a drought tolerator (51, 52). This adaptive divergence in flowering time results in strong premating isolation in sympatry—preventing 79 to 96% of gene flow (31). However, flowering time variation along environmental gradients is often caused by plasticity, not adaptation (53). Much like adaptive responses, a given environmental variable can lead to differences in the direction and magnitude of flowering time change among species with similar life histories. For example, serpentine soils can induce both earlier and later flowering across 17 annual California wildflowers (9). As such, when abiotic environments fluctuate through time, the extent of overlap in flowering time may vary in concert because of species' differing responses to shared environments. Results from this study suggest that *C. xantiana* contact zones with greater variance in spring precipitation experience greater temporal fluctuations in phenological reproductive isolation, leading to greater opportunities for gene flow and introgression. Past documentation of flowering time in these contact zones, along with 20 y of personal observations, have suggested that flowering times overlap more in years with low spring precipitation (*SI Appendix*, Table S4) (34). Moreover, transplant experiments have shown that flowering times of the two taxa are more similar in some environments than others (54), and a resurrection study found that a multi-year period of intense drought resulted in genetically based shifts toward earlier flowering in the outcrosser (55), which may also cause greater overlap in flowering time. Taken together, environmental fluctuations may be pivotal for determining the degree of reproductive isolation between incipient species because of their influence on the expression of traits that mediate premating isolation.

Mating system divergence between the outcrosser and selfer has resulted in asymmetric introgression upon secondary contact. In all contact zones (except two with negligible levels of introgression), there was substantially more introgression into the outcrosser than into the selfer. Our observations of chloroplast genome variation support the hypothesis that pollen from the outcrosser sires F1 offspring on the selfer—multiple outcrosser individuals from three contact zones were nested within the clade of selfer chloroplast sequences. Moreover, our results suggest that consistent backcrossing to the outcrosser often involves the outcrosser siring offspring on F1 and early-generation hybrids. There are several key phenotypic and ecological differences between outcrossers and selfers, including the taxon pair in this study, that likely mediate this direction of hybridization and, ultimately, introgression (56). Outcrossers are typically more abundant (31, 57) and produce considerably more pollen (up to two orders of magnitude) (58) than selfers, causing the outcrosser pollen pool to be much larger than the selfer pollen pool. In addition, outcrossers are typically more attractive

to bee pollinators (31) which further increases the probability that pollinator foraging transitions occur from outcrossers to selfers. Last, outcrosser pollen tubes typically outcompete selfer pollen tubes when growing in the same style (31, 59–61). In the *Clarkia* taxa used in this study, an asymmetric postzygotic barrier (F1 seed viability) in this system also makes it more likely for hybrids to successfully form on a selfer parent (31). Ultimately, these factors overwhelm countervailing factors that might increase gene flow in the opposite direction (e.g., self-fertilization preempts ovules from fertilization by outcrosser pollen). Other studies have found similar asymmetric introgression between selfers and outcrossers (23, 24, 26, 27, 62–64), suggesting that these phenotypic and ecological factors associated with mating system shifts have consistent consequences on hybridization and introgression.

Although most studies have not found introgression from an outcrosser into a selfer (26, 27, 63) [but see Brandvain et al. (24)], we detected such introgression, albeit at relatively low levels (admixture proportions < 0.07, which is approximately that observed in sympatric outcrosser populations with the lowest levels of introgression). However, our capacity to detect this direction of introgression depended on the method: genome-wide methods (e.g., ABBA-BABA, D_{xy}-based tests of introgression, and f_d) did not find evidence of introgression into the selfer. Only the local ancestry analysis, which uses a HMM to detect tracts of ancestry across the genome, detected introgression in this direction. The discrepancy between these types of analyses is likely due to a detection bias in observing introgression into a recently derived taxon with low diversity, such as into the selfer in this study. The detection bias occurs when divergence between the taxa is comparable to diversity within the outcrosser (as is the case here, Table 1). Such comparable levels of divergence and diversity cause the time to coalescence between an outcrosser and selfer allele to be the same when either 1) the selfer allele is introgressed from the outcrosser (coalescence time is proportional to diversity within the outcrosser) or 2) the selfer allele is not introgressed (coalescence time is proportional to interspecific divergence), making it hard to distinguish introgression from unsorted ancestral polymorphisms (see *SI Appendix, SI Text 2* for details). We advocate for consideration of this detection bias and for extra validation when investigating introgression in a system with recently diverged taxa that differ in levels of diversity, as may generally be true in cases of budding speciation (65–67).

While there are many ways to delimit species boundaries, species delimitation is best served with multiple lines of evidence. The two taxa in this study were originally described as subspecies (33) based on fertility of hybrids in a greenhouse (35). However, the data presented here adds to a growing body of evidence that the two *C. xantiana* taxa are as distinct as many described species. They are phylogenetically (34, 52, 68, 69) and phenotypically (34, 52, 68, 69) distinct, and have high (>98%) but incomplete levels of reproductive isolation (70). This leaky reproductive isolation has resulted in variable levels of hybridization and introgression across the zone of sympatry, in both past and contemporary times. However, the contact zones in this study are generally bimodal [i.e., predominantly parental types and low frequency hybrids (56)] with little intergradation in morphology and genetics between the species, supporting species status (71, 72). Moreover, the divergence time we estimated between the taxa [56,000 y, see also Pettengill and Moeller (25)] is on par with other pairs of outcrosser-selfer sister taxa that are designated as sister species [*Capsella rubella-C. grandiflora*, 50 kya (73); *Mimulus nasutus-M. guttatus*, ~200 kya (24); *Collinsia parryi-C. concolor*, <200 kya (74)]. Population admixture proportions—ranging from zero to thirty-five percent—are within the range often reported in studies of hybridizing species. For example, a recent review reported admixture proportions nearing ten percent in admixed populations of butterflies, birds, plants, and fish (75, 76), and a study in two swordtail fish hybrid zones found admixture proportions between ten and twenty-four percent (77, 78). We predict these incipient species will continue to maintain their species integrity in sympatry. However, given that there are weak intrinsic postzygotic barriers between the taxa, there is always the possibility that boundaries between taxa could collapse if environmental change decreases reproductive isolation (79). It will be useful to continue investigating contact zones that depart from strong bimodality and characterize how introgression dynamics change in future years.

Secondary contact in *C. xantiana* has resulted in variable levels of introgression across a small spatial scale that is nearly fully explained by mating system divergence and spatial variation in interannual spring precipitation. We add the caveat that interannual variance in spring precipitation may not be the causal factor affecting introgression but instead is correlated with the causal factor. We suspect that if this is the case, the causal factor is also temporal variation in some aspect of the environment. To our knowledge, this is the first study to demonstrate a quantitative link between temporal variance in the environment and introgression. Temporal variance in the environment can cause plastic shifts in traits mediating reproductive isolation (e.g., flowering time) that increase the probability of hybridization. As environments continue to rapidly change and become more variable due to climate change, we may expect to see more opportunities for hybridization (80–83). Reduced reproductive isolation and increased introgression have already been observed in other systems in response to pollution (84), disturbance (85), and drought (86, 87) [but see Rivest et al. (34)]. An important next step will be to determine whether the introgression resulting from changes in the rates of hybridization with global change have adaptive (e.g., refs. 86, 87) or maladaptive consequences in natural settings.

Materials and Methods

Study System. *C. xantiana* A. Gray (Onagraceae) is a self-compatible, annual plant native to the foothills of the southern Sierra Nevada of California, USA (34, 52, 68, 69). *C. xantiana* is composed of two subspecies–the primarily outcrossing *C. xantiana ssp. xantiana* (i.e., the outcrosser or *xantiana*) and the primarily selfing *C. xantiana ssp. parviflora* (i.e., the selfer or *parviflora*) (33–35, 68, 88, 89). Given that these taxa are phylogenetically distinct (36), phenotypically distinct (88), and > 98% reproductively isolated (89, 90), they are likely incipient species as opposed to their original taxonomic description as subspecies. The outcrosser occurs in the western foothills of the Sierra Nevada in low-elevation grasslands, middle-elevation oak-pine savanna, and high-elevation chaparral. The selfer occurs in low-productivity scrub habitats with occasional junipers, pines, and Joshua trees in the eastern foothills of the Sierra Nevada. While their ranges are largely allopatric, they overlap in a narrow region centered on Isabella Lake. In the region of overlap, the two taxa co-occur at multiple discrete sites, which are separated by 0.4 to 27.5 km. When sympatric, individuals occur within meters of one another at the boundary of different microhabitats. As such, they have parapatric distributions at both geographic and local scales.

Constructing the First *Clarkia* Reference Genome. We assembled and annotated the first *Clarkia* reference genome. The genome was derived from one selfer individual that was selfed for five generations. We chose this individual from the Long Valley population (LV on Fig. 1*A*) because prior analyses indicated that this was the earliest diverging, extant selfer population (91) and thus should minimize reference bias when mapping sequences from the selfer and the outcrosser. We combined PacBio sequencing and Bionano optical mapping to generate a 695.6 Mbp genome assembled on 59 scaffolds with an N50 length of 32.5 Mbp. We created a genome annotation by combining gene predictions from transcriptome data (from leaf, root, stem, flower, and fruit tissue) and a suite of ab initio gene predictor programs using the funannotate pipeline (v. 1.8.9) (90), resulting in

44,998 total gene models. We assembled a reference genome for the chloroplast by aligning raw PacBio reads to the chloroplast genome of a species from a closely related genus (*Oenothera villarica*, Onagraceae; NCBI accession GI 1034702878). We used Canu (v. 2.1) (91, 92) to assemble aligned raw reads, resulting in one continuous contig 161,015 nucleotides long. Full details about how the reference genome was assembled and annotated are provided in *SI Appendix, SI Text 1*.

WGS of the Outcrosser and Selfer. To quantify the distribution of genetic diversity within and between taxa and examine patterns of introgression in the region of secondary sympatry, we collected range-wide samples of each taxon (Fig. 1*A* and *SI Appendix*, Table S1). In the summer of 2019, we collected seeds from 13 outcrosser populations (six in contact zones and seven allopatric) and 16 selfer populations (six in contact zones and ten allopatric). We also collected seeds from four congeneric species for use as outgroups in analyses (*SI Appendix*, Table S1). We germinated field-collected seeds and collected leaf tissue for WGS. We sampled an average of 8.5 and 5 individuals from contact zone and allopatric populations, respectively. In total, we sequenced 200 individuals.

DNA was extracted using the cetyltrimethylammonium bromide (CTAB) method and libraries were prepared with the Illumina TruSeq NanoDNA library preparation kit. Samples were split into two unique sequencing pools, each containing 100 samples. Each pool was sequenced across two lanes of an Illumina NovaSeq S4 2 × 150-bp run, with a targeted depth of 13×.

We removed adaptor sequences and low-quality reads with scythe (93). We used BWA mem to map trimmed reads to the masked reference genome. We used Samtools (94) to sort reads, Picard to add read groups, and Samtools to index alignments. We used GATK (v 3.8.0) (95) to call genotypes. GATK's HaplotypeCaller function was used to generate per-individual genotypes, which were subsequently grouped for joint genotyping with the "GenotypeGVCF" function. Because joint genotyping with GenotypeGVCF assumes random mating among samples, we grouped our samples differently for the outcrosser and the selfer. For the selfer, we conducted joint genotyping for only individuals within a population because there is much higher differentiation among populations (F_{ST}) in the selfer than the outcrosser (95). We grouped all samples of the outcrosser together for joint genotyping, except for individuals from the southernmost population THPO, which is geographically disjunct from the other samples. The three individuals from each of the four outgroup species were grouped together for joint genotyping. We ran GenotypeGVCF over each scaffold and merged all VCFs for a given scaffold (i.e., from the different groups) into one VCF.

We filtered our VCFs in multiple steps. We first surveyed average depth and percent missing data across all individuals and scaffolds using the BCFtools (94) stat function and custom R scripts. We flagged seven scaffolds for which the vast majority of individuals had low (<5) or high (>30) depth; these same seven scaffolds were flagged for having the vast majority of individuals with >75% missing data. We excluded these seven scaffolds, eliminating 1.2% of the genome. We also flagged and removed 13 individuals that had >50% missing data across the genome. Paralogs can create mapping issues by having multiple paralogs map to the same regions. We identified and removed regions with potential paralogs by identifying regions with high heterozygosity in the selfer. We flagged sites at which ≥50 of 100 selfer individuals were heterozygous (hereafter high heterozygosity sites), binned high heterozygosity sites into 1 kb windows, and removed windows for which there were >15 high heterozygosity sites. We then filtered out all sites with "LowQual" scores in the FILTER column and changed all genotypes to missing for any individual/site combination where read depth <4 or >30. We then removed any site that was missing data for over 50% of the outcrosser individuals and 50% of the selfer individuals. After these filtering steps, we reduced the genome by an additional 73%, resulting in 171 Mb.

To ensure that we were using high-quality sites for analyses, we further filtered our dataset to fourfold degenerate sites in coding sequences and to biallelic SNPs within that dataset.

Quantifying Genetic Diversity within and between Taxa. We first compared the extent of genetic divergence within and between the two taxa using PCA. We calculated a genomic PCA using a custom R script to account for missing genotype data by calculating a pairwise genotypic covariance matrix across SNPs at fourfold degenerate sites and finding components as the eigenvectors of this matrix. To better visualize population structure within each taxon, we generated within taxon PCAs as above, discarding geographically distant outlier populations from the southern edge of the range for each taxon (VALY in the selfer, and OC and THPO in the outcrosser), which had undue influence on PC space (*SI Appendix*, Fig. S2).

To determine how genetic variation is structured across the range of each taxon, we characterized the mean number of pairwise sequence differences at fourfold degenerate sites (an estimate of π within populations and D_{xy} between populations) and population differentiation (measured as F_{ST}) with the program pixy (96).

We asked whether patterns of interspecific D_{xy} were consistent with an allopatric origin of the selfer, wherein the selfer arose from a historic outcrosser population that was more closely related to one of the allopatric (southern) outcrosser populations than one of the sympatric (northern) populations, as proposed by Pettengill and Moeller (36). We found that the allopatric outcrosser population LCW was the least diverged with the selfer. We used LCW as a proxy for the ancestral outcrosser population from which the selfer arose to estimate a population split time. We used the method described in Hudson et al. (40), which uses the difference in divergence between taxa and diversity within this putative ancestral population divided by two times the mutation rate (standard assumption of 1.5×10^{-8} mutations per base pair mutation per generation), to estimate a population split time.

We also used interspecific D_{xy} to ask whether introgression occurs in sympatry. We expected gene flow to decrease divergence (interspecific D_{xy}) between a sympatric outcrosser population with a history of introgression and all selfer populations, but such gene flow would not affect divergence between the source selfer populations and allopatric outcrosser populations. In principle, the same logic should apply for cases of gene flow from the outcrosser to the selfer. However, because divergence between outcrosser populations is comparable to divergence between the taxa, this approach is underpowered for detecting gene flow in this direction (*SI Appendix, SI Text 2*).

Tests of Introgression. We formally tested for the presence of introgression by designing ABBA–BABA tests (97) for introgression into each taxon at each contact zone. In ABBA–BABA tests, the four taxa are labeled P1, P2, P3, and O for outgroup. P1 and P2 are populations of the hypothesized recipient species, where P1 does not experience introgression but P2 does. P3 is a population of the hypothesized donor species. Among discordant trees that do not match population history, introgression results in more (P2, P3)(P1, P4) trees than (P1, P3)(P2, P4) trees, while no such excess is predicted without introgression. We conducted 12 such ABBA–BABA tests–in each test, one of the six contact zone populations of each taxon was designated as the putative recipient of gene flow, P2, and the outgroup in all tests was set as the congener, *Clarkia unguiculata*. For tests of introgression into the outcrosser, P1 was the allopatric outcrosser population DLE and P3 was the allopatric selfer population FAY. For tests of introgression into the selfer, we swapped P1and P3. By using allopatric populations as P1 and P3, we can separately test for introgression in both directions. For all tests, we calculated the number of discordant trees across the genome at coding sequence sites, calculated the imbalance in discordant tree topologies using the D-statistic and evaluated significance based on 1,000 bootstrap replicates and whether the bootstrapped 95% CI overlapped zero.

We then tested the null hypothesis that the observed introgression across contact zones represents one historical introgression event in which introgressed ancestry spread across the region of sympatry or independent occurrences of introgression at each contact zone (or some combination of the two). To do so, we used $\binom{6}{2} = 15$ ABBA–BABA tests, each of which consisted of outcrosser and selfer samples from two contact zones. We tested for an excess of discordant trees that grouped populations by contact zone site (*SI Appendix*, Fig. S3, tree cartoon, A-C).

We investigated patterns of relatedness among chloroplast sequences of the two taxa to determine whether there was a consistent direction of gene flow occurring between the taxa. For example, if one taxon always acted as the seed parent in the F1 cross and backcrossing consistently occurred though pollen flow of the other taxon, we would expect to find the donor taxon chloroplast in the recipient taxon. We extracted the chloroplast genome, using the coordinates identified when blasting our chloroplast assembly to the reference genome, from our WGS dataset. We removed the 13 individuals that had high missing data (as above), removed LowQuality sites from the FILTER column of the VCF, turned genotypes with sample depth < 10 to missing, removed sites with >20% missing data for either taxon and subsetted to only biallelic SNPs. We converted our filtered VCF to a fasta (vcf2phylip.py script from https://github.com/edgardomortiz/vcf2phylip)

and inferred the phylogenetic relationships among taxa with RAxML version 8.2.11 (98) using the GTR model and rapid bootstrapping. We used the four congeneric *Clarkia* species (*SI Appendix*, Table S1) as outgroups.

Quantifying Admixture Proportions across Contact Zone Sites. In addition to testing for the presence and independence of introgression, we quantified the extent to which introgression was occurring across the region of sympatry. We used a HMM to infer ancestry across the genome and to estimate the genome-wide admixture proportion for each individual (38). These models use a reference panel containing allele frequencies at every site for each taxon and a transition matrix based on recombination rates to estimate posterior distributions of three ancestry genotypes at each site: Homozygous for the conspecific ancestry, heterozygous for both ancestry types and homozygous for the heterospecific ancestry. To construct reference panels for each taxon, we selected individuals from three allopatric populations of each taxon that were spatially proximate to the contact zone. We used FAY, KV, and WB for the selfer and CF, DLE, and KYE for the outcrosser. We ran the ancestry HMM on each taxon/contact zone combination separately, for both taxa in the six contact zones, using a SNP matrix of fourfold synonymous coding region sites.

To parameterize the emission probabilities, the HMM requires an estimate of global (population-level) admixture proportions in the sample. As we did not know this value a priori, we initially assumed it to be 10% and then used the Baum–Welch expectation-maximization algorithm to update this estimate, as suggested by Corbett-Detig and Nielsen (38). We reran the HMM until the input and output admixture proportions differed by less than 0.001. Because we do not have a high-resolution genetic map, we used a uniform recombination rate of 1×10^{-8} bp per generation to infer genetic distances among markers.

At the end of the final HMM run for each taxon/contact zone combination, we calculated individual admixture proportions in two ways. First, we used only sites that had high confidence ancestry calls (posterior probability of ancestry > 0.8). Second, we used all sites, with ancestry calls weighted by their posterior probabilities. These estimates are highly correlated ($r = 0.9$; *SI Appendix*, Fig. S7), and we present the estimates based on all sites in the main text.

Because we found introgression into the selfer with the HMM local ancestry analysis but not with genome-wide ABBA-BABA analyses, we confirmed that the HMM-inferred introgression rather than unsorted ancestral polymorphism using D_{xy}-based logic (*SI Appendix, SI Text 3*). We compared our results from the HMM with two additional estimates of the admixture proportion–the f_d statistic (41) and a D_{xy}-based calculation (*SI Appendix, SI Text 4*). Because all estimates were strongly correlated (*SI Appendix*, Figs. S6 and S7), we only discuss results from the HMM in the main text.

Explaining Variation in Admixture Proportion among Populations. We found substantial variation in the magnitude of introgression into the outcrosser across contact zones. Guided by these results, we tested whether the outcrosser's admixture proportion across the six contact zones, inferred from the HMM, was correlated with a set of eight environmental variables. These environmental variables were collected over 15 y at each contact zone except S7 (37). For S7, we used weather station data from a spatially proximate (1.1 km) outcrosser population with a similar slope and aspect, BR. In our models, we included mean and variance of temperature and precipitation in the winter (November to January) and spring (February to May), because plants primarily germinate in winter and primarily grow and flower in spring. Prior demographic studies have also shown that environmental variance can be pivotal in determining population dynamics (37). To overcome the challenges of modeling fewer observed responses ($n = 6$) than potential explanatory variables ($n = 8$), we used a LOO validation method to select an appropriate model. The LOO validation compares eight regression models (one for each environmental variable) to a model with just the intercept. For each environmental variable, we built six linear models, each one "holding out" one data point. We then summarized model performance for a given environmental variable as the square root of the mean squared difference between the model's prediction for the missing data point and its observed value (i.e., the rmse). An rmse exceeding that of the "intercept-only" model indicates that the environmental variable predicts the admixture proportion. We report *P*-values from both a simple linear regression of our best model and from a permutation comparing the observed rmse from LOO validation to its distribution under the 720 ways we could permute the order of the six observations.

We found that only one variable–variance in spring precipitation (*Results*)–significantly explained variance in the outcrosser's admixture proportion. We estimated the effect of this environmental variable on admixture proportions by finding the standardized regression coefficient (i.e., regression coefficient when X and Y are Z-transformed), which can take values from −1 to 1. By convention, values between 0.3 and 0.5 are considered modest effects, and values greater than 0.5 are considered strong effects.

Finally, we specifically asked how much of the variance in admixture proportions across contact zones for both taxa was explained by variance in spring precipitation and mating system (i.e., outcrosser vs. selfer). We ran a linear model on the HMM-inferred average population admixture proportion with mating system and variance in spring precipitation as interacting fixed effects.

Data, Materials, and Software Availability. Reference genome raw sequences, assembly, and whole-genome resentencing data have been deposited in NCBI (PRJNA906640) (99). Processed genomic data, environmental data, and data analysis code have been deposited in figshare (https://doi.org/10.6084/m9.figshare.23961792.v1) (100).

ACKNOWLEDGMENTS. We thank V. Eckhart for collecting the weather station data, A. Raduski for help with genome assembly and genotype calling, and Z. Radford and M. Geber for help with seed collections in the field. We thank the University of Minnesota Genomics Center for their guidance and for performing DNA extraction, library preparation, and sequencing. The Minnesota Supercomputing Institute at the University of Minnesota provided resources that contributed to the research results reported within this article. Funding for this project was provided by the NSF award DEB-1754246 to D.A.M. and Y.B.

1. B. M. Moran *et al.*, The genomic consequences of hybridization. *Elife* **10**, e69016 (2021).
2. A. Le Moan, P.-A. Gagnaire, F. Bonhomme, Parallel genetic divergence among coastal–marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. *Mol. Ecol.* **25**, 3187–3202 (2016).
3. Y. Y. Yamasaki *et al.*, Genome-wide patterns of divergence and introgression after secondary contact between Pungitius sticklebacks. *Philos. Trans. R. Soc. B* **375**, 20190548 (2020).
4. D. R. Schield *et al.*, Insight into the roles of selection in speciation from genomic patterns of divergence and introgression in secondary contact in venomous rattlesnakes. *Ecol. Evol.* **7**, 3951–3966 (2017).
5. E. G. Mandeville, T. L. Parchman, D. B. McDonald, C. A. Buerkle, Highly variable reproductive isolation among pairs of Catostomus species. *Mol. Ecol.* **24**, 1856–1872 (2015).
6. A. L. Lewanski, J. Golcher-Benavides, J. A. Rick, C. E. Wagner, Variable hybridization between two Lake Tanganyikan cichlid species in recent secondary contact. *Mol. Ecol.* **31**, 5041–5059 (2022).
7. R. G. Harrison, E. L. Larson, Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Mol. Ecol.* **25**, 2454–2466 (2016).
8. L. Zhang, G. R. Hood, I. Carroo, J. R. Ott, S. P. Egan, Context-dependent reproductive isolation: Host plant variability drives fitness of hybrid herbivores. *Am. Nat.* **197**, 732–739 (2021).
9. S. A. Sianta, K. M. Kay, Parallel evolution of phenological isolation across the speciation continuum in serpentine-adapted annual wildflowers. *Proc. Biol. Sci.* **288**, 20203076 (2021).
10. T. Hatfield, D. Schluter, Ecological speciation in sticklebacks: Environment-dependent hybrid fitness. *Evolution* **53**, 866–873 (1999).
11. G. M. Hewitt, Hybrid zones-natural laboratories for evolutionary studies. *Trends Ecol. Evol.* **3**, 158–167 (1988).
12. N. H. Barton, G. M. Hewitt, Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.* **16**, 113–148 (1985).
13. Z. Gompert, E. G. Mandeville, C. A. Buerkle, Analysis of population genomic data from hybrid zones. *Annu. Rev. Ecol. Evol. Syst.* **48**, 207–229 (2017).
14. H. E. A. MacGregor, G. M. While, J. Barrett, Experimental contact zones reveal causes and targets of sexual selection in hybridizing lizards. *Funct. Ecol.* **31**, 742–752 (2017).
15. T. J. Parsons, S. L. Olson, M. J. Braun, Unidirectional spread of secondary sexual plumage traits across an avian hybrid zone. *Science* **260**, 1643–1646 (1993).
16. L. V. Clark *et al.*, Genetic structure of Miscanthus sinensis and Miscanthus sacchariflorus in Japan indicates a gradient of bidirectional but asymmetric introgression. *J. Exp. Bot.* **66**, 4213–4225 (2015).
17. K. S. Burgess, M. Morgan, L. Deverno, B. C. Husband, Asymmetrical introgression between two Morus species (M. alba, M. rubra) that differ in abundance. *Mol. Ecol.* **14**, 3471–3483 (2005).
18. J. E. Zalapa, J. Brunet, R. P. Guries, Patterns of hybridization and introgression between invasive Ulmus pumila (Ulmaceae) and native U. rubra. *Am. J. Bot.* **96**, 1116–1128 (2009).
19. L. H. Rieseberg, H. C. Choi, D. Ham, Differential cytoplasmic versus nuclear introgression in helianthus. *J. Hered.* **82**, 489–493 (1991).
20. A. L. Sweigart, J. H. Willis, Patterns of nucleotide diversity in two species of Mimulus are affected by mating system and asymmetric introgression. *Evolution* **57**, 2490–2506 (2003).
21. L. E. Wallace *et al.*, Asymmetrical gene flow in a hybrid zone of Hawaiian Schiedea (Caryophyllaceae) species with contrasting mating systems. *PLoS One* **6**, e24845 (2011).
22. V. Ducarme, J. Vrancken, R. A. Wesselingh, Hybridization in annual plants: Patterns and dynamicsduring a four-year study in mixed rhinanthus populations. *Folia Geobot.* **45**, 387–405 (2010).

23. C. Palma-Silva *et al.*, Sympatric bromeliad species (Pitcairnia spp.) facilitate tests of mechanisms involved in species cohesion and reproductive isolation in Neotropical inselbergs. *Mol. Ecol.* **20**, 3185–3201 (2011).
24. Y. Brandvain, A. M. Kenney, L. Flagel, G. Coop, A. L. Sweigart, Speciation and introgression between Mimulus nasutus and Mimulus guttatus. *PLoS Genet.* **10**, e1004410 (2014).
25. J. B. Pettengill, D. A. Moeller, Tempo and mode of mating system evolution between incipient Clarkia species. *Evolution* **66**, 1210–1225 (2012).
26. J. L. Rifkin, A. S. Castillo, I. T. Liao, M. D. Rausher, Gene flow, divergent selection and resistance to introgression in two species of morning glories (Ipomoea). *Mol. Ecol.* **28**, 1709–1729 (2019).
27. M. Ruhsam, P. M. Hollingsworth, R. A. Ennos, Early evolution in a hybrid swarm between outcrossing and selfing lineages in Geum. *Heredity* **107**, 246–255 (2011).
28. G. L. Stebbins, *Flowering Plants: Evolution Above the Species Level* (Harvard University Press, 1974).
29. D. Grossenbacher, R. D. Briscoe Runquist, E. E. Goldberg, Y. Brandvain, No association between plant mating system and geographic range overlap. *Am. J. Bot.* **103**, 110–117 (2016).
30. N. H. Martin, J. H. Willis, Ecological divergence associated with mating system causes nearly complete reproductive isolation between sympatric Mimulus species. *Evolution* **61**, 68–82 (2007).
31. R. D. Briscoe Runquist, E. Chu, J. L. Iverson, J. C. Kopp, D. A. Moeller, Rapid evolution of reproductive isolation between incipient outcrossing and selfing Clarkia species. *Evolution* **68**, 2885–2900 (2014).
32. K. Christie, L. S. Fraser, D. B. Lowry, The strength of reproductive isolating barriers in seed plants: Insights from studies quantifying premating and postmating reproductive barriers over the past 15 years. *Evolution* **76**, 2228–2243 (2022).
33. H. Lewis, P. H. Raven, New combinations in the genus Clarkia (Onagraceae). *Madroño* **39**, 163–169 (1992).
34. V. M. Eckhart, M. A. Geber, Character variation and geographic distribution of Clarkia xantiana A. Gray (Onagraceae): Flowers and phenology distinguish two subspecies. *Madroño* **46**, 117–125 (1999).
35. D. M. Moore, H. Lewis, The evolution of self-pollination in Clarkia xantiana. *Evolution* **19**, 104–114 (1965).
36. J. B. Pettengill, D. A. Moeller, Phylogeography of speciation: Allopatric divergence and secondary contact between outcrossing and selfing *Clarkia*. *Mol. Ecol.* **21**, 4578–4592 (2012).
37. J. B. Pettengill, R. D. Briscoe Runquist, D. A. Moeller, Mating system divergence affects the distribution of sequence diversity within and among populations of recently diverged subspecies of Clarkia xantiana (Onagraceae). *Am. J. Bot.* **103**, 99–109 (2016).
38. V. M. Eckhart *et al.*, The geography of demography: Long-term demographic studies and species distribution models reveal a species border limited by adaptation. *Am. Nat.* **178**, S26–S43 (2011).
39. R. R. Hudson, M. Kreitman, M. Aguadé, A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159 (1987).
40. R. Corbett-Detig, R. Nielsen, A Hidden Markov Model approach for simultaneously estimating local ancestry and admixture time using next generation sequence data in samples of arbitrary ploidy. *PLoS Genet.* **13**, e1006529 (2017).
41. S. H. Martin, J. W. Davey, C. D. Jiggins, Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244–257 (2015).
42. G. Aldridge, Variation in frequency of hybrids and spatial structure among Ipomopsis (Polemoniaceae) contact sites. *New Phytol.* **167**, 279–288 (2005).
43. O. Lepais *et al.*, Species relative abundance and direction of introgression in oaks. *Mol. Ecol.* **18**, 2228–2242 (2009).
44. L. Natola, S. M. Billerman, M. D. Carling, S. S. Seneviratne, D. Irwin, Geographic variability of hybridization between red-breasted and red-naped sapsuckers. *Evolution* **77**, 580–592 (2023).
45. P. Zieliński *et al.*, Differential introgression across newt hybrid zones: Evidence from replicated transects. *Mol. Ecol.* **28**, 4811–4824 (2019).
46. K. A. Thompson *et al.*, Analysis of ancestry heterozygosity suggests that hybrid incompatibilities in threespine stickleback are environment dependent. *PLoS Biol.* **20**, e3001469 (2022).
47. K. B. Bartlett, M. W. Austin, J. B. Beck, A. E. Zanne, A. B. Smith, Beyond the usual climate? Factors determining flowering and fruiting phenology across a genus over 117 years *Am. J. Bot.* **110**, e16188 (2023).
48. M. A. Zettlemoyer, K. Renaldi, M. D. Muzyka, J. A. Lau, Extirpated prairie species demonstrate more variable phenological responses to warming than extant congeners. *Am. J. Bot.* **108**, 958–970 (2021).
49. H. Ganjurjav *et al.*, Warming and precipitation addition interact to affect plant spring phenology in alpine meadows on the central Qinghai-Tibetan Plateau. *Agric. For. Meteorol.* **287**, 107943 (2020).
50. E. R. Matthews, S. J. Mazer, Historical changes in flowering phenology are governed by temperature × precipitation interactions in a widespread perennial herb in western North America. *New Phytol.* **210**, 157–167 (2016).
51. S. J. Mazer, L. S. Dudley, A. A. Hove, S. K. Emms, A. S. Verhoeven, Physiological performance in Clarkia sister taxa with contrasting mating systems: Do early-flowering autogamous taxa avoid water stress relative to their pollinator-dependent counterparts? *Int. J. Plant Sci.* **171**, 1029–1047 (2010).
52. T. E. Burnette, V. M. Eckhart, Evolutionary divergence of potential drought adaptations between two subspecies of an annual plant: Are trait combinations facilitated, independent, or constrained? *Am. J. Bot.* **108**, 309–319 (2021).
53. T. Ramirez-Parada *et al.*, Plasticity and not adaptation is the primary source of temperature-mediated variation in flowering phenology in North America. *Nat. Ecol. Evol.*, 10.1038/s41559-023-02304-5 (2024).
54. V. M. Eckhart, M. A. Geber, C. M. McGuire, Experimental studies of adaptation in Clarkia xantiana. I. Sources of trait variation across a subspecies border. *Evolution* **58**, 59–70 (2004).
55. J. W. Benning, A. Faulkner, D. A. Moeller, Rapid evolution during climate change: Demographic and genetic constraints on adaptation to severe drought. *Proc. Biol. Sci.* **290**, 20230336 (2023).
56. M. Pickup *et al.*, Mating system variation in hybrid zones: Facilitation, barriers and asymmetries to gene flow. *New Phytol.* **224**, 1035–1047 (2019).
57. S. C. H. Barrett, R. Arunkumar, S. I. Wright, The demography and population genomics of evolutionary transitions to self-fertilization in plants. *Philos. Trans. R Soc. B* **369**, 20130344 (2014).
58. R. W. Cruden, Pollen-ovule ratios: A conservative indicator of breeding systems in flowering plants. *Evolution* **31**, 32–46 (1977).
59. Å. Lankinen, H. G. Smith, S. Andersson, J. A. Madjidian, Selection on pollen and pistil traits during pollen competition is affected by both sexual conflict and mixed mating in a self-compatible herb. *Am. J. Bot.* **103**, 541–552 (2016).
60. C. R. Darwin, *The Effects of Cross and Self Fertilisation in the Vegetable Kingdom* (J. Murray, 1876).
61. J. H. Williams, S. J. Mazer, Pollen–tiny and ephemeral but not forgotten: New ideas on their ecology and evolution. *Am. J. Bot.* **103**, 365–374 (2016).
62. M. Ruhsam, P. M. Hollingsworth, R. A. Ennos, Patterns of mating, generation of diversity, and fitness of offspring in a Geum hybrid swarm. *Evolution* **67**, 2728–2740 (2013).
63. A. M. Kenney, A. L. Sweigart, Reproductive isolation and introgression between sympatric Mimulus species. *Mol. Ecol.* **25**, 2499–2517 (2016).
64. L. E. Wallace, Spatial genetic structure and frequency of interspecific hybridization in Platanthera aquilonis and P. dilatata (Orchidaceae) occurring in sympatry. *Am. J. Bot.* **93**, 1001–1009 (2006).
65. D. L. Grossenbacher, S. D. Veloz, J. P. Sexton, Niche and range size patterns suggest that speciation begins in small, ecologically diverged populations in North American monkeyflowers (Mimulus spp.). *Evolution* **68**, 1270–1280 (2014).
66. D. J. Crawford, Progenitor-derivative species pairs and plant speciation. *Taxon* **59**, 1413–1423 (2010).
67. L. D. Gottlieb, Rethinking classic examples of recent speciation in plants. *New Phytol.* **161**, 71–82 (2004).
68. D. A. Moeller, Geographic structure of pollinator communities, reproductive assurance, and the evolution of self-pollination. *Ecology* **87**, 1510–1522 (2006).
69. C. J. Runions, M. A. Geber, Evolution of the self-pollinating flower in Clarkia xantiana (Onagraceae). I. Size and development of floral organs. *Am. J. Bot.* **87**, 1439–1451 (2000).
70. D. M. Hillis, E. A. Chambers, T. J. Devitt, Contemporary methods and evidence for species delimitation. *Ichthyol. Herpetol.* **109**, 895–903 (2021).
71. Y. Brandvain, T. Slotte, K. M. Hazzouri, S. I. Wright, G. Coop, Genomic identification of founding haplotypes reveals the history of the selfing species Capsella rubella. *PLoS Genet.* **9**, e1003754 (2013).
72. Y.-L. Guo *et al.*, Recent speciation of Capsella rubella from Capsella grandiflora, associated with loss of self-incompatibility and an extreme bottleneck. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 5246–5251 (2009).
73. A. Salcedo, S. Kalisz, S. I. Wright, Limited genomic consequences of mixed mating in the recently derived sister species pair, Collinsia concolor and Collinsia parryi. *J. Evol. Biol.* **27**, 1400–1412 (2014).
74. Q. K. Langdon *et al.*, Genome evolution is surprisingly predictable after initial hybridization. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.12.21.572897 (Accessed 7 January 2024).
75. A. M. Kearns *et al.*, Genomic evidence of speciation reversal in ravens. *Nat. Commun.* **9**, 906 (2018).
76. O. Seehausen, G. Takimoto, D. Roy, J. Jokela, Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Mol. Ecol.* **17**, 30–44 (2008).
77. S. A. Taylor, E. L. Larson, R. G. Harrison, Hybrid zones: Windows on climate change. *Trends Ecol. Evol.* **30**, 398–406 (2015).
78. K. C. Grabenstein, S. A. Taylor, Breaking barriers: Causes, consequences, and experimental utility of human-mediated hybridization. *Trends Ecol. Evol.* **33**, 198–212 (2018).
79. D. L. Powell, A. D. Rose, G. G. Rosenthal, A widely-used pollutant causes reversal of conspecific mate preference in a freshwater fish. bioRxiv [Preprint] (2022). https://doi.org/10.1101/2022.09.07.507014 (Accessed 9 August 2023).
80. S. E. Carney, K. A. Gardner, L. H. Rieseberg, Evolutionary changes over the fifty-year history of a hybrid population of sunflowers (Helianthus). *Evolution* **54**, 462–474 (2000).
81. B. B. Lamont, T. He, N. J. Enright, S. L. Krauss, B. P. Miller, Anthropogenic disturbance promotes hybridization between Banksia species by altering their biology. *J. Evol. Biol.* **16**, 551–557 (2003).
82. O. Seehausen *et al.*, Speciation through sensory drive in cichlid fish. *Nature* **455**, 620–626 (2008).
83. E. B. Taylor *et al.*, Speciation in reverse: Morphological and genetic evidence of the collapse of a three-spined stickleback (Gasterosteus aculeatus) species pair. *Mol. Ecol.* **15**, 343–355 (2005).
84. S. J. Franks, A. E. Weis, Climate change alters reproductive isolation and potential gene flow in an annual plant. *Evol. Appl.* **2**, 481–488 (2009).
85. S. Rivest, B. D. Inouye, J. R. K. Forrest, Warmer springs increase potential for temporal reproductive isolation among habitat patches in subalpine flowering plants. *J. Ecol.* **111**, 2257–2268 (2023).
86. E. M. Oziolor *et al.*, Adaptive introgression enables evolutionary rescue from extreme environmental pollution. *Science* **364**, 455–457 (2019).
87. C. J. Brauer *et al.*, Natural hybridization reduces vulnerability to climate change. *Nat. Clim. Chang.* **13**, 282–289 (2023).
88. D. A. Moeller, M. A. Geber, V. M. Eckhart, P. Tiffin, Reduced pollinator service and elevated pollen limitation at the geographic range limit of an annual plant. *Ecology* **93**, 1036–1048 (2012).
89. H. Lewis, M. E. Lewis, The genus Clarkia. *Univ. Calif. Publ. Bot.* **20**, 241–392 (1955).
90. J. M. Palmer, Funannotate: Pipeline for genome annotation (2017).
91. S. Koren *et al.*, Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
92. S. Nurk *et al.*, HiCanu: Accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res.* **30**, 1291–1305 (2020).
93. V. Buffalo, Scythe–A Bayesian Adapter Trimmer (v. 0.994 BETA, Github, 2011). https://github.com/vsbuffalo/scythe. Accessed 8 October 2020.
94. P. Danecek *et al.*, Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
95. G. A. Van der Auwera *et al.*, From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinf.* **43**, 11.10.1–11.10.33 (2013).
96. K. L. Korunes, K. Samuk, pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol. Ecol. Resour.* **21**, 1359–1368 (2021).
97. E. Y. Durand, N. Patterson, D. Reich, M. Slatkin, Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011).
98. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
99. S. A. Sianta, D. A. Moeller, Y. Brandvain, Clarkia xantiana reference genome and population genomics. NCBI. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA906640. Deposited 16 August 2023.
100. S. A. Sianta, D. A. Moeller, Y. Brandvain, Data for "The extent of introgression between incipient Clarkia species is determined by temporal environmental variation and mating system." figshare. https://doi.org/10.6084/m9.figshare.23961792.v1. Deposited 15 August 2023.