Received: 8 January 2018 | Revised: 3 September 2018 | Accepted: 1 November 2018

DOI: 10.1111/mec.14945

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Gene flow, divergent selection and resistance to introgression in two species of morning glories (*Ipomoea*)

Joanna L. Rifkin | Allan S. Castillo | Irene T. Liao | Mark D. Rausher

Department of Biology, Duke University, Durham, North Carolina

Correspondence
Joanna L. Rifkin, Department of Biology, Duke University, Durham, NC.
Email: joanna.rifkin@alumni.duke.edu

Present Address
Joanna L. Rifkin, Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada

Funding information
NSF, Grant/Award Number: DEB 1542387; NSF DDIG, Grant/Award Number: DEB 1501954

Abstract

Gene flow is thought to impede genetic divergence and speciation by homogenizing genomes. Recent theory and research suggest that sufficiently strong divergent selection can overpower gene flow, leading to loci that are highly differentiated compared to others. However, there are also alternative explanations for this pattern. Independent evidence that loci in highly differentiated regions are under divergent selection would allow these explanations to be distinguished, but such evidence is scarce. Here, we present multiple lines of evidence that many of the highly divergent SNPs in a pair of sister morning glory species, *Ipomoea cordatotriloba* and *I. lacunosa*, are the result of divergent selection in the face of gene flow. We analysed a SNP data set across the genome to assess the amount of gene flow, resistance to introgression and patterns of selection on loci resistant to introgression. We show that differentiation between the two species is much lower in sympatry than in allopatry, consistent with interspecific gene flow in sympatry. Gene flow appears to be substantially greater from *I. lacunosa* to *I. cordatotriloba* than in the reverse direction, resulting in sympatric and allopatric *I. cordatotriloba* being substantially more different than sympatric and allopatric *I. lacunosa*. Many SNPs highly differentiated in allopatry have experienced divergent selection, and, despite gene flow in sympatry, resist homogenization in sympatry. Finally, five out of eight floral and inflorescence characteristics measured exhibit asymmetric convergence in sympatry. Consistent with the pattern of gene flow, *I. cordatotriloba* traits become much more like those of *I. lacunosa* than the reverse. Our investigation reveals the complex interplay between selection and gene flow that can occur during the early stages of speciation.

KEYWORDS
divergent selection, gene flow, introgression, *Ipomoea*, morning glory, speciation

1 | INTRODUCTION

Gene flow is generally believed to impede genetic divergence and speciation. The tension between these forces has led to recent interest in the extent to which speciation can proceed in the face of gene flow (Feder, Egan, & Nosil, 2012; Martin et al., 2013; Nosil, 2008; Papadopulos et al., 2011). While allele frequencies at neutral loci are expected to be homogenized between populations with only a small amount of gene flow (on the order of one successful migrant individual per generation; Wright, 1969), greater gene flow is needed to counteract divergent selection on individual loci. When divergent selection is intense, gene flow will not prevent allele-frequency divergence, but can prevent alternate alleles from fixing in different populations. In particular, allele-frequency divergence will reach an equilibrium reflecting a balance of selection tending to increase divergence and gene flow tending to reduce it. If selection is strong and there is little (but nonzero) gene flow, this equilibrium will be manifested by large differences in allele frequencies. By contrast, if

selection is weak and there is substantial gene flow, allele-frequency differences may be small or absent (Bulmer, 1972; Haldane, 1930; Wright, 1931).

These considerations suggest that incipient species connected by gene flow should exhibit substantial variation in divergence across the genome, with genetic divergence maintained at loci subject to divergent selection but homogenization at neutral loci (Noor, Grams, Bertucci, & Reiland, 2001; Noor, Grams, Bertucci, & Reiland, et al., 2001; Nosil et al., 2009; Rieseberg, 2001; Wu & Ting, 2004). There is considerable evidence to support this suggestion. For example, studies in hybrid zones have commonly found variation in the degree to which different loci introgress (Larson, White, Ross, & Harrison, 2014; Maroja, Andrés, Harrison, 2009; Payseur, Krenz, & Nachman, 2004; Teeter et al., 2008), a pattern which has been interpreted as due to variation in the strength of divergent selection opposing gene flow. Many studies have also used genome scans to document the spatial pattern of variation in divergence and have found genomic regions that show substantial divergence compared to the remainder of the genome (Carneiro et al., 2014; Ellegren et al., 2012; Gagnaire, Normandeay, Pavey, & Bernatchez, 2013; Harr, 2006; Hohenlohe, Bassham, Currey, & Cresko, 2012; Hohenlohe et al., 2010; Nadeau et al., 2012; White, Cheng, Simard, Costantini, & Besansky, 2010). However, this pattern is not universal (Garrigan et al., 2012; Kenney & Sweigart, 2016; Michel et al., 2010). Theory predicts that such highly divergent regions result from strong divergent selection on specific loci, which effectively shields closely linked loci from gene flow (Charlesworth, Nordborg, & Charlesworth, 1997; Feder & Nosil, 2010; Via, Conte, Mason-Foley, & Mills, 2012), while the remainder of the genome that is not subject to divergent selection is homogenized by gene flow.

Although variation in divergence is consistent with variation in the strength of divergent selection, other explanations are possible. One is that at least some of this variation is due to chance effects. Such effects may be especially likely if one or both incipient species have gone through recent bottlenecks, possibly associated with speciation. Studies that attribute divergent selection to outlier loci are particularly vulnerable to this criticism (Bierne, Roze, & Welch, 2013). A second possibility is that regions of high divergence may result from a combination of background selection and low rates of recombination, which results in reduced within-species variation in regions of low recombination, and thus inflation of divergence measures such as Fst that are commonly used to quantify divergence (Charlesworth, Charlesworth, & Morgan, 1995; Cruickshank & Hahn, 2014; Noor & Bennett, 2009). Finally, a similar effect can result from global fixation of an adaptive allele, the opposite of divergent selection (Bierne, 2010).

Ideally, inferences that regions of high divergence are due to divergent selection would be supported by independent evidence that loci in those regions are subject to such selection. However, only a limited amount of direct evidence of this type exists. Some studies have reported that QTLs suspected to be subject to divergent selection coincide with genomic regions of elevated divergence (Carling & Brumfield, 2009; Janoušek et al., 2012; Kenney & Sweigart, 2016; Via et al., 2012), although a few studies have failed to observe such coincidence (Eckert et al., 2010; Yatabe, Kane, Scotti-Saintagne, & Rieseberg, 2007). Other studies have documented a negative correlation between absolute divergence and recombination rate, pointing to variation across the genome in the strength of purifying selection against introgressing loci (Brandvain, Kenney, Flagel, Coop, & Sweigart, 2014; Kenney & Sweigart, 2016).

One type of evidence that would strongly support the hypothesis that highly divergent loci are subject to strong divergent selection would be a strong correlation between divergence in allopatry and resistance to introgression in sympatry (Harrison & Larson, 2016; Noor & Bennett, 2009). However, investigations that have conducted this type of analysis find only partial or weak correspondence of this type (Gompert et al., 2012; Hamilton, Lexer, & Aitken, 2013; Larson, Andrés, Bogdanowicz, & Harrison, 2013; Parchman et al., 2013; Taylor, Curry, White, Ferretti, & Lovette, 2014). While evidence suggests that some regions of elevated divergence are likely to reflect strong divergent selection, this may not be true generally. More information is thus needed on the extent to which regions of elevated differentiation in incipient species experiencing gene flow correspond to regions experiencing strong divergent selection.

Gene flow is expected to affect phenotypic divergence between species. For characters that have diverged in allopatry by genetic drift, the underlying loci are expected to be largely neutral and not resistant to homogenization. Consequently, gene flow upon secondary contact is expected to act to reduce differentiation in these types of characters and, over long periods, even eliminate phenotypic differences. Whether the change in the character is symmetric (equal in both directions) or asymmetric (one species changes more than the other) will be influenced by any asymmetry in gene flow. It may also be influenced by genetic architecture. For example, if there is directional dominance in the character, change is expected to be greater in the species with the recessive alleles. By contrast, for characters that have diverged due to selection, gene flow may reduce character differences compared to what they would be in the absence of gene flow, but not eliminate those differences, as long as selection in sympatry is similar to selection in allopatry. Despite these expectations, we are unaware of any previous studies that have tested them.

In this report, we investigate the relationship between divergent selection and genetic differentiation by examining gene flow, resistance to introgression and patterns of selection on loci that have diverged in frequency between two sister species of morning glory, *Ipomoea cordatotriloba* and *I. lacunosa*. To characterize patterns of divergence, gene flow and introgression, we compare allele-frequency differences between allopatric and sympatric samples of the two species at single nucleotide polymorphisms (SNPs) from transcriptome sequence data and ask whether SNPs that are highly differentiated in allopatry are also resistant to homogenization in sympatry. We then use an extension of the standard McDonald–Kreitman (M-K) test (McDonald & Kreitman, 1991) to ascertain the extent to which divergent selection contributes to highly divergent SNPs. In addition, we characterize phenotypic divergence in allopatry and

sympatry for floral and inflorescence traits to determine the extent to which gene flow reduces character divergence in sympatry compared with allopatry.

2 | METHODS

2.1 | Study system

Ipomoea cordatotriloba and *I. lacunosa* are annual weeds in the Batatas section of the genus *Ipomoea* (Convolvulaceae) and have overlapping ranges (Figure 1). *Ipomoea lacunosa*, which is highly selfing (Duncan & Rausher, 2013a), extends north into Canada and west to Texas, while *I. cordatotriloba*, a mixed-mater with an average selfing rate of 0.5 (Duncan & Rausher, 2013a), is found from North Carolina to Mexico (USDA, NRCS 2019; Figure 1). The geographical extent of range overlap is large, spanning across the Southeast from Texas to North Carolina. At some localities within the range of overlap, the two species occur sympatrically, with individuals of the two species often growing intertwined. At other sites within the range of overlap, only one of the species occurs. This geographical arrangement allows us to compare genetic differences between allopatric samples to determine the effects of gene flow on divergence.

In keeping with its high selfing rate, *I. lacunosa* exhibits many characteristics of the "selfing syndrome" (Ornduff, 1969; Sicard & Lenhard, 2011) compared with *I. cordatotriloba*, including smaller flowers with less nectar and pollen (McDonald, Hansen, McDill, & Simpson, 2011; Duncan & Rausher, 2013a; Rifkin, 2017; also see below). Vegetatively, the two species appear very similar. Both species are visited primarily by bumblebees. In areas of sympatry, we have observed movements of individual bees between species, suggesting the possibility of pollen transfer. The two species exhibit partial cross-incompatibility in both directions (Duncan & Rausher, 2013b). In addition, there is at least some postzygotic incompatibility (Rifkin, unpublished data). Both of these phenomena would tend to restrict, but not eliminate, gene flow between species in sympatric populations.

We collected or obtained 31 accessions of *I. lacunosa* and 30 accessions of *I. cordatotriloba*, where an accession is either an individual plant or a single seed (Supporting Information Table S1, Appendix S1). From all except one site, with four *I. lacunosa* accessions, we used no more than 3 accessions of each species, and from many sites, there was only a single accession. Accessions were categorized into one of three categories (Figure 1, Supporting Information Table S1): (a) Sympatric. These accessions were from sites where we observed the two species growing within 10 m. of each other and often intertwined. Accessions of both species were collected at these sites. (b) Allopatric. These accessions were from sites where we only observed one of the species; the other species did not grow within 1 km. Accessions were categorized as allopatric if we collected them and searched for the other species in the vicinity, or if accessions were outside the geographic range of the other species. (c) Unknown. These accessions were from inside the area of range overlap, but detailed information about the collecting site was absent (e.g., accessions obtained from USDA). Under this categorization, there were 13 sympatric, 16 allopatric and 2 unknown *I. lacunosa* accessions and 11 sympatric, 14 allopatric and 4 unknown *I. cordatotriloba* accessions (Figure 1, Supporting Information Table S1). Because of the limited number of samples per site and apparent

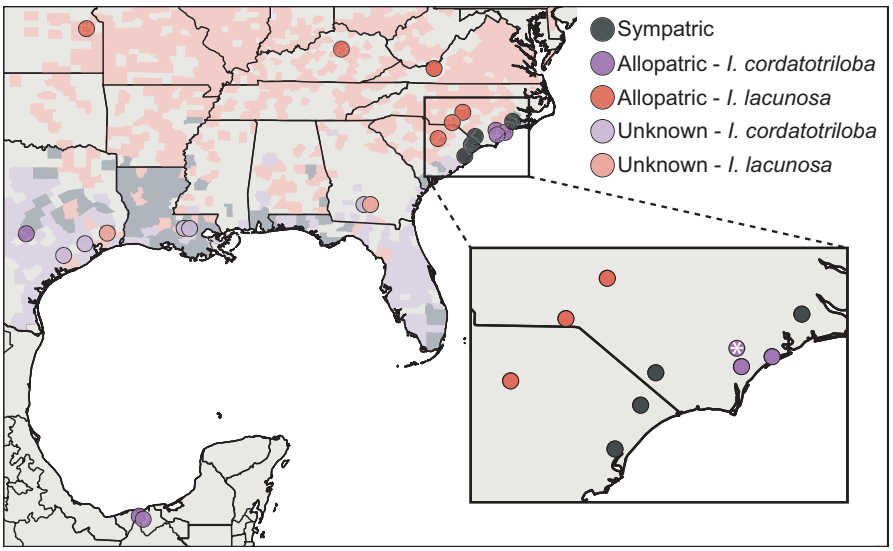

FIGURE 1 Distribution map of *Ipomoea cordatotriloba* and *Ipomoea lacunosa* by county in the southeastern United States with locations of sites used in the study. *I. cordatotriloba's* distribution extends south into Mexico, but detailed locality records are not available. *I. lacunosa* extends north into Canada, but we only used samples from as far north as Kansas. Background colors indicate range data for the two species (pink: *I. lacunosa*; lavender: *I. cordatotriloba*; gray: both species) taken from the most recent USDA NRCS distribution records (https://plants.usda.gov/core/profile?symbol=IPCO8 and https://plants.usda.gov/core/profile?symbol=IPLA). Inset indicates the sympatric and "close" allopatric sites. The asterisk (*) in inset indicates the location of the *I. cordatotriloba* SOS site. A list of all samples is found in Supporting Information Table S1

lack of genetic structure among sites within species, we perform all analyses on individual samples.

For some of our analyses, we contrast allopatric and sympatric samples. In order to account for geographic divergence within species, we do this in two ways, which differ in what samples are included in the allopatric category: (a) all samples known to be allopatric ("known allopatric samples"); (b) allopatric samples that are in the region of the sympatric sites (i.e., NC and SC, "close allopatric samples") (Figure 1, Supporting Information Table S1).

We have classified 3 accessions of *I. cordatotriloba* from the SOS site as allopatric (Figure 1, with an asterisk) even though there are indications that *I. lacunosa* may have recently been present. This population contains some *I. cordatotriloba* individuals that have white flowers like *I. lacunosa*. Unpublished evidence from our laboratory indicates that variation in a single gene is responsible for the difference in flower colour between the two species (*I. cordatotriloba* typically has purple flowers). All the white-flowered *I. cordatotriloba* individuals we have investigated carry an allele of that gene that is identical to that in *I. lacunosa*, consistent with recent introgression between the two species at this location. The sympatric populations we sampled from CCR, Site 1, CHAD and POL also contained white-flowered *I. cordatotriloba* individuals, but none of the known allopatric populations besides SOS did, a pattern consistent with introgression at SOS. However, because we cannot yet prove introgression of this gene, we conservatively classify these accessions as allopatric because they satisfy our criterion. The classification is conservative because it would tend to make sympatric accessions appear genetically more similar to allopatric accessions when testing for divergence and gene flow. However, we also performed all analyses with SOS samples classified as sympatric. These analyses yielded similar conclusions and are reported in the Supporting Information Appendix S1.

2.2 | Genome sequencing and assembly

For this investigation, we used a draft assembly of the *I. lacunosa* genome that our laboratory has produced. Here, we provide a brief description of this genome; a fuller description will be published elsewhere. The assembled genome has a total length of 431 Mb, comparable to the estimated genome size of 497 Mb determined by flow cytometry (Duncan & Rausher, 2013b). It consists of 2064 contigs, an N50 of approximately 550 Kb, and the length of the longest contig is approximately 4.7 Mb. Annotation with MAKER (Cantarel et al., 2008) yielded 32,757 unigenes composed of a total of 191,171 coding sequence features (i.e., exons). Preliminary comparisons indicate it is largely colinear with the draft genomes of *I. trifida* and *I. triloba*, two other species in the Batatas section of *Ipomoea* (http://sweetpotato.plantbiology.msu.edu/index.shtml; Wu et al., 2018).

2.3 | Transcriptome sequencing and SNP identification

One individual of each accession was grown in the Duke University Greenhouses from December 2014 to June 2016. For each accession, we extracted RNA from a single young leaf (0.5–2 cm) using a modified version of the standard TRI Reagent (Sigma-Aldrich) protocol. The modifications include an additional TRI Reagent:chloroform clean-up step, adding the suggested amount of glycogen and three ethanol washes before drying the RNA pellet. RNA was resuspended in 30 μl of RNase-free water. RNA quality was assessed using the 2200 TapeStation system (Agilent Technologies); all samples had an RNA integrity score of 7 or higher. A complete extraction protocol is available on https://github.com/joannarifkin/IpomoeaSNPCalling. We generated RNA libraries starting with 4 μg of total RNA using the KAPA Stranded mRNA-Seq Kit (KAPA Biosystems) and multiplexed using NEBNext Multiplex Oligos for Illumina (New England BioLabs). The libraries were quality checked using the Bioanalyzer Agilent High Sensitivity DNA kit (Agilent Technologies) and Qubit Fluorometer (Thermo Fisher Scientific). We pooled 20 or 21 libraries per sequencing lane; the samples were sequenced using three lanes on an Illumina HiSeq 4000 v4 platform running 150-bp paired-end reads at the Duke Sequencing & Genomic Technologies Shared Resource.

We identified SNPs using a modified version of the GATK best practices for RNAseq (Van der Auwera et al., 2013). Specifically, after aligning reads to the *I. lacunosa* genome using STAR 2-pass (Dobin & Gingeras, 2015) and cleaning the alignments with Picard Tools (http://broadinstitute.github.io/picard), we used the GATK Joint Genotyper in -erc GVCF mode to identify SNPs. SNPs were hard-filtered (Fisher Strand bias <30, quality-by-depth <2, SNP clustering) and kept only if the minimum depth was 10 reads. We eliminated all nonvariant sites and all SNPs not called in at least 60 individuals. This filtering produced 66,729 SNPs that were used in the analyses. Although the minimum depth cut-off may eliminate some singletons, and thus bias our estimates of genetic diversity downwards, it should not bias estimates of differences in measures of genetic diversity or divergence between species or sets of samples. All SNP-calling scripts are available at https://github.com/joannarifkin/IpomoeaSNPCalling.

SNPs were categorized as synonymous, nonsynonymous, noncoding or unknown by comparing SNP positions to assembled transcripts in the annotated *I. lacunosa* draft genome using an APL (Iverson, 1962) script written by MDR (scripts available in .pdf format at DRYAD Digital Depository, https://doi.org/10.5061/dryad.f6qb7c5). This categorization resulted in 27,079 synonymous SNPs and 21,746 nonsynonymous SNPs. Of the remaining SNPs, 11,281 were contained in transcripts that aligned to annotated genes, but lay outside the coding regions of those genes. These SNPs, which we designate as noncoding, are potentially regulatory in function since they occur in the 5′ untranslated or 3′ untranslated regions that could contain regulatory sequences.

2.4 | Data analysis

For PCA and population structure analyses, we conducted LD-based SNP pruning on the SNP data set to remove SNPs that were under high LD to avoid correlated SNPs from strongly influencing these analyses. We performed the pruning in the R package SNPRelate

using the snpgdsLDpruning function (Zheng et al., 2012) with a threshold of 0.3, which resulted in a pruned SNP data set of 10,173 SNPs. The snpgdsPCA function was used to perform principal component analysis. We then took a Bayesian clustering approach using STRUCTURE (Pritchard, Stephens, & Donnelly, 2000) and INSTRUCT (Gao, Williamson, & Bustamante, 2007) to find the genetic structure of the two species. INSTRUCT uses a similar algorithm to STRUCTURE, but allows equilibria to differ from Hardy–Weinberg within clusters. This is important because these populations are known to be moderately to highly selfing, which can lead to spurious results if inbreeding is not accounted for. Both STRUCTURE and INSTRUCT were run similarly, with a burn-in step of 20^5 and simulated for 10^6 iterations for clusters K = 1 to K = 10. The STRUCTURE and INSTRUCT results were visualized using CLUMPAK (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015). Results from STRUCTURE and INSTRUCT were similar. We therefore report the latter in the main text, while those from the former are reported in the online Supporting Information Appendix S1.

All other statistics were calculated using custom APL (Iverson, 1962) scripts written by MDR (scripts available in .pdf format at DRYAD Digital Depository, https://doi.org/10.5061/dryad.f6qb7c5). We calculated admixture proportions for sympatric samples of both species using a modified version of the approach described by Hanis, Chakraborty, Ferrell, and Schull (1986). In particular, the original method assumes complete outcrossing and estimates *m*, the proportion of alleles in sympatric samples that are derived from *I. cordatotriloba* allopatric samples. Our modification additionally allows for a proportion *s* of offspring to be produced by selfing. It calculates the *log*-likelihood for different combinations of *m* and *s* (each ranging between 0.01 and 0.99) and determines the values that yield the maximum *log*-likelihood (see Appendix). Sympatric samples of each species were analysed separately. Based on our finding that a substantial fraction of SNPs with allopatric frequency difference ≥0.9 ("highly divergent SNPs") were subject to divergent selection but no selection was detected on SNPs with a frequency difference <0.9 ("less-divergent SNPs"), admixture proportions were calculated for two groups of SNPs separately: those for which the allele-frequency difference in allopatry was <0.9, and those for which this difference was greater than or equal to 0.9. We excluded loci at which the same allele was fixed in allopatric samples of both species because these loci are uninformative. For comparison with sympatric samples, selfing rates were estimated for allopatric samples of each species as described in the Appendix. In all of these likelihood analyses, we assessed differences in the values of *s* and *m* among groups by comparing 99% confidence intervals between pairs of estimates. Such confidence intervals are normally calculated as the values that are 3.3 *ln*-likelihood units on either side of the mean. In our maximum-likelihood analyses, intervals between successive values of *s* and *m* were 0.1. In all cases, the *ln*-likelihoods between successive values differed by much more than 3.3 units. We therefore take (mean − 0.01, mean + 0.01) as a conservative estimate of the 99% confidence interval. In other words, if s_1 and s_2 are maximum-likelihood estimates of *s* for two groups, then we concluded the estimates were significantly different at $p < 0.01$ if $|s_1 - s_2| > 0.02$.

We calculated admixture linkage disequilibrium using Eq. (5) from Loh et al. (2013). Specifically, we calculated admixture LD, $\hat{\alpha}(d)$, for SNPs within a given distance bin $S(d)$, as

$$\hat{\alpha}(d) = \sum_{S(d)} \widehat{\mathrm{cov}(X,Y)} \left(p_A(x) - p_B(x)\right)\left(p_A(y) - p_B(y)\right) \,/\, |S(d)|$$

where $\widehat{\mathrm{cov}(X,Y)}$ is the covariance between SNPs *X* and *Y* in the admixed (sympatric) sample (either *I. lacunosa* or *I. cordatotriloba*), $p_i(j)$ is the estimated frequency of the reference allele at SNP *j* in allopatric samples of species *i*, and $|S(d)|$ is the number of SNPs in bin $S(d)$.

We employed the McDonald–Kreitman (M-K) test (McDonald & Kreitman, 1991) to determine whether divergent selection contributed to fixed or nearly fixed differences between the two species. In particular, we performed two types of M-K analyses. The first type asked whether the set of fixed or nearly fixed SNPs is enriched with nonsynonymous SNPs. We defined "fixed" SNPs as those that exhibited a frequency difference between species equal to 1, whereas "nearly fixed" SNPs were those with a frequency difference greater than or equal to 0.9 but <1. Preliminary analyses provided no evidence that any SNPs with smaller frequency differences (i.e., <0.9) were subject to selection. For each of these categories, we tabulated the number of nonsynonymous and synonymous SNPs that were fixed, nearly fixed, or polymorphic. The latter category excludes fixed or nearly fixed SNPs. Separate G tests were then performed for fixed vs. polymorphic and nearly fixed vs. polymorphic SNPs using an APL program written by MDR. We estimate the proportion of nonsynonymous SNPs fixed or nearly fixed by selection, α, using the heuristic approach of Messer and Petrov (2013). Because this analysis requires that frequencies of derived alleles be known, we polarized the alleles at our loci by mapping our SNPs to the *I. triloba* and *I. trifida* genomes. To do this, we used the whole genome aligner Progressive Cactus (https://github.com/glennhickey/progressiveCactus; Paten, Diekhans, et al., 2011; Paten, Earl, et al., 2011) and extracted the allele from the corresponding position in the other species based on the .maf alignment file. Because the species in our study are more closely related to *I. triloba* than to *I. trifida* (Muñoz-Rodríguez et al., 2018), we used the *I. triloba* allele as the ancestral allele when alleles for *I. triloba* and *I. trifida* differed, and the *I. trifida* allele when the *I. triloba* allele was unidentified. This mapping yielded ancestral alleles for approximately 55% of our SNPs, primarily because many of our SNPs did not map uniquely to either genome. However, as long as the SNPs with missing ancestral alleles are a random subset, our estimates of α should not be biased. 95% confidence intervals for α were calculated by bootstrapping (1,000 replicates) over SNPs as in Messer and Petrov (2013).

Traditionally, the M-K test has been used to determine whether nonsynonymous sites are over-represented compared to synonymous sites in sites that exhibit fixed or nearly fixed differences between species. However, as Egea, Casillas, and Barbadilla (2008) point out, the same test can be applied to different classes of sites,

such as noncoding sites. We therefore performed a second type of M-K analysis, which asked whether the set of fixed or nearly fixed SNPs is enriched with noncoding SNPs, compared to synonymous SNPs, as would occur if divergent selection acted on regulatory SNPs. These analyses were performed in exactly the same manner as described above, except the set of noncoding SNPs was substituted for the set of nonsynonymous SNPs. We interpret significant enrichment of fixed or nearly fixed SNPs with noncoding SNPs as indicative of selection. We believe this interpretation is appropriate because if all noncoding SNPs are neutral, the ratio of fixed to polymorphic noncoding SNPs should equal the ratio of fixed to polymorphic synonymous SNPs, based on an argument analogous to the justification for the standard M-K test. An excess of fixed noncoding SNPs thus implies that some non-neutral process, that is, selection, has operated.

As a measure of divergence between the two species, we used π_{LC}, the average of pairwise π values, where pairs included all combinations of samples from one species with all samples from the other species (π_{XY} of Nei & Li, 1979). For each pair, we calculated a difference value, d, for each SNP, which was 0 if the two samples were homozygous for the same allele, 1 if homozygous for different alleles, and 0.5 otherwise. These values are the probabilities that an allele drawn randomly from each sample will be different. The values for each SNP were then summed over all SNPs and divided by the total number of bp in the transcriptome (30,036,768) to yield the π value for that pair of samples. The π values were then averaged over all pairs to yield π_{LC}. Separate values of π_{LC} were calculated for sympatric and allopatric samples, and the significance of the difference between these values was determined by bootstrapping over samples. We calculated the divergence between allopatric and sympatric samples within each species, π_{AS}, in similar fashion.

We also quantified divergence between species using allele frequencies. For each SNP, we first determined the allele with the greatest frequency in the combined samples from the two species. We then calculated Δp for this allele as $p_C - p_L$, where p_i is the frequency of this allele in species *i*. To calculate the average allele-frequency difference between allopatric samples, we calculated the average absolute frequency difference over SNPs, $D = \sum_j |\Delta p_j|/n$, where n is the number of SNPs, and j represents an individual SNP. To calculate the corresponding average allele-frequency difference between sympatric samples of the two species, D, we used the formula $D = \sum_j \delta \Delta p_j/n$, where $\delta = 1$ if $\Delta p_j > 0$ in the corresponding allopatric comparison and $\delta = -1$ if $\Delta p_j < 0$ in the corresponding allopatric comparison, which allows for the difference in sympatry to be in the opposite direction from the difference in allopatry. Differences in frequencies between allopatric and sympatric sites within a species were calculated similarly. The significance of allele-frequency differences was determined by bootstrapping over samples 500 or 1,000 times depending on the analysis.

Differences in genetic composition between populations (sites) within a species can complicate comparisons of differentiation both between species and between allopatric vs. sympatric populations within species. In such circumstances, a possible strategy is to analyse only one sample per population. Unfortunately, this type of subsampling restricts the power of analyses. As will be seen below, however, there is little evidence that populations (sites) are genetically differentiated within *I. lacunosa* or within allopatric or sympatric populations in *I. cordatotriloba* in genetic composition. Consequently, in our analyses, we use all samples from each population (site).

2.5 | Simulating the effects of gene flow

We report data that suggest that while gene flow in sympatry homogenizes most less-divergent SNPs, highly divergent SNPs appear to be resistant to homogenization. We used simulations to determine whether this pattern can be explained by gene flow alone as opposed to requiring divergent selection. Our logic is as follows: If highly divergent SNPs are not subject to divergent selection in sympatry, then the effective amount of introgression should be the same for both categories of SNPs. By contrast, if highly divergent SNPs are subject to divergent selection, the magnitude of introgression should be lower for highly divergent SNPs. This simulation involved two steps (Supporting Information Figure S1). First, we created initial "newly sympatric" samples for each species by replacing the sympatric genotypes of that species with the allopatric genotypes of that species at each locus. This was meant to model the genotypes in sympatric populations upon secondary contact, that is, those genotypes should reflect the extant genotypes from allopatric sites.

Second, at each locus, we replaced a fraction *f* of genotypes (randomly chosen) in the newly sympatric *I. cordatotriloba* samples with randomly chosen genotypes from the newly sympatric *I. lacunosa* samples. This was meant to mimic the effects of one-way gene flow from *I. lacunosa* to *I. cordatotriloba* in homogenizing those loci. We employed one-way gene flow because our results indicate that gene flow from *I. cordatotriloba* to *I. lacunosa* is negligible. This second replacement involved genomic blocks rather than individual loci. The assembled genome was broken into a set of blocks of approximately 100 kb in length, informed by the extent of admixture linkage disequilibrium (see below). If a genome contig was <100 kb in length, it was considered a block. If a contig was >100 kb, it was broken into successive approximately 100-kb blocks: The first SNP in the contig was combined with all SNPs within 100 kb of it to form the first block. The next SNP not included in the first block then formed the second block, along with all other SNPs within 100 kb of it, and so forth. Preliminary analyses with different block sizes yielded similar results. For all SNPs in a block, genotypes from the same *I. lacunosa* samples were substituted for the same *I. cordatotriloba* samples.

We show below that gene flow results in a decreased between-species π in sympatry compared with allopatry. To estimate the effective amount of introgression for the two SNP classes, we calculated between-species π in sympatry for different values of ***f***. We took the value that corresponds to the observed value of π (***f****) to indicate the effective amount of introgression that occurred.

Both the observed and predicted values of π have error associated with them. We estimated the error for the observed value by bootstrapping over samples to produce a 95% credible interval.

To estimate the error in the predicted π associated with a particular value of ***f***, we ran 20 replicate simulations for that value of ***f*** and determined the 95% confidence interval.

For these analyses, we calculated π in a different way from above because there are different numbers of highly and less-divergent SNPs. In particular, we calculated π^*, which is the average value of π for all combinations of sympatric samples for the two species.

For each replicate array, we assessed how well the array matched the corresponding array constructed from the actual data. For corresponding cells in the two arrays, we calculated the squared differences in number of loci, then summed these over all cells to obtain a Sum of Squares (SS) associated with the replicate simulated array. We then averaged these SS's over all replicates for a given value of *f*. The *f* with the smallest mean SS was chosen as the appropriate *f* value for further analyses.

To determine whether loci showing fixed or nearly fixed differences (frequency difference ≥0.9) in allopatry were resistant to homogenization, we compared plots of proportion of loci with sympatric frequency differences in bins of width 0.1 for the actual and simulated data. One pair of plots was constructed for each allopatric frequency difference bin of width 0.1.

2.6 | Phenotypic divergence

We grew selfed seeds derived from the same accessions used for genotyping in a glasshouse and measured eight traits known or suspected to differ between these two species: corolla length, corolla width, cyme length (length of inflorescence from stem to flower base), herkogamy (position of anthers relative to the stigma, which determines selfing rate in these species; Duncan & Rausher, 2013a), nectar volume, nectar sugar concentration, pollen grains per ovule and flowers produced per day.

Corolla and inflorescence traits were measured using a digital calliper ("Mitutoyo Digimatic CD6″ CS). Herkogamy was measured as the number of anthers below and not touching the stigma; in highly outcrossing populations, the stigma is exserted well above the anthers whereas in selfing populations, it is nested within them (Duncan & Rausher, 2013a). To quantify nectar volume, the day before a flower opened, the bud was capped with a plastic straw covered with parafilm. The next morning, all nectar was extracted from the base of the flower with a 2-μl microcapillary tube (Drummond Scientific) and the height of the nectar in the tube was measured with the digital calliper. Because each tube is 32 mm long and holds 2 μl in total, this measurement was converted to volume with the formula V = 2 μl*(height of nectar in tube/32 mm). Nectar sugar concentration was quantified by expelling all of the nectar from the microcapillary tube onto a Master-53M ATAGO refractometer. The refractometer was standardized with water at the start of each day's measurements. Because refractometer readings are often imprecise with low volumes, two sugar concentration measurements were taken: undiluted nectar and nectar diluted with 2.5 μl water. Refractometer readings, in weight/weight (w/w) percentages, were converted into concentrations according to the recommendations of Bolten, Feinsinger, Baker, and Baker (1979) as follows: using the table "Concentrative Properties of Aqueous Solutions: Density, Refractive Index, Freezing Point Depression, and Viscosity" for sucrose solutions from Handbook of Chemistry & Physics (Rumble, 2018), sucrose solute values were converted into mg/ml by multiplying the molarity (mol/L) values by the molecular weight of sucrose (342.2964 g/mol). A plot of values between mass (w/w) and mg/ml was generated, and a polynomial line of best fit was created to convert w/w to mg/ml ($mg/ml = 0.0524(w/w)^2 + 9.6554(w/w) + 1.3904$). For nectar diluted with 2.5 μl water, the diluted sugar concentration was first converted into mg/ml and then multiplied by the ratio: (actual nectar amount + 2.5 μl)/actual nectar amount. The diluted and undiluted nectar sugar concentration values in mg/ml were averaged to produce the nectar sugar concentration used in our analyses. To quantify pollen per ovule, anthers were removed the day before anthesis, dried overnight in an open tube and resuspended in 500 μl 70% ethanol. We manually counted all pollen grains in a 100 μl aliquot from each sample under a dissecting microscope, multiplied by five (500 μl/100 μl) and divided by 4 (the number of ovules in both species; McDonald et al., 2011). Nectar and pollen measurements were taken from 1 to 3 flowers per individual and averaged by individual using the R function aggregate (R Core Team, 2016).

To test whether phenotypic differences between species were different in allopatry and sympatry, we performed a two-factor ANOVA in JMP using the "Fit model" platform. In our model, species and location (sympatry vs. allopatry) were crossed fixed effects and accession was included, nested within species, as a random effect. A significant interaction effect between species and location indicates that the difference between species was not the same in allopatry and sympatry. In all cases in which this effect was significant, this difference was smaller in sympatry. We further tested for asymmetry using the "Effect Details" function to test the contrast (allopatric *I. cordatotriloba* − sympatric *I. cordatotriloba*) − (sympatric *I. lacunosa* − allopatric *I. lacunosa*) = 0. This tests whether in sympatry the character in one species changed more towards the mean than in the other species, compared to the value in allopatry. In addition, we performed contrasts testing whether for either species there was a significant difference between allopatry and sympatry. In particular, if gene flow is substantial from *I. lacunosa* to *I. cordatotriloba*, but minimal in the opposite direction, we would expect *I. cordatotriloba* to show significant differences between allopatry and sympatry, but would not expect *I. lacunosa* to do so. For these contrasts, the mean square for accession was used as the denominator means square in *F* tests. All tests were corrected for multiple comparison using either corrections for false discovery rate (FDR; Benjamini & Hochberg, 1995) or a Binomial test.

3 | RESULTS

3.1 | Patterns of genetic divergence

Ipomoea cordatotriloba and *I. lacunosa* are unambiguously genetically differentiated. In the INSTRUCT and STRUCTURE analyses, the optimal number of genetic groups corresponds to K = 3. One group

consists of all *I. lacunosa* samples, while *I. cordatotriloba* consists of two differentiated groups (Figure 2a, Supporting Information Figure S2). Samples from five known allopatric *I. cordatotriloba* sites fall into one group (orange in Figure 2a), while samples from the four sympatric *I. cordatotriloba* sites and the SOS allopatric site fall into the other group (purple in Figure 2a). A Fisher exact test indicates that this association of allopatry vs. sympatry site category with the two genetic groups is statistically significant ($p = 0.0476$, two-tailed test), suggesting that the two groups represent populations with different histories of gene flow. As described above, the presence of white-flowered *I. cordatotriloba* at SOS suggests that this site may have recently been sympatric. If SOS is treated as sympatric, then this association becomes even more significant ($p = 0.00476$).

Samples where it is unknown whether they are sympatric or allopatric ("U" in Figure 2a) fall into both genetic groups. Interestingly, when K = 2, the two *I. cordatotriloba* groups fuse to form a single group. Within this group, the sympatric and SOS samples exhibit a greater contribution from the *I. lacunosa* group (blue) than the known allopatric samples, consistent with introgression from *I. lacunosa* in sympatry (Figure 2a, Supporting Information Figure S2).

Principal components analysis yields a similar pattern. In the PC1–PC2 plane, *I. lacunosa* samples form a tight cluster that is separated from the *I. cordatotriloba* samples (Figure 2b; see also the interactive version of this figure, available at https://plot.ly/~joannarifkin/8 and in the Supporting Information Appendix S1 as a downloadable zip archive of an .html folder, which identifies each point by site and sample),

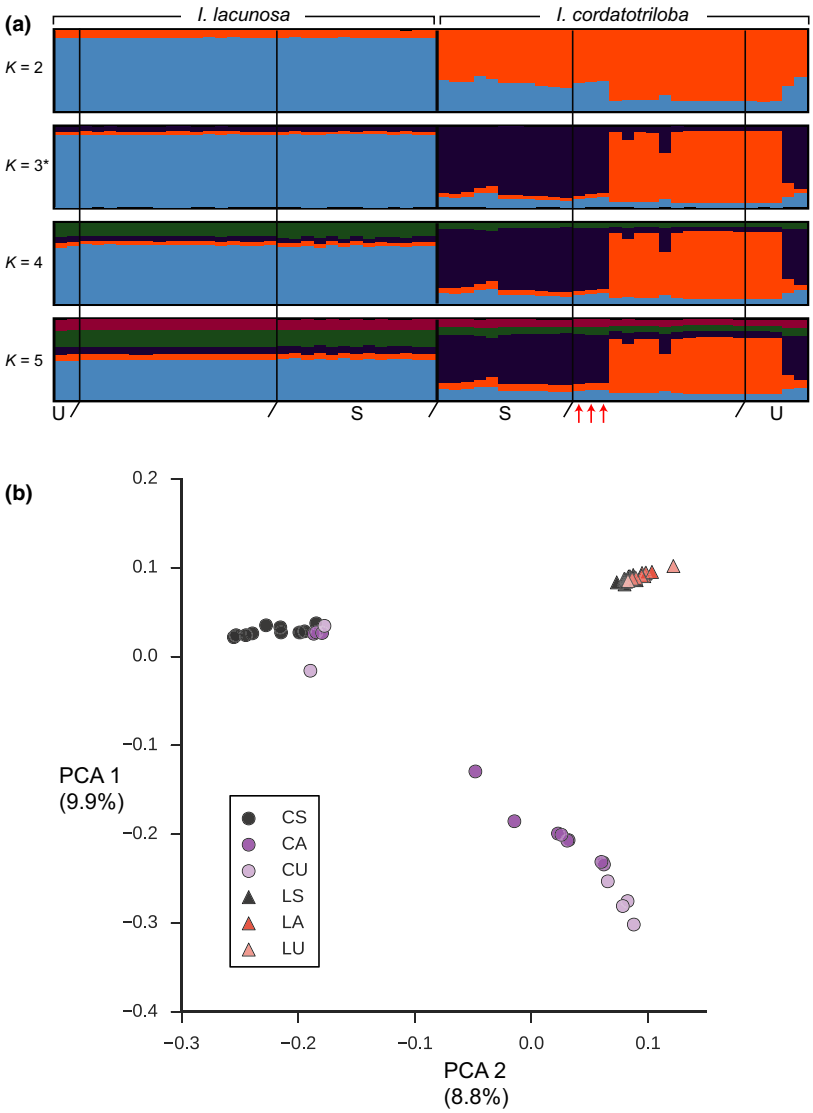

FIGURE 2 INSTRUCT and PCA plots. (a) INSTRUCT results from simulating populations K = 2 to K = 5. U = unknown; A = allopatric; S = sympatric. Red arrows indicate the *Ipomoea cordatotriloba* samples from SOS. Asterisk (*) indicates model with best DIC score. (b) PCA showing the genetic divergence of the samples. CS = *I. cordatotriloba* sympatric samples; LS = *Ipomoea lacunosa* sympatric samples; CA = *I. cordatotriloba* allopatric samples; LA = *I. lacunosa* allopatric samples; CU = *I. cordatotriloba* unknown samples LA = *I. lacunosa* allopatric samples. *Ipomoea lacunosa* (triangles) cluster together in the upper right-hand corner while *I. cordatotriloba* is separated into 2 clusters that generally follows the pattern of sympatry (upper left) and allopatry (lower right). An interactive version of this figure, with points labelled with the identities of the sites, is available at https://plot.ly/~joannarifkin/8 and in the Supporting Information Appendix S1

while *I. cordatotriloba* forms two clusters, one consisting of the sympatric and SOS samples, the other consisting of the remaining samples.

Although the two species are clearly differentiated, allele-frequency differences between the species are moderate, with the average allele-frequency difference being only 0.234. Generally, frequency differences were below 0.5, but approximately 10% of loci exhibited larger frequency differences, including 2372 (3.6%) for which different alleles were fixed or nearly fixed (frequency difference ≥ 0.9) (Figure 3).

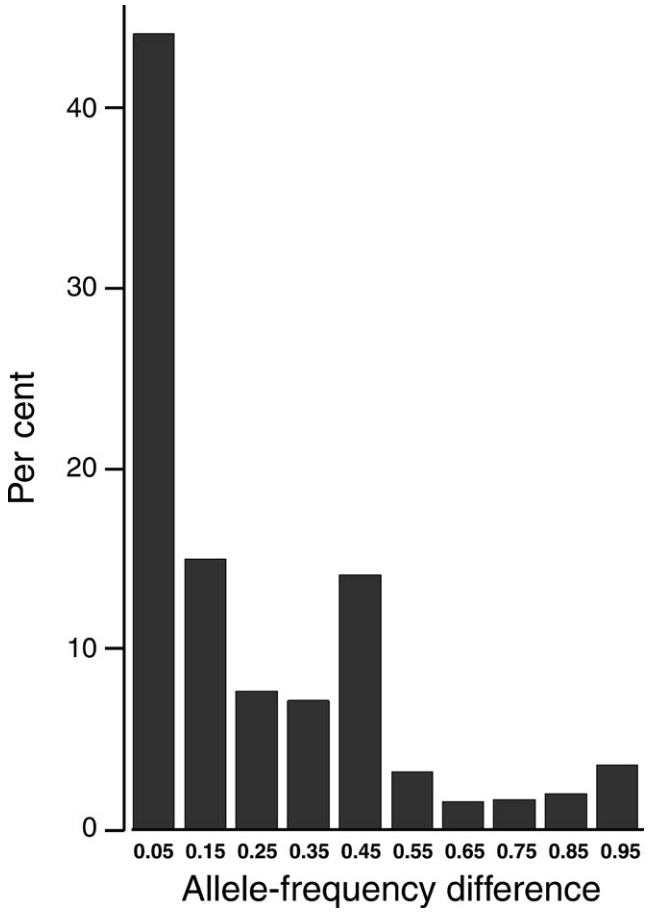

FIGURE 3 Frequency histogram of allele-frequency differences between *Ipomoea cordatotriloba* and *I. lacunosa* for 66,729 SNPs

3.2 | Positive selection

We performed a McDonald–Kreitman analysis on all samples to determine whether selection contributed to fixed or nearly fixed differences in nonsynonymous SNP frequencies between the two species. For both fixed and nearly fixed SNPs, there was a significant excess of nonsynonymous SNPs (Table 1), indicating the occurrence of divergent selection. The estimated proportions of fixed and nearly fixed SNPs subjected to divergent selection were 0.55 and 0.45 respectively, while the estimated numbers of such SNPs were 105 and 290 (Table 1, Supporting Information Table S2).

We also performed an analogous test comparing noncoding and synonymous SNPs. For SNPs with fixed differences and nearly fixed differences, there was a significant excess of noncoding SNPs (Table 1). By analogy with a standard M-K test, we interpret this excess as indicating that positive selection contributed to fixation or near fixation of regulatory SNPs. Approximately 30% of fixed noncoding differences, or 32 noncoding SNPs, are attributable to selection, whereas approximately 7.5%, or 25 nearly fixed noncoding SNPs, are attributable to selection (Supporting Information Table S2).

3.3 | Gene flow

Our approach to ascertaining whether there is ongoing gene flow between the two species is to ask whether genetic differentiation is lower between sympatric samples of the two species than between allopatric and samples (Kulathinal, Stevison, & Noor, 2009; Martin et al., 2013; Noor & Bennett, 2009). As measured by π_{LC}, divergence between allopatric samples was approximately 1.65–1.85 times greater than between sympatric samples, depending on whether known allopatric samples or close allopatric samples were used. In both cases, the difference was highly significant (Table 2). The

TABLE 1 McDonald-Kreitman test table for all samples. (a) Loci are considered fixed if allele- frequency difference = 1. (b) Loci are considered nearly fixed if allele-frequency difference is greater than or equal to 0.9 but <1. G = G-statistic of association. Prob = the probability of no association. α = estimated proportion of nonsynonymous fixed differences that were fixed by selection. "Fixed by selection" is the estimated number of fixed differences that were fixed by selection

	Nonsynonymous	Synonymous	Noncoding	Synonymous
(a) Sympatric frequency difference = 1				
Fixed differences	190	169	107	169
Polymorphisms	21,556	26,910	11,153	26,910
G (Prob)	10.219 (=0.00139)		11.236 (=0.000802)	
α (Fixed by selection)	0.55 (105)		0.30 (32)	
(b) Sympatric frequency difference ≥ 0.9, < 1				
Fixed differences	644	699	333	699
Polymorphisms	20,912	26,211	10,820	26,211
G (Prob)	6.729 (=0.00949)		4.423 (=0.035)	
α (Fixed by selection)	0.45 (290)		0.075 (25)	

divergence between allopatric and sympatric *I. cordatotriloba* was also about 6–7 times greater than the analogous divergence for *I. lacunosa*. This difference was significant regardless of whether known or close allopatric samples were used (Table 2). These results are consistent with gene flow occurring in sympatry and with greater gene flow occurring from *I. lacunosa* to *I. cordatotriloba* than in the reverse direction.

Analysis of allele-frequency differences yields a similar conclusion. Average allele-frequency differences between the species are more than twice as large when allopatric samples are compared than when sympatric samples are compared, with these effects being highly significant when either known or close allopatric samples are used (Table 3, Figure 4). Within both species, the average difference in allele frequency between allopatric and sympatric samples is significantly >0, indicating genetic differentiation consistent with gene flow in sympatry (Table 3). However, this difference is much smaller for *I. lacunosa* than for *I. cordatotriloba* (Table 3), again suggesting greater gene flow from *I. lacunosa* to *I. cordatotriloba* than *vice versa*.

3.4 | Resistance to gene flow

We examined the degree to which gene flow in sympatry homogenized allele frequencies by examining the relationship between allele-frequency divergence between the species in allopatry and allele-frequency divergence in sympatry (Figure 5). There appear to be two categories of SNPs: one in which allele-frequency differences in allopatry are greater than or equal to 0.9 ("highly divergent" SNPs), and one in which those frequencies are <0.9 ("less-divergent" SNPs). These two categories differ markedly in the proportion that are homogenized. Most less-divergent SNPs exhibit frequency differences in sympatry that are near 0, indicating substantial homogenization (Figure 5a,b). By contrast, there was substantially less homogenization of highly divergent SNPs, with the modal frequency difference in sympatry being >0.9 (Figure 5a,b) and with most of these SNPs having a frequency in sympatry >0.5, suggesting that these SNPs are resistant to homogenization. This pattern is exhibited when either known allopatric samples or close allopatric samples are compared with sympatric samples.

The previous analyses of divergent selection, which found significant positive selection for highly divergent nonsynonymous and noncoding SNPs, are consistent with the interpretation that highly diverged SNPs are subject to divergent selection. These analyses, however, used all samples, whereas the data indicating resistance to homogenization used only known (or close) allopatric and sympatric samples. We therefore performed a second set of M-K analyses that omitted unknown samples, asking whether nonsynonymous or noncoding SNPs were subject to selection. For analyses using known allopatric samples and analyses using close allopatric samples, both fixed and nearly fixed nonsynonymous SNPs show evidence of selection (Table 4). The estimated proportion of nonsynonymous SNPs subject to selection range from 0.39 to 0.56, with an estimated 303 fixed and nearly fixed SNPs fixed by selection in the known allopatric analysis and an estimated and 227 fixed and nearly fixed SNPs fixed by selection in the close allopatric analyses (Table 4, Supporting Information Table S2). The

TABLE 2 Analysis of π between and within species. Between species values (π_{LC}) are average π values of pairwise comparisons between a sample from *Ipomoea lacunosa* (*L.*) and a sample from *Ipomoea cordatotriloba* (*C.*). Within species values (π_{AS}) are average π values of pairwise comparisons between samples from allopatry and sympatry. *p* values determined from 1,000 bootstrap samples. Reported π values are 1,000 times actual value

Between species	π_{LC}	*p*	Hypothesis tested
a. Known allopatric samples	0.651	—	—
b. Close allopatric samples	0.580	—	—
c. Sympatric samples	0.352	—	—
d. a. – c.	0.299	<0.001	Sympatric divergence = Allopatric divergence (a. – c. = 0)
e. b. – c.	0.228	<0.001	Sympatric divergence = Allopatric divergence (b. – c. = 0)
Within species, between sample categories	π_{AS}		
f. *C.* (Known allopatric vs. sympatric)	0.592	—	—
g. *L.* (Known allopatric vs. sympatric)	0.083	—	—
h. *C.* (Close allopatric vs. sympatric)	0.518	—	—
i. *L.* (Close allopatric vs. sympatric)	0.083	—	—
j. f. – g.	0.509	<0.001	Equal divergence for *C.* and *L.* (f. – g. = 0)
k. h. – i.	0.435	=0.005	Equal divergence for *C.* and *L.* (h. – i. = 0)

TABLE 3 Analysis of average allele-frequency differences between species for allopatric and sympatric samples, *D*. Between species values are average *D* comparing the two species. Within species values are average *D* values for comparisons between allopatric and sympatric samples for a given species. *C.*: *Ipomoea cordatotriloba*. *L.*: *Ipomoea lacunosa*. *p* values determined from 1000 bootstrap samples. Note: c1. and c2 differ slightly because in the analysis with close allopatric samples, some SNPs dropped out because they were no longer variable

Between species	*D*	*p*	Hypothesis tested
a. Known allopatric samples	0.311	—	—
c1. Sympatric samples	0.148	—	—
b. Close allopatric samples	0.313	—	—
c2. Sympatric samples	0.165	—	—
d. a. – c1.	0.164	=0.004	Sympatric divergence = Allopatric divergence (a. – c. = 0)
e. b. – c2.	0.148	=0.006	Sympatric divergence = Allopatric divergence (b. – c. = 0)
Within species, between sample categories			
f. *C.* (Known allopatric — sympatric)	0.154	<0.001	Allele frequency difference = 0
g. *L.* (Known allopatric — sympatric)	0.008	=0.001	Allele frequency difference = 0
h. *C.* (Close allopatric — sympatric)	0.135	=0.004	Allele frequency difference = 0
i. *L.* (Close allopatric — sympatric)	0.013	=0.242	Allele frequency difference = 0
j. f. – g.	0.146	=0.004	Equal divergence for *C.* and *L.* (f. – g. = 0)
k. h. – i.	0.121	=0.006	Equal divergence for *C.* and *L.* (h. – i. = 0)

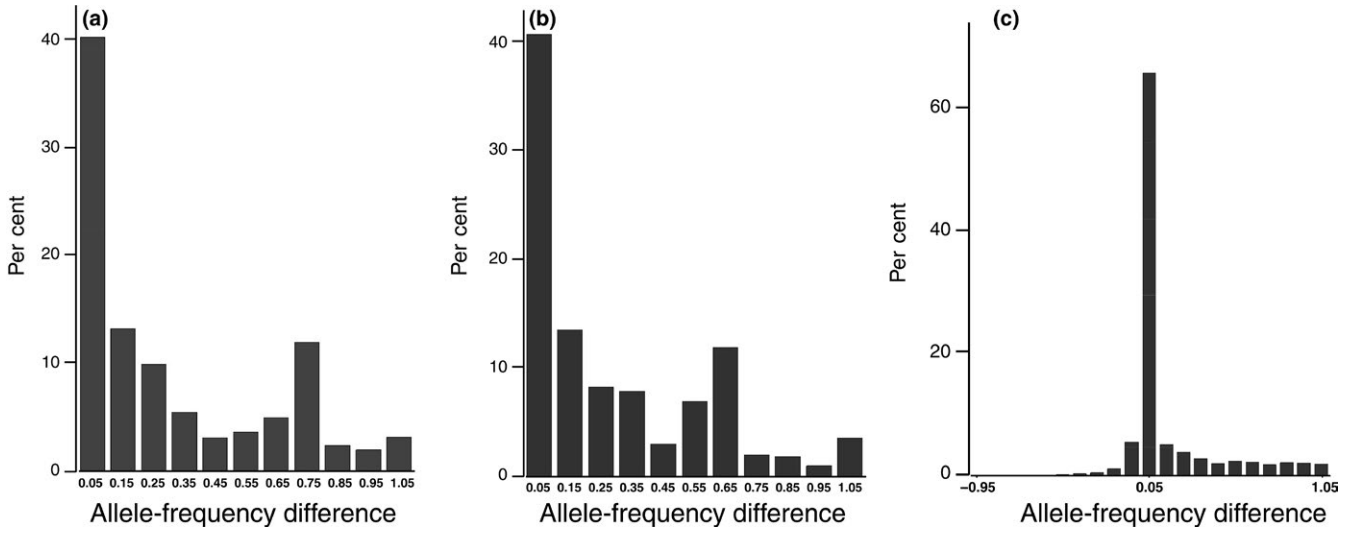

FIGURE 4 Frequency histograms of SNP allele-frequency differences between species, D, in allopatry (a, b) and in sympatry (c). (a) Differences for known allopatric samples. (b) Differences for close allopatric samples. (c) Differences for sympatric samples. Negative values arise if the more-frequent allele in allopatry is the less-frequent allele in sympatry. Frequency histograms of known allopatric samples (a) and close allopatric samples (b) are very similar. Bin labelled 1.05 corresponds to an allele-frequency difference of 1.0

remainder of the fixed and nearly fixed SNPs are presumably neutral but resist homogenization because they are linked to the selected SNPs.

The set of SNPs resistant to selection also include noncoding SNPs that have experienced divergent selection. However, estimates of the proportion of such SNPs subject to selection are <10%, with the estimated number of such SNPs being 17 or less (Table 5, Supporting Information Table S2).

3.5 | Asymmetric admixture

We further examined the hypothesis that highly divergent genes in allopatry are resistant to homogenization in sympatry in two ways. One involved an admixture analysis, while the second involved simulating gene flow. If this hypothesis is true, then we would expect that loci that are highly divergent in allopatry would exhibit less admixture than loci that are less divergent in allopatry. We therefore compared admixture of highly divergent loci with

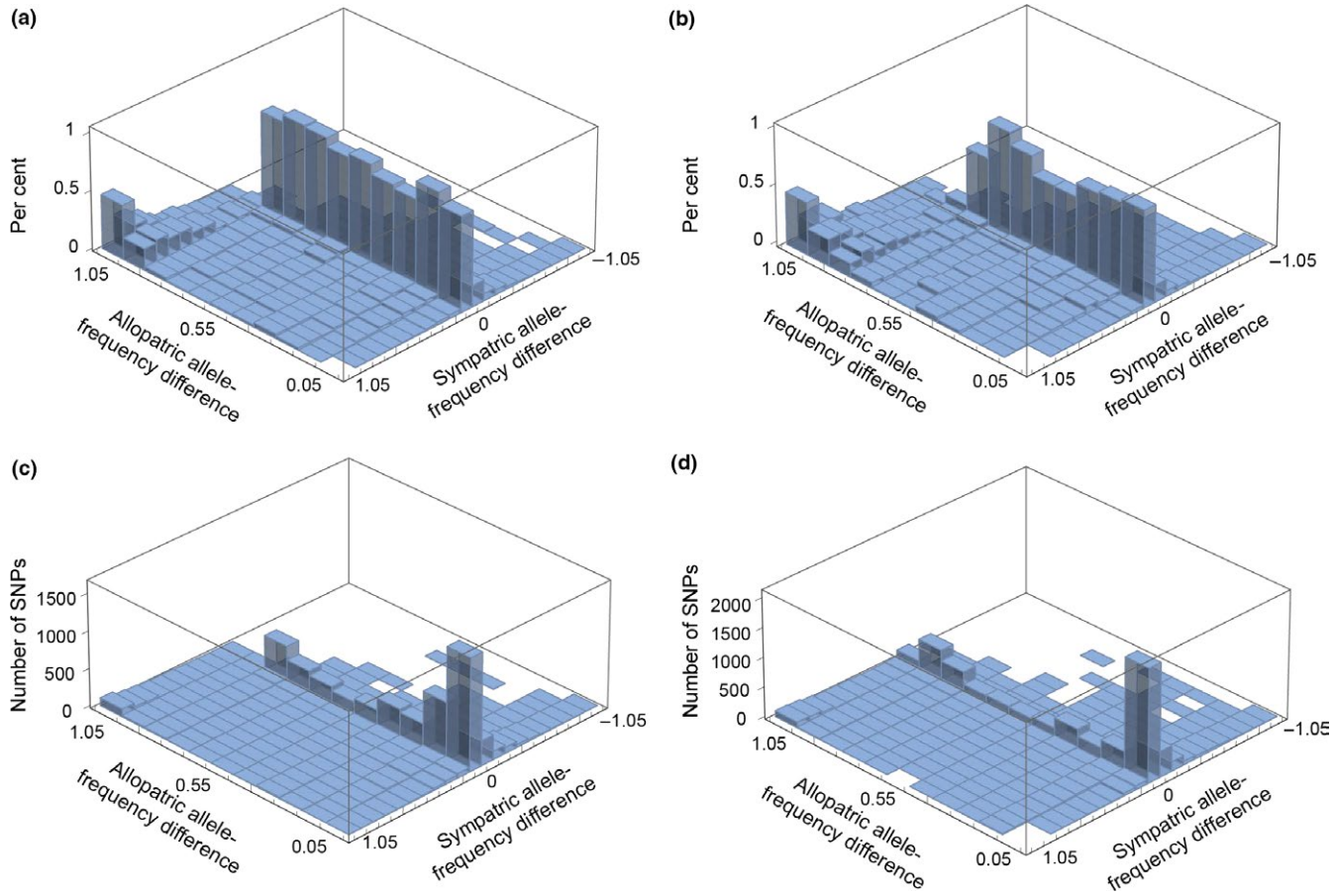

FIGURE 5 Frequency histogram of SNP allele-frequency differences between species in sympatry vs. differences in allopatry. (a and b) Percentages within each allopatric frequency difference category were normalized to sum to 1. (c and d) Numbers of SNPs. Bin midpoints are labelled. Bins labelled 1.05 and −1.05 indicate fixed differences between the species. (a and c) Data from known allopatric and sympatric samples. (b and) Data from close allopatric and sympatric samples [Colour figure can be viewed at wileyonlinelibrary.com]

	Known allopatric samples		Close allopatric samples	
	Nonsynonymous	**Synonymous**	**Nonsynonymous**	**Synonymous**
(a) Sympatric frequency difference = 1				
Fixed differences	314	274	292	265
Polymorphisms	19,773	25,301	17,270	22,510
G (Prob)	21.44 (<0.00001)		17.96 (=0.000023)	
α (Fixed by selection)	0.56 (176)		0.47 (137)	
(b) Sympatric frequency difference ≥0.9, <1				
Fixed differences	255	255	232	226
Polymorphisms	19,478	25,046	17,038	22,284
G (Prob)	7.947 (=0.0048)		9.799 (=0.00175)	
α (Fixed by selection)	0.50 (127)		0.39 (90)	

TABLE 4 McDonald-Kreitman test table for positive selection on SNPs for which the allele- frequency difference in allopatry is greater than or equal to 0.9 and the corresponding difference in sympatry is as indicated. G is G statistic of association. Prob is the probability of no association. α is estimated proportion of nonsynonymous focal SNPs that were fixed by selection. "Fixed by selection" is estimated number of nonsynonymous SNPs that were fixed by selection. Tests for sympatric allele-frequency differences <0.9 were all non-significant

less-divergent loci. For the highly divergent loci, the *I. cordatotriloba* sympatric samples exhibit less admixture (1 − *m* = proportional contribution from *I. lacunosa* = 0.17 and 0.16 for known and close allopatric analyses, respectively) than less-divergent loci (1 − *m* = 0.69 and 0.62, respectively; Figure 6, Supporting Information Figure S3), consistent with highly divergent loci being resistant

to homogenization by gene flow. By contrast, for both highly and less-divergent loci, there is no evidence of introgression from *I. cordatotriloba* into the sympatric *I. lacunosa* samples (*m* = proportional contribution from *I. cordatotriloba* = 0.01 for both known and close allopatric analyses) (Supporting Information Figures S3 and S4). Gene flow thus seems to be highly asymmetric, with most occurring from *I. lacunosa* to *I. cordatotriloba*, and little in the reverse direction.

Admixture creates linkage disequilibrium within an admixed sample. To determine the scale of this LD, we calculated admixture LD as a function of SNP separation distance. For both the *I. cordatotriloba* and *I. lacunosa* sympatric samples, admixture LD decayed almost completely within 100 kb (Supporting Information Figure S5). At low distances, admixture LD was about 2.5 times higher for *I. cordatotriloba* than for *I. lacunosa*, a pattern consistent with evidence above suggesting that introgression is greater from *I. lacunosa* to *I. cordatotriloba* than in the reverse direction.

3.6 | Simulating the effects of gene flow on differentiation

Our second approach for testing the hypothesis that highly divergent loci are resistant to homogenization was to use simulation to ask whether effective introgression from newly sympatric *I. lacunosa* samples into newly sympatric *I. cordatotriloba* samples (estimated by f^*) was less for highly divergent SNPs than for less-divergent SNPs.

In our simulations, the best fitting values of effective introgression, f^*, for highly divergent SNPs were 0.19 and 0.18 for analyses using known and close allopatric samples, respectively. By contrast, the comparable values for less-divergent SNPs were both 0.59. Both sets of values are very similar to the admixture proportions estimated by the admixture analysis (Figure 6). The 95% credible intervals for f^* for highly divergent SNPs were 0.15 – 0.23 for both analyses, while for less-divergent SNPs, they were 0.50 – 0.64 and 0.49 – 0.63 for analyses using known and close allopatric samples, respectively (Figure 7). Since the credible intervals for

TABLE 5 McDonald-Kreitman test tables for all positive selection on noncoding SNPs for which the allele-frequency difference in allopatry is greater than or equal to 0.9 and the corresponding difference in sympatry is 1. G is G statistic of association. Prob is the probability of no association. α is estimated proportion of nonsynonymous focal SNPs that were fixed by selection. "Fixed by selection" is estimated number of noncoding SNPs that were fixed by selection. Tests for sympatric allele-frequency differences <1 were all non-significant

	Known allopatric samples		Close allopatric samples	
	Noncoding	Synonymous	Noncoding	Synonymous
Fixed differences	172	274	165	265
Polymorphisms	10,318	25,301	9,018	22,510
G (Prob)	18.65 (=0.000016)		18.133 (=0.000021)	
α (Fixed by selection)	0.096 (17)		0.0 (0)	

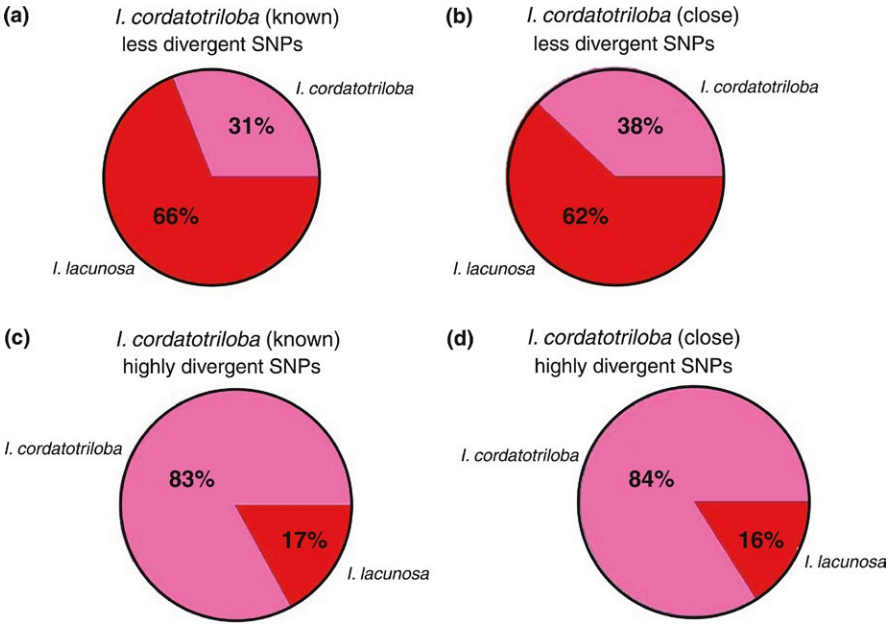

FIGURE 6 Admixture contributions to sympatric samples of *Ipomoea cordatotriloba*. Numbers are percentages and correspond to m (*I. cordatotriloba* contribution) and 1 – m (*Ipomoea lacunosa* contribution). (a, b) contributions for less-divergent SNPs (allopatric allele-frequency difference between species <0.9). (c, d) contributions for highly divergent SNPs (allopatric frequency difference ≥0.9). (a, c) analyses using known allopatric samples. (b, d) analyses using close allopatric samples. Differences between less divergent and highly divergent samples are highly significant ($p < 0.001$) for both known and close allopatric samples [Colour figure can be viewed at wileyonlinelibrary.com]

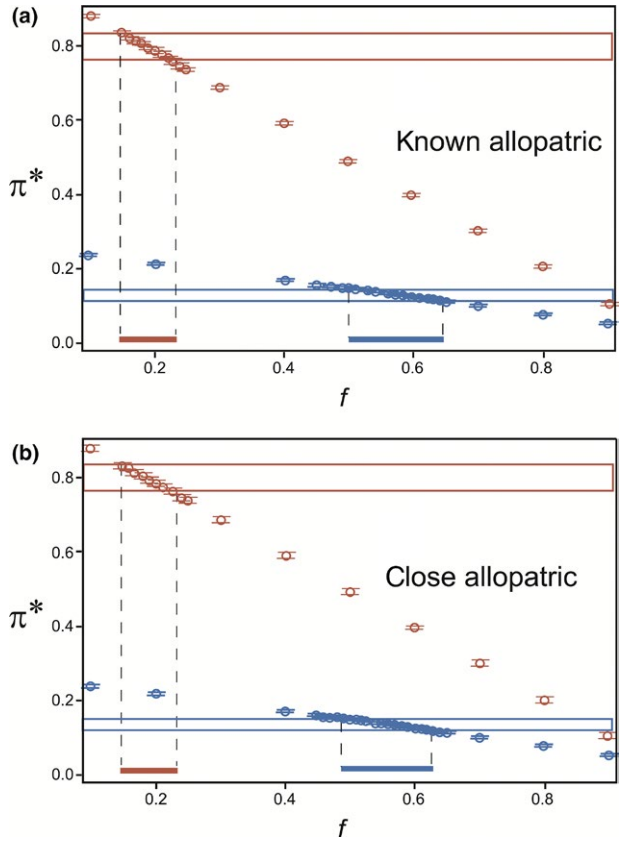

FIGURE 7 Determination of effective introgression rate for highly divergent (red) and less-divergent (blue) SNPs. Points (circles) indicate average predicted value of π^* for a given value of f (20 replicates). 95% confidence intervals are indicated by error bars (most within the circles). Rectangular boxes indicate the 95% credible interval for the observed value of π^*. Dashed lines indicate overlap between π^* credible interval and predicted values of π^*, which represents the credible values of f^* (thick bars at base of dashed lines)

the two types of SNPs fail to overlap, we can reject the hypothesis that effective introgression was similar for highly diverged and less-diverged SNPs. In particular, effective introgression was substantially lower for the highly diverged SNPs, consistent with the operation of divergent selection in preventing homogenization of these SNPs in sympatry.

3.7 | Phenotypic divergence and convergence

In an analysis using known allopatric samples, all eight characters examined exhibited a highly significant difference between species ($p < 0.0003$ in all cases; Supporting Information Table S3). Six characters (corolla length, corolla width, herkogamy, nectar sugar concentration, nectar volume and cyme length) exhibited nominally significant ($p < 0.05$) species x location (allopatric vs. sympatric) interactions, and all of these except corolla width remained significant after correcting for false discovery rate (Supporting Information Table S4). For these characters, species differed less in sympatry than in allopatry. For these five characters, the asymmetry contrast was nominally significant (Supporting Information Table S3, Figure 8), with two remaining significant after FDR correction (Supporting Information Table S3). In addition, in *I. cordatotriloba*, all five characters were significantly smaller in sympatry than in allopatry after correction for multiple comparisons, whereas in *I. lacunosa*, none of the characters were significantly different in allopatry vs. sympatry (Figure 8, Supporting Information Table S4).

These patterns are consistent with gene flow partially reducing phenotypic differences in these characters. Moreover, the direction of asymmetry is consistent with the asymmetry in gene flow: Phenotypically, *I. cordatotriloba* changes more (becomes more like *I. lacunosa*) than *I. lacunosa* (Figure 8). An analysis using only close allopatric sites produced a qualitatively similar result: All characters exhibited differences between species in allopatry; five characters exhibited nominally significant species x location interactions, with three remaining significant after correction for multiple comparisons; and four of the five exhibited significant asymmetric convergence, with *I. cordatotriloba* converging substantially more towards *I. lacunosa* than the reverse (Supporting Information Figure S6 and Tables S3, S4).

The reduction of herkogamy in sympatric *I. cordatotriloba* samples compared to allopatric samples suggests there might also be an increase in selfing rate. Our estimates of selfing rates confirm this expectation (Supporting Information Table S5). Selfing rate in sympatric *I. cordatotriloba* is significantly increased ($p < 0.01$) compared to allopatric samples. By contrast, selfing rates are essentially the same in sympatric and allopatric samples of *I. lacunosa*, which does not experience gene flow from *I. cordatotriloba* in sympatry.

4 | DISCUSSION

Ipomoea cordatotriloba and *I. lacunosa* appear to be at an intermediate stage in the process of speciation. Allele-frequency divergence is generally low. Although some degree of both prezygotic isolation and postzygotic isolation have evolved (Duncan & Rausher, 2013b; Rifkin unpublished data), hybrids that are generally healthy and fertile can form, allowing for the possibility of gene flow. Our investigation has revealed four key findings that are relevant for understanding the processes that contribute to divergence and speciation in this system: (a) Natural selection appears to have caused divergence between the two species at a small number of loci; (b) substantial, but asymmetric, gene flow occurs at sites at which the two species are sympatric; (c) gene flow is sufficient to homogenize allele frequencies at apparently neutral loci, but insufficient to homogenize divergently selected loci; and (d) the two species are phenotypically more similar in sympatry than in allopatry, with asymmetry in phenotypic convergence being consistent with the asymmetry in gene flow. In the following sections, we discuss each of these findings in turn.

4.1 | Divergence and selection

Although allele-frequency divergence between *I. cordatotriloba* and *I. lacunosa* is generally low, for approximately 12 per cent of SNPs

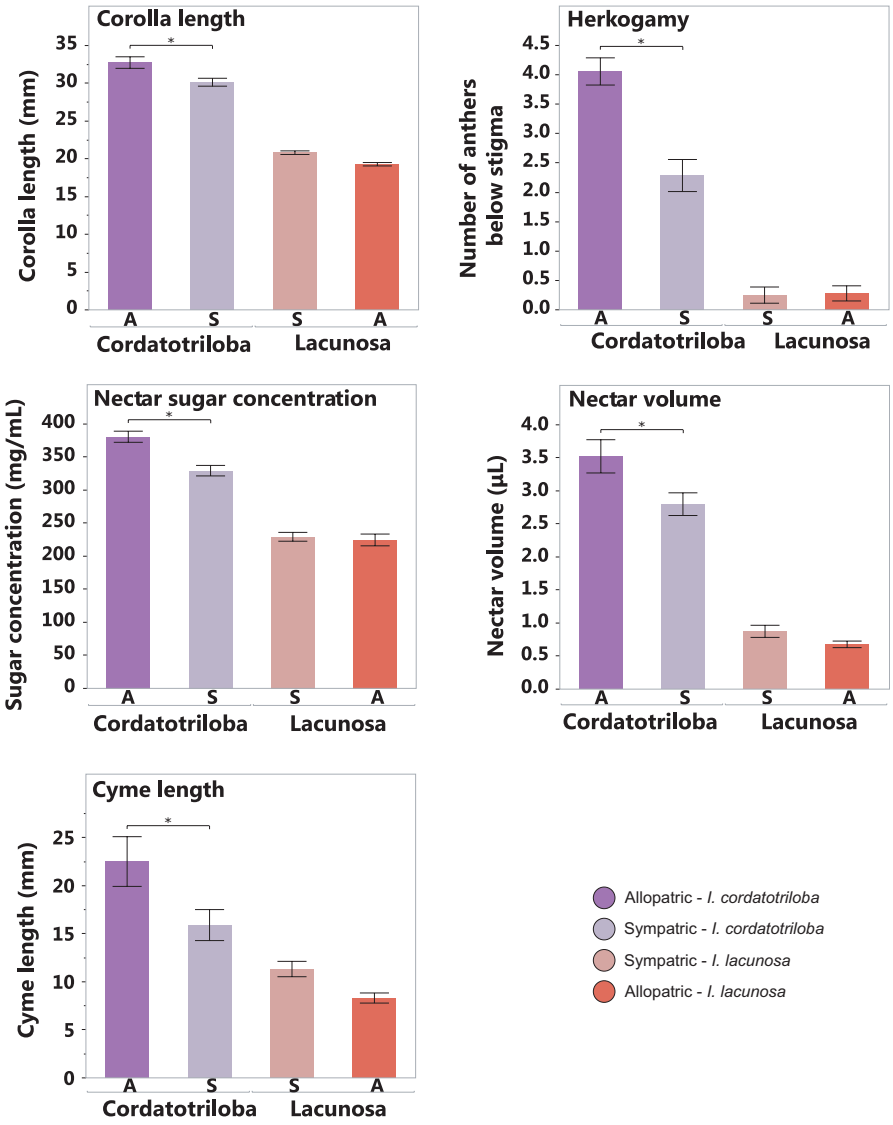

FIGURE 8 Comparison of mean trait values between *Ipomoea cordatotriloba* and *Ipomoea lacunosa* in allopatry (outside bars) and sympatry (inside bars) for traits showing significantly smaller differences in sympatry than in allopatry. Error bars indicate standard error. Asterisk indicates difference significant at overall level of $p < 0.05$ after correction for multiple comparisons [Colour figure can be viewed at wileyonlinelibrary.com]

frequency divergence is >0.5. While most of this divergence appears to be caused by genetic drift, our M-K analyses indicate that divergence at a small number (between 200 and 450) of highly divergent SNPs was likely caused by selection. This is likely an underestimate of the number of SNPs involved in adaptive divergence because the M-K test examined only SNPs in transcripts. While we detected divergent selection acting on a small number of noncoding, and presumably regulatory, SNPs, our transcriptome-based approach likely failed to detect many SNPs in regulatory regions, particularly those upstream of the transcription start site, downstream of the polyadenylation site and in introns. Additionally, because we only sampled leaf tissue, our analyses do not include possibly divergent genes expressed in other tissues. Nevertheless, our results indicate that selection has played a role in genetic divergence of the two incipient species. This conclusion is consistent with Qst-Fst analyses of character divergence, which have indicated that selection has contributed to divergence in floral characters between the two species (Duncan & Rausher, 2013a; Rifkin, 2017), although we are currently unable to associate selected SNPs with particular traits.

4.2 | Gene flow

Our approach to determining whether ongoing or recent gene flow has occurred between the two species was to ask whether the two species are less differentiated when growing sympatrically than when they grow allopatrically. Three different types of evidence exhibit this pattern. First, in our STRUCTURE and INSTRUCT analyses with two genetic groups, the two species separated unambiguously into the two groups. However, sympatric *I. cordatotriloba* samples exhibited evidence of greater admixture with *I. lacunosa*, consistent with gene flow from *I. lacunosa* to *I. cordatotriloba*. Similarly, in our PCA analysis, sympatric *I. cordatotriloba* samples were intermediate between allopatric *I. cordatotriloba* and *I. lacunosa*. Second, divergence, as measured both by between-species π and allele-frequency differences, was significantly less between sympatric samples than between allopatric samples of the two species. Finally, our admixture analysis indicated substantial admixture in sympatric *I. cordatotriloba* samples.

One possible alternative explanation for these patterns is that variation within each species is structured geographically. If, for example, populations of both species far from the sympatric sites have diverged from those near sympatric sites because of environmental differences, the average divergence for all known allopatric sites could be elevated. However, this explanation is not supported by the analysis using close allopatric sites, which shows the same extent of reduced divergence in sympatry as the analysis using all known allopatric sites (Supporting Information Table S1). We thus believe the data are best interpreted as indicative of recent gene flow in sympatry.

Our observation of flower-colour variation supports this interpretation. In particular, we have found white-flowered *I. cordatotriloba* plants in all of the sympatric populations we examined, but in none of the known allopatric populations. Given that preliminary results of complementation tests indicate that the same alleles at the same locus cause white flowers in both species, this pattern is consistent with introgression of this allele from *I. lacunosa* into *I. cordatotriloba*, and thus gene flow, in all sympatric populations. Our failure to ever observe a purple-flowered *I. lacunosa* is also consistent with our inference of very low gene flow from *I. cordatotriloba* into *I. lacunosa*.

As has been found in other closely related species pairs involving a highly selfing species and a more outcrossing species (Brandvain et al., 2014; Kenney & Sweigart, 2016; Palma-Silva et al., 2011; Ruhsam, Hollingsworth, & Ennos, 2011; Sweigart & Willis, 2003), we observed asymmetric gene flow, with greater introgression from the selfer *I. lacunosa* into the mixed-mater *I. cordatotriloba* than in the opposite direction. Common explanations for this pattern include the following: 1) Highly selfing species tend to produce less pollen than outcrossing species (McDonald et al., 2011); 2) pollen from the selfing species is likely less competitive (Diaz & Macnair, 1999); 3) pollinators tend to visit outcrossing flowers more than selfing flowers (Brandvain et al., 2014); and 4) novel alleles spread more easily in outcrossing than in selfing populations (Morjan & Rieseberg, 2004). Under these circumstances, F1 hybrids are more likely to be produced by an outcrosser pollinating a selfer than *vice versa* because pollen from the outcrosser is highly competitive when placed on the selfer's stigma, while pollen from the selfer is not very competitive when placed on an outcrosser's stigma (Brandvain & Haig, 2005). Regardless of how an F1 hybrid is produced, however, it is more likely to successfully pollinate an outcrossing parent than a selfing parent because the selfer has a greater tendency to self-pollinate before it is visited by a pollinator. This asymmetry in pollen flow would contribute to an asymmetry in introgression. To the extent that pollinators are more likely to visit the outcrossing parent than the selfing parent, an F1 individual is also more likely to be pollinated by an outcrossing parent than by a selfing parent. In addition, the production of more pollen and more competitive pollen by the outcrossing parental species will tend to exacerbate this trend (Sweigart & Willis, 2003; Palma-Silva et al., 2011; Ruhsam et al., 2011; Brandvain et al., 2014). These patterns would tend to create backcross offspring that are more like the outcrossing parent. As in the F1, any tendency of these hybrids to mate more readily with the outcrossing parent than the selfing parent would also contribute to asymmetric introgression from the selfer to the outcrosser.

Features of *I. cordatotriloba* and *I. lacunosa* are consistent with this explanation. Flowers of the mixed-mater *I. cordatotriloba* produce about 2.99 times as much pollen (McDonald et al., 2011), and pollen that is 1.17 times larger (Rifkin unpublished data), than the highly selfing *I. lacunosa* (Rifkin unpublished data). And while comparative pollination studies of these two species have not been performed, *I. lacunosa* has smaller flowers that produce substantially less nectar than *I. cordatotriloba*, presumably making them less attractive to pollinators and thus visited less frequently (Rifkin, 2017). In addition, *I. lacunosa* anthers are more tightly clustered around the stigma (Duncan & Rausher, 2013b), which may constitute an impediment to outcross pollen that is not present in *I. cordatotriloba*.

Despite the likelihood that these factors contribute to the observed asymmetry in gene flow, other factors may also be involved. While Duncan and Rausher (2013b) found that prezygotic or very early acting postzygotic incompatibilities appear to be symmetric, this study was based on only a few crosses. It is thus possible that these incompatibilities may more generally be asymmetric, which could also contribute to the asymmetry in gene flow.

4.3 | Homogenization and resistance to introgression

When there is gene flow between genetically differentiated incipient species, two patterns are expected. First, as long as effective migration rates are greater than about one individual per generation, homogenization of allele frequencies is expected at neutral loci (Wright, 1969). Second, for loci subject to divergent selection, divergence in allele frequencies between species is expected, with the difference increasing with the strength of selection (Haldane, 1930; Wright, 1931). Together, these patterns can produce substantial variation in the degree of divergence exhibited across the genome. In particular, strong divergent selection in the face of gene flow is expected to create local regions of divergence within the genome (Charlesworth et al., 1997; Harr, 2006; Feder & Nosil, 2010; White et al., 2010; Carneiro et al., 2014; Nadeau et al., 2012; Via et al., 2012; Hohenlohe et al., 2010, 2012; Ellegren et al., 2012; Gagnaire et al., 2013; Delmore et al., 2015).

Our analysis demonstrates both of these patterns. For SNPs that exhibit a between-species allele-frequency difference <0.9 in allopatry (less-divergent SNPs), almost complete homogenization takes place in sympatry. Homogenization occurs for synonymous, nonsynonymous and noncoding SNPs. The synonymous SNPs are most likely subject almost entirely to drift; we infer the latter are largely neutral or nearly neutral because we failed to detect any evidence of selection on them in our M-K analyses. Thus, these neutral variants appear to be effectively homogenized by gene flow.

By contrast, SNPs that exhibit an allele-frequency difference in allopatry of greater than or equal to 0.9 (fixed and nearly fixed

SNPs) appear to be resistant to homogenization due to divergent selection. Three types of evidence point to this conclusion. First, our M-K analysis indicates that a substantial fraction of nonsynonymous and noncoding SNPs in this category experience divergent selection. Second, admixture in sympatric *I. cordatotriloba* was much less for highly diverged SNPs than for less-diverged SNPs. And finally, our simulations indicate that substantially less introgression occurred for highly diverged SNPs than for less-diverged SNPs, consistent with the admixture analysis. We take this evidence to mean that divergent selection is strong enough to resist the tendency of gene flow to eliminate allele-frequency differences at these SNPs. However, we also observed that the between-species allele-frequency differences for these loci are generally smaller in sympatry than in allopatry, pointing to an effect of gene flow short of homogenization.

There are certainly limitations to our simulation analyses. For example, our approach does not allow for different loci to experience different rates of gene flow for stochastic reasons. It also fails to take into account genetic drift that may occur after gene flow. Nevertheless, we believe that the similarity of the simulation results to those of the admixture analyses indicate that they capture the general features of differential introgression between highly divergent and less-divergent SNPs and are showing a true effect of divergent selection on resistance to homogenization.

This pattern raises the question as to why synonymous SNPs, as well as the majority of presumably neutral nonsynonymous and noncoding SNPs, exhibiting fixed or nearly fixed allopatric frequency differences are not homogenized. Such SNPs could be protected from homogenization if they are linked to SNPs subject to divergent selection: Theory indicates a neutral SNP will be resistant to homogenization as long as it is close enough to a selected SNP that the recombination rate is substantially lower than the selection coefficient (Barton & Bengtsson, 1986; Nordborg, 1997). This effect is enhanced with high selfing rates (Charlesworth et al., 1997; Nordborg, 1997), such as those of the species in this study. But if this explanation is true, it raises yet another question: Why would a large majority of neutral highly divergent SNPs become disproportionately linked to divergently selected SNPs, rather than scattered throughout the genome?

One possible answer to this question is that the demographic and geographic history of the two species may have been complex. In particular, the following scenario can explain this pattern: Initial divergence between the two species occurred in allopatry. During this phase, divergent selection led to the fixation or near fixation of perhaps a couple hundred nonsynonymous and noncoding SNPs. During the same period, genetic drift also fixed or nearly fixed neutral variants throughout the genome. Subsequently, species ranges shifted, perhaps caused by Pleistocene climatic changes, causing secondary contact with extensive gene flow between the species. This gene flow would have homogenized the frequencies at neutral SNPs unlinked to the selected SNPs. The only neutral SNPs remaining with fixed or nearly fixed differences would be those that were protected by linkage to selected SNPs (Barton & Bengtsson, 1986)—those that today in allopatry exhibit large divergence. Finally, if ranges shifted again so that they were largely allopatric, drift would cause frequency divergence at SNPs unlinked to the selected SNPs. If this last phase was short enough, few of these diverging SNPs would have diverged sufficiently to become fixed or nearly fixed differences. At this stage, fixed or nearly fixed neutral SNPs would be linked to the selected SNPs, while neutral SNPs exhibiting less divergence would be unlinked. In populations that regained secondary contact—those that correspond to the sympatric sites in our study—gene flow would homogenize the unlinked SNPs, but not the linked SNPs. In other words, gene flow would homogenize neutral SNPs with low between-species frequency differences, but not neutral SNPs that exhibited fixed or nearly fixed differences. This pattern is exactly what we have observed.

4.4 | Phenotypic divergence

Gene flow is expected to reduce between-species phenotypic divergence for two reasons. First, for characters that diverged due to drift, the variation at the underlying loci is neutral and is expected to be homogenized by gene flow. Second, for characters that have diverged due to selection, divergence at the underlying loci will be reduced by gene flow, at least to some extent. Our analysis of phenotypic divergence provides evidence consistent with gene flow reducing phenotypic divergence: Five of eight characters examined exhibit smaller differences in sympatry than in allopatry. Moreover, reduction in divergence was asymmetric and consistent with the asymmetry in gene flow: Gene flow was greater from *I. lacunosa* to *I. cordatotriloba* than in the reverse direction, and there was greater phenotypic change in *I. cordatotriloba*.

Finally, gene flow from the highly selfing *I. lacunosa* appears to have increased the selfing rate in the sympatric *I. cordatotriloba* samples, consistent with the observed reduction in herkogamy. However, selfing rate in sympatric *I. cordatotriloba* remains below that in *I. lacunosa*, which is the same in both sympatric and allopatric samples. Because increased selfing can reduce gene flow from heterospecifics, it can serve as a prezygotic isolating mechanism (Hu, 2015). This suggests that gene flow may have strengthened reproductive isolation between the two species in sympatry. Increased selfing in sympatry may be a simple consequence of gene flow causing introgression of alleles that reduce herkogamy. Alternatively, or in addition, reinforcement may have contributed to this increase because there is some reduction in hybrid fitness (Rifkin, 2017). Additional studies will be necessary to determine the relative contributions of these two processes.

The failure of any of the characters in *I. cordatotriloba* to completely converge on those of *I. lacunosa*, given that gene flow appears to homogenize neutral loci, indicates that these characters, and the loci underlying them, are likely subject to strong divergent selection, even in sympatry. Because in sympatry the two species grow in what appear to be identical environments (they often grow intertwined), it seems unlikely that this divergent selection is caused directly by environmental factors. Instead, we suggest that divergent selection may arise from selection to maintain a functioning

suite of developmentally integrated floral traits to ensure successful pollination. If true, this suggests that the two species occupy separate peaks in the phenotypic adaptive landscape with respect to floral form and function. In this situation, gene flow in sympatry would be pulling *I. cordatotriloba* down from its adaptive peak, but is not strong enough to make it cross the adaptive valley separating it from *I. lacunosa*'s peak.

5 | CONCLUSIONS

Our investigation reveals the complex interplay between selection and gene flow that can occur during the early stages of speciation. In our system, selection appears to have driven frequency divergence at a number of loci. To the extent that selection on these loci represents adaptation to different environments, they constitute extrinsic isolating mechanisms (Rundle & Nosil, 2005; Seehausen et al., 2014). At the same time, gene flow has prevented overall genomic divergence and has reduced phenotypic divergence, at least at locations where both species are present. In these locations, gene flow appears to homogenize loci that have not diverged to the extent of near fixation. That gene flow fails to homogenize the frequencies of presumably neutral loci that have diverged to fixation, or near fixation, suggests that a complex history of secondary contact, separation and recontact between the species has occurred. Our results thus suggest that explaining the genomic pattern of divergence between closely related species may require further exploration of the historical dynamics of species population sizes and range overlap. Finally, our results support suggestions that the species boundaries may be maintained in the face of gene flow because divergent selection prevents homogenization of loci contributing to those boundaries (Noor, Grams, Bertucci, & Reiland, 2001; Noor, Grams, Bertucci, & Reiland, 2001; Noor, Grams, Bertucci, & Reiland, et al., 2001; Rieseberg, 2001; Wu & Ting, 2004).

ACKNOWLEDGEMENTS

We thank Carol Baskin for providing seeds, Lena Hileman and Craig Freeman for helping locate samples in Kansas, Robert Pitman and Michael Flessner for helping locate samples in Virginia and the officers of the USDA GRIN network for providing seeds. Amanda Lea provided advice for generating the RNA libraries. Steve Jones of Laptop GPS World assisted in converting population location data into a readable format for our collections. Funding was provided by NSF grant DEB 1542387 to MDR, by NSF DDIG grant DEB 1501954 to JLR, by a Sigma Xi grant to JLR and by Duke University for sequencing funds. We thank three anonymous reviewers for helpful suggestions for improving the manuscript.

AUTHOR CONTRIBUTION

J.L.R., A.S.C., I.T.L., and M.D.R. designed the study, analysed the data and wrote the manuscript. A.S.C., J.L.R., and I.T.L. collected and grew the specimens. J.L.R. and I.T.L. measured the phenotypic traits. A.S.C. and I.T.L. performed RNA library preparation. J.L.R. and A.S.C. performed the SNP calling.

DATA ACCESSIBILITY

Ipomoea lacunosa genome: Available at https://doi.org/10.5061/dryad.f6qb7c5. *I. cordatotriloba* and *I. lacunosa* SNP data set: Available at https://doi.org/10.5061/dryad.f6qb7c5 will be submitted to Dryad upon acceptance of manuscript. APL Scripts: Available at https://doi.org/10.5061/dryad.f6qb7c5. SNP-calling and ANOVA scripts are available on Github at https://github.com/joannarifkin/IpomoeaSNPCalling and https://github.com/joannarifkin/IpomoeaQstFst/ANOVA.

REFERENCES

Barton, N., & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity*, *57*(3), 357–376. https://doi.org/10.1038/hdy.1986.135

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, *57*(1), 289–300. https://doi.org/10.2307/2346101

Bierne, N. (2010). The distinctive footprints of local hitchhiking in a varied environment and global hitchhiking in a subdivided population. *Evolution*, *64*(11), 3254–3272. https://doi.org/10.1111/j.1558-5646.2010.01050.x

Bierne, N., Roze, D., & Welch, J. J. (2013). Pervasive selection or is it…? Why are FST outliers sometimes so frequent? *Molecular Ecology*, *22*(8), 2061–2064. https://doi.org/10.1111/mec.12241

Bolten, A. B., Feinsinger, P., Baker, H. G., & Baker, I. (1979). On the calculation of sugar concentration in flower nectar. *Oecologia*, *41*, 301–304. https://doi.org/10.1007/BF00377434

Brandvain, Y., & Haig, D. (2005). Divergent mating systems and parental conflict as a barrier to hybridization in flowering plants. *The American Naturalist*, *166*(3), 330–338. https://doi.org/10.1086/432036

Brandvain, Y., Kenney, A. M., Flagel, L., Coop, G., & Sweigart, A. L. (2014). Speciation and introgression between *Mimulus nasutus* & *Mimulus guttatus*. *PLoS Genetics*, *10*(6), e1004410. https://doi.org/10.1371/journal.pgen.1004410

Bulmer, M. G. (1972). Multiple niche polymorphism. *The American Naturalist*, *106*(948), 254–257. https://doi.org/10.1086/282765

Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., … Yandell, M. (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, *18*(1), 188–196. https://doi.org/10.1017/S0305004100015450

Carling, M. D., & Brumfield, R. T. (2009). Speciation in Passerina buntings: Introgression patterns of sex-linked loci identify a candidate gene region for reproductive isolation. *Molecular Ecology*, *18*(5), 834–847. https://doi.org/10.1111/j.1365-294X.2008.04038.x

Carneiro, M., Albert, F. W., Afonso, S., Pereira, R. J., Burbano, H., Campos, R., … Ferrand, N. (2014). The genomic architecture of population divergence between subspecies of the European rabbit. *PLOS Genetics*, *10*(8), e1003519. https://doi.org/10.1371/journal.pgen.1003519

Charlesworth, D., Charlesworth, B., & Morgan, M. T. (1995). The pattern of neutral molecular variation under the background selection model. *Genetics*, *141*(4), 1619–1632.

Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations.

Genetics Research, *70*(2), 155–174. https://doi.org/10.1017/S0016672397002954

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*(13), 3133–3157. https://doi.org/10.1111/mec.12796

Delmore, K. E., Hübner, S., Kane, N. C., Schuster, R., Andrew, R. L., Câmara, F., … Irwin, D. E. (2015). Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. *Molecular Ecology*, *24*(8), 1873–1888. https://doi.org/10.1111/mec.13150

Diaz, A., & Macnair, M. R. (1999). Pollen tube competition as a mechanism of prezygotic reproductive isolation between *Mimulus nasutus* and its presumed progenitor *M. guttatus*. *The New Phytologist*, *144*(3), 471–478. https://doi.org/10.1046/j.1469-8137.1999.00543.x

Dobin, A., & Gingeras, T. R. (2015). Mapping RNA-seq reads with STAR. *Current Protocols in Bioinformatics*, *51*, 11.14.1–11.14.19. https://doi.org/10.1002/0471250953.bi1114s51

Duncan, T. M., & Rausher, M. D. (2013a). Evolution of the selfing syndrome in *Ipomoea*. *Frontiers in Plant Science*, *4*. https://doi.org/10.3389/fpls.2013.00301

Duncan, T. M., & Rausher, M. D. (2013b). Morphological and genetic differentiation and reproductive isolation among closely related taxa in the *Ipomoea* series *Batatas*. *American Journal of Botany*, *100*(11), 2183–2193. https://doi.org/10.3732/ajb.1200467

Eckert, A. J., van Heerwaarden, J., Wegrzyn, J. L., Nelson, C. D., Ross-Ibarra, J., González-Martínez, S. C., & Neale, D. B. (2010). Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics*, *185*(3), 969–982. https://doi.org/10.1534/genetics.110.115543

Egea, R., Casillas, S., & Barbadilla, A. (2008). Standard and generalized McDonald–Kreitman test: A website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Research*, *36*(Suppl_2), W157–W162. https://doi.org/10.1093/nar/gkn337

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., … Wolf, J. B. W. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, *491*, 756–760. https://doi.org/10.1038/nature11584

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, *28*(7), 342–350. https://doi.org/10.1016/j.tig.2012.03.009

Feder, J. L., & Nosil, P. (2010). The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*, *64*(6), 1729–1747. https://doi.org/10.1111/j.1558-5646.2009.00943.x

Gagnaire, P.-A., Normandeay, É., Pavey, S. A., & Bernatchez, L. (2013). The genetic architecture of reproductive isolation during speciation-with-gene-flow in Lake Whitefish species pairs assessed by RAD sequencing. *Evolution*, *67*(9), 2483–2497. https://doi.org/10.1111/evo.12075

Gao, H., Williamson, S., & Bustamante, C. D. (2007). A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics*, *176*(3), 1635–1651. https://doi.org/10.1534/genetics.107.072371

Garrigan, D., Kingan, S. B., Geneva, A. J., Andolfatto, P., Clark, A. G., Thornton, K. R., & Presgraves, D. C. (2012). Genome sequencing reveals complex speciation in the *Drosophila simulans* clade. *Genome Research*, *22*(8), 1499–1511. https://doi.org/10.1101/gr.130922.111

Gompert, Z., Lucas, L. K., Nice, C. C., Fordyce, J. A., Forister, M. L., & Buerkle, C. A. (2012). Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution*, *66*(7), 2167–2181. https://doi.org/10.1111/j.1558-5646.2012.01587.x

Haldane, J. B. S. (1930). A mathematical theory of natural and artificial selection. (Part VI, Isolation). *Mathematical Proceedings of the Cambridge Philosophical Society*, *26*(2), 220–230.

Hamilton, J. A., Lexer, C., & Aitken, S. N. (2013). Differential introgression reveals candidate genes for selection across a spruce (*Picea sitchensis*× *P. glauca*) hybrid zone. *New Phytologist*, *197*(3), 927–938. https://doi.org/10.1111/nph.12055

Hanis, C. L., Chakraborty, R., Ferrell, R. E., & Schull, W. J. (1986). Individual admixture estimates: Disease associations and individual risk of diabetes and gallbladder disease among Mexican-Americans in Starr County, Texas. *American Journal of Physical Anthropology*, *70*, 433–441. https://doi.org/10.1002/(ISSN)1096-8644

Harr, B. (2006). Genomic islands of differentiation between house mouse subspecies. *Genome Research*, *16*(6), 730–737. https://doi.org/10.1101/gr.5045006

Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, *25*(11), 2454–2466. https://doi.org/10.1111/mec.13582

Hohenlohe, P. A., Bassham, S., Currey, M., & Cresko, W. A. (2012). Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philosophical Transactions of the Royal Society B*, *367*(1587), 395–408. https://doi.org/10.1098/rstb.2011.0245

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLOS Genetics*, *6*(2), e1000862. https://doi.org/10.1371/journal.pgen.1000862

Holsinger, K. (2017). Lecture notes in population genetics. 2017 (8). Retrieved from https://figshare.com/articles/Lecture_notes_in_population_genetics/100687. https://doi.org/10.6084/m9.figshare.100687.v2

Hu, X.-S. (2015). Mating system as a barrier to gene flow. *Evolution*, *69*(5), 1158–1177. https://doi.org/10.1111/evo.12660

Iverson, K. E. (1962). A programming language. *Proceedings of the May 1–3, 1962, spring joint computer conference*, 345–351.

Janoušek, V., Wang, L., Luzynski, K., Dufková, P., Vyskočilová, M. M., Nachman, M. W., … Tucker, P. K. (2012). Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between *Mus musculus musculus* & *M. m. domesticus*. *Molecular Ecology*, *21*(12), 3032–3047.

Kenney, A. M., & Sweigart, A. L. (2016). Reproductive isolation and introgression between sympatric *Mimulus* species. *Molecular Ecology*, *25*(11), 2499–2517. https://doi.org/10.1111/mec.13630

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

Kulathinal, R. J., Stevison, L. S., & Noor, M. A. (2009). The genomics of speciation in *Drosophila*: Diversity, divergence, and introgression estimated using low-coverage genome sequencing. *PLoS Genetics*, *5*(7), e1000550.

Larson, E. L., Andrés, J. A., Bogdanowicz, S. M., & Harrison, R. G. (2013). Differential introgression in a mosaic hybrid zone reveals candidate barrier genes. *Evolution*, *67*(12), 3653–3661. https://doi.org/10.1111/evo.12205

Larson, E. L., White, T. A., Ross, C. L., & Harrison, R. G. (2014). Gene flow and the maintenance of species boundaries. *Molecular Ecology*, *23*(7), 1668–1678. https://doi.org/10.1111/mec.12601

Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J. K., Reich, D., & Berger, B. (2013). Inferring admixture histories of human populations using linkage disequilibrium. *Genetics*, *193*(4), 1233–1254. https://doi.org/10.1534/genetics.112.147330

Maroja, L., Andrés, J., & Harrison, R. (2009). Genealogical discordance and patterns of introgression and selection across a cricket hybrid zone. *Evolution*, *63*(11), 2999–3015.

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., … Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, *23*(11), 1817–1828. https://doi.org/10.1101/gr.159426.113

McDonald, J. A., Hansen, D. R., McDill, J. R., & Simpson, B. B. (2011). A phylogenetic assessment of breeding systems and floral morphology of North American *Ipomoea* (Convolvulaceae). *Journal of the Botanical Research Institute of Texas*, *5*(1), 159–177.

McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, *351*(6328), 652–654. https://doi.org/10.1038/351652a0

Messer, P. W., & Petrov, D. A. (2013). Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology and Evolution*, *28*(11), 659–669. https://doi.org/10.1016/j.tree.2013.08.003

Michel, A. P., Sim, S., Powell, T. H., Taylor, M. S., Nosil, P., & Feder, J. L. (2010). Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences*, *107*(21), 9724–9729. https://doi.org/10.1073/pnas.1000939107

Morjan, C. L., & Rieseberg, L. H. (2004). How species evolve collectively: Implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology*, *13*, 1341–1356. https://doi.org/10.1111/j.1365-294X.2004.02164.x

Muñoz-Rodríguez, P., Carruthers, T., Wood, J. R. I., Williams, B. R. M., Weitemier, K., Kronmiller, B., … Scotland, R. W. (2018). Reconciling conflicting phylogenies in the origin of sweet potato and dispersal to Polynesia. *Current Biology*, *28*(8), 1246–1256. https://doi.org/10.1016/j.cub.2018.03.020

Nadeau, N. J., Whibley, A., Jones, R. T., Davey, J. W., Dasmahapatra, K. K., Baxter, S. W., … Jiggins, C. D. (2012). Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B*, *367*(1587), 343–353. https://doi.org/10.1098/rstb.2011.0198

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, *76*(10), 5269–5273. https://doi.org/10.1073/pnas.76.10.5269

Noor, M. A., & Bennett, S. M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, *103*(6), 439–444. https://doi.org/10.1038/hdy.2009.151

Noor, M. A. F., Grams, K. L., Bertucci, L. A., Almandarez, Y., Reiland, J., & Smith, K. R. (2001). The genetics of reproductive isolation and the potential for gene exchange between *Drosophila pseudoobscura* & *D. persimilis* via backcross hybrid males. *Evolution*, *55*(3), 512–521.

Noor, M. A. F., Grams, K. L., Bertucci, L. A., & Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences*, *98*(21), 12084–12088. https://doi.org/10.1073/pnas.221274498

Nordborg, M. (1997). Structured coalescent processes on different time scales. *Genetics*, *146*(4), 1501–1514.

Nosil, P. (2008). Speciation with gene flow could be common. *Molecular Ecology*, *17*(9), 2103–2106. https://doi.org/10.1111/j.1365-294X.2008.03715.x

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, *18*(3), 375–402. https://doi.org/10.1111/j.1365-294X.2008.03946.x

Ornduff, R. (1969). Reproductive biology in relation to systematics. *Taxon*, *18*(2), 121–133. https://doi.org/10.2307/1218671

Palma-Silva, C., Wendt, T., Pinheiro, F., Barbará, T., Fay, M. F., Cozzolino, S., & Lexer, C. (2011). Sympatric bromeliad species (*Pitcairnia* spp.) facilitate tests of mechanisms involved in species cohesion and reproductive isolation in Neotropical inselbergs. *Molecular Ecology*, *20*(15), 3185–3201. https://doi.org/10.1111/j.1365-294X.2011.05143.x

Papadopulos, A. S., Baker, W. J., Crayn, D., Butlin, R. K., Kynast, R. G., Hutton, I., & Savolainen, V. (2011). Speciation with gene flow on Lord Howe Island. *Proceedings of the National Academy of Sciences*, *108*(32), 13188–13193. https://doi.org/10.1073/pnas.1106085108

Parchman, T. L., Gompert, Z., Braun, M. J., Brumfield, R. T., McDonald, D. B., Uy, J. A. C., … Buerkle, C. A. (2013). The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular Ecology*, *22*(12), 3304–3317. https://doi.org/10.1111/mec.12201

Paten, B., Diekhans, M., Earl, D., St. John, J., Ma, J., Suh, B., & Haussler, D. (2011). Cactus graphs for genome comparisons. *Journal of Computational Biology*, *18*(3), 169–481.

Paten, B., Earl, D., Nguyen, N., Diekhans, M., Zerbino, D., & Haussler, D. (2011). Cactus: Algorithms for genome multiple sequence alignment. *Genome Research*, *21*(9), 1512–1528. https://doi.org/10.1101/gr.123356.111

Payseur, B. A., Krenz, J. G., & Nachman, M. W. (2004). Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. *Evolution*, *58*(9), 2064–2078. https://doi.org/10.1111/j.0014-3820.2004.tb00490.x

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends in Ecology and Evolution*, *16*(7), 351–358. https://doi.org/10.1016/S0169-5347(01)02187-5

Rifkin, J. L. (2017). Population genetics, natural selection and genetic architecture of the selfing syndrome in the morning glory Ipomoea lacunosa (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (1935570052). https://doi.org/10.1534/genetics.106.064469

Ruhsam, M., Hollingsworth, P. M., & Ennos, R. A. (2011). Early evolution in a hybrid swarm between outcrossing and selfing lineages in *Geum*. *Heredity*, *107*(3), 246–255. https://doi.org/10.1038/hdy.2011.9

Rumble, J. R. (Ed.) (2018). *CRC handbook of chemistry and physics*, 98th edn. Boca Raton, FL: CRC Press/Taylor & Francis.

Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology letters*, *8*(3), 336–352. https://doi.org/10.1111/j.1461-0248.2004.00715.x

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., … Brelsford, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, *15*(3), 176–192. https://doi.org/10.1038/nrg3644

Sicard, A., & Lenhard, M. (2011). The selfing syndrome: A model for studying the genetic and evolutionary basis of morphological adaptation in plants. *Annals of Botany*, *107*(9), 1433–1443. https://doi.org/10.1093/aob/mcr023

Sweigart, A. L., & Willis, J. H. (2003). Patterns of nucleotide diversity in two species of *Mimulus* are affected by mating system and asymmetric introgression. *Evolution*, *57*(11), 2490–2506. https://doi.org/10.1111/j.0014-3820.2003.tb01494.x

Taylor, S. A., Curry, R. L., White, T. A., Ferretti, V., & Lovette, I. (2014). Spatiotemporally consistent genomic signatures of reproductive isolation in a moving hybrid zone. *Evolution*, *68*(11), 3066–3081. https://doi.org/10.1111/evo.12510

Teeter, K. C., Payseur, B. A., Harris, L. W., Bakewell, M., Thibodeau, L. M., O'Brien, J. E., … Tucker, P. T. (2008). Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Research*, *18*(1), 67–76.

USDA, NRCS (2019). *The PLANTS Database*. Greensboro, NC: National Plant Data Team. Retrieved from. http://plants.usda.gov (accessed 16 January 2019).

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., … Banks, E. (2013). From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*(11.10), 1–33. https://doi.org/10.1002/0471250953.bi1110s43

Via, S., Conte, G., Mason-Foley, C., & Mills, K. (2012). Localizing FST outliers on a QTL map reveals evidence for large genomic regions of reduced gene exchange during speciation-with-gene-flow. *Molecular Ecology*, *21*(22), 5546–5560. https://doi.org/10.1111/mec.12021

White, B. J., Cheng, C., Simard, F., Costantini, C., & Besansky, N. J. (2010). Genetic association of physically unlinked islands of genomic divergence in incipient species of *Anopheles gambiae*. *Molecular Ecology*, *19*(5), 925–939. https://doi.org/10.1111/j.1365-294X.2010.04531.x

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*(2), 97–159.

Wright, S. (1969). *Evolution and the genetics of populations: Vol. 2. The theory of gene frequencies*. Chicago, IL: University of Chicago Press.

Wu, S., Lau, K. H., Cao, Q., Hamilton, J. P., Sun, H., Zhou, C., … Fei, Z. (2018). Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement. *Nature Communications*, *9*(4580), 1–12. https://doi.org/10.1038/s41467-018-06983-8

Wu, C.-I., & Ting, C.-T. (2004). Genes and speciation. *Nature Reviews Genetics*, *5*, 114–122. https://doi.org/10.1038/nrg1269

Yatabe, Y., Kane, N. C., Scotti-Saintagne, C., & Rieseberg, L. H. (2007). Rampant gene exchange across a strong reproductive barrier between the annual sunflowers. *Helianthus annuus* & *H. petiolaris*. *Genetics*, *175*(4), 1883–1893.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328.

SUPPORTING INFORMATION

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

How to cite this article: Rifkin JL, Castillo AS, Liao IT, Rausher MD. Gene flow, divergent selection and resistance to introgression in two species of morning glories (*Ipomoea*). *Mol Ecol*. 2019;28:1709–1729. https://doi.org/10.1111/mec.14945

APPENDIX

ESTIMATING ADMIXTURE PROPORTIONS

Hanis et al. (1986) describe a method for estimating admixture proportions in a population. This approach assumes complete outcrossing. Here, we show how this procedure can be modified to incorporate the possibility of selfing.

Following Hanis et al. (1986), let there be a population C that represents admixture of two other populations, A and B. Define *m* as the proportional representation of alleles from population A in population C. For a given locus *i*, at which there are two alleles, let the frequency of allele *1* in population A be p_{iA} and the frequency in population B be p_{iB}. Then, the corresponding frequencies of allele *2* are $q_{iA} = (1 - p_{iA})$ and $q_{iB} = (1 - p_{iB})$. The frequency of allele *1* in population C is

and the frequency of allele *2* is

$$p_{iC} = m\,p_{iA} + (1-m)p_{iB}$$

$$q_{iC} = m\,q_{iA} + (1-m)q_{iB}.$$

If population C is completely outcrossing, then the probability that individual *j* is homozygous for allele *1* is

$$P_{ij}(11) = (p_{iC})^2$$

However, if the selfing rate in population C is *s*, then the probability that an individual is homozygous for allele *1* is

$$P_{ij}(11) = (p_{iC})^2 + \mathrm{K},$$

where

$$\mathrm{K} = \frac{s\,p_{iC}\,(1-p_{iC})}{2\left(1-\frac{s}{2}\right)}$$

Holsinger (2017). The analogous probabilities for heterozygous individuals and individuals homozygous for allele *2* are

$$P_{ij}(12) = 2p_{iC}\,q_{iC} - 2\,\mathrm{K}$$

and

$$P_{ij}(22) = (q_{iC})^2 + \mathrm{K}.$$

The likelihood of the data is then

$$L = \prod_i \prod_j P_{ij}\,(G),$$

where $G \in (11, 12, 22)$. The log-likelihood is then

$$\ln L = \sum_i \sum_j \ln P_{ij}\,(G) \tag{1}$$

To find the maximum *ln L*, Equation 3 was evaluated for different combinations of *m* and *s*, with both parameters running between 0.01 and 0.99 using and APL script written by MDR.

ESTIMATING SELFING RATES IN ALLOPATRIC SAMPLES

As above, in a population with sefing rate *s*, at a given locus the probability that an individual *j* will be a particular genotype at locus *i* is given by

$$P_{ij}(11) = (pi)^2 + \mathrm{K}$$

and

$$P_{ij}(12) = 2p_i\,q_i - 2\mathrm{K}$$

$$P_{ij}(22) = (q_i)^2 + \mathrm{K},$$

where p_i is the frequency of allele *1*, $q_i = 1 - p_i$ is the frequency of allele *2*, and K is as defined above. The log-likelihood of the data is then

$$\ln L = \sum_i \sum_j \ln P_{ij}\,(G)$$

as above. To find the maximum-likelihood estimate of *s*, this equation was evaluated for different values of *s* between 0.01 and 0.99 using an APL script written by MDR.