The Genomic Outcomes of Recent and Ancient Hybridization Across a Monkeyflower Radiation

by

Aidan William Short

A dissertation accepted and approved in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in Biology

Dissertation Committee:

Prof. Bill Cresko, Chair

Prof. Matt Streisfeld, Advisor

Prof. Yaniv Brandvain, Core Member

Prof. Andy Kern, Core Member

Prof. Peter Ralph, Institutional Representative

University of Oregon

Fall 2024

© 2024 Aidan William Short
This work is openly licensed via CC BY 4.0.

DISSERTATION ABSTRACT

Aidan William Short

Doctor of Philosophy in Biology

Title: **The Genomic Outcomes of Recent and Ancient Hybridization Across a Monkeyflower Radiation**

New species can form when populations diverge without the homogenizing effects of gene flow. In these cases, genetic and phenotypic differences are able to accumulate, leading to the evolution of reproductive isolation. However, recent genomic analyses have shown that speciation can proceed even with frequent hybridization. This is often due to the effects of natural selection acting against alleles at loci that lead to reduced hybrid fitness. Prior to the evolution of reproductive isolation, hybridization can occur throughout the history of diverging lineages, sometimes facilitating the transfer of beneficial alleles. Thus, by investigating the mechanisms that influence gene flow among hybridizing taxa at various stages of divergence, we can better understand how speciation occurs despite gene flow.

In this thesis, I explore the genomic outcomes of hybridization among taxa in the *Mimulus aurantiacus* species complex, a diverse radiation of wildflowers exhibiting extensive floral and ecological diversity. Leveraging this diversity, I conducted multiple tests for admixture to reconstruct the history of hybridization within this complex. In Chapter 2, I show that ancient hybridization has led to the recurrent evolution of red flowers within the most diverse clade of this radiation. In at least one instance, this ancient hybridization led to strong selection favoring red flowers, which has contributed to the evolution of reproductive isolation. Chapter 3 examines the history of introgression between two hybridizing taxa endemic to the Channel Islands of California, revealing evidence for both recent and ancient hybridization, with stronger selection

against recent introgression. Chapter 4 investigates a contemporary hybrid zone between two recently diverged taxa in southern California, finding that selection against gene flow at barrier loci contributes to their differentiation. Collectively, these findings demonstrate how hybridization occurring at different points in time can result in divergence that can fuel diversification or can lead to the maintenance of species barriers.

This dissertation includes previously published and unpublished coauthored material.

ACKNOWLEDGMENTS

I would like to acknowledge everyone who has helped me throughout my academic journey. My mother for first introducing me to science. My dad for instilling in me a love of books and learning. My brother for always encouraging me to pursue my dreams. My wife, Jingbo Luo for supporting me through my low points. My son, Rowan for helping me to unplug from my research when I need to take a break. Professor David Carlon for convincing me that I had a talent for research and for introducing me to population genetics. Professor Alison Kim Shan Wee for giving me a chance when no one else would and for showing me how to navigate a career in academia. Professor Matt Streisfeld for introducing me to speciation and hybridization genomics and encouraging me to fully pursue my academic interests. Professor Carrie Wessinger, for helping me to plan out my next steps and apply for funding to support my postdoc. My committee for providing me with valuable feedback on my research and progress throughout my doctoral journey. I would like to especially thank Professor Peter Ralph for always being willing to chat with me about research and provide me with valuable suggestions about how I can expand on my current work.

DEDICATION

I would like to dedicate this thesis to my cousin, Donovan X. Brown, who didn't have the chance to achieve his dreams.

Table of Contents

CHAPTER 1: INTRODUCTION

Speciation is the process by which diverging populations become distinct species. Historically, it was believed that speciation occurred primarily in the absence of gene flow. However, the availability of genomic data has revealed that interspecies gene flow (i.e., introgressive hybridization) is a common occurrence during divergence (Green et al. 2010; Gokcumen 2020). While the transfer of foreign genetic material across species boundaries is often deleterious, there is also evidence that introgressive hybridization can, in some cases, be advantageous (e.g., Zhang et al. 2021). Indeed, ancient hybridization has been shown to fuel diversification of entire lineages (Meier et al. 2017). These observations suggest that the outcomes of hybridization will likely vary across the genome and between pairs of species, depending on the specific environmental niches and demographic histories of each species (e.g., Schumer et al. 2018). For diverging lineages to ultimately become distinct species, reproductive barriers must accumulate, eventually resulting in the cessation of gene flow. This raises the question: how can reproductive barriers accumulate when gene flow persists between diverging lineages?

Studying the outcomes of hybridization between lineages at various stages of divergence can provide insight into how reproductive isolation increases throughout the speciation process and how genomic barriers to introgression accumulate over time. For example, laboratory hybridization studies in *Drosophila* and various plant genera have revealed that reproductive isolation generally increases with divergence time. Additionally, investigating hybridization between species at early levels of divergence can help identify the evolutionary factors that affect the outcomes of this hybridization (Roux et al. 2016). Demographic model-based estimates of gene flow in taxa at various levels of divergence, for instance, have demonstrated that taxa at similar divergence levels can exhibit significant variation in gene flow, suggesting that

hybridization outcomes can differ even among closely related pairs. In this thesis, I examine the relationships among hybridization, adaptation, and speciation to study how the impacts of introgression vary over time. Specifically, I investigated natural cases of hybridization between parapatric members of the *Mimulus aurantiacus* species complex.

The *M. aurantiacus* species complex is a radiation of seven closely related woody shrub species distributed throughout California, exhibiting extensive variation in floral traits and ecological habitats. Despite these differences, many of these taxa come into contact in the wild and frequently hybridize, sometimes resulting in the formation of hybrid zones. Much of the research in the *M. aurantiacus* complex has focused on the red and yellow ecotypes of subspecies *puniceus*, which exhibit a phenotypic transition in floral traits along a west-to-east gradient (Streisfeld & Kohn 2005). Specifically, flowers vary extensively in color, with red flowers to the west and yellow flowers to the east. In addition, red ecotype flowers are smaller and have exerted stigmas compared to yellow flowers, which leads to differences in pollinator preference (Streisfeld & Kohn 2007). Beyond floral differences, the ecotypes also vary in traits associated with adaptation to the abiotic environment, specifically drought tolerance and climate (Sobel et al 2019). These trait differences lead to extensive, but incomplete premating isolation (Sobel & Streisfeld 2015). Floral color variation is determined by allelic differences at the *MaMyb2* locus (Streisfeld et al. 2013). The *MaMyb2* allele that confers red flower color was introgressed deep in the evolutionary history of this radiation between ancestral lineages (Stankowski & Streisfeld 2015; Short & Streisfeld 2023). Indeed, this has led to the repeated evolution of red flowers among lineages in this radiation, which I describe in Chapter 2 (Short & Streisfeld 2023). Evidence of widespread selection against gene flow has been identified

between these ecotypes (Stankowski et al. 2023) and among most pairs of taxa across the radiation (Stankowski et al 2019).

In the second chapter of this dissertation, I examine the long-term outcomes of hybridization between subspecies *aridus* and the common ancestor of the most diverse clade within the species complex. This analysis revealed that ancient hybridization led to the transfer of a major effect allele contributing to red flowers, which resulted in the repeated evolution of red flowers in this clade. In at least one case, this ancient hybridization led directly to the divergence between the red and yellow ecotypes. This chapter has been published in Evolution Letters with Matt Streisfeld and I as coauthors.

In the third chapter, I explore how the outcomes of hybridization between two subspecies that are endemic to the Channel Islands of California have varied over time. By conducting multiple tests for introgression between sister taxa at varying levels of divergence, I identified hybridization events occurring throughout the divergence history of the radiation. I then examined the relationship between these introgression events and local recombination rates, revealing stronger selection against recent introgression compared to ancient introgression, likely due to the accumulation of reproductive barriers over time. This chapter is available as a preprint on biorxiv with Matt Streisfeld and I as coauthors.

In the fourth chapter, I investigate the outcomes of hybridization between the red and yellow ecotypes of subspecies *puniceus* in San Diego to identify reproductive barriers between these incipient species. Unlike the previous two chapters, these ecotypes are much more closely related. Therefore, distinguishing between gene flow and shared ancestral polymorphism across the genome requires different approaches. I sequenced individuals collected from allopatric populations of both ecotypes and along a narrow west-to-east transect in San Diego County.

Using the allopatric populations as unadmixed references, I analyzed samples from the hybrid zone to identify genomic regions displaying reduced introgression relative to genome-wide levels. This analysis revealed selection against gene flow in several genomic regions, including one containing the *MaMyb2* locus. These findings illustrate how selection against gene flow can accumulate rapidly between recently diverged taxa and how a few genomic regions under selection can contribute to species maintenance. This chapter will be published with Matt Streisfeld and I as coauthors.

Together, the results of the work completed for this dissertation underscore how the outcomes of hybridization vary with divergence time between hybridizing taxa, how a few reproductive barriers can maintain species differentiation, and how beneficial alleles can be shared between taxa despite widespread selection against gene flow.

1.1 References

1. Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., ... & Pääbo, S. (2010). A draft sequence of the Neandertal genome. *science*, *328*(5979), 710-722.
2. Gokcumen, O. (2020). Archaic hominin introgression into modern human genomes. *American journal of physical anthropology*, *171*, 60-73.
3. Zhang, D., Rheindt, F. E., She, H., Cheng, Y., Song, G., Jia, C., ... & Lei, F. (2021). Most genomic loci misrepresent the phylogeny of an avian radiation because of ancient gene flow. *Systematic Biology*, *70*(5), 961-975.
4. Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature communications*, *8*(1), 14363.

5. Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., ... & Przeworski, M. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, *360*(6389), 656-660.
6. Roux, C., Fraisse, C., Romiguier, J., Anciaux, Y., Galtier, N., & Bierne, N. (2016). Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS biology*, *14*(12), e2000234.
7. Streisfeld, M. A., & Kohn, J. R. (2005). Contrasting patterns of floral and molecular variation across a cline in Mimulus aurantiacus. *Evolution*, *59*(12), 2548-2559.
8. Streisfeld, M. A., & Kohn, J. R. (2007). Environment and pollinator-mediated selection on parapatric floral races of Mimulus aurantiacus. *Journal of Evolutionary Biology*, *20*(1), 122-132.
9. Sobel, J. M., Stankowski, S., & Streisfeld, M. A. (2019). Variation in ecophysiological traits might contribute to ecogeographic isolation and divergence between parapatric ecotypes of Mimulus aurantiacus. *Journal of Evolutionary Biology*, *32*(6), 604-618.
10. Sobel, J. M., & Streisfeld, M. A. (2015). Strong premating reproductive isolation drives incipient speciation in Mimulus aurantiacus. *Evolution*, *69*(2), 447-461.
11. Streisfeld, M. A., Young, W. N., & Sobel, J. M. (2013). Divergent selection drives genetic differentiation in an R2R3-MYB transcription factor that contributes to incipient speciation in Mimulus aurantiacus. *PLoS genetics*, *9*(3), e1003385.
12. Stankowski, S., & Streisfeld, M. A. (2015). Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1814), 20151666.

13. Short, A. W., & Streisfeld, M. A. (2023). Ancient hybridization leads to the repeated evolution of red flowers across a monkeyflower radiation. *Evolution Letters*, *7*(5), 293-304.
14. Stankowski, S., Chase, M. A., McIntosh, H., & Streisfeld, M. A. (2023). Integrating top-down and bottom-up approaches to understand the genetic architecture of speciation across a monkeyflower hybrid zone. *Molecular Ecology*, *32*(8), 2041-2054.
15. Stankowski, S., Chase, M. A., Fuiten, A. M., Rodrigues, M. F., Ralph, P. L., & Streisfeld, M. A. (2019). Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers. *PLoS biology*, *17*(7), e3000391.

CHAPTER 2: Ancient Hybridization Leads to the Repeated Evolution of Red Flowers Across a Monkeyflower Radiation

This work has been published in Evolution Letters. The experimental question and design for this project was developed by me and Professor Matt Streisfeld. Prof. Streisfeld and I performed all the analyses for this work and wrote the manuscript together.

2.1 Introduction

The rapid diversification rates and extraordinary levels of phenotypic variation found in evolutionary radiations provide excellent opportunities to study the processes of adaptation and speciation. Many classic models suggest that adaptation and reproductive isolation evolve due to the accumulation of new mutations within isolated lineages (Barton, 2001; Orr, 1995, 1998). However, the extensive phenotypic diversity found in radiations often arises too quickly to be explained by the long waiting times between new mutations (Hedrick, 2013). Therefore, rapid diversification is likely to depend on pre-existing genetic variation. This variation either can originate in an ancestral population (e.g., Choi et al., 2021), or it can be transferred across species boundaries during divergence due to natural hybridization and introgression (e.g., Han et al., 2017; Meier et al., 2017). Indeed, the recognition that gene flow is common during evolutionary radiations has led some to propose that introgressive hybridization may be an important contributor of rapid speciation (Marques et al., 2019).

Determining the history of hybridization in evolutionary radiations can help us to understand the role that gene flow has played in generating biodiversity. Recent or ongoing hybridization between lineages can lead to the transfer of adaptive genetic variation that directly contributes to reproductive isolation between diverging populations (Pardo-Diaz et al., 2012). Alternatively, hybridization can occur much earlier in a radiation, which can facilitate rapid and

repeated diversification (Meier et al., 2019). The re-use of old genetic variation that predates species splitting events immediately generates large amounts of polymorphism in populations, which can lead to novel phenotypes and drive adaptive divergence and reproductive isolation (Marques et al., 2019). In addition, introgressed alleles will tend to be at higher frequencies than new mutations, making them less likely to be lost to drift (Barrett & Schluter, 2008). Moreover, when these events coincide with novel ecological opportunity, rapid diversification becomes possible. For example, hybridization early in the divergence history of African cichlid fishes (Meier et al., 2017), Darwin's finches (Han et al., 2017), and the Hawaiian silverswords (Barrier et al., 1999) has played a critical role in the rapid diversification of these radiations.

Despite the potential for hybridization to lead to rapid diversification, in most cases, introgressed genetic variation will be neutral or deleterious when shared between species (Mallet, 2005). Deleterious effects arise because of genetic incompatibilities that occur in the recipient's genetic background or due to maladaptation of the transferred genetic material in a divergent ecological environment (Dobzhansky, 1982; Nosil, 2012). Natural selection operating against deleterious introgressed ancestry will result in its removal from populations, the extent of which will be modulated by variation in the local recombination rate (Schumer et al., 2018). Specifically, selection will be highly efficient at removing blocks of introgressed ancestry when they occur in regions of low recombination, as deleterious alleles are less likely to be separated from neutral or beneficial variation (Aeschbacher et al., 2017; Brandvain et al., 2014). This is expected to be more pronounced under highly polygenic architectures of reproductive isolation, as more of the genome acts as a barrier to gene flow (Martin et al., 2019). As a result, selection against gene flow will result in a heterogeneous pattern of introgression across the genome (Ellegren et al., 2012; Liu et al., 2022; Nelson et al., 2021).

Repeated phenotypic transitions within recent radiations provide excellent opportunities to study the evolutionary history of the genetic variation responsible for adaptive trait differences. Over the past decade, an increasing amount of evidence suggests that introgressive hybridization can be an important source of the genetic variation used for repeated trait evolution (e.g., De-Kayne et al., 2022). In this study, we take advantage of the independent evolution of red flower color among the members of the *Mimulus aurantiacus* species complex to determine the genome-wide effects and evolutionary history of hybridization that may have contributed to adaptation and speciation within this radiation.

The *M. aurantiacus* species complex (Phrymaceae) consists of seven closely related, woody shrub subspecies that radiated across California after they diverged from their sister species *M. clevelandii* roughly one million years ago (Stankowski et al., 2019). The subspecies inhabit a diverse range of environments and show remarkable phenotypic differentiation, primarily in their flowers (Figure 1) (Chase et al., 2017). However, all of the taxa are interfertile to varying degrees (McMinn, 1951), and there is evidence for gene flow between many of them (Stankowski et al., 2019). However, the specific patterns and history of gene flow among the members of this complex and their role in reproductive isolation have not been explored.

Most work in this complex has focused on the red- and yellow-flowered ecotypes of subspecies *puniceus*, which display a phenotypic transition along a west to east gradient in San Diego County (hereafter, referred to as the red and yellow ecotypes, respectively). The primary trait that distinguishes the ecotypes is flower color, with allelic variation in the *MaMyb2* gene largely responsible for the derived, anthocyanin pigment found in the flowers of the red, but not yellow, ecotype (Streisfeld et al., 2013). Divergence between these ecotypes is maintained by pollinator-mediated selection, with hummingbirds preferring red flowers and hawkmoths

preferring yellow flowers (Sobel & Streisfeld, 2015; Streisfeld & Kohn, 2007). Moreover, previous sequence data across a portion of the *MaMyb2* gene allowed Stankowski & Streisfeld (2015) to conclude that red flowers arose in the red ecotype because of introgression with a distantly related subspecies in the complex (*M. aurantiacus* subspecies *aridus*). Subspecies *aridus* inhabits rock crevices in desert areas of southern California and northern Baja California and currently has yellow flowers (Beeks, 1962). Previous phylogenetic analyses indi- cate that *aridus* is sister to the red-flowered subspecies *parviflorus* that is endemic to the Channel Islands of California (Chase et al., 2017; Stankowski et al., 2019). This suggests a complex history surrounding the evolution of this important trait and implies that ancestral populations of *aridus* may have had red flowers that were subsequently lost (Stankowski & Streisfeld, 2015).

In addition to the red ecotype, transitions from yellow to red flowers also occurred in other lineages in this radiation. For example, there is a rare, red-colored form that occurs in some populations of the predominantly yellow-flowered subspecies *longiflorus*. In addition, red and yellow flowers are found in a separate population series of subspecies *puniceus* to the north of San Diego (Beeks, 1962) (Figure 1). The red variant of *longiflorus* was described previously as *M. rutilus*, but genetic data reveal that it is nearly indistinguishable from the yellow-flowered form of *longiflorus* (Chase et al., 2017). There is little known about the red- and yellow-flowered variants of *puniceus* to the north, but Beeks (1962) used morphological evidence to suggest that northern and southern *puniceus* should be divided into two population series. Similarly, Streisfeld and Kohn (2005) found that northern and southern *puniceus* were genetically differentiated at two chloroplast DNA markers. The presence of red flowers in these distinct populations of *puniceus* and *longiflorus* thus raises questions about the origin of the genetic variation responsible for red flowers. Specifically, is the same variant shared among these taxa,

which could suggest that a single hybridization event deeper in their evolutionary history may have fueled the repeated evolution of this diversity? Or have there been independent genetic origins of red flowers in these groups?

In this study, we examined the history and genomic consequences of introgression within this radiation to reveal its impact on the repeated evolution of red flowers. We identified evidence of both recurrent and ancient hybridization between lineages, but selection against gene flow appears to have erased much of the genomic signal of this introgression. Nevertheless, we found that introgression and positive selection were responsible for the transfer and subsequent maintenance of a common haplotype in the *MaMyb2* gene between lineages, where it led to the repeated evolution of red flowers. These results reveal that the accumulation of reproductive barriers between divergent taxa can reduce genomic signatures of ancient hybridization, but introgression can still provide the functional variation that facilitates the repeated transitions of phenotypic traits and adaptive divergence between lineages.

2.2 Results and discussion

2.2.1 Multiple transitions from yellow to red flowers among lineages

Previous phylogenetic analyses using both reduced representation and whole genome sequencing revealed four monophyletic clades that delineated the members of the *M. aurantiacus* species complex (Chase et al., 2017; Stankowski & Streisfeld, 2015; Stankowski et al., 2019). Clade A consisted entirely of individuals of *M. a. ssp. grandiflorus* from northeastern California. Clade B included individuals from *M. a. ssp. aridus* from southeastern California and *M. a. ssp. parviflorus* that is endemic to the Channel Islands off the California coast. Clade C comprised samples from *M. a. ssp. aurantiacus* in central and northern California. The highly diverse Clade D from southern California included *M. a. ssp. calycinus, M. a. ssp. longiflorus*, and the red and

yellow ecotypes of *M. a. ssp. puniceus*. Using 10 more whole genome sequences generated here (Supplementary Table S1), we inferred patterns of relatedness among additional red-flowered samples. Specifically, we sequenced four individuals from the red-flowered *M. a. ssp. longiflorus* and three individuals each from red and yellow-flowered populations of the northern *M. a. ssp. puniceus* collected from Orange County. These newly generated sequences were combined with 37 whole genomes from the 7 subspecies and 2 ecotypes (n = 4–5 per taxon) and their sister taxon *M. clevelandii* (n = 3) that were described in Stankowski et al. (2019). After reads were aligned to the *M. aurantiacus* reference genome, we performed variant calling, filtering, and haplotype phasing as described previously (Stankowski et al., 2019), which resulted in a final dataset containing 12,730,568 SNPs. Hereafter, we refer to the taxa only by their subspecies name.

To determine the relationships of these additional samples, we generated a consensus species tree and neighbor-joining splits network that included all 47 samples. Both analyses revealed pat- terns consistent with the four primary clades that were defined previously (Figure 1B, Supplementary Figure S1). The relationships among the 27 samples from clade D were also confirmed using principal components analysis (PCA), which revealed five groups that corresponded to the five taxa of clade D referred to in this study (Figure 1C). Of note, these analyses revealed a clear separation between the northern and southern *puniceus* populations, suggesting that *puniceus* from Orange County is a distinct lineage that likely diverged prior to the split between the red and yellow ecotypes. In addition, even though the Orange County *puniceus* contains populations with either red or yellow flowers, none of the analyses identified separation between flower color morphs from Orange County. Thus, the red and yellow forms of the northern *puniceus* do not appear to be as diverged as the red and yellow ecotypes in San

Diego, but additional work is needed to quantify levels of reproductive isolation between them. In addition, the red and yellow forms of *longiflorus* were interdigitated with each other (Figure 1B), confirming that the red-flowered variants represent a polymorphism within *longiflorus*, rather than a distinct taxonomic entity. Hereafter, the samples from the northern *puniceus* will be referred to as *OC*, for Orange County *puniceus*.

Consistent with previous ancestral state reconstructions that indicated the independent derivation of red flowers (Stankowski & Streisfeld, 2015), the red variants of *longiflorus* and *puniceus* do not group with the other red-flowered taxa (*parviflorus* or the red ecotype). This demonstrates that the red variants in *longiflorus* and *OC* represent additional, independent phylogenetic transitions to red flowers from their yellow-flowered ancestors. However, all red-flowered taxa produce anthocyanin pigments in their flowers, and the red ecotype and *parviflorus* share a common genetic basis controlling the production of red flowers (Stankowski & Streisfeld, 2015). This implies either multiple, convergent origins of red flowers, or that there was a single gain that occurred early in the evolutionary history of clade D, followed by multiple losses.

2.2.2 Genome-wide evidence of gene flow

Although there is evidence that red flowers arose in the red ecotype due to prior introgression of *MaMyb2* with *aridus* (Stankowski & Streisfeld, 2015), there is currently no information on the genomic extent of this gene flow between these taxa. Moreover, these two taxa do not currently come into geographic contact with one another. By contrast, the yellow ecotype does contact populations of *aridus* in certain parts of its range (Thompson, 2005), suggesting that *aridus* may have hybridized with the yellow ecotype more recently than the red ecotype. We tested for genome-wide evi- dence of introgression among these taxa using *D*-statistics (Green et al.,

2010). Patterson's *D* measures asymmetry between the numbers of sites with ABBA and BABA patterns (where A and B are ancestral and derived alleles, respectively) across a phylogeny with three ingroup taxa and an outgroup that have the relation- ship (((P1, P2), P3) O). A significant excess of either pattern gives a nonzero value of *D*, which is taken as evidence that gene flow has occurred between P3 and one of the ingroup taxa.

We calculated Patterson's *D* among all possible triplets of ingroup taxa (Supplementary Table S2, Supplementary Figure S2), but we focus here on differences that occurred when the red or yellow ecotypes were set as P2. Specifically, we set different members of clades C and D as P1, the red or yellow ecotypes as P2, *aridus* as P3, and *M. clevelandii* as the outgroup. In all cases, we found significant evidence for gene flow between *aridus* and the red or yellow ecotypes (Table 1). In addition, we found that values of *D* were consistently higher when the yellow ecotype was P2, regardless of which taxon was used as P1. To the extent that values of *D* are informative of the magnitude of gene flow, this suggests more extensive or more recent introgression has occurred between *aridus* and the yellow ecotype, which would be consistent with their current parapatric distribution. By contrast, even though *parviflorus* also has red flowers, there was no significant evidence for introgression when we repeated these tests using *parviflorus* as P3 (Supplementary Table S2). These findings are consistent with the current geographic distribution of *parviflorus*, which is endemic to the Channel Islands off the coast of California and is isolated from the taxa on the mainland.

2.2.3 Ancient and recurrent hybridization throughout the history of this radiation

Although powerful for detecting genome-wide evidence of gene flow, the *D*-statistic can only identify introgression that has occurred after the two sister taxa used in the test have diverged from each other. This is because genetic variation that was transferred into the ancestor of these

sister taxa would be inherited by both modern-day taxa, resulting in them both sharing alleles with the third taxon (Figure 2A). By taking advantage of the diver- sity of taxa in the *M. aurantiacus* species complex, we ran multiple tests for introgression using successively more distantly related taxa as P1, which allowed us to identify allelic differences that were shared between the red or yellow ecotypes (P2), and *aridus* (P3) that occurred after the split between the P1 and P2 taxa (Figure 2). Thus, by gradually increasing the phylogenetic distance between the two sister taxa used in these tests, we should be able to identify introgression that occurred at various points further back in time.

To obtain estimates of the variation in introgression along the genome, we calculated (*fd*) in 50 kb windows. Patterson's *D* is intended for genome-wide estimates of gene flow, but *fd* provides a measure of the admixture proportion that has been modified specifically for use in genomic windows (Martin et al., 2015). We calculated *fd* repeatedly, each time using one of the five taxa that occurred at different levels of sequence divergence from the red or yellow ecotype as P1. In all tests, the red or yellow ecotypes were used as P2, *aridus* was used as P3, and *M. clevelandii* was the outgroup. This allowed us to determine how levels of introgression varied across the genome at different periods throughout the history of this radiation. Consistent with our calculations of Patterson's *D* (Table 1), the mean *fd* among windows was greater than zero in all tests, regardless of which taxon was used as P1. This suggests that introgression with *aridus* or its ancestor has occurred at multiple points throughout the divergence history of clade D. In addition, gene flow increases significantly with levels of sequence divergence between the P1 taxon and the red or yellow ecotypes (measured as *da*), with mean *fd* nearly doubling when *aurantiacus* is P1 compared to when the red or yellow ecotype is P1 (Figure 3, Supplementary Table S3). This provides clear evidence that introgression dates back to at least the ancestor of

clade D after it diverged from clade C. Although additional taxonomic diversity would be needed to determine if gene flow occurred even further back in the radiation, these results reveal the historical presence of hybridization between lineages.

2.2.4 Widespread selection against gene flow has removed much of the signal of introgression

To determine the genomic consequences of this hybridization, we plotted how the admixture proportion (*fd*) varied across the genome between the parapatric yellow ecotype and *aridus* (see Supplementary Figure S3 for similar results between the red ecotype and *aridus*). This revealed extensive heterogeneity in introgression (Figure 4A), with broad regions of elevated *fd* on each chromosome that were interspersed with regions of lit- tle to no admixture. This same pattern was found regardless of which taxon was set as P1, though the peaks of *fd* tended to vary in height based on levels of divergence with the yellow ecotype (see also Figure 3). Moreover, we found that the distribution of *fd* values among windows was highly skewed, with much of the den- sity near zero (Supplementary Figure S4), indicating that despite a few regions where introgression has been maintained in the genome, little signal of introgression remains.

Given that the entire genome is admixed initially upon hybridization, it is important to determine why some regions of the genome maintain evidence for introgression, and others do not. By characterizing the regions of the genome that fail to introgress between taxa, we can identify loci that act as barriers to gene flow. Until recently, this was commonly attempted by scanning genomes for locally-elevated patterns of genetic divergence, as these regions were expected to contain barrier loci (Ellegren et al., 2012; Turneret al., 2005). However, within-species processes (i.e., background selection and positive selection) that impact genome- wide patterns of effective population size (*Ne*) also can result in similar heterogeneous patterns along

the genome, particularly for taxa that are more diverged and rarely exchange genes (Burri et al., 2015; Cruickshank and Hahn, 2014). This affects our use of divergence statistics that are dependent on *Ne*. By contrast, *fd* directly measures localized variation in admixture across the genome (Martin et al., 2015), so local reductions in *fd* between taxa can indicate the presence of selection against gene flow without the confounding effects of linked selection.

Specifically, natural selection can result in heterogeneous pat- terns of introgression across the genome, because introgressed genetic variants that are deleterious will tend to be eliminated from a population following admixture (Ellegren et al., 2012; Liu et al., 2022; Nelson et al., 2021). It is well known that local rates of recombination also can modulate whether introgressed tracts are maintained, particularly as the number of loci responsible for reproductive isolation increases (Aeschbacher et al., 2017; Brandvain et al., 2014; Martin et al., 2019; Schumer et al., 2018). This is because deleterious variants are expected to become decoupled from neutral variants more rapidly in regions of higher recombination, allowing a more pronounced signal of introgression to be maintained. By contrast, introgressed deleterious and neutral variants will remain in linkage disequilibrium in regions of low recombination, which will make it easier for selection to purge introgressed variation from the population (Aeschbacher et al., 2017; Brandvain et al., 2014; Schumer et al., 2018). Thus, under scenarios where a polygenic architecture contributes to reproductive isolation, we expect a heterogeneous pattern of admixture across the genome, such that the admixture proportion is correlated with levels of genetic divergence and recombination rate (Martin et al., 2019).

We began by examining the relationship between *fd* and genetic divergence (*FST*). *FST* is a relative measure of divergence that reflects differences in allele frequencies between populations. Because gene flow opposes divergence, *FST* will be reduced in genomic regions of

elevated admixture (Martin et al., 2013). Indeed, the elevated *fd* that we detected in certain areas of the genome suggests that there is (or was) gene flow between the yellow ecotype and *aridus*. Given that *fd* is proportional to the effective migration rate (Martin et al., 2015), we expect that selection against gene flow should result in locally elevated *FST* but reduced *fd*. Consistent with these predictions, we found a strong negative correlation between *fd* and *FST*, regardless of which taxon was used as P1, with the strength of the correlation ranging from −0.606 to −0.634 (Figure 4B, Supplementary Figure S3). We also found a strong positive correlation between *fd* and levels of nucleotide diversity (π) (Supplementary Figure S5), which is to be expected given the negative relationship between *FST* and π (Stankowski et al., 2019). In addition, we examined the relationship between recombination rate and admixture by sorting the *fd* values into quantile bins of recombination rate, which were calculated in 500 kb windows based on a genetic linkage map described in Stankowski et al. (2019). We then calculated the mean and 95% confidence intervals for *fd* within each recombination rate quantile bin. We found that, regardless of which taxon was set as P1 for the calculation of *fd*, the two bins with the highest recombination rate (2.33–3.66 cM/Mb and 3.66–10.9 cM/Mb) had significantly higher mean values of *fd* than the three bins with lower recombination rates (0–0.824 cM/Mb, 0.824–1.51 cM/Mb and 1.51–2.33 cM/ Mb) (Figure 4C, Supplementary Figure S3; Supplementary Table S4). Thus, regions of high recombination appear to maintain greater levels of introgressed genetic variation than regions with low recombination, contributing to the heterogeneous pattern that we detected (Figure 4A).

In sum, we found a genome-wide negative correlation between *fd* and *FST*, as well as a positive relationship between *fd* and recombination rate (and π). These results are consistent with previous simulation studies in this system demonstrating that background selection alone was not sufficient to generate these correlations (Stankowski et al., 2019). Rather, the simulations

revealed that selection against gene flow (either due to genetic incompatibilities or local adaptation) was a plausible explanation for the heterogeneous genomic landscapes found in these *Mimulus* taxa. Thus, combined with evidence of introgression between these taxa, the observed relationships suggest that the widespread accumulation of barriers to gene flow has resulted in the removal of introgressed variation across broad swaths of the genome. Although more detailed studies of ecological divergence between these taxa are warranted, reductions in F1 hybrid seed viability and pollen fertility exist in crosses between *aridus* and several members of clade D, indicating the presence of postzygotic barriers (J.M. Sobel, personal communication). Similar patterns have been observed in swordtail fish (Schumer et al., 2018) and *Heliconius* butterflies (Martin et al., 2019), further revealing the importance of polygenic architectures underlying both prezygotic and postzygotic forms of reproductive isolation.

2.2.5 Ancient hybridization leads to repeated transitions of red flowers

Despite the barriers to gene flow that are likely in place now, the current findings indicate that hybridization was possible between ancestral individuals of *aridus* and members of clade D, thus raising the possibility that the *MaMyb2* allele responsible for red flowers was present in the ancestor of this clade. This conclusion would be supported if the same allele was responsible for the repeated transitions to red flowers in the different lineages of clade D. Moreover, we would expect to find common signatures of admixture and genetic divergence among the red-flowered taxa across the genomic region containing *MaMyb2*, and these signatures should not be found in their closely-related yellow-flowered counterparts. Specifically, we would expect to see elevated *fd* but reduced genetic divergence (*dxy*) when the red-flowered samples are compared to samples from *aridus*. By contrast, we predict that there will be elevated divergence in this region of the

genome when red-flowered samples are compared to their most closely related yellow-flowered partners.

We scanned the genomes of these taxa and calculated *fd* and *dxy* in shorter, overlapping 10 kb windows (step size of 100 bp). We calculated *fd* separately for each of the three red-flowered groups in clade D (i.e., red ecotype, red *OC*, and red *longiflorus*), which were each set as P2, and we set *aurantiacus* as P1 and *aridus* as P3. Across a roughly 30 kb region that contains the *MaMyb2* gene, we see a substantial increase in *fd* that is considerably larger than the genome-wide average *fd* values for these comparisons, indicating that numerous sites across this region are admixed between *aridus* and the red taxa (Figure 5A). In addition, *dxy* between red-flow- ered samples from clade D and *aridus* is reduced in this same region relative to their most closely related yellow-flowered partners (i.e., yellow ecotype, yellow-*OC*, and yellow *longiflorus*) (Figure 5B). However, just downstream of this region, the red- and yellow-flowered samples from clade D all show elevated divergence with *aridus*, which is consistent with the phylogenetic pat- terns of relatedness between these taxa. These results indicate that this region corresponds to a common block of introgression found in each of the red-flowered taxa that is absent from their yellow-flowered counterparts. The position along the chromosome where this block ends on the 5′ end remains unclear, as divergence is unusually low for all samples. Additional long-read sequencing will likely be necessary to determine the precise borders of this introgressed block. Finally, we detected elevated divergence between each of the red-flowered taxa and their most closely related yellow-flowered partner in this same region (Figure 5C), consistent with this block of introgression being shared in the red-flowered taxa but not in the yellow-flowered plants.

To demonstrate that the red-flowered taxa share a common sequence at *MaMyb2*, we compared haplotype variation across the entire *MaMyb2* gene in members of the radiation (Figure 5D). Consistent with findings from Stankowski & Streisfeld (2015), all 12 of the red-flowered individuals from clade D included in this study had shared sequences that grouped together with the distantly related *aridus*. Indeed, there was a single common haplotype that contained 9 of 10 sequenced chromosomes from the red ecotype, 5 of 6 red *OC* chromosomes, and 4 of 8 red *longiflorus* chromosomes. By contrast, all yellow-flowered samples from clades C and D grouped together but were positioned distantly from the red-flowered haplotypes. The only exception was a single red-flowered *longiflorus* individual. While one of the two chromosomes grouped with the red-flowered samples, the other chromosome from this individual grouped with haplotypes from yellow *longiflorus*, implying that despite having red flowers, this individual was heterozygous for multiple SNPs within *MaMyb2*. Finally, the common red haplotype was separated by only two mutational steps to the closest *aridus* haplotype, rather than its sister subspecies, the red-flowered *parviflorus*.

Despite the red ecotype and *parviflorus* sharing a common genetic basis for red flowers (Stankowski & Streisfeld 2015), the current data reveal genome-wide, heterogeneous signatures of introgression between *aridus* and members of clade D. Moreover, the *MaMyb2* sequences in the red-flowered taxa are more similar to *aridus* than they are to the red-flowered *parviflorus*, indicating that the introgressed block was donated from an ancestor of *aridus* after the split with *parviflorus.* These findings rule out two possibilities for the origins of red flowers in clade D: (a) that the red allele was introgressed from the ancestor of *aridus* and *parviflorus* into the ancestor of clade D, and (b) that the red allele at *MaMyb2* represents an ancient polymorphism that formed prior to the divergence between clades B and D and has sorted differentially into

descendant lineages. It is also unlikely that *MaMyb2* was recently introgressed between *aridus* and one of the red-flowered taxa followed by ongoing gene flow among populations. The red-flowered populations are largely geographically isolated from each other and are surrounded by yellow-flowered populations, but no evidence of introgression was found in this genomic region in any of the yellow-flowered individuals we sequenced. Moreover, we find that sequence divergence at *MaMyb2* between the red-flowered samples is lower than divergence between the red-flowered taxa and *aridus*, and it is similar to levels of divergence between red-flowered plants across the genome (Supplementary Figure S6), suggesting that introgression in this region did not occur recently. Thus, the most likely explanation for the similarity between *MaMyb2* haplo- types from red-flowered samples and yellow-flowered *aridus* is that red-flowered plants present in ancestral populations of *aridus* hybridized with the ancestor of clade D but have since gone extinct. Although distinguishing between this hypothesis and others for the origins of red flowers will ultimately require finding the causal variant(s) responsible for red flowers in each of these taxa, the current results deepen our understanding of the origins of a haplotype associated with divergence and the early stages of speciation. Specifically, this haplotype originated in an ancestor of *aridus*, was shared with an ancestor of clade D, and has since been maintained in some populations and lost in others.

An additional question related to the maintenance of red flowers in these lineages is whether this shared haplotype shows consistent evidence of natural selection among taxa. Multiple lines of evidence support the presence of geographically widespread and recent positive selection in the red ecotype in San Diego. Specifically, despite ongoing gene flow between the ecotypes, there is a steep geographic cline in both allele frequency and ancestry at this locus (Stankowski et al., 2023; Streisfeld et al., 2013) that shows a unique pattern of divergence

relative to the rest of the genome (Stankowski et al., 2017, 2019). However, it is unclear whether the other red-flowered samples also show evidence of positive selection following introgression.

We explored patterns of recent positive selection in the red-flowered taxa by calculating the site-specific extended haplotype homozygosity (EHH) statistic. Due to a recent selective sweep, a rapid increase in the frequency of a beneficial mutation will result in elevated linkage disequilibrium, leading to extended patterns of homozygosity in haplotypes (Sabeti et al., 2002; Smith & Haigh, 1974; Voight et al., 2006). However, in the absence of selection, we expect haplotypes to break down over time due to new mutations and recombination. Indeed, we detected a broader region of EHH surrounding the *MaMyb2* gene in the red ecotype relative to the yellow ecotype (Figure 5E). Individuals from populations of *OC* with red flowers have similar EHH to the red ecotype, but the haplotypes from red *longiflorus* are considerably shorter (Figure 5F). These findings are consistent with a similar strength or timing of natural selection in the two *puniceus* red-flowered taxa, but the shorter haplotypes in red *longiflorus* suggest that consistent, directional selection on red flowers is weak, absent, or occurred deeper in the past in these populations relative to the red-flowered populations of *puniceus.* Although it is possible this red-flowered variant in *longiflorus* may be lost in the future, as may have been the case in the numerous populations across southern California that currently produce only yellow flowers, some form of balancing selection also may maintain the different flower color variants in these polymorphic populations (Schemske & Bierzychudek, 2001). Future work directly connecting the effects of flower color with fitness differences in the wild will be necessary to evaluate these alternate hypotheses. Regardless, these findings make it clear that the ancient intro- duction of this haplotype into the ancestor of clade D has had important consequences for the evolution of floral diversity in this group.

2.3 Conclusions

The results from this study demonstrate how hybridization can occur early in the history of a radiation, but that widespread selection against gene flow can reduce that signal over time. However, hybridization also can introduce beneficial variation that promotes divergence by facilitating adaptation to new eco- logical niches. The introgression of the *MaMyb2* gene early in the divergence of clade D has fueled the repeated evolution of red flowers in this radiation, which has led to a partial barrier to gene flow between the red and yellow ecotypes. In addition, there also appear to be numerous barriers to gene flow in place between *aridus* and members of clade D, which likely limit ongoing hybridization between these taxa in nature. More broadly, the re-use of old genetic variation has become a leading explanation for rapid diversification in evolutionary radiations (Marques et al., 2019). In some cases, introgression was not the source of this variation, as in threespine sticklebacks that repeatedly colonized freshwater lakes due to the presence of extensive standing variation in the ancestral, marine population (Nelson & Cresko, 2018). Alternatively, radiations of cichlid fish and Hawaiian silverswords appear to have benefited directly from ancient hybridization, which led to the remarkable morphological diversity we see today in those groups (Barrier et al., 1999; Meier et al., 2017). In *Mimulus*, we have shown that recent natural selection has pre- served introgressed haplotypes in the red-flowered taxa, which have been lost in yellow-flowered plants. Thus, hybridization can be a creative force in evolution, but low fitness in hybrids also can lead to the evolution of reproductive barriers, both of which reveal gene flow's important role in promoting and maintaining diversity.

2.4 Materials and methods

2.4.1 Genome resequencing and variant calling

Leaf tissue was collected from the wild (Supplementary Table S1), and DNA from the 10 new samples was extracted, prepared, and sequenced as described in Stankowski et al. (2019). Variant calling, filtering, and haplotype phasing were performed as described in Stankowski et al. (2019) after mapping reads from all 47 samples to the *M. aurantiacus* reference genome. We then further filtered the VCF file for biallelic SNPs and removed sites with missing gen- otype data. The final data set contained 12,730,568 SNPs across all 47 samples.

2.4.2 PCA, network, and phylogenetic analyses

Due to evidence of previous gene flow among taxa in this radiation (Stankowski et al., 2019), we generated a neighbor-joining splits network using *SplitsTree4 v.4.17*.1 (Husonet al., 2008) from all 47 samples. A splits network provides a way to visualize historical events, such as hybridization, in a tree-like framework with- out requiring a fully bifurcating tree. We filtered SNPs for linkage disequilibrium in *Plink* version 1.90 (Chang et al., 2015) to remove variant sites with an *r2* greater than 0.2 in 50 kb windows, sliding by 10 kb, resulting in a data set containing 1,640,850 SNPs. For comparison, we also used IQ-TREE v1.6.12 (Nguyen et al. 2015) to generate a maximum likelihood consensus tree from a con- catenated dataset of all 12,730,568 biallelic SNPs across the 47 samples. Support for each node was generated from 1,000 boot- strap replicates, visualized using FigTree v1.4.4 (Rambaut, 2009), and rooted on the branch leading to the *M. clevelandii* samples. To assess clustering patterns among more closely related samples, we ran a principal components analysis (PCA) in *Plink* using only the 27 samples from clade D.

2.4.3 Tests for genome-wide admixture

We calculated Patterson's *D* for all possible groups of three in-group taxa using the Dtrios command in the *Dsuite* software package (Malinskyet al., 2021) based on the relationships inferred from genome-wide data in Stankowski et al. (2019), with *M. clevelandii* used as the outgroup for all tests. The four samples from subspecies *grandiflorus* were removed prior to this analysis, as the calculation of *D* using these samples does not provide information about the history of introgression between clade D and *aridus.* We corrected for multiple tests using a Bonferroni-corrected alpha of 0.0009.

fd was calculated in 50 kb nonoverlapping genomic windows using the ABBABABAwindows.py Python script (https://github. com/simonhmartin/genomics_general) and smoothed across the genome using the *loess* function from the *scales* package (Wickham & Seidel, 2022) in R with a span of 0.02. To determine the average level of admixed genetic variation between *aridus* and the red or yellow ecotypes using P1 taxa at different levels of taxonomic divergence, we estimated the mean *fd* from each test across 50 kb windows. To estimate the average level of sequence divergence between the red or yellow ecotypes and the taxa from clades C and D, we calculated average *da* in 50 kb windows for each pair of taxa (Nei & Li, 1979). *da* accounts for genetic divergence that predated species divergence by subtracting the aver- age intraspecific pairwise differences (π in both species) from the observed interspecific value (*dxy*). π and *dxy* were calculated using PIXY version 1.2.5 (Korunes & Samuk, 2021) with both variant and invariant sites included. To test for significant differences among the mean *fd* values, we fit a linear model with *fd* as the dependent variable and the P1 taxon as the independent variable using the *lm* function in R and then used the *emmeans* package

(Lenth, 2019) to perform pairwise comparisons of the estimated marginal mean *fd* values for the different P1s from the linear model.

We estimated the Spearman's rank correlation coefficient between *fd*, *FST*, and π using the *cor.test* function in R. *FST* was calculated in each 50 kb window between the red or yellow ecotype and one of the five remaining taxa from clades C and D using the *popgenwindows.py* Python script (https://github.com/simonhmartin/genomics_general), with only variant sites included. Finally, we sorted the *fd* values calculated in 50 kb windows into quantile bins based on recombination rates estimated in 500 kb windows in Stankowski et al. (2019). We then calculated the mean and 95% confidence intervals of the *fd* values within each recombination rate quantile bin. To test for differences in *fd* among recombination bins, we fit a linear model with *fd* as the dependent variable and recombination bin as the independent variable and then performed pairwise comparisons of the esti- mated marginal mean *fd* values for the different recombination bins from the linear model.

2.4.4 Admixture and genetic divergence across the *MaMyb2* region

To determine the history of the red allele at *MaMyb2* in the different lineages of clade D, we calculated *fd* and *dxy* as described above, but in overlapping 10 kb windows with 100 bp steps. *fd* was calculated three times, each time with *aurantiacus* as P1 and *aridus* as P3, but varying the different red samples (red ecotype, red-OC, and red *longiflorus*) as P2. *dxy* was calculated between each of the red taxa and *aridus*, each of their closest yellow-flowered partners and *aridus*, between each of the red-flowered taxa, and between each of the red taxa and their most closely related yellow partners.

2.4.5 Haplotype network

We used VCFx version: 2.0.6b to generate a FASTA file of the entire *MaMyb2* gene (positions 12,317,113 to 12,318,500 on chromosome 4) from the VCF file from all 47 samples. This FASTA file was then used to construct a haplotype network from all recovered haplo- types based on an infinite site model and uncorrected distances using *Pegas* version 1.1 (Paradis, 2010) in R version 3.6.3.

2.4.6 Extended haplotype homozygosity

We calculated the site-specific extended haplotype homozygosity statistic for a SNP in the middle of the *MaMyb2* gene (position 12,317,808) using *rehh* version 3.2.2 (Gautier & Vitalis, 2012). Separate VCFs were created that contained samples from the red ecotype, the yellow ecotype, red *OC*, or red *longiflorus*. Haplotype and marker information was extracted from each VCF file using the *data2haplohh* function of the *rehh* package. EHH was calculated from the marker and haplotype information for each of the taxa using the *calc_ehh* function in *rehh*.

2.5 Bridge

In this chapter, I explored the outcomes of ancient hybridization resulting in the repeated evolution of red flowers across the most diverse clade of taxa in the species complex. Despite the evidence for widespread selection against gene flow between these lineages, I found that ancient hybridization also introduced an allele contributing to the production of red flowers that has been maintained in at least three distinct lineages. Furthermore, this allele shows evidence for strong, recent positive selection, likely due to differences in pollinator preferences between the red and yellow ecotypes of subspecies *puniceus*. The results of this chapter reveal how the diversity present within species radiations can be used to identify introgression that occurred throughout an evolutionary radiation's divergence history. In the following chapter, I use the same logic to

identify evidence of introgression from different time points and then expand upon the results to estimate how the fitness consequences of hybridization between two lineages have changed over time.

2.6 Figures and Tables

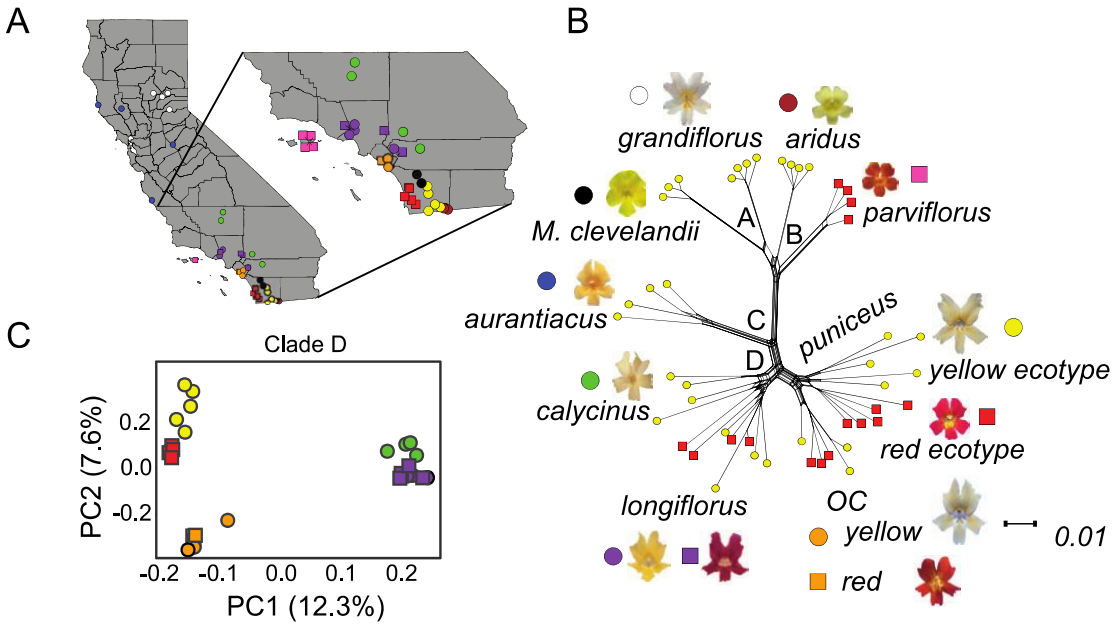

Figure 1. Geographic sampling and patterns of relatedness among members of the M. aurantiacus species complex. (A) Locations of the samples used for whole genome sequencing from the seven subspecies of the *M. aurantiacus* species complex and its sister species, *M. clevelandii* across California. Colors correspond to the different taxa, with *M. clevelandii* in black, *grandiflorus* in white, *aridus* in brown, *aurantiacus* in blue, *calycinus* in green, *longiflorus* in purple, *parviflorus* in pink, Orange County *puniceus* in orange, the red ecotype of *puniceus* from San Diego in red and the yellow ecotype of *puniceus* from San Diego in yellow. Circles correspond to plants with yellow flowers and squares correspond to plants with red flowers. (B) Splitstree network showing the evolutionary relationships of the taxa, with representative photographs of their flowers. The four major clades of the radiation are labelled with black letters. Symbols next to each photograph match those in panel A. Tips of the tree are colored according to the flower color of the individual plant, with yellow circles denoting yellow flowers and red squares corresponding to red flowers. (C) Plot of the first two principal component axes from taxa in Clade D. The percent variation explained by each axis is reported. Shapes and colors match those in (A).

Table 1. *D*-statistics, with associated *Z*-scores and *p*-values calculated using either the red or yellow ecotypes as P2, *aridus* as P3, *M. clevelandii* as the outgroup, and various taxa from clades C and D as P1.

P1	P2	P3	D-Statistic	Z-score	p-value
aurantiacus	red ecotype	*aridus*	0.072	5.571	2.5×10^{-8}
calcyinus	red ecotype	*aridus*	0.052	4.889	1.0×10^{-8}
longiflorus	red ecotype	*aridus*	0.074	6.305	2.9×10^{-10}
OC	red ecotype	*aridus*	0.044	6.578	4.8×10^{-11}
aurantiacus	yellow ecotype	*aridus*	0.098	7.929	2.2×10^{15}
calcyinus	yellow ecotype	*aridus*	0.082	7.924	2.3×10^{-15}
longiflorus	yellow ecotype	*aridus*	0.103	9.502	0
OC	yellow ecotype	*aridus*	0.074	11.646	0
red ecotype	yellow ecotype	*aridus*	0.035	5.366	8.1×10^{-8}

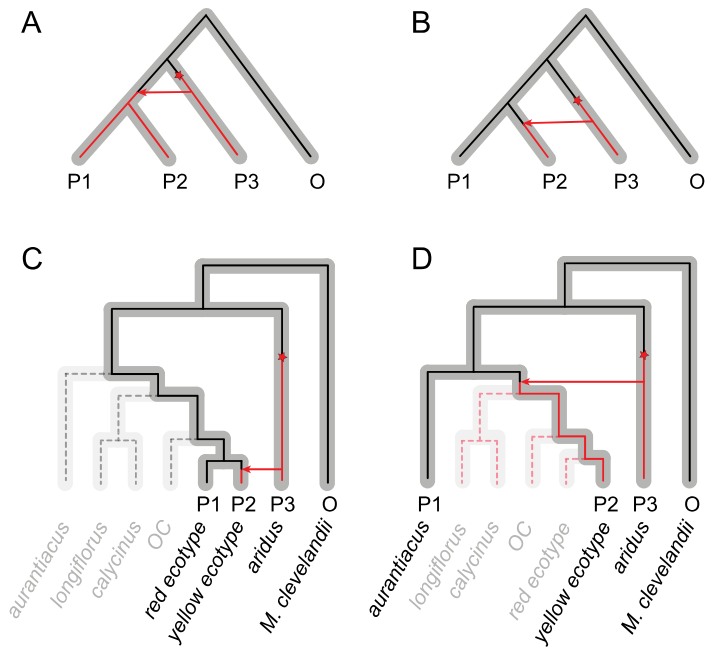

Figure 2. Evolutionary radiations can help evaluate the history of introgression. (A–B) In a four taxon tree, with three ingroups (P1–P3) and an outgroup (O), introgression between P2 and P3 can only be identified if gene exchange occurred after the split between the two sister taxa (P1 and P2). The species trees are in gray, and the ancestral (black) and derived (red) alleles at a locus are indicated. (A) A mutation (red star) that occurs in the lineage leading to P3 that is transferred via introgression into an ancestor of P1 and P2 will result in the sharing of that genetic information in both P1 and P2. As a consequence, this deeper history of introgression will be undetectable using tests, such as D-statistics. (B) However, introgression between P2 and P3 can be detected if the genetic exchange occurred after P1 and P2 diverged from each other, as P1 would retain the ancestral sequence, but P2 would contain the derived variant that is shared with P3. (C–D) The species tree from the *M. aurantiacus* radiation is presented in gray, but tips are grayed out to show only the four taxa used in the tests of introgression (denoted by P1–P3, and O). By using different P1 taxa at increasing levels of divergence with the yellow ecotype, we can track introgression that occurred deeper in time. (C) Only genetic variation that was introgressed after the divergence of the red and yellow ecotypes can be identified when the red ecotype is used as P1 and the yellow ecotype is used as P2. The black dashed line represents the retention of the ancestral allele in the taxa not being used in the test. (D) When a more distant taxon, such as *aurantiacus*, is used as P1, genetic variation that was introgressed in all of clade D can be identified.

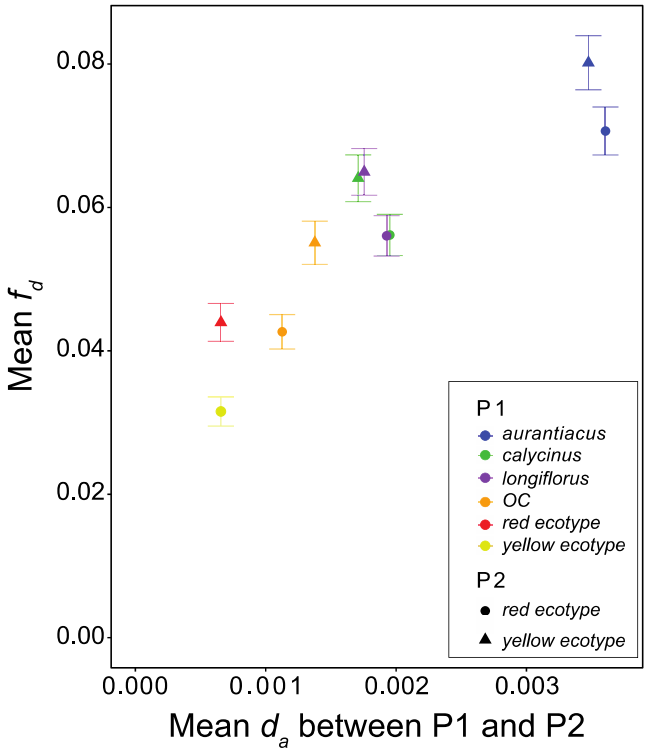

Figure 3. A long history of gene flow in the radiation. Mean and 95% confidence intervals of the admixture proportion (f*d*) calculated in 50 kb windows are plotted against mean levels of sequence divergence (*da*) between the red or yellow ecotypes and taxa in clades C and D. *fd* was calculated repeatedly, with one of five different taxa as P1, the red or yellow ecotype as P2, *aridus* as P3, and *M. clevelandii* as the outgroup. The increase in mean *fd* with divergence time indicates that admixture with the lineage leading to *aridus* has occurred throughout the radiation, and as far back as the ancestor of clade D.

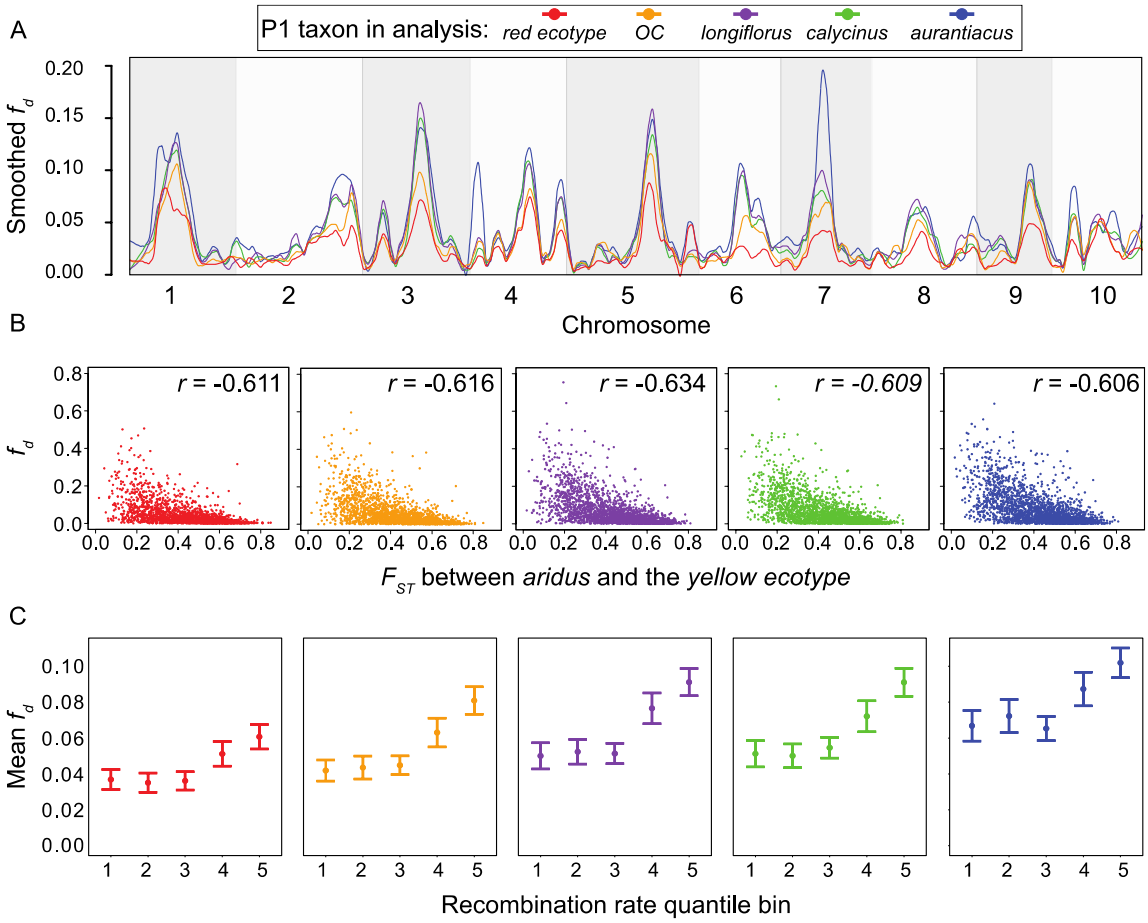

Figure 4. Heterogeneous introgression across the genome indicates selection against gene flow. (A) Loess-smoothed f*d* values plotted across the 10 chromosomes of the *M. aurantiacus* genome, with the yellow ecotype as P2, *aridus* as P3, and one of five different taxa as P1. Colors indicate the taxa from clades C and D used as P1. (B) Scatterplots showing the relationship between *FST* and *fd* in these same 50 kb windows, with the correlation coefficient between the statistics in the upper right hand corner of each plot. (C) Mean and 95% confidence intervals of *fd* in different quantile bins of recombination rate. Quantile bins of recombination rate are as follows: 1 = 0–0.824 cM/Mb; 2 = 0.824–1.51 cM/Mb; 3 = 1.51–2.33 cM/Mb; 4 = 2.33–3.66 cM/Mb; 5 = 3.66–10.9 cM/Mb.

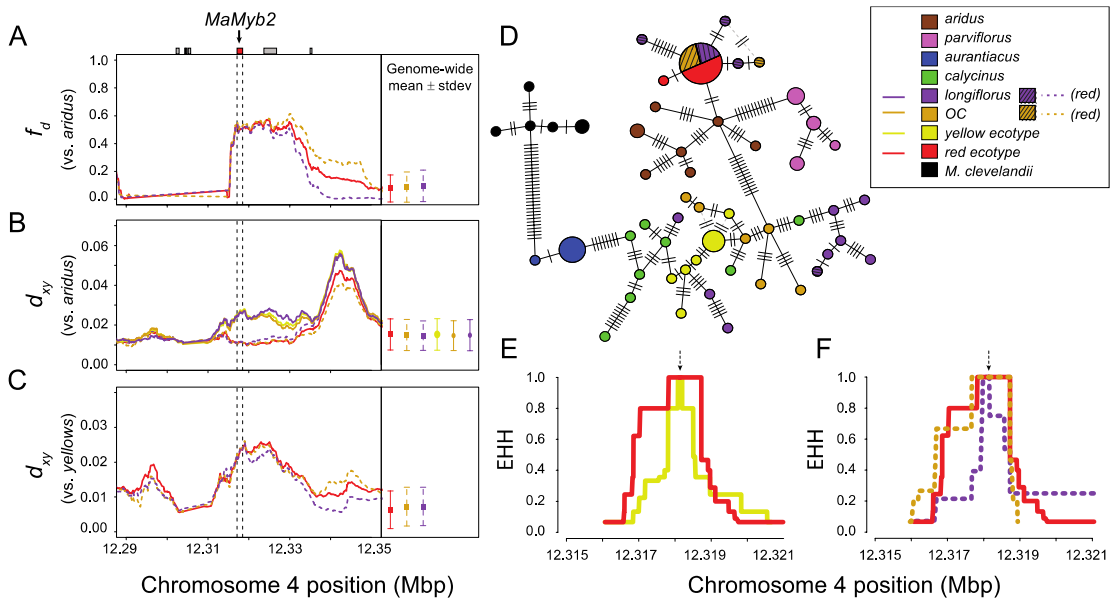

Figure 5. Ancient hybridization leads to the repeated evolution of red flowers. (A–C) On the left, the admixture proportion (f*d*) and genetic divergence (*dxy*) are plotted across a 60 kb region surrounding the *MaMyb2* gene on chromosome 4. The dashed vertical line represents the start and stop positions of *MaMyb2*, and gray boxes at the top correspond to the position of additional genes in this region. On the right of each plot is the genome-wide mean ± standard deviation of each statistic, calculated in 10 kb windows and 100 bp steps across all 10 chromosomes. (A) *fd* is calculated using the red-flowered samples from clade D as P2 (red ecotype, red *OC*, and red *longiflorus*), *aurantiacus* as P1, and *aridus* as P3. (B) *dxy* between the red-flowered samples and *aridus* (dotted lines) is low across the *MaMyb2* region, but *dxy* between *aridus* and each of their closely related yellow samples (yellow ecotype, yellow *OC*, and yellow *longiflorus*) is elevated in this same region (solid lines). (C) *dxy* calculated between the three sets of red-flowered samples and their most closely related yellow-flowered partners is elevated across this same region compared to genome-wide averages. (D) A haplotype network showing the number of sequence differences among unique *MaMyb2* haplotypes in all taxa. Each haplotype is represented by a circle, the size of which is proportional to its observed frequency. Black dashes show the number of mutational steps separating haplotypes. Alternate connections among haplotypes are shown as dashed, gray bars. (E) Site–specific extended haplotype homozygosity (EHH) for a site in the middle of *MaMyb2* (black arrow) decays more slowly for the red ecotype compared to the yellow ecotype. (F) EHH also decays more slowly for the red ecotype and red *OC* samples than for the red *longiflorus* samples.

Table S1. Sample information for the 47 sequenced individuals used in this study. The table includes taxon identity, sampling location, percent read alignment, and average sequencing depth. Samples in red text were sequenced as a part of this study, while those in black were described in Stankowski et al (2019).

Sample	Taxon	Latitude	Longitude	% Reads aligned	Seq-Depth
159_83	ssp. *aridus*	32.6630	-116.2230	91.7	21.12
159_84	ssp. *aridus*	32.6630	-116.2230	89.3	21.98
195_1	ssp. *aridus*	32.6300	-116.1429	92.6	20.20
T84	ssp. *aridus*	32.6526	-116.2449	87.2	21.75
T102	ssp. *aurantiacus*	39.0424	-122.7727	94.9	23.74
T104	ssp. *aurantiacus*	39.2045	-123.7646	94.6	25.09
T50	ssp. *aurantiacus*	35.9865	-121.4928	88.3	24.36
T92	ssp. *aurantiacus*	37.8459	-120.6110	94.0	15.16
T144	ssp. *calycinus*	34.1929	-117.2784	93.2	26.00
T150	ssp. *calycinus*	33.8564	-116.8481	94.7	24.02
T90	ssp. *calycinus*	35.5918	-118.5052	91.3	19.97
T91	ssp. *calycinus*	35.3172	-118.5871	95.5	27.91
T101	ssp. *grandiflorus*	39.5536	-121.4301	92.0	16.05
T61	ssp. *grandiflorus*	39.5590	-120.8243	91.6	17.31
T96	ssp. *grandiflorus*	39.0122	-120.7552	92.0	28.21
T99	ssp. *grandiflorus*	39.4376	-121.0599	91.4	23.84
DPR_Y3	ssp. *longiflorus*, yellow	33.7459	-117.4485	96.0	26.88
SS_Y16	ssp. *longiflorus*, yellow	34.2722	-118.6100	94.2	30.86
T33_1	ssp. *longiflorus*, yellow	34.3438	-118.5099	94.6	18.87
T8_8	ssp. *longiflorus*, yellow	34.1347	-118.6452	82.6	25.11
DPR_R12	ssp. *longiflorus*, red	33.7459	-117.4485	89.8	18.52
MBR_R3	ssp. *longiflorus*, red	34.2056	-117.6765	94.4	23.20
SS_R19	ssp. *longiflorus*, red	34.2722	-118.6100	91.3	23.89
WTF_R1	ssp. *longiflorus*, red	34.2174	-117.7630	94.2	23.45
KK168	ssp. *parviflorus*	34.0180	-119.6730	91.8	23.66
KK161	ssp. *parviflorus*	34.0180	-119.6730	92.0	19.11
KK180	ssp. *parviflorus*	34.0180	-119.6730	92.4	18.18
KK182	ssp. *parviflorus*	34.0193	-119.6802	91.3	19.46
MTH17	ssp. *puniceus*, *OC*, red	33.6414	-117.8252	90.6	21.23
MTH15	ssp. *puniceus*, *OC*, red	33.6414	-117.8252	91.1	18.85
RPR16	ssp. *puniceus*, *OC*, red	33.6058	-117.8020	95.1	21.68
LFP13	ssp. *puniceus*, *OC*, yellow	33.6532	-117.6579	90.3	22.23
VCR14	ssp. *puniceus*, *OC*, yellow	33.4871	-117.6489	92.2	19.97
VCR17	ssp. *puniceus*, *OC*, yellow	33.4871	-117.6489	90.7	21.38
ELF	ssp. *puniceus*, red	33.0860	-117.1453	93.0	18.20
JMC	ssp. *puniceus*, red	32.7373	-116.9541	93.8	19.06
LH	ssp. *puniceus*, red	33.0609	-117.1188	87.1	19.77

MT	ssp. *puniceus*, red	32.8210	-117.0618	93.7	20.85
UCSD	ssp. *puniceus*, red	32.8894	-117.2362	87.0	18.23
BCRD	ssp. *puniceus*, yellow	32.9496	-116.6380	94.6	20.85
INJ	ssp. *puniceus*, yellow	33.0979	-116.6643	93.1	18.83
LO	ssp. *puniceus*, yellow	32.6767	-116.3312	93.4	18.04
PCT	ssp. *puniceus*, yellow	32.7326	-116.4698	92.3	19.68
POTR	ssp. *puniceus*, yellow	32.6038	-116.6339	90.5	19.27
CLV_GH	*M. clevelandii*	33.1589	-116.8122	92.3	21.31
CLV_11	*M. clevelandii*	33.3391	-116.9325	84.4	15.52
CLV_4	*M. clevelandii*	33.3391	-116.9325	89.3	17.31

Table S2. Patterson's D statistics, estimated using the *Dtrios* function in the *Dsuite* package, for all trios of taxa included in this study. In all cases, *M. clevelandii* is used as the outgroup, and the three ingroup taxa are indicated (P1, P2, P3). The identity of the P1 and P2 taxon is adjusted, so that the D-statistics are always positive. Counts of each site category (BBAA, ABBA, and BABA) are indicated.

P1	P2	P3	D stat	Z-score	p-value	f4-ratio	BBAA	ABBA	BABA
aurantiacus	*calycinus*	*aridus*	0.0247	3.5809	3.42E-04	0.0085	503634	144599	137624
longifloru s	*calycinus*	*aridus*	0.0231	5.6748	1.39E-08	0.0066	585489	120190	114762
aurantiacus	*longifloru s*	*aridus*	0.0055	0.7555	4.50E-01	0.0019	500695	140659	139113
aurantiacus	*OC*	*aridus*	0.0393	3.4859	4.91E-04	0.014	484366	152419	140898
calycinus	*OC*	*aridus*	0.0169	2.0655	3.89E-02	0.0056	531413	136476	131930
longifloru s	*OC*	*aridus*	0.0389	4.2424	2.21E-05	0.0122	545247	133213	123239
aurantiacus	*Red ecotype*	*aridus*	0.072	5.5712	2.53E-08	0.0264	476747	161375	139700
calycinus	*Red ecotype*	*aridus*	0.0524	4.8891	1.01E-06	0.018	515736	147655	132955
longifloru s	*Red ecotype*	*aridus*	0.0743	6.3053	2.88E-10	0.0245	525314	145558	125430
OC	*Red ecotype*	*aridus*	0.044	6.5784	4.76E-11	0.0125	590243	120491	110336
aurantiacus	*Yellow ecotype*	*aridus*	0.0976	7.9287	2.21E-15	0.0366	472730	169291	139178
calycinus	*Yellow ecotype*	*aridus*	0.0817	7.9242	2.30E-15	0.0284	517873	153126	129988
longifloru s	*Yellow ecotype*	*aridus*	0.1033	9.5022	0.00E+00	0.0348	524357	152595	124029
OC	*Yellow ecotype*	*aridus*	0.0741	11.6458	0.00E+00	0.0229	563239	134736	116144
Red ecotype	*Yellow ecotype*	*aridus*	0.0354	5.3659	8.06E-08	0.0105	586759	123454	115016
longifloru s	*calycinus*	*aurantiacus*	0.0142	3.0375	2.39E-03	0.0152	264814	158551	154123
OC	*calycinus*	*aurantiacus*	0.0646	12.9041	0.00E+00	0.0728	226218	185771	163228
Red ecotype	*calycinus*	*aurantiacus*	0.081	13.8921	0.00E+00	0.0916	217043	193298	164335
Yellow ecotype	*calycinus*	*aurantiacus*	0.0914	16.1788	0.00E+00	0.1016	222638	193789	161330
OC	*longifloru s*	*aurantiacus*	0.0542	7.6557	1.92E-14	0.0585	238017	176044	157929

Red ecotype	*longifloru s*	*aurantiac us*	0.071	9.0906	1.08E-19	0.077 6	22481 3	18496 3	16042 9
Yellow ecotype	*longifloru s*	*aurantiac us*	0.081	12.616 7	0.00E+0 0	0.087 8	22742 4	18713 0	15910 0
aridus	*parvifloru s*	*aurantiac us*	0.196 1	14.021 3	0.00E+0 0	0.134 8	28531 4	26836 0	18037 3
Red ecotype	*OC*	*aurantiac us*	0.023	4.3617	1.29E-05	0.020 3	29056 0	14260 1	13618 1
Yellow ecotype	*OC*	*aurantiac us*	0.033	4.9649	6.87E-07	0.031	27035 4	15520 5	14528 9
Yellow ecotype	*Red ecotype*	*aurantiac us*	0.012 6	2.2007	2.78E-02	0.010 9	29720 6	14083 4	13733 9
aridus	*parvifloru s*	*calycinus*	0.174 8	10.941 7	0.00E+0 0	0.157 9	28580 7	26741 5	18784 1
Red ecotype	*OC*	*calycinus*	0.055 5	9.6966	0.00E+0 0	0.138 4	24372 7	15875 8	14205 5
Yellow ecotype	*OC*	*calycinus*	0.035 6	6.1916	5.96E-10	0.100 2	22090 4	16874 6	15714 8
Red ecotype	*Yellow ecotype*	*calycinus*	0.017 1	3.8015	1.44E-04	0.042 1	25057 0	15201 1	14690 7
aridus	*parvifloru s*	*longifloru s*	0.185	12.901 8	0.00E+0 0	0.157 2	28749 1	26769 1	18409 7
Red ecotype	*OC*	*longifloru s*	0.072 3	14.543 3	0.00E+0 0	0.167 8	23192 8	16405 6	14193 2
Yellow ecotype	*OC*	*longifloru s*	0.065 6	13.047	0.00E+0 0	0.165	21126 9	17620 9	15452 9
Red ecotype	*Yellow ecotype*	*longifloru s*	0.001 5	0.327	7.44E-01	0.003 4	24255 6	15101 3	15056 9
calycinus	*aurantiac us*	*parvifloru s*	0.004 8	0.6611	5.09E-01	0.002 4	42815 4	15013 0	14869 2
longifloru s	*aurantiac us*	*parvifloru s*	0.009 5	1.0433	2.97E-01	0.004 7	42517 8	15158 3	14873 7
OC	*aurantiac us*	*parvifloru s*	0.021 3	2.2775	2.28E-02	0.010 7	41260 4	15712 3	15055 6
Red ecotype	*aurantiac us*	*parvifloru s*	0.021 4	2.5212	1.17E-02	0.010 8	40698 5	15792 4	15130 6
Yellow ecotype	*aurantiac us*	*parvifloru s*	0.019 9	2.3536	1.86E-02	0.010 2	40574 4	16017 9	15393 3
longifloru s	*calycinus*	*parvifloru s*	0.005 7	0.9576	3.38E-01	0.002 3	50981 8	12409 3	12268 5
OC	*calycinus*	*parvifloru s*	0.018 6	2.9785	2.90E-03	0.008 4	46016 3	14025 4	13512 5
Red ecotype	*calycinus*	*parvifloru s*	0.018 3	3.2225	1.27E-03	0.008 4	44707 8	14387 1	13869 1
Yellow ecotype	*calycinus*	*parvifloru s*	0.017	3.2455	1.17E-03	0.007 8	45194 9	14363 8	13883 0
OC	*longifloru s*	*parvifloru s*	0.013 9	1.8662	6.20E-02	0.006 1	47369 4	13528 0	13155 8

Red ecotype	*longifloru s*	*parvifloru s*	0.013 7	2.5911	9.57E-03	0.006 1	45635 5	14006 5	13629 2
Yellow ecotype	*longifloru s*	*parvifloru s*	0.012 2	2.2038	2.75E-02	0.005 5	45798 8	14125 4	13785 4
Red ecotype	*OC*	*parvifloru s*	0.000 2	0.0429	9.66E-01	0.000 1	52494 6	11493 9	11488 8
aurantiac us	*Yellow ecotype*	*parvifloru s*	-0.019 9	2.3536	1.86E-02	0.010 2	40574 4	16017 9	15393 3
calycinus	*Yellow ecotype*	*parvifloru s*	-0.017	3.2455	1.17E-03	0.007 8	45194 9	14363 8	13883 0
longifloru s	*Yellow ecotype*	*parvifloru s*	-0.012 2	2.2038	2.75E-02	0.005 5	45798 8	14125 4	13785 4
OC	*Yellow ecotype*	*parvifloru s*	0.001 3	0.2585	7.96E-01	0.000 5	50063 7	12376 2	12344 1
OC	*Yellow ecotype*	*parvifloru s*	0.001 3	0.2585	7.96E-01	0.000 5	50063 7	12376 2	12344 1
Red ecotype	*Yellow ecotype*	*parvifloru s*	0.001 6	0.6136	5.40E-01	0.000 6	52721 8	11554 0	11516 8
Red ecotype	*Yellow ecotype*	*parvifloru s*	0.001 6	0.6136	5.40E-01	0.000 6	52721 8	11554 0	11516 8
calycinus	*longifloru s*	*OC*	0.052 1	6.8952	5.38E-12	0.092 1	22124 9	17253 1	15543 4
aridus	*parvifloru s*	*OC*	0.153 4	8.9739	3.25E-19	0.120 4	28631 2	26279 0	19289 2
Yellow ecotype	*Red ecotype*	*OC*	0.100 6	13.849 2	0.00E+00	0.245 7	18141 1	17942 0	14660 9
calycinus	*longifloru s*	*Red ecotype*	0.036 8	4.7774	1.78E-06	0.072 5	23525 5	16441 3	15273 7
aridus	*parvifloru s*	*Red ecotype*	0.13	7.232	4.76E-13	0.112 9	28307 1	25949 8	19980 5
calycinus	*longifloru s*	*Yellow ecotype*	0.022	3.1743	1.50E-03	0.057 6	23365 0	16325 1	15623 6
aridus	*parvifloru s*	*Yellow ecotype*	0.111 4	6.1979	5.72E-10	0.106 2	28077 0	25757 0	20594 2

Table S3. The *t*-ratio from a linear mixed-effects model showing the difference in admixture proportion (*fd*) using different taxa as P1, with the yellow ecotype as P2, *aridus* as P3 and *M. clevelandii* as the outgroup. Statistical significance is denoted as: *** for $p < 0.001$, ** for $p < 0.01$ and * for $p < 0.05$.

Factor Level	**Estimate**	***df***	***t*-ratio (Fd)**
aurantiacus vs *calycinus*	0.016113	11102	6.945***
aurantiacus vs *longiflorus*	0.015227	11102	6.65***
aurantiacus vs *OC*	0.025113	11102	10.757***
aurantiacus vs red ecotype	0.036214	11102	15.363***
calycinus vs *longiflorus*	-0.000885	11102	-0.387
calycinus vs *OC*	0.009	11102	3.862**
calycinus vs red ecotype	0.020101	11102	8.542***
longiflorus vs OC	0.009885	11102	4.298***
longiflorus vs *OC*	0.020987	11102	9.033***
OC vs red ecotype	0.011102	11102	4.69***

Table S4. The *t*-ratio from a linear mixed-effects model showing the difference in admixture proportion (*fd*) due to recombination rate. Quantile bins of recombination rate (in cM/Mb) are presented. Statistical significance is denoted as: *** for p < 0.001, ** for p < 0.01, and * for p < 0.05.

Factor Level	**Estimate**	***df***	***t*-ratio (Fd)**
[0,0.824] vs (0.824,1.51]	-0.001286	10577	-0.526
[0,0.824] vs (1.51,2.33]	-0.001071	10577	-0.446
[0,0.824] vs (2.33,3.66]	-0.020672	10577	-8.687***
[0,0.824] vs (3.66,10.9]	-0.035771	10577	-15.431***
(0.824,1.51] vs (1.51,2.33]	0.000215	10577	0.089
(0.824,1.51] vs (2.33,3.66]	-0.019386	10577	-8.089***
(0.824,1.51] vs (3.66,10.9]	-0.034486	10577	-14.766***
(1.51,2.33] vs (2.33,3.66]	-0.019601	10577	-8.33***
(1.51,2.33] vs (3.66,10.9]	-0.034700	10577	-15.149***
(2.33,3.66] vs (3.66,10.9]	-0.015099	10577	-6.654***

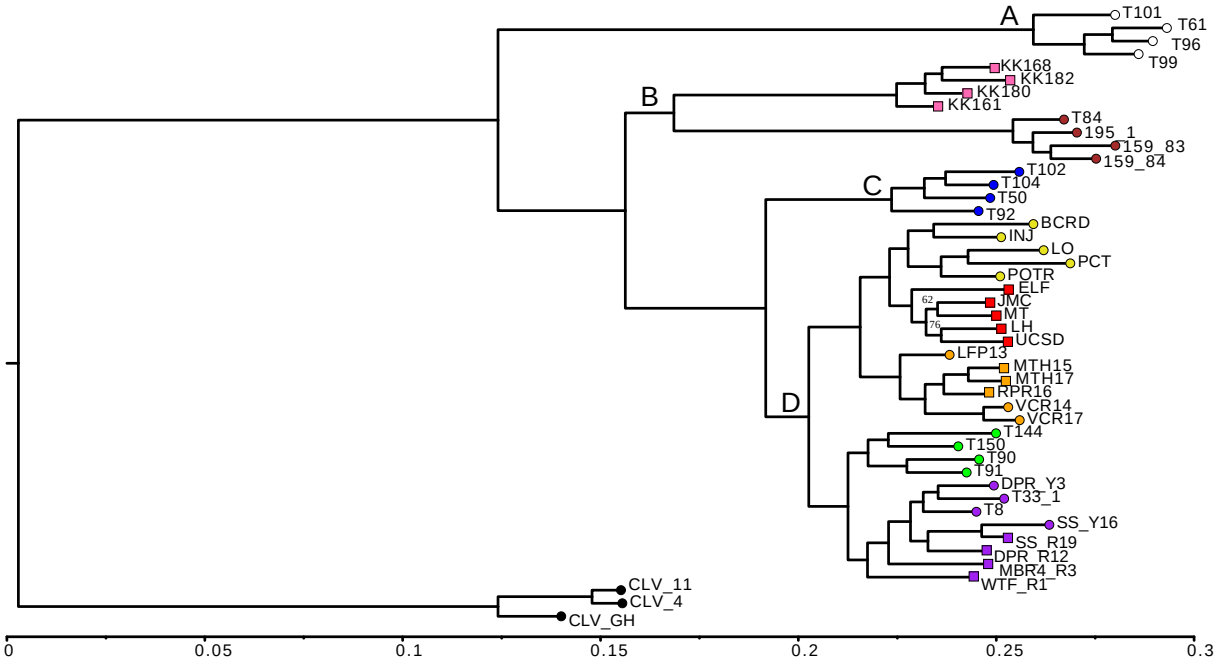

Fig. S1. Maximum likelihood consensus species tree generated from all variant sites among the 47 samples and estimated using IQ-TREE v1.6.12. Colors at the tips vary by taxon, and are the same as used in Fig 1 of the main text. Circles correspond to plants with yellow flowers and squares denote plants with red flowers. The four primary clades are noted with letters A-D. Nodes show 100% support from 1000 bootstrap replicates, except where indicated.

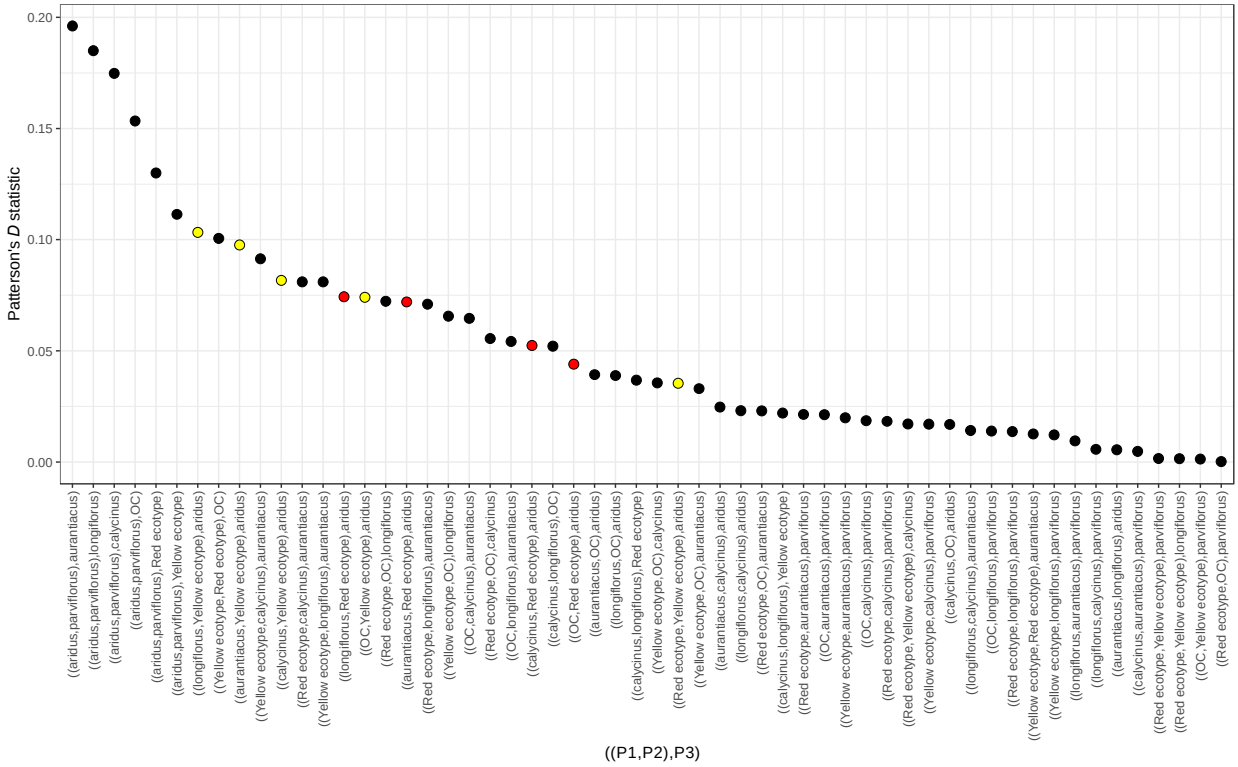

Fig. S2. The distribution of Patterson's D statistic among all calculated trios. The tests that are included in Table 1 are colored red or yellow, depending on whether the red ecotype or yellow ecotype was set as P2.

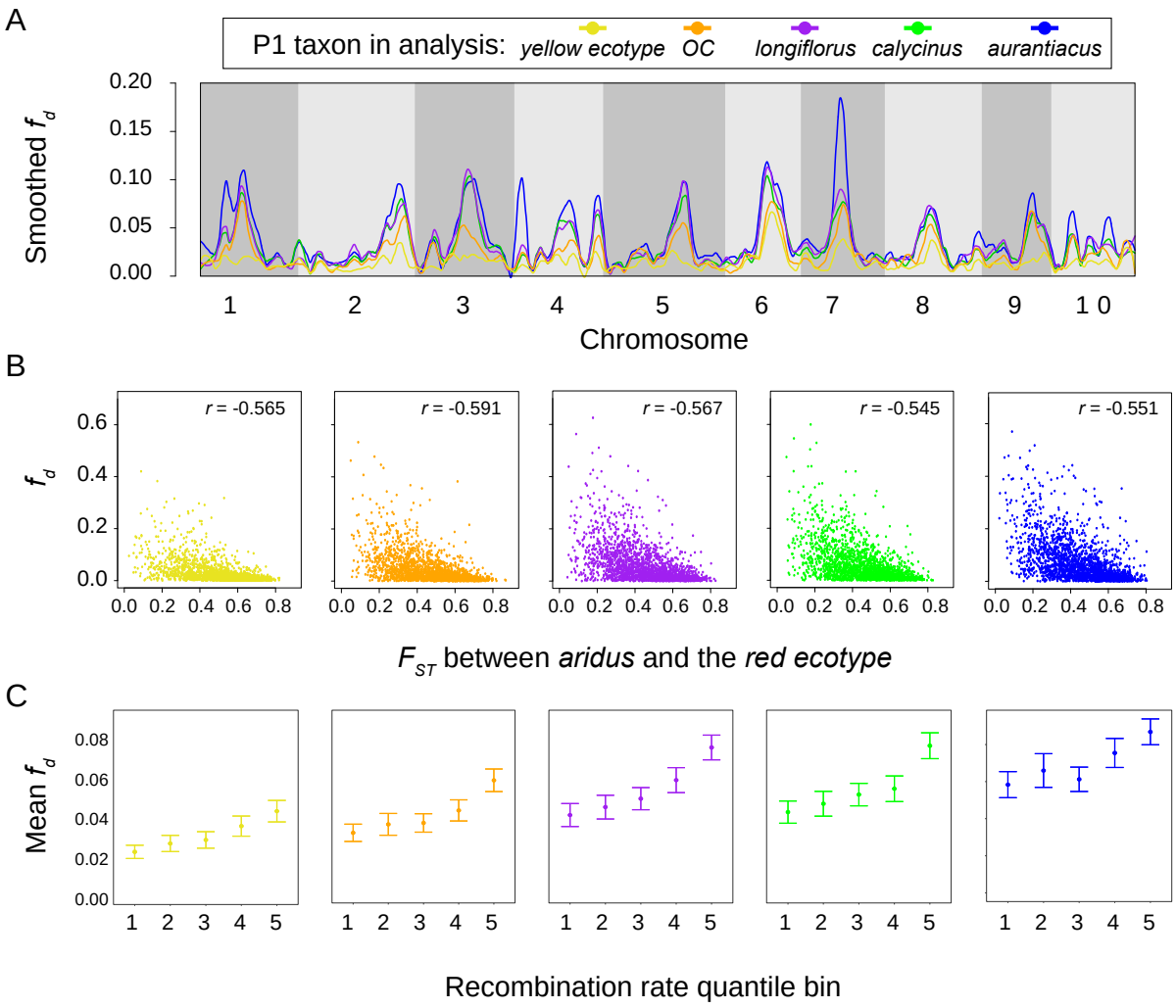

Fig. S3. (A) Loess-smoothed *fd* values plotted across the 10 chromosomes of the *Mimulus aurantiacus* genome, with the red ecotype as P2, *aridus* as P3, and one of five different taxa as P1. Colors indicate the taxa from clades C and D used as P1. (B) Scatterplots showing the relationship between Fst and *fd* in these same 50 kb windows, with the correlation coefficient between the statistics in the upper right-hand corner of each plot. (C) Mean and 95% confidence intervals of *fd* in different quantile bins of recombination rate. Quantile bins of recombination rate are as follows: 1 = 0 - 0.824 cM/Mb; 2 = 0.824 - 1.51 cM/Mb; 3 = 1.51 - 2.33 cM/Mb; 4 = 2.33 - 3.66 cM/Mb; 5 = 3.66 - 10.9 cM/Mb.

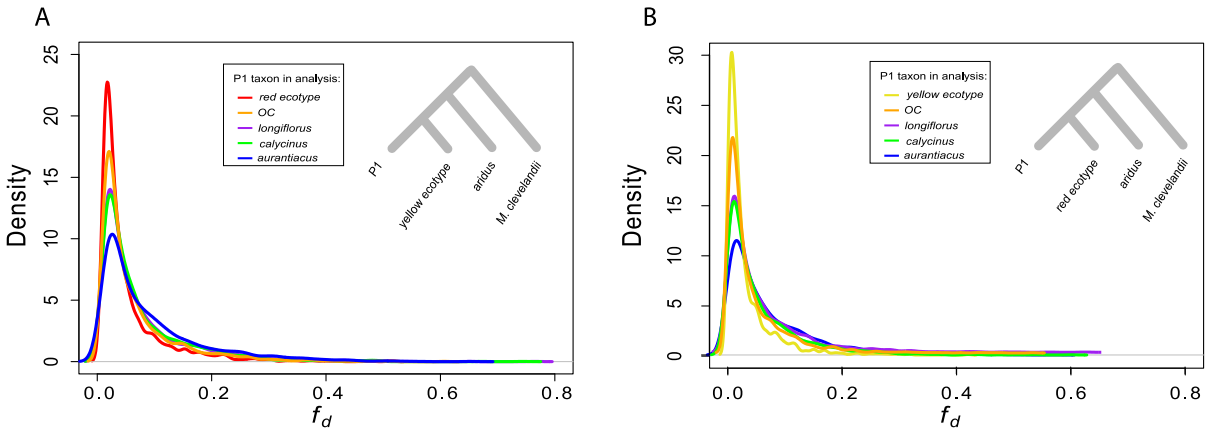

Fig. S4. Density plots of *fd*, calculated in 50 kb windows across the genome. A) The P1 taxon varied in each test, with P2 set as the yellow ecotype, P3 set as *aridus*, and the outgroup is *M. clevelandii*. B) The P1 taxon varied in each test, with P2 set as the red ecotype, P3 as *aridus*, and the outgroup is *M. clevelandii*.

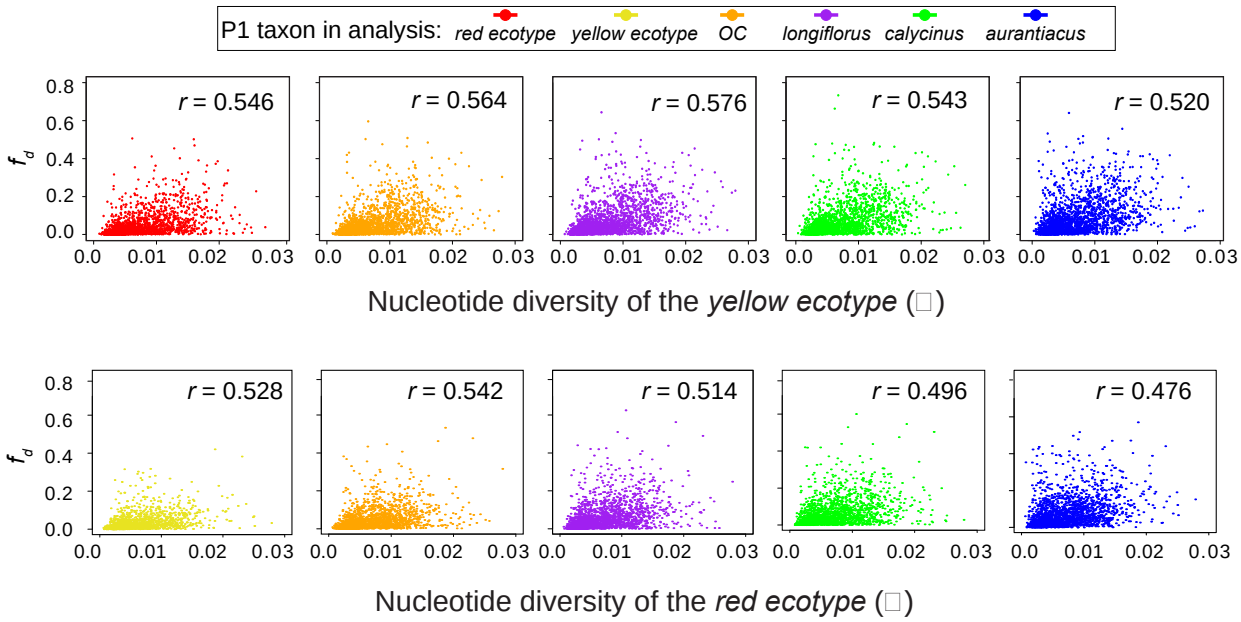

Fig. S5. Scatterplots and correlations between *fd* and π in 50 kb windows. Top, *fd* is calculated with different P1 taxa, the yellow ecotype as P2, *aridus* as P3, and *M. clevelandii* as the outgroup. Bottom, *fd* is calculated with different P1 taxa, the red ecotype as P2, *aridus* as P3, and *M. clevelandii* as the outgroup.

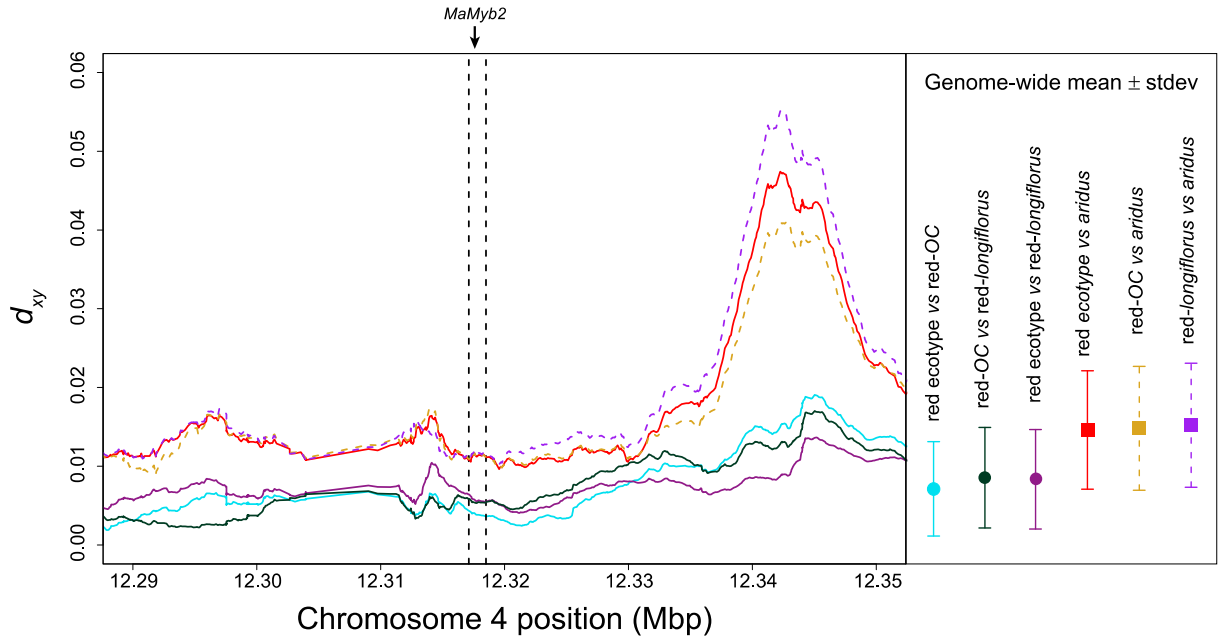

Fig. S6. Scans of *dxy* along chromosome 4, in the vicinity of *MaMyb2*. Divergence is calculated in overlapping 10 kb windows (100 bp steps) either between each of the three red-flowered taxa and *aridus*, or between each of the red-flowered taxa. The genome-wide mean values of *dxy* (error bars = standard deviations) are presented to the right, with the color scheme used the same as the scans.

2.7 References

1. Aeschbacher, S., Selby, J. P., Willis, J. H., & Coop, G. (2017). Population-genomic inference of the strength and timing of selection against gene flow. *Proceedings of the National Academy of Sciences*, **114**(27), 7061–7066. https://doi.org/10.1073/pnas.1616755114
2. Barrett, R. D., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, **23**(1), 38–44. https://doi. org/10.1016/j.tree.2007.09.008
3. Barrier, M., Baldwin, B. G., Robichaux, R. H., & Purugganan, M. D. (1999). Interspecific hybrid ancestry of a plant adaptive radiation: Allopolyploidy of the Hawaiian silversword alliance (*Asteraceae*) inferred from floral homeotic gene duplications. *Molecular Biology and Evolution*, **16**(8), 1105–1113. https://doi. org/10.1093/oxfordjournals.molbev.a026200
4. Barton, N. H. (2001). The role of hybridization in evolu- tion. *Molecular Ecology*, **10**(3), 551–568. https://doi. org/10.1046/j.1365-294x.2001.01216.x
5. Beeks, R. M. (1962). Variation and hybridization in southern California populations of *Diplacus* (Scrophulariaceae). *Aliso A Journal of Systematic and Floristic Botany*, **5**(2), 83–122.
6. Brandvain, Y., Kenney, A. M., Flagel, L., Coop, G., & Sweigart, A. L. (2014). Speciation and introgression between *Mimulus nasutus* and *Mimulus guttatu*s. *PLoS Genetics*, **10**(6), e1004410. https://doi. org/10.1371/journal.pgen.1004410
7. Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., Suh, A., Dutoit, L., Bureš, S., Garamszegi, L. Z., Hogner, S., Moreno, J., Qvarnström, A., Ružić, M., Sæther, S.A., Sætre, G.P., Török, J., & Ellegren, H. (2015). Linked selection and

recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Research*, **25**(11), 1656–1665. https://doi. org/10.1101/gr.196485.115

8. Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, **4**(1), s13742–s13015.
9. Chase, M. A., Stankowski, S., & Streisfeld, M. A. (2017). Genomewide variation provides insight into evolutionary relationships in a monkeyflower species complex (*Mimulus* sect. *Diplacus*). *American Journal of Botany*, **104**(10), 1510–1521. https://doi. org/10.3732/ajb.1700234
10. Choi, J. Y., Dai, X., Alam, O., Peng, J. Z., Rughani, P., Hickey, S., Harrington, E., Juul, S., Ayroles, J. F., Purugganan, M. D., & Stacy, E. A. (2021). Ancestral polymorphisms shape the adaptive radi- ation of *Metrosideros* across the Hawaiian Islands. *Proceedings of the National Academy of Sciences*, **118**(37), e2023801118.
11. 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.
12. De-Kayne, R., Selz, O. M., Marques, D. A., Frei, D., Seehausen, O., & Feulner, P. G. (2022). Genomic architecture of adaptive radiation and hybridization in Alpine whitefish. *Nature Communications*, **13**(1), 1–13.
13. Dobzhansky, T. (1982). *Genetics and the origin of species (no. 11)*. Columbia University Press.

14. Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska- Brzyska, K., Qvarnström, A., Uebbing, S., & Wolf, J. B. W. (2012). The genomic landscape of species divergence in *Ficedula* fly- catchers. *Nature*, **491**(7426), 756–760. https://doi.org/10.1038/ nature11584

15. Gautier, M., & Vitalis, R. (2012). rehh: An R package to detect foot- prints of selection in genome-wide SNP data from haplo- type structure. *Bioinformatics*, **28**(8), 1176–1177. https://doi. org/10.1093/bioinformatics/bts115

16. Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M....Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science*, **328**(5979), 710–722.

17. Han, F., Lamichhaney, S., Grant, B. R., Grant, P. R., Andersson, L., & Webster, M. T. (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of diver- gence among Darwin's finches. *Genome Research*, **27**(6), 1004–1015. https://doi.org/10.1101/gr.212522.116

18. Hedrick, P. W. (2013). Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, **22**(18), 4606– 4618. https://doi.org/10.1111/mec.12415

19. Huson, D. H., Kloepper, T., & Bryant, D. (2008). SplitsTree 4.0-Computation of phylogenetic trees and networks. *Bioinformatics*, **14**, 68–73.

20. Korunes, K. L., & Samuk, K. (2021). pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Molecular Ecology Resources*, **21**(4), 1359–1368. https://doi. org/10.1111/1755-0998.13326

21. Lenth, R. (2019). *Emmeans: Estimated marginal means, aka least-squares means. R package version 1.4.2*. https://CRAN.R-project.org/ package=emmeans.

22. Liu, S., Zhang, L., Sang, Y., Lai, Q., Zhang, X., Jia, C., Long, Z., Wu, J., Ma, T., Mao, K., Street, N. R., Ingvarsson, P. K., Liu, J., & Wang, J. (2022). Demographic history and natural selection shape patterns of deleterious mutation load and barriers to introgression across *Populus* genome. *Molecular Biology and Evolution*, **39**(2), msac008. https://doi.org/10.1093/molbev/msac008

23. Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite-Fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*, **21**(2), 584–595. https://doi. org/10.1111/1755-0998.13265

24. Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, **20**(5), 229–237. https://doi.org/10.1016/j. tree.2005.02.010

25. Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A combina- torial view on speciation and adaptive radiation. *Trends in Ecology & Evolution*, **34**(6), 531–544. https://doi.org/10.1016/j. tree.2019.02.008

26. Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., & 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

27. Martin, S. H., Davey, J. W., & Jiggins, C. D. (2015). Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Molecular Biology and Evolution*, **32**(1), 244–257. https://doi.org/10.1093/ molbev/msu269

28. Martin, S. H., Davey, J. W., Salazar, C., & Jiggins, C. D. (2019). Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biology*, **17**(2), e2006288. https:// doi.org/10.1371/journal.pbio.2006288

29. McMinn, H. E. (1951). Studies in the genus *Diplacus*, Scrophulariaceae. *Madroño*, **11**(2), 33–128.

30. Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, **8**(1), 1–11.

31. Meier, J. I., Stelkens, R. B., Joyce, D. A., Mwaiko, S., Phiri, N., Schliewen, U. K., Selz, O. M., Wagner, C. E., Katongo, C., & Seehausen, O. (2019). The coincidence of ecological opportunity with hybrid- ization explains rapid adaptive radiation in Lake Mweru cichlid fishes. *Nature Communications*, **10**(1), 5391. https://doi. org/10.1038/s41467-019-13278-z

32. 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

33. Nelson, T. C., & Cresko, W. A. (2018). Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evolution Letters*, **2**(1), 9–21. https://doi.org/10.1002/ evl3.37

34. Nelson, T. C., Stathos, A. M., Vanderpool, D. D., Finseth, F. R., Yuan, Y. W., & Fishman, L. (2021). Ancient and recent introgression shape the evolutionary history of pollinator adaptation and speciation in a model monkeyflower radiation (*Mimulus* section *Erythranthe*). *PLoS Genetics*, **17**(2), e1009095. https://doi. org/10.1371/journal.pgen.1009095

35. Nosil, P. (2012). *Ecological speciation*. Oxford University Press.

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast

and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*, **32**(1), 268–274.

Orr, H. A. (1995). The population genetics of speciation: The evolution of hybrid incompatibilities. *Genetics*, **139**(4), 1805–1813. https://doi.org/10.1093/genetics/139.4.1805

36. *Evolution Letters* (2023), Vol. XX | **11**
37. Orr, H. A. (1998). The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. *Evolution*, **52**(4), 935–949. https://doi.org/10.1111/j.1558-5646.1998.tb01823.x
38. Paradis, E. (2010). pegas: An R package for population genetics with an integrated–modular approach. *Bioinformatics*, **26**(3), 419–420. https://doi.org/10.1093/bioinformatics/btp696
39. Pardo-Diaz, C., Salazar, C., Baxter, S. W., Merot, C., Figueiredo-Ready, W., Joron, M., McMillan, W. O., & Jiggins, C. D. (2012). Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genetics*, **8**(6), e1002752. https://doi.org/10.1371/journal. pgen.1002752
40. Rambaut, A. (2009). FigTree. Tree figure drawing tool. http://tree.bio. ed.ac.uk/software/figtree/.
41. Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., Gabriel, S. B., Platko, J. V., Patterson, N. J., McDonald, G. J., Ackerman, H. C., Campbell, S. J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., & Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, **419**(6909), 832–837. https://doi. org/10.1038/nature01140

42. Schemske, D. W., & Bierzychudek, P. (2001). Perspective: Evolution of flower color in the desert annual *Linanthus parryae*: Wright revisited. *Evolution*, **55**(7), 1269–1282. https://doi. org/10.1111/j.0014-3820.2001.tb00650.x

43. Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., Blazier, J. C., Sankararaman, S., Andolfatto, P., Rosenthal, G. G., & Przeworski, M. (2018). Natural selection interacts with recom- bination to shape the evolution of hybrid genomes. *Science*, **360**(6389), 656–660. https://doi.org/10.1126/science.aar3684

44. Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, **23**(1), 23–35.

45. Sobel, J. M., & Streisfeld, M. A. (2015). Strong premating reproduc- tive isolation drives incipient speciation in *Mimulus aurantiacus*. *Evolution*, **69**(2), 447–461. https://doi.org/10.1111/evo.12589

46. Stankowski, S., Chase, M. A., Fuiten, A. M., Rodrigues, M. F., Ralph, P. L., & Streisfeld, M. A. (2019). Widespread selection and gene flow shape the genomic landscape during a radiation of monkey- flowers. *PLoS Biology*, **17**(7), e3000391. https://doi.org/10.1371/ journal.pbio.3000391

47. Stankowski, S., Chase, M. A., McIntosh, H., & Streisfeld, M. A. (2023). Integrating top-down and bottom-up approaches to understand the genetic architecture of speciation across a monkeyflower hybrid zone. *Molecular Ecology*, **32**, 2041–2054.

48. Stankowski, S., Sobel, J. M., & Streisfeld, M. A. (2017). Geographic cline analysis as a tool for studying genome-wide variation: A case study of pollinator-mediated divergence in a monkeyflower. *Molecular Ecology*, **26**(1), 107–122. https://doi.org/10.1111/mec.13645

49. Stankowski, S., & Streisfeld, M. A. (2015). Introgressive hybridiza- tion facilitates adaptive divergence in a recent radiation of monkeyflowers. *Proceedings of the Royal Society B: Biological Sciences*, **282**(1814), 20151666. https://doi.org/10.1098/rspb.2015.1666

50. Streisfeld, M. A., & Kohn, J. R. (2007). Environment and pollinator- mediated selection on parapatric floral races of Mimulus aurantiacus. *Journal of Evolutionary Biology*, **20**(1), 122–132.

Streisfeld, M. A., Young, W. N., & Sobel, J. M. (2013). Divergent selec- tion drives genetic differentiation in an R2R3-MYB transcription factor that contributes to incipient speciation in *Mimulus auran- tiacus*. *PLoS Genetics*, **9**(3), e1003385. https://doi.org/10.1371/journal.pgen.1003385

51. Thompson, D. M. (2005). Systematics of *Mimulus* subgenus *Schizoplacus*. *American Society of Plant Taxonomists*, **75**, 1–213. Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, **3**(9), e285. https://

52. doi.org/10.1371/journal.pbio.0030285

53. Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biology*, **4**(3), e72. https://doi.org/10.1371/journal.pbio.0040072

54. Wickham, H., & Seidel, D. (2022). *scales: Scale functions for visualization*. https://scales.r-lib.org; https://github.com/r-lib/scales

CHAPTER 3: The Genomic Outcomes of Hybridization Vary Over Time Within a Monkeyflower Radiation

This work is available as a preprint on biorxiv. The experimental question and design for this project was developed by me and Professor Matt Streisfeld. Prof. Streisfeld and I performed all the analyses for this work and wrote the manuscript together.

3.1 Introduction

Phenotypic and genetic differences that evolve between populations can lead to the accumulation of reproductive barriers and the origin of new species (Mayr 1942, Coyne & Orr 2004). Determining how reproductive isolation develops through time is critical for understanding how variation within populations translates to divergence between species (Matute & Cooper 2021). Although previously believed to be rare, hybridization between emerging species is now known to be common (Roux et al. 2016, Martin & Jiggins 2017, Malinsky et al. 2018, Stankowski et al. 2019, Liu et al. 2022). Depending on when it occurs, hybridization and subsequent gene flow between diverging lineages (i.e., introgression) can have various evolutionary consequences. For example, introgression can expose genetic incompatibilities that promote speciation (Coughlan & Matute 2020), it can swamp out divergence (Todesco et al. 2016), or in some cases, it can lead to the transfer of beneficial alleles across taxonomic boundaries (Suarez-Gonzalez et al. 2018). Thus, understanding the evolutionary history and genomic consequences of introgressive hybridization can help reveal the processes contributing to divergence and speciation (Harrison & Larson 2016, Ravinet et al. 2017).

In their classic study, Coyne & Orr (1989) found that reproductive isolation accumulated with levels of sequence divergence between multiple pairs of *Drosophila* species, a finding that has been supported in various other groups of divergent taxa (Moyle et al. 2004; Merot et al.

2017).Given that reproductive isolation takes time to arise, hybridization can occur at multiple points throughout the divergence of any pair of species (e.g. Martin et al. 2013, Malinsky et al. 2018, Meier et al. 2019, 2023). Although various modeling and simulation approaches have attempted to reveal the timing of past gene flow events, there remain serious challenges to accurately parameterize the models (Momigliano et al. 2021). Another option is to take advantage of the taxonomic diversity available in evolutionary radiations. By performing tests for introgression among the taxa in a radiation, it is possible to estimate the relative timing of introgression. Martin et al. (2013) used the diversity among *Heliconius* butterflies to demonstrate a history of continuous hybridization during the divergence of currently sympatric species pairs. Malinsky et al. (2018) expanded on this approach by performing tests for introgression between all sets of taxa in the Lake Malawi cichlid radiation, enabling researchers to evaluate which lineages showed evidence of hybridization and when it was likely to have occurred across the phylogeny of the group. They identified evidence of extensive ancient hybridization between the ancestors of modern cichlid lineages, as well as recent hybridization among closely related species found in similar ecological niches. These studies highlight how the taxonomic diversity present in evolutionary radiations can be used to identify the relative timing of past gene flow events. In this study, we expand on these previous examples to develop tests for the timing of introgression between island and mainland taxa in a radiation of monkeyflowers.

In addition to characterizing the evolutionary history of gene flow, understanding the fitness consequences of this hybridization can provide clues about the accumulation of reproductive isolation. In most cases, when functionally-relevant genetic variation is transferred into a foreign genetic background, it will decrease fitness, resulting in the removal of these alleles due to negative selection (Mallet 2005). These introgression-resistant loci are termed

"barrier loci," and as time proceeds, they should accumulate across the genome, leading to increased reproductive isolation (Wu 2001, Feder et al. 2012). However, positive selection following introgression can also result in the preservation and eventual fixation of these alleles (i.e. adaptive introgression) (Suarez-Gonzalez et al. 2018). For example, introgression of genetic material into a species that recently experienced a population bottleneck can introduce the necessary genetic variation to mitigate the negative effects of any deleterious genetic load that has accumulated (Schumer et al. 2018, Liu et al. 2022). Alternatively, the introgression of ecologically beneficial alleles can lead to adaptation, but this is expected to produce a more localized signature across the genome. By contrast, post-divergence gene flow can have no fitness consequences, in which case allele frequencies will be influenced primarily by genetic drift (Schumer et al. 2018).

To distinguish among these evolutionary processes, we expect certain genome-wide relationships to emerge. For example, the recombination rate determines how quickly introgressed alleles will be separated from resident alleles. Thus, we expect selection to rapidly remove deleterious, immigrant alleles in regions of low recombination where they remain linked with resident alleles. This will result in a positive relationship between introgression and recombination rate, provided the effects of reproductive isolation are widespread across the genome (Brandvain et al. 2014, Schumer et al. 2018, Martin et al. 2019). Similarly, because gene flow opposes divergence, we would expect a negative relationship between introgression and genetic divergence (Martin et al. 2013, Martin et al. 2019). In contrast, the repeated fixation of adaptively introgressed alleles will result in a genome-wide, negative relationship between introgression and recombination rate, such that longer, introgressed haplotypes are preserved in regions of low recombination (Duranton & Pool 2022, Feng et al. 2024). While this pattern is

expected in scenarios where introgression acts to mitigate the widespread, deleterious effects of genetic load, introgression that results in local adaptation to a particular ecological environment will not lead to this genome-wide relationship. Finally, no relationship between introgression and recombination rate would imply the action of neutral processes (Schumer et al. 2018).

In this study, we take advantage of the diversity in the *Mimulus aurantiacus* species complex to investigate the evolutionary history and genomic consequences of introgression between a pair of taxa restricted to the Channel Islands off the coast of California. The *M. aurantiacus* complex is a radiation of seven closely related, woody shrub subspecies distributed throughout California that display extensive phenotypic variation in their floral and vegetative traits and diverged from their sister species roughly one million years ago (Tulig 2000, Tulig & Nesom 2012, Chase et al. 2017; Stankowski et al. 2019). However, despite extensive phenotypic differentiation, there is evidence of hybridization between many of the taxa (Streisfeld & Kohn 2005, Stankowski et al. 2019, Short & Streisfeld 2023). Indeed, ancient introgression resulted in the repeated evolution of red flowers, a trait that has been shown to contribute to both pollinator adaptation and speciation in this group (Stankowski & Streisfeld 2015, Short & Streisfeld 2023, Sobel & Streisfeld 2015, Stankowski et al. 2017). The Channel Islands are currently inhabited by two subspecies of *M. aurantiacus* that are known to hybridize (Wells 1980, Chase et al. 2017). The red-flowered subspecies *parviflorus* is endemic to many of the islands (Tulig 2000, Tulig & Nesom 2012) but is listed as rare by the California Native Plant Society (2023). Subspecies *longiflorus* has larger, yellow flowers and is found both on the Channel Islands and throughout mainland southern California (Tulig 2000, Tulig & Nesom 2012). Both taxa occur in close contact on the islands, where they tend to inhabit dry, hillside habitats in the chaparral (Beeks 1962).

We used whole genome sequence data from individuals sampled across the radiation to address hree primary objectives. First, we developed an approach that incorporates multiple tests of introgression to estimate the history of hybridization among taxa and determine the relative timing of past gene flow events. Second, we assessed the genome-wide impacts of this introgression. Finally, we examined the relationship between introgression and recombination rate at different points throughout the history of the radiation to determine which processes have likely contributed to these patterns through time. Our results reveal that both recent and ancient introgression have shaped the genomic landscapes of these island taxa. Moreover, we found that the fitness effects of introgressed alleles can vary through time, such that the effects of selection against gene flow appear to increase with time, implying the accumulation of reproductive isolation.

3.2 Materials and Methods

3.2.1 Genome sequencing and variant calling

Leaf tissue was collected from 27 samples from four locations on Santa Cruz Island, California, USA (Table S1), consisting of red-flowered *parviflorus*, yellow-flowered *longiflorus*, and their putative hybrids. Tissue was dried in silica in the field, and DNA was isolated using the Zymo Plant and Seed DNA kit following the manufacturer's instructions. Sequencing libraries were prepared according to Gaio et al (2022), with slight modifications. Bead-Linked Transposase from the Illumina Nextera XT Kit was used for initial tagmentation, generating insert sizes in the range of 400 – 1200 bp. Multiplexed libraries were sequenced on the Illumina Novaseq 6000 using paired-end 150 bp reads at the University of Oregon's Genomics Core Facility.

New sequences from Santa Cruz Island were combined with previously generated whole genome sequences from various *M. aurantiacus* subspecies from Stankowski et al. (2019) and Short &

Streisfeld (2023), resulting in a final dataset containing 74 individuals. Raw reads were filtered using *fastp* to remove reads with uncalled bases or poor quality scores (Chen et al. 2018). The retained reads were then aligned to the reference assembly (Stankowski et al 2019) using *BWA* version 0.7.17 (Li & Durbin 2009). An average of 90.56% of reads aligned (range: 75.45% – 95.33%), and the average sequencing depth was 10× per individual (range: 6–18×). PCR duplicates were marked using *Picard* (https://broadinstitute.github.io/picard/). Variant calling was performed following Stankowski et al. (2019). We then phased the VCF using *BEAGLE* (Browning & Browning 2007), and further filtered the VCF file for biallelic SNPs using *vcftools* (Danecek et al. 2011). The final data set contained 12,749,566 SNPs across all 74 samples. Finally, we ran *UnifiedGenotyper* with the EMIT_ALL_CONFIDENT_SITES option to output all variant and invariant genotyped sites.

3.2.2 Admixture, PCA, and phylogenetic analysis

To determine how the island samples were related to the other taxa in the complex, we used *Admixture* (Alexander et al. 2009) to estimate ancestry proportions from all samples, with the number of clusters (K) set from 2 to 11. To further estimate ancestry among the admixed island samples and assess the relationship between samples collected on Santa Cruz Island and the mainland, we re-ran *Admixture* at K=2, but only using samples of *parviflorus*, *calycinus*, and Island and mainland *longiflorus*. Samples from the island that showed no evidence of admixture at K = 2 were used to assign individuals to subspecies. Admixed individuals were removed from further analysis, as they likely represented contemporary hybrids. To further assess clustering patterns among these samples, we also performed a principal components analysis (PCA) in *Plink* using the 31 samples from the island and the 7 mainland *longiflorus*. To determine the phylogenetic relationships among samples from the island and mainland, we

generated a maximum likelihood consensus tree using *IQ-TREE* v1.6.12 (Nguyen et al. 2015). We used a concatenated dataset consisting of 12,749,566 biallelic SNPs that did not include island samples that showed evidence of admixture in the above analyses (resulting in 63 individuals, see Results).

3.2.3 Tests for the timing of admixture

To test for genome wide evidence of introgression, we used *dsuite* (Malinsky et al. 2021) to calculate Patterson's D (Green et al. 2010) and the *f4*-ratio (Reich et al. 2009) for all possible trios of ingroup taxa, using *M. clevelandii* as the outgroup. Patterson's D and the *f4*-ratio measure asymmetries in the numbers of sites with ABBA and BABA patterns (where A and B are ancestral and derived alleles, respectively) across a phylogeny with three ingroup taxa and an outgroup that have the relationship (((P1, P2), P3) O). A significant excess of either pattern gives a nonzero value of D, which is taken as evidence that gene flow has occurred between P3 and one of the sister taxa. We then calculated the fbranch statistic to identify recent and historical signatures of hybridization (Malinsky et al. 2018, 2021). Given a set of *f4*-ratios calculated for all possible trios among a set of closely related taxa, *fbranch* uses the inferred phylogenetic relationship among these taxa, as well as variation in the phylogenetic distance between the sister taxa used as P1 and P2, to assign introgression to specific branches on a phylogenetic tree.

To estimate how introgression varied across the genome, we calculated the admixture proportion (*fd*) (Martin et al 2015) in 100 kb non-overlapping windows using the *ABBABABAwindows.py* Python script (https://github. com/simonhmartin/genomics_general). Similar to Patterson's D and the *f4*-ratio, *fd* searches for asymmetries in the number of ABBA and BABA sites across the genome, but it has been optimized specifically for use in genomic windows.

In addition to examining how *fd* varies across the genome, we took advantage of the diversity of

taxa in the *M. aurantiacus* species complex to estimate the relative timing of introgression between the taxa on Santa Cruz Island (Fig 1). Specifically, alleles that were introgressed into the common ancestor of Island and mainland *longiflorus* should be present in roughly equal proportions in both descendent lineages, making it difficult for *fd* to distinguish shared ancestral variation from introgression. However, by gradually increasing the phylogenetic distance between the two sister taxa used in these tests, we should be able to identify introgression that occurred at various points further back in time (Martin et al 2013; Short and Streisfeld 2023). This can be identified as an increase in mean *fd* with levels of sequence divergence. Alternatively, if introgression occurred only after the divergence between Island and mainland *longiflorus*, then we would not expect *fd* to increase with levels of sequence divergence.

To examine this, we ran two distinct tests for introgression by varying the set of taxa used in each test (Fig 1). To identify the potential for recent introgression, we calculated mean *fd* among 100 kb windows, setting Island *longiflorus* as P2, *parviflorus* as P3, and varying the taxon used as P1. Hereafter, we refer to these as Test 1. Then, to determine whether there was evidence for recent and ancient introgression, we took advantage of the more extensive divergence between *parviflorus* and *aridus*, which occurred prior to the split between taxa in clades C and D (Stankowski et al 2019). If introgression predated the divergence between Island and mainland *longiflorus*, then we would expect *fd* to increase when these more distantly related taxa were included as P1 and P2. Thus, we set *aridus* as P1, *parviflorus* as P2, and Island *longiflorus* as P3. These analyses will be referred to as Test 2. *M. clevelandii* was used as the outgroup for all calculations of *fd*.

To further demonstrate whether there was recent introgression between *parviflorus* and Island *longiflorus*, we re-calculated Test 2 to obtain the mean *fd* among windows using *aridus* as

P1, *parviflorus* as P2, and the various taxa from clades C and D as P3. Among these tests, if mean *fd* is greatest when Island *longiflorus* is P3, this would indicate both ancient introgression that predated the divergence of clades C and D, as well as recent introgression between Island *longiflorus* and *parviflorus*. However, if introgression occurred recently between *parviflorus* and any of the other taxa in clades C and D, then we would expect *fd* to be equal to or greater than the *fd* value calculated using Island *longiflorus* as P3.

To test for significant differences among the mean *fd* values, we fit linear models, with *fd* as the dependent variable and sequence divergence (da) between P1 and P2 as the independent variable. We then used the *emmeans* package in R (Lenth, 2019) to perform pairwise comparisons of the estimated marginal mean *fd* values from the linear model. To estimate levels of sequence divergence, we calculated da, which describes the difference in divergence between taxa (*dxy*) relative to the mean diversity within taxa (π). We calculated *dxy* and π in 100 kb windows using *PIXY* version 1.2.5 (Korunes & Samuk, 2021), with both variant and invariant sites included.

3.2.4 The relationship between genome wide phylogenetic discordance and introgression

To further describe the history of introgression with *parviflorus*, we explored patterns of phylogenetic discordance across the genome using *TWISST* (Martin & Van Belleghem 2017). Given a set of samples from multiple ingroup taxa, *TWISST* builds trees in genomic windows and then calculates the support for that topology among all possible topologies, which is referred to as the topology weighting. If all samples from each taxon are reciprocally monophyletic, then that topology is given a weighting of 1.0 for that window. Topological discordance among samples within a taxon will result in lower topology weightings for that window. We used *TWISST* to identify variation in the relationships among *aridus*, *parviflorus*, Island *longiflorus*,

and mainland *longiflorus* in 100 kb genomic windows, with *M. clevelandii* as the outgroup. With four ingroup taxa, this resulted in 15 possible topologies, which were then partitioned into those that support: *a*) the species tree, where *aridus* and *parviflorus* were sister and reciprocally monophyletic relative to Island *longiflorus* and mainland *longiflorus*; *b*) the introgression tree, where *parviflorus* and Island *longiflorus* were sister to one another relative to *aridus* and mainland *longiflorus*; and *c*) the ancient introgression tree, where *parviflorus* was sister to mainland and Island *longiflorus* relative to *aridus*. By quantifying variation in the support for these topologies across the genome, we can identify how often ancient and recent introgression contributed to phylogenetic discordance.

However, topological discordance can also be influenced by incomplete lineage sorting. Thus, to confirm that regions with high support for the introgression or ancient introgression trees were consistent with admixture, we first identified 100 kb windows with topology weightings of 1.0 for the species tree, introgression tree, or the ancient introgression tree, and then investigated the distribution of *fd* values within these windows. We calculated *fd* from Test 1 (with mainland *longiflorus* as P1) to identify the effects of recent introgression, while *fd* from Test 2 (with Island *longiflorus* as P3) was used to identify the combined effects of recent and ancient introgression. By taking the difference between these two sets of *fd* values, we should be able to identify only the signal of ancient introgression, which we refer to as "ancient *fd*." For example, when mainland and Island *longiflorus* are set as P1 and P2 in Test 1, we can only identify introgression with *parviflorus* that has occurred since they diverged from each other (Fig 1). However, when *aridus* and *parviflorus* are set as P1 and P2 in Test 2, the more ancient divergence between them allows us to identify older introgression with the ancestor of *parviflorus* that occurred prior to the split between clades C and D, as well as recent

introgression that occurred after the divergence between mainland and Island *longiflorus*. Thus, the difference between the two provides us with the signal of ancient introgression that occurred prior to the split of clades C and D. We determined how ancient and recent introgression impacted the relationships among these taxa by quantifying the topology weightings for the species tree, introgression tree, and ancient introgression tree within the top 5% of *fd* windows.

3.2.5 The genomic consequences of introgression

Genome wide heterogeneity in admixture can be caused by one (or a combination) of several processes: selection against gene flow, adaptive introgression, or genetic drift. To determine which of these processes may be contributing to genome wide variation in introgression, we estimated the relationship between the recombination rate, relative genetic divergence (*FST*), and *fd* due to both recent and ancient introgression. Under a model of polygenic selection against gene flow, we expect a negative relationship between *fd* and *FST*, but a positive relationship between *fd* and recombination rate (Brandvain et al. 2014, Schumer et al. 2018, Liu et al. 2022). Alternatively, if there is a genome-wide signal of adaptive introgression, which could occur to mitigate the effects of deleterious genetic load induced by hybridization (Feng et al. 2024), then we would expect a negative relationship between *fd* and recombination rate (Duranton & Pool 2022). In the absence of selection, we expect no relationship between *fd* and recombination rate. Finally, as reproductive isolation accumulates through time, we expect a stronger positive relationship between recombination and *fd* for estimates of recent introgression but a weaker relationship with ancient introgression. This is because fewer reproductive barriers would be in place if gene flow occurred deeper in the past.

FST was estimated between *parviflorus* and Island *longiflorus* in 100 kb non-overlapping windows using the *popgenwindows.py* Python script (https://github.com/simonhmartin/genomics_general), with only variant sites included. *fd* values from Test 1, Test 2, and "ancient *fd*" were correlated with *FST* using Spearman's correlation coefficient in R. To estimate the relationship between *fd* and recombination rate, we partitioned the *fd* values calculated in 100kb windows into quantile bins of recombination rates that were calculated previously in 500 kb windows by Stankowski et al. (2019). We then calculated the mean and 95% confidence intervals for *fd* within each recombination rate quantile bin and fit a linear model with *fd* as the dependent variable and the recombination rate quantile bins as the independent variable. We then used the *emmeans* package (Lenth, 2019) to perform pairwise comparisons of the estimated marginal mean *fd* values for the different quantile bins from the linear model.

It has also been hypothesized that introgression between closely related island endemic taxa may contribute to the maintenance of high levels of genetic diversity among island taxa that are geographically isolated from other populations of the same taxon (Carlquist 1966). If differences in genetic load between mainland and island endemics affect the outcomes of introgression between these taxa, we would expect to find reduced genetic diversity and a lower effective population size in the island taxa as compared to their mainland relatives (Schumer et al. 2018, Liu et al. 2022). To determine whether there is evidence of reduced genetic diversity among the hybridizing island taxa, we asked whether the mean π in both island subspecies differed among the other taxa in the complex. To identify differences in the effective population sizes of these species, we used PSMC (Li and Durbin 2011) to estimate the effective population size through time for all samples from the island taxa and their most closely related mainland

relatives. In accordance with Stankowski et al. (2023), we assumed a generation time of 2 years and a mutation rate of 7x10-9 for these calculations.

3.2.6 Identifying signatures of adaptive introgression

Extensive heterogeneity in admixture across the genome raises the possibility that adaptive introgression has maintained foreign ancestry in particular regions. To determine whether regions of elevated *fd* display signatures of adaptive introgression between *parviflorus* and island *longiflorus*, we calculated the frequency of Q95 sites found in windows across the genome (Racimo et al. 2016, Feng et al. 2024). Q95 sites are defined as those that are fixed (at a frequency of 1.0) in the donor taxon, near fixation (at a frequency greater than the 95th percentile) in the recipient taxon, and absent in the taxon that is sister to the recipient. Thus, any derived alleles that are present at high frequency in a recipient taxon that are also absent in the unadmixed sister taxon but are fixed in a more distantly related donor taxon, are candidates for adaptive introgression. Genomic regions that display an excess of these putatively adaptively introgressed alleles are even more likely to have been targets of positive selection, because selection will have correlated effects on linked sites. Thus, by identifying an elevated frequency of Q95 sites in 100 kb windows across the genome, we should be able to identify genomic regions that have been adaptively introgressed.

To identify Q95 sites, we first identified all derived SNPs that were polymorphic among samples. Derived SNPs were defined as those that were fixed in *M. clevelandii* and variant in at least one chromosome from samples of mainland *longiflorus*, Island *longiflorus*, *parviflorus*, or *aridus*. We then determined the number of derived sites that were fixed in *parviflorus*, absent in mainland *longiflorus*, and present in at least one haplotype in Island *longiflorus*. These represented sites that were introgressed from *parviflorus* into Island *longiflorus*. The 95th

percentile of the allele frequency distribution among introgressed sites in Island *longiflorus* (=0.875) was used as a cutoff to identify adaptively introgressed sites (i.e., Q95 sites) (Racimo et al. 2016, Feng et al. 2024). For each 100 kb window, we then divided the number of Q95 sites by the total number of derived SNPs that were present on at least one chromosome in Island *longiflorus* to estimate the frequency of Q95 sites in each window.

To identify signatures of adaptive introgression that were transferred from Island *longiflorus* into *parviflorus*, we identified derived SNPs that were fixed in Island *longiflorus*, present on at least one haplotype in *parviflorus*, but were absent in mainland *longiflorus*. In this case, the 95th percentile of the allele frequency distribution in *parviflorus* included only sites fixed in *parviflorus*. We then divided the number of Q95 sites by the total number of derived SNPs that were present on at least one chromosome in *parviflorus* in 100 kb windows.

To further explore signatures of adaptive introgression for the region with the highest frequency of Q95 sites across the genome, we calculated *fd*, *dxy*, *FST*, π, and the frequency of Q95 sites in 10kb windows with 1000 bp steps (see Results).

3.3 Results

3.3.1 Evolutionary relationships and hybridization on Santa Cruz Island

At K=8, *Admixture* assigned samples to ancestry groups that were consistent with previous analyses of the phylogenetic history of these subspecies (Fig 2A, Fig S1) (Chase et al. 2017, Stankowski et al. 2019, Short and Streisfeld 2023). The newly sequenced Santa Cruz Island samples contained a mixture of individuals with both pure and admixed ancestry, suggesting the occurrence of ongoing hybridization between *parviflorus* and Island *longiflorus* (Fig 2A). By re-running Admixture at K=2, including only the island samples, as well as mainland *longiflorus*

and *calcyinus,* we identified 12 “unadmixed” *parviflorus* and 8 “unadmixed” Island *longiflorus* (with q-scores greater than 0.99; Fig 2B). We also identified 11 admixed samples from Santa Cruz Island, which were removed from subsequent analyses. Results from the PCA were largely consistent with Admixture, with “unadmixed” Island *longiflorus* and *parviflorus* samples clustering separately along PC1 and admixed samples distributed between them (Fig 2C).

The consensus species tree was consistent with previous analyses that revealed four primary clades (Stankowski and Streisfeld 2015, Chase et al. 2017, Short & Streisfeld 2023). Clade A consisted entirely of subspecies *grandiflorus*, clade B showed *parviflorus* and *aridus* as sister taxa, clade C consisted entirely of *aurantiacus*, and the remaining taxa from southern California comprised the diverse clade D (Fig 2D). Within clade D, Island *longiflorus* was sister to both mainland *longiflorus* and *calycinus*. To confirm this relationship, an additional run of Admixture was performed at K=3 (Fig S2), revealing three distinct ancestry groups that corresponded to Island *longiflorus*, mainland *longiflorus*, and *calycinus*.

3.3.2 Evidence of recent and ancient hybridization

Consistent with Stankowski et al (2019), our calculations of the *fbranch* statistic revealed evidence of widespread introgression, particularly among the very recently diverged taxa within clades C and D (Fig 3A). In addition, we identified evidence of hybridization between *aurantiacus* and *grandiflorus,* as well as between *aridus* and the clade D taxa, confirming previous results by Short and Streisfeld (2023). We observed a strong signal of introgression between *parviflorus* and all the taxa from clades C and D, but this signal was the strongest between *parviflorus* and Island *longiflorus* (Fig 3A). This raises the possibility that there may have been recent introgression between *parviflorus* and Island *longiflorus*, as well as historical hybridization between *parviflorus* and the ancestor of clades C and D.

To test these hypotheses, we performed calculations of *fd* using taxa at various levels of sequence divergence from one another. This allowed us to identify recent and historical signals of introgression (Fig 1). From Test 1, we identified no significant increase in mean *fd* with sequence divergence (da) when the different taxa from clades C and D were set as P1 (Fig 3B, Table S2). This indicates introgression was recent and occurred only after the split between Island and mainland *longiflorus*. However, when the more diverged *aridus* and *parviflorus* were set as P1 and P2 in Test 2, we found a significantly higher mean *fd* (Fig 3B, Table S2), indicating both recent hybridization between Island *longiflorus* and *parviflorus*, as well as ancient hybridization between the ancestor of modern *parviflorus* and the ancestor of clades C and D. In addition, mean *fd* was greatest when island *longiflorus* was set as P3 (Fig S3, Table S3), confirming that Test 2 identified the signals of both recent and ancient introgression. In addition to the presence of ongoing hybridization between the island taxa, these results reveal that introgression with *parviflorus* has occurred at two time points, once between its ancestor and the ancestor of clades C and D, and then again, in the more recent past with Island *longiflorus*.

3.3.3 Ancient and recent hybridization contribute to genome wide phylogenetic discordance

Using *TWISST*, we defined tree topologies that supported the 'species tree,' a set of 'introgression trees' where Island *longiflorus* and *parviflorus* were sister, and an 'ancient introgression' topology where *parviflorus* was sister to both mainland and Island *longiflorus* (Fig 4A). By scanning the genome for the distribution of these tree topologies, we found widespread evidence of phylogenetic discordance (Fig 4B). We identified 553 windows with topology weightings of 1.0 for the species tree, 61 windows that supported the introgression tree, and 419 trees that matched the ancient introgression tree (Fig 4C) (a remaining 915 windows did not

match any of these topologies or had topology weightings less than 1.0, which we refer to as 'unresolved'). Despite there being nearly as many topologies that supported a history of introgression as the species tree, this phylogenetic discordance also may have been caused by incomplete lineage sorting.

To determine if discordance could be attributed to introgression or incomplete lineage sorting, we tested for an association between these topologies and *fd* across the genome. Consistent with a primary role for introgression, we found that windows supporting the 'introgression tree' displayed higher Test 1 and Test 2 *fd* values than the other topologies (Figs 4D and 4E). However, there were some windows with high support for the 'introgression tree' that had low values of *fd*, suggesting that some of the discordance may have been caused by incomplete lineage sorting. In windows with complete support for the 'ancient introgression tree,' *fd* values from Test 2 were higher than for Test 1 (compare Fig 4D and 4E), again consistent with Test 2 identifying signals of both recent and ancient introgression. When we calculated the difference between *fd* values from Test 2 and Test 1, the windows with the highest 'ancient *fd*' values largely corresponded to those with complete support for the 'ancient introgression' topology (Fig 4F). By focusing only on the top 5% of windows with the highest *fd* for both Test 1 and Test 2, we found greater support for the 'introgression' and 'ancient introgression' topologies (Figs 4G-H), with more evidence of ancient introgression for Test 2. This pattern became even more clear when we plotted the distribution of topologies for the top 5% of 'ancient *fd*' vales (Fig 4I). Thus, while there is some evidence for incomplete lineage sorting, much of the phylogenetic discordance appears to be caused by recent and ancient introgression.

3.3.4 The genomic consequences of recent and ancient hybridization

We identified extensive variation in the extent of introgression across the genome (Fig 5, S4). Raw values of *fd* in 100 kb windows ranged from 0.0 to around 0.9 (Fig S4). To determine the evolutionary processes responsible for this extreme heterogeneity in introgression across the genome, we compared *fd* with levels of genetic divergence and recombination rates. We identified a strong negative relationship between *fd* and *FST* for both Test 1 and Test 2 ($r > -0.70$; Fig. 5D, 5E), implying that negative selection has likely purged deleterious foreign genetic variation across the genome. By contrast, although the relationship between *FST* and *fd* remained negative when we examined the impacts of ancient introgression, it was considerably weaker ($r = -0.26$, Fig 5F).

We found an overall positive relationship between recombination rate and mean *fd* for Tests 1 and 2 (Fig 5G, 5H). Specifically, the recombination rate quantile bin with the lowest recombination (0–0.824 cM/Mb) had significantly lower mean *fd* values than the highest recombination quantile (3.66–10.9 cM/Mb) for both Tests 1 and 2 (Table S4, S5). In addition, for Test 1, mean *fd* in the lowest recombination rate quantile was significantly lower than all four of the remaining higher quantiles, but this was not the case for Test 2, which showed no significant differences in mean *fd* among intermediate recombination quantile bins. The slightly weaker relationship for Test 2 suggests the possibility that the effects of selection may have varied through time. Consistent with this hypothesis, we identified no significant difference in ancient *fd* values between any of the recombination rate bins (Figure 5I, Table S6).

We observed no difference in average nucleotide diversity (π) among taxa, including the island endemics (Fig S5). Furthermore, we found no evidence for recent or historical reductions

in effective population size in the island taxa (Fig S6). However, we did observe a small increase in effective population size in *parviflorus* and Island *longiflorus* around 100,000 years ago, suggesting that hybridization between these subspecies may have started around that time.

3.3.5 Evidence of localized adaptive introgression

To investigate evidence for adaptive introgression in regions of elevated *fd*, we calculated the frequency of Q95 sites in 100 kb windows across the genome (Fig S7). As might be expected, most windows displayed no increases in the frequency of Q95 sites. However, we did identify clear pileups of elevated Q95 ratios in a few regions (Fig S7), with one region in particular showing more than 40% of derived sites within two adjacent windows that matched the Q95 pattern (chromosome 5, 24.7-24.9 Mbp). These same windows had *fd* values greater than 0.85 for Test 2 and about 0.7 for Test 1 (Fig S4). Such an elevated local signal of introgression surrounded by regions of considerably higher admixture suggests the possibility of adaptive introgression. In addition, although we found an overall positive relationship between *fd* and π, these two windows have the highest *fd* for both Tests 1 and 2 but some of the lowest nucleotide diversity across the genome (Fig S8).

By zooming in on this region on chromosome 5, we identified a strong signal of introgression that spanned approximately 300 kb (Fig 6A). Furthermore, we also identified reduced *dxy* and *FST* between *parviflorus* and Island *longiflorus* and a clear increase in the frequency of Q95 sites transferred from Island *longiflorus* into *parviflorus* (Fig 6). In contrast, we did not see a signal of elevated Q95 when considering introgression in the other direction. In addition, we also identified elevated *FST* and *dxy* between *parviflorus* and *aridus*, but no clear increase in *FST* or *dxy* between Island and mainland *longiflorus* (Figure 6B, 6C), providing

further support that adaptive introgression in this region occurred from Island *longiflorus* into *parviflorus*. However, we observed no clear decrease in π for either Island taxon (Figure 6D).

3.4 Discussion

Hybridization can occur throughout the history of a radiation, but the evolutionary consequences of this mixing can vary over time. In this study, we took advantage of the diversity of taxa present in the *Mimulus aurantiacus* species complex to demonstrate that introgressive hybridization occurred at different points during the history of the group. We identified evidence of recent introgression between a pair of island endemic taxa, as well as ancient hybridization between their ancestors. Moreover, by comparing the relationship between introgression and recombination rate at these different time points, we found that selection against recent gene flow on the island is more prevalent than following ancient hybridization, likely due to the accumulation of reproductive barriers through time. Although it is possible that time has eroded previous signatures of selection against ancient introgression, we identified widespread phylogenetic discordance associated with ancient introgression, which suggests that much of the signal of ancient introgression has been preserved over time.

3.4.1 Recent and ancient hybridization contribute to distinct signals of introgression between a pair of island endemic monkeyflower subspecies

By examining the history of hybridization between these island taxa, we demonstrate how the diversity of taxa present within evolutionary radiations can be used to estimate the relative timing of past hybridization events. Martin et al. (2013) first demonstrated that it was possible to detect introgression that occurred further back in time by increasing the level of sequence divergence between the sister pair of ingroup taxa used in a four-taxon test of introgression. Malinsky et al. (2018) advanced this approach by developing a statistic to estimate the relative

timing of hybridization across a phylogenetic tree. In the current study, we further expand on these methods by performing two tests that can: 1) estimate levels of recent introgression in a genomic window, and 2) identify levels of ancient and recent introgression. These tests rely on the variation in the divergence times between a pair of hybridizing taxa and their respective sister taxa.

Specifically, we used Test 1 to demonstrate that hybridization between Island *longiflorus* and *parviflorus* began after Island and mainland *longiflorus* diverged from each other. This is because average levels of *fd* remain similar regardless of which taxon is used as P1 in the test (Fig 3B). We then performed additional tests for introgression using the sister pair of *aridus* and *parviflorus* as P1 and P2 to identify introgression that occurred prior to the divergence of clades C and D (Test 2). Because *parviflorus* and *aridus* are more diverged from each other than any of the taxa within clades C and D, Test 2 can detect recent hybridization between *parviflorus* and Island *longiflorus*, as well as any ancient introgression that occurred between the ancestor of *parviflorus* and the common ancestor of clades C and D. The difference between these two tests can be used to obtain an estimate of ancient introgression.

Although *parviflorus* is currently endemic to the Channel Islands, these results suggest that the ancestor of *parviflorus* was likely also present on the mainland of California, where it hybridized with the common ancestor of clades C and D at some point in the past. It has been argued that all plant species currently found on the Channel Islands are descended from mainland California ancestors (Axelrod 1965, Thorne 1969, Schoenherr et al. 2003). Furthermore, many of the woody plant species restricted to the Channel Islands were found on the mainland in the past, where they appear to have been driven to extinction by changing climate conditions (Axelrod 1965). Many of the Channel Island's endemic species are believed to have migrated from the

mainland during the Pleistocene glacial period when sea levels were lower (Johnson 1978, Muhs et al. 2015, Mychajliw et al. 2020). During this period, Santa Cruz Island was connected to Santa Rosa and Anacapa Islands, and the distance between these islands and the mainland was only between 6-10 km (Johnson 1978, Muhs et al. 2015), making dispersal to the islands possible. Furthermore, consistent with what we report here in *Mimulus*, there is also evidence of ancient and recent introgression between a pair of Channel Island endemic oak species (Ortego et al. 2018, Mead 2023, Mead et al. 2024). One of the island species appears to be a relic of a now extinct species that was present on the mainland, and the other is widely distributed on both the mainland and the Channel Islands. This raises the possibility that hybridization may be a common feature among closely related Channel Island endemics.

Indeed, hybridization between plant species on the Channel Islands appears to be common (Thorne 1969), which is consistent with a general finding of increased hybridization between closely related island endemics (Carlquist 1966, Reatini & Vision 2023). One proposed explanation for this is that hybridization can mitigate the effects of deleterious genetic load that accumulates in geographically isolated, small populations (Carlquist 1966). However, we observed no reductions in genetic diversity or effective population size in the island taxa, suggesting that there is no evidence of greater genetic load in either island taxon. Nevertheless, given that the samples of *parviflorus* and Island *longiflorus* sequenced here continue to display evidence of admixture, it is possible that these patterns of diversity and population size may be attributable to their shared history of introgression. Additional sampling from the other Channel Islands where the two taxa do not occur in sympatry will be necessary to test this hypothesis further.

3.4.2 Variation in the fitness effects of introgression over time

Speciation is a continuous process that involves the accumulation of reproductive isolation (Stankowski and Ravinet 2021). Multiple factors will determine the rate at which isolation evolves, but the time since divergence has been shown to be highly relevant (Coyne and Orr 1989). Thus, at any point along the speciation continuum before reproductive isolation is complete, hybrids can form, leading to the potential for gene flow between emerging species. The consequences of this gene flow for fitness will depend on the extent of isolation that has evolved and the environments that the taxa inhabit. For example, if a pair of taxa has only recently diverged, many of the variants exchanged between them will likely have few deleterious effects on fitness. However, as divergence times increase and more reproductive isolation becomes established, selection against hybrids may be stronger. Alternatively, variants transferred between taxa may be beneficial in a shared ecological environment. By identifying variation in the timing of introgression between taxa, we were able to reveal that the consequences of introgression for fitness also varied over time.

By building phylogenetic trees in windows across the genome, we found widespread phylogenetic discordance due to a combination of recent and ancient hybridization and incomplete lineage sorting. Surprisingly, we identified nearly equal numbers of windows that supported the species tree and the ancient introgression tree, suggesting that the signal of ancient introgression is widespread throughout the genome. In contrast, we identified a much smaller number of windows that supported the introgression tree, suggesting that recent introgression was limited to fewer regions throughout the genome.

To estimate the fitness effects of introgression, we compared the relationship between recombination rate and recent and ancient introgression. Selection against introgression should

result in a positive relationship between introgression and local levels of recombination, because the recombination rate determines how quickly introgressed alleles will be separated from resident alleles. If selection acts against foreign variants, it should rapidly remove deleterious alleles in regions of low recombination where they remain linked with resident alleles (Brandvain et al. 2014, Schumer et al. 2018, Martin et al. 2019). Consistent with this expectation, we found that recent introgression between *parviflorus* and Island *longiflorus* was positively related to recombination rate, suggesting that admixed variants were often deleterious. By contrast, there was no relationship between ancient introgression and recombination rate. These findings imply that reproductive isolation has accumulated between *parviflorus* and Island *longiflorus*, such that more recent gene exchange between these diverged taxa resulted in selection against gene flow. By contrast, shortly after the split between their ancestors, there were likely fewer reproductive barriers in place, allowing free exchange of genetic information and a corresponding signal of neutral introgression. However, reproductive isolation is not complete between *parviflorus* and Island *longiflorus*, as numerous contemporary hybrids were detected in our samples (Fig 2B, 2C). Thus, additional work characterizing the components and extent of reproductive isolation, as well as the fitness of hybrids between these taxa will be needed to confirm these conclusions.

Another explanation for these findings is that the signatures of selection against ancient introgression have eroded over time. Although difficult to assess directly, we found widespread evidence of ancient introgression across the genome (Fig 4C), suggesting that much of this signal has been maintained through time. Indeed, the presence of numerous genomic windows with high support for the ancient introgression topology is consistent with anciently introgressed alleles being largely neutral. Although some windows likely display high support for the ancient

introgression topology due to incomplete lineage sorting, windows of elevated "ancient *fd*" display strong support for the ancient introgression topology, suggesting that ancient introgression is at least partially contributing to this pattern. Moreover, many cases of widespread phylogenetic discordance (Nelson et al. 2021, Zhang et al. 2021a) and adaptive introgression (Meier, et al. 2017, Malinsky et al. 2018, Ma et al. 2019, Zhang et al. 2021b, Short & Streisfeld 2023) appear to be the result of ancient introgression, further implying that selection against introgression was weaker earlier in the divergence history of these taxa.

Finally, by examining *fd* values from Test 1 and Test 2 across the genome, we found extensive variation in genome wide levels of introgression, with some windows being nearly fixed for introgressed alleles. This raises the possibility that adaptation has contributed to the localized maintenance of introgressed alleles. Despite evidence that selection against recent gene flow was the primary factor shaping the heterogeneous patterns of introgression between *parviflorus* and Island *longiflorus*, we also found evidence for localized increases in the frequency of Q95 sites.

Specifically, we identified a region on chromosome 5 in *parviflorus* that was nearly fixed for Island *longiflorus* alleles. Thus, although most alleles exchanged between these taxa likely decreased fitness, some introgressed alleles appear to have increased fitness. Both taxa inhabit similar environments and often occur in sympatry, implying that they experience many of the same selection pressures on Santa Cruz Island. Thus, the transfer of alleles from Island *longiflorus* into *parviflorus* may have been facilitated by their shared environments, but future studies on the ecology and physiology of these taxa will be needed to identify potential selective agents contributing to adaptation.

In conclusion, we identified evidence of selection against recent introgression between

parviflorus and Island *longiflorus*, potentially because more reproductive barriers are currently in place. However, we found no current signal of selection between their ancestors. Thus, this study reveals that hybridization can occur at multiple points throughout the divergence history of a radiation, but the processes that shape their genomes can change over time.

3.5 Bridge

In this chapter, I identified evidence of introgression from different time points between a pair of actively hybridizing monkeyflower subspecies that are endemic to the Channel Islands of California and then examined how the fitness consequences of introgression have changed over time. This revealed evidence that the strength of selection against introgression appears to have increased over time, potentially due to the accumulation of reproductive barriers. This suggests that there may be less selection against gene flow earlier on in the divergence history of a species pair. However, the results from this chapter are based on signals of ancient hybridization that were identified using phylogenetic tests for introgression. To determine whether more recently diverged species display less selection against gene flow, we examined the outcomes of hybridization between a pair of recently diverged ecotypes, the red and yellow ecotypes of *puniceus* from San Diego. In this next chapter, I use samples from a geographic hybrid zone between these ecotypes to identify genome wide variation in gene flow between them and then to characterize regions that act as barriers to gene flow.

3.6 Figures and Tables

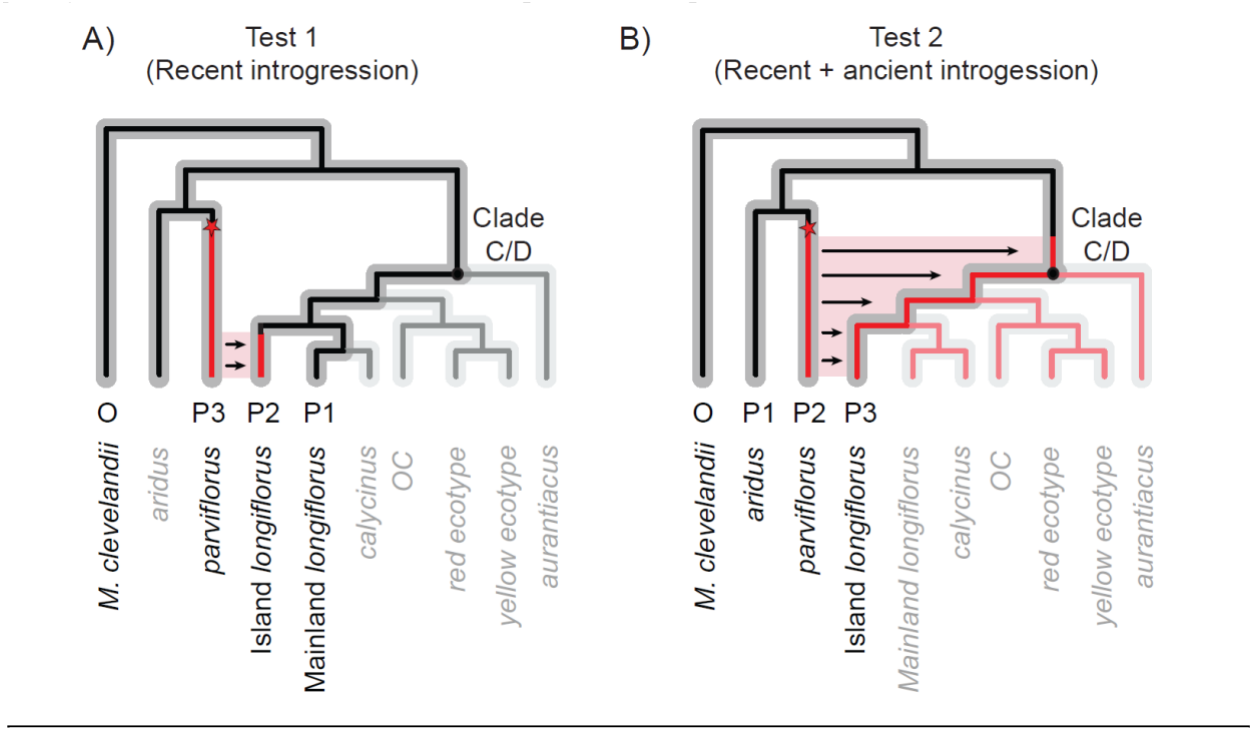

Figure 1. Evolutionary radiations can be used to estimate the relative timing of introgression. (A-B) In a four-taxon tree, with three ingroup taxa (P1–P3) and an outgroup (O), introgression between P2 and P3 only can be identified if gene flow occurred after the split between the two sister taxa (P1 and P2). The species tree from the *M. aurantiacus* radiation is presented in gray, with tips grayed out to show only the four taxa used in the tests for introgression (denoted by P1–P3, and O). The ancestral (black) and derived (red) alleles at a locus are indicated as lines. Mutations are indicated by red stars. By using different pairs of taxa with increasing levels of divergence from one another as P1 and P2, we can track introgression that occurred further back in time. (A) To estimate recent introgression (i.e., the genetic variation that was introgressed after the divergence of Island and mainland *longiflorus*), we performed Test 1, which includes mainland *longiflorus* as P1, Island *longiflorus* as P2, and *parviflorus* as P3. (B) To estimate recent ancient introgression (i.e., any genetic variation that was exchanged between the ancestor of *parviflorus* and the common ancestor of clades C and D), we performed Test 2, which includes the more distantly related *aridus* and *parviflorus* as P1 and P2, and Island *longiflorus* as P3. The tests performed do not specify the direction of introgression. For simplicity, the examples shown in the figure do indicate directionality, but the same conclusion would be drawn about the timing of introgression if the direction was reversed.

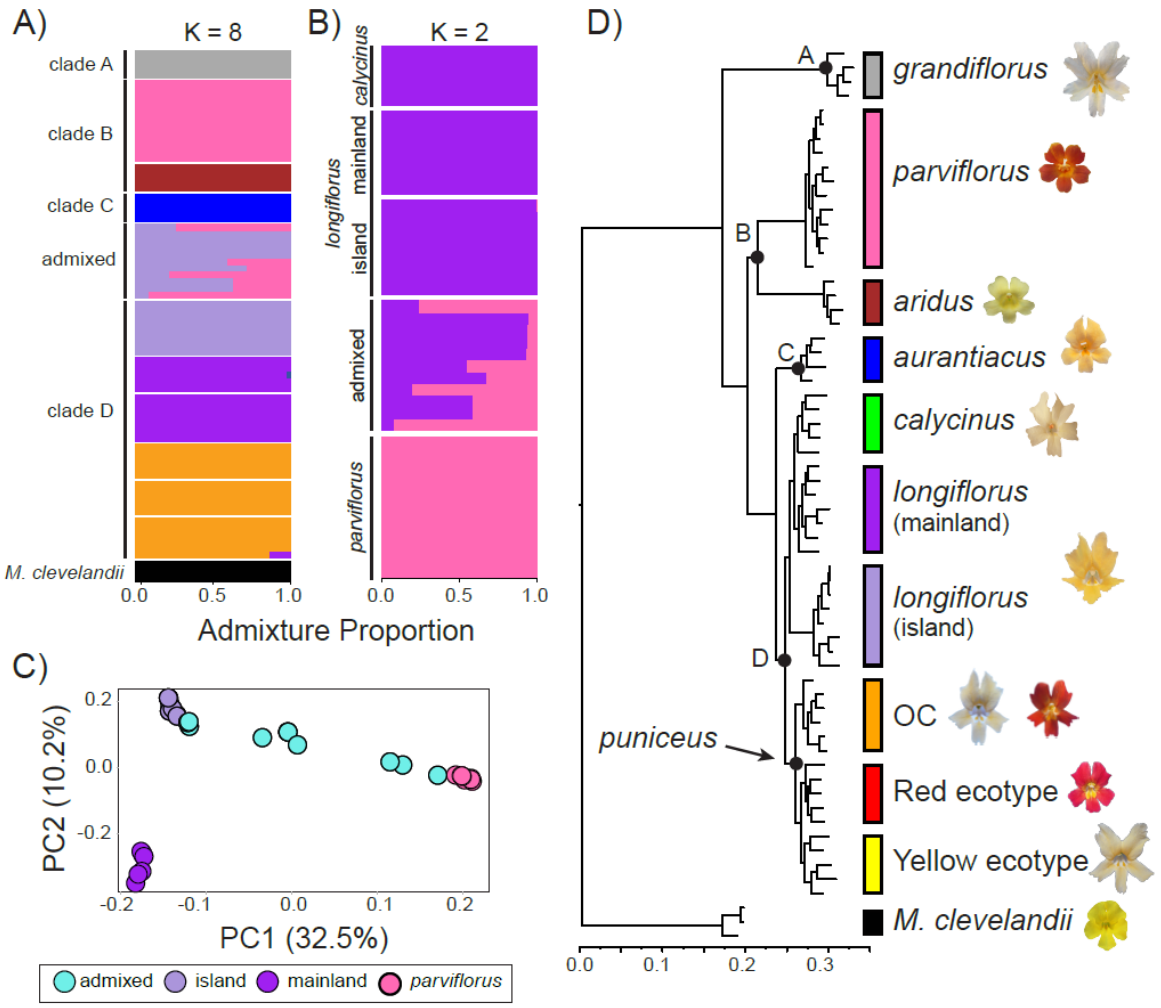

Figure 2. The relationships among the taxa in the *M. aurantiacus* species complex. (A) The ancestry proportions from *Admixture* at K = 8 for all the subspecies and their sister species, *M. clevelandii*. (B) Ancestry proportions at K = 2 from a run that included only the island samples, mainland *longiflorus*, and *calycinus*. (C) Plot of the first two principal components from the island samples and mainland *longiflorus*. The percent variation explained by each axis is reported. (D) Phylogenetic tree showing evolutionary relationships, with representative photographs of each taxon's flower. The four major clades of the radiation are labelled with black letters. The tips corresponding to each taxon are indicated by a colored bar.

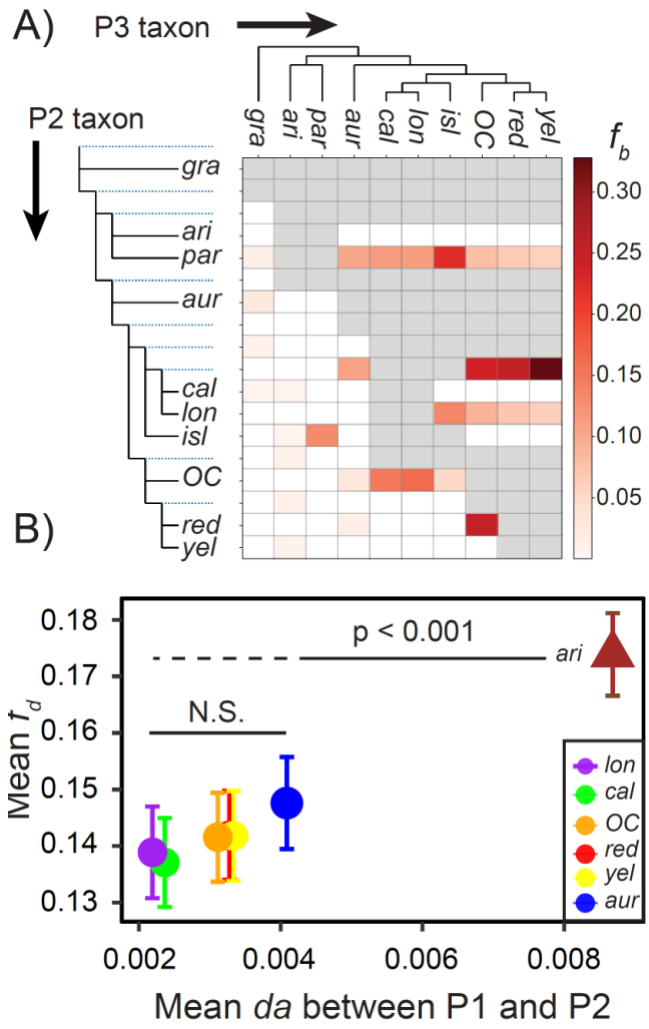

Figure 3. Hybridization history in the *Mimulus aurantiacus* species complex. (A) Genome-wide patterns of introgression using the f-branch statistic (fb). This analysis uses knowledge of the relationships among taxa and the results of the *f4*-ratio (calculated using all possible trios of ingroup taxa) to estimate the relative position of hybridization on a species tree. The color gradient represents the values of fb, with darker colors indicating greater evidence of introgression. (B) Mean and 95% confidence intervals for Test 1and Test 2 *fd* values calculated in 100 kb windows are plotted against mean sequence divergence (da) between the taxa used as P1 and P2 for the calculation of *fd*. Test 1 calculations of *fd* were performed using the various clade C and D taxa as P1, Island *longiflorus* as P2, and *parviflorus* as P3. Test 2 calculations of *fd* were performed using *aridus* as P1, *parviflorus* as P2, and Island *longiflorus* as P3. Colors indicate the P1 taxon used to calculate *fd* for the specific test, with mainland *longiflorus* in purple (*lon*), *calycinus* in green (*cal*), Orange County *puniceus* in orange (*OC*), the red ecotype in red (*red*), the yellow ecotype in yellow (*yel*), *aurantiacus* in blue (*aur*), and *aridus* in brown (*ari*). Shapes indicate which test was performed, with circles indicating Test 1 and triangles indicating Test 2.

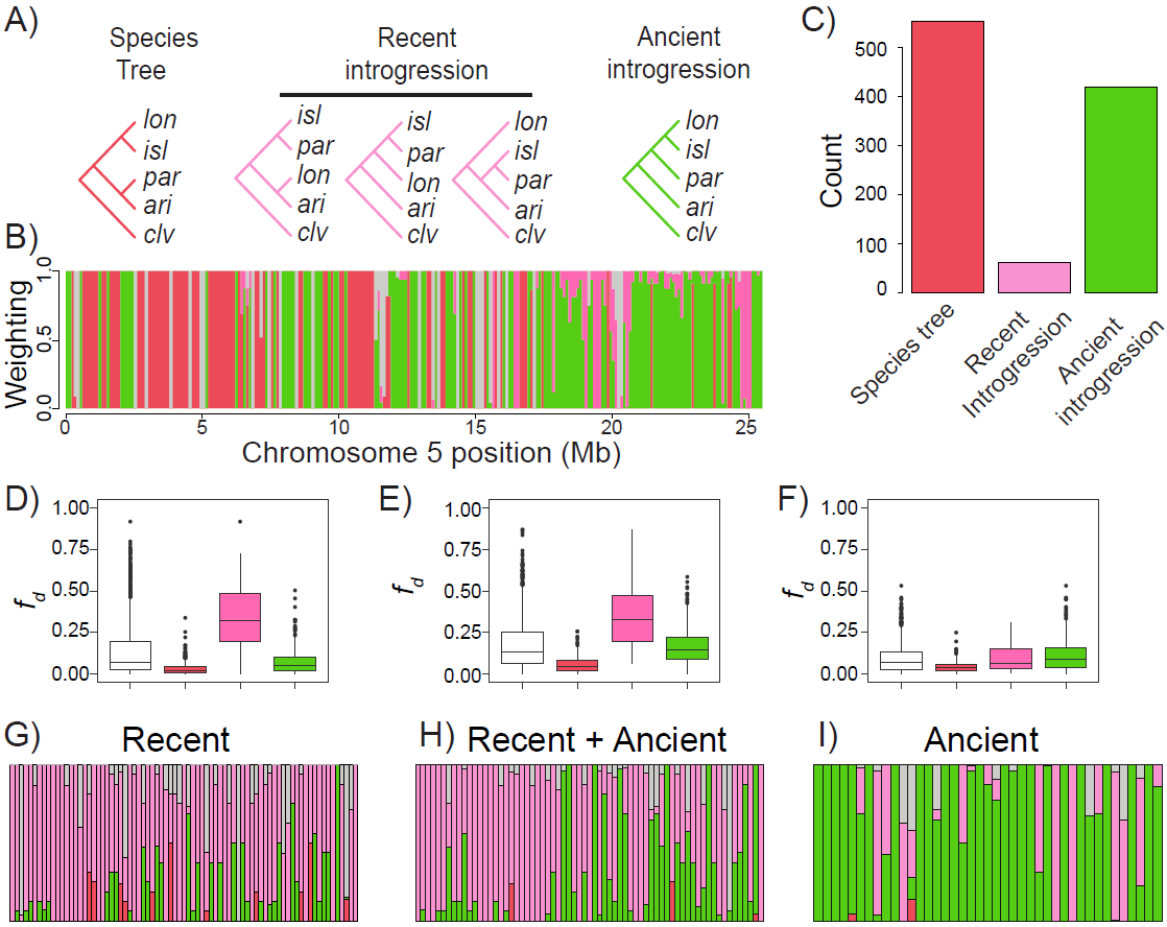

Figure 4. Genome wide variation in levels of phylogenetic discordance. (A) Topologies that represent the species tree (red), recent introgression trees (pink), and the ancient introgression tree (green). (B) Representative variation in topology weighting across chromosome 5. Red bars indicate support for the species tree, pink bars indicate support for the recent introgression tree, green bars indicate support for the ancient introgression tree, and gray bars indicate no support for these topologies. (C) The number of 100 kb windows across the genome containing a topology weighting of 1.0 for the species tree, recent introgression tree, and the ancient introgression tree. (D-F) Distribution of *fd* values among windows with a topology weighting of 1.0 for either the species tree (red), recent introgression tree (pink), or the ancient introgression tree (green). The distribution for the entire genome is provided in white. (D) Test 1 *fd* values. (E) Test 2 *fd* values. (F) Ancient *fd* values, calculated by taking the difference between Test 2 and Test 1 *fd* values. (G-I) Distribution of topology weightings among the 5% of windows with the highest *fd* values. (G) Test 1 *fd* values. (H) Test 2 *fd* values. (I) Ancient *fd* values.

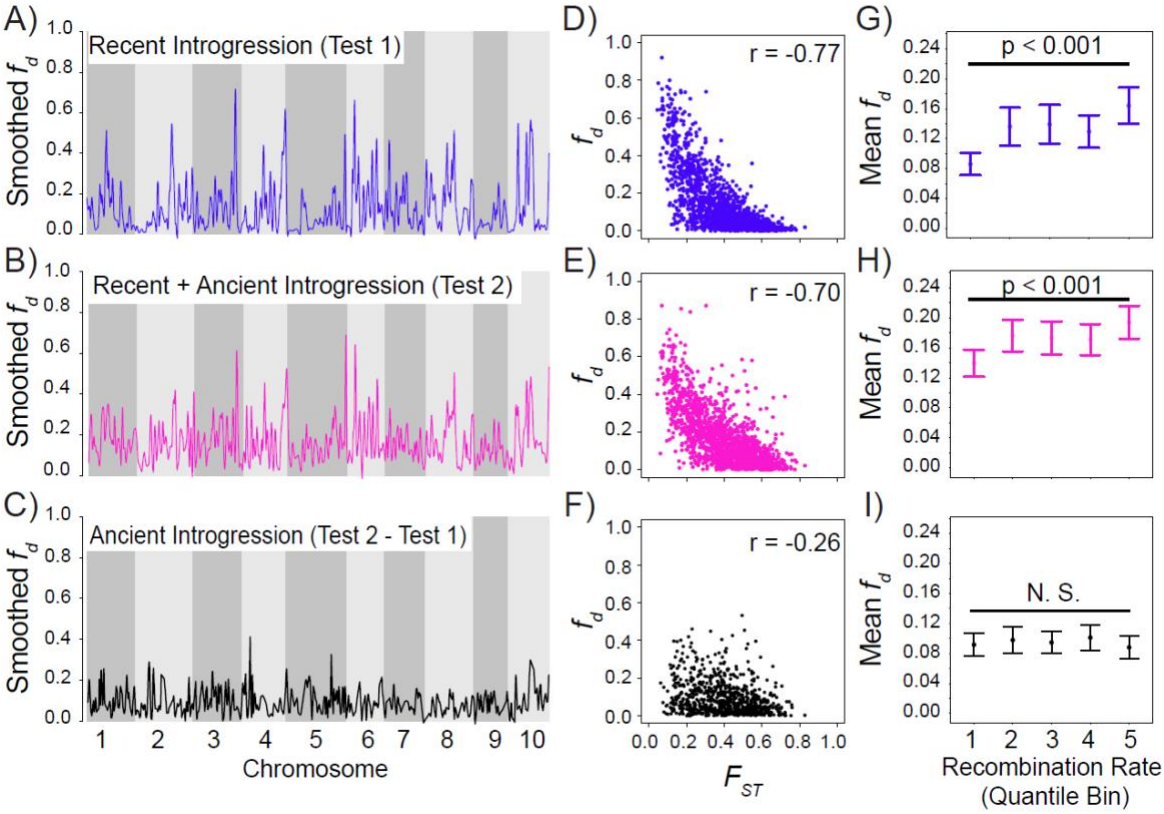

Figure 5. Genome wide variation in introgression. (A-C) Loess-smoothed *fd* values plotted across the 10 chromosomes of the *M. aurantiacus* genome. Colors indicate how the *fd* values were calculated: (A) Test 1 *fd* values, (B) Test 2 *fd* values, (C) Ancient *fd*. (D-F) Scatterplots showing the relationship between *fd* and genetic differentiation (*FST*) between Island *longiflorus* and *parviflorus*, with the correlation coefficient between the statistics in the upper right hand corner of each plot. (D) The relationship between recent introgression (i.e., Test 1 *fd*) and *FST*. (E) The relationship between both recent and ancient introgression (i.e., Test 2 *fd*) and *FST*. (F) The relationship between ancient introgression and *FST*. (G-I) Mean and 95% confidence intervals of *fd* calculated among quantile bins of recombination rate. Quantile bins of recombination rate are as follows: 1 = 0–0.824 cM/Mb; 2 = 0.824–1.51 cM/Mb; 3 = 1.51–2.33 cM/Mb; 4 = 2.33–3.66 cM/Mb; 5 = 3.66–10.9 cM/Mb. (G) Test 1 *fd* values, (H) Test 2 *fd* values, and (I) ancient *fd* values.

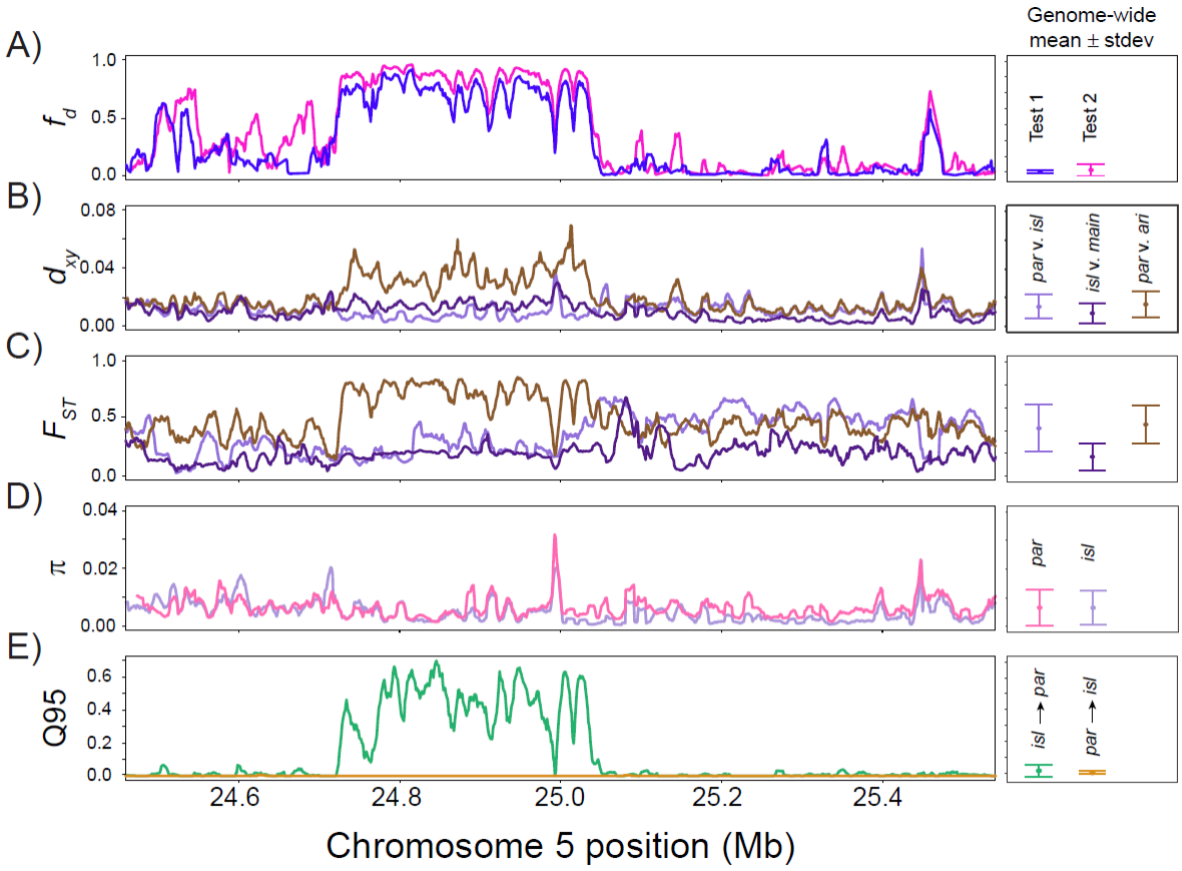

Figure 6. Hybridization leads to adaptive introgression from Island *longiflorus* into *parviflorus*. On the left, the admixture proportion (*fd*), genetic divergence (*dxy*), genetic differentiation (*FST*), genetic diversity (π), and Q95 are plotted across an 800 kb region on chromosome 5 identified as a candidate region using the Q95 statistic (Fig S7). On the right of each plot is the genome-wide mean ± standard deviation of each statistic, calculated in 10 kb windows and 1000 bp steps across all 10 chromosomes. (A) Both Test 1 and Test 2 *fd* values indicate a defined region of elevated introgression. (B) *dxy* between *parviflorus* and *aridus* (brown) is clearly elevated across this region. However, *dxy* between *parviflorus* and Island *longiflorus* (light purple) and *dxy* between Island and mainland *longiflorus* (purple) are nearly equal to genome wide levels. (C) *FST* between *parviflorus* and Island *longfilorus* (light purple) is reduced in this region, and *FST* between Island and mainland *longiflorus* (purple) is nearly equal to genome wide levels. *FST* between *parviflorus* and *aridus* (brown) displays a clear increase in this region as compared to genome wide levels. (D) No clear decrease in π was observed for either *parviflorus* (pink) or Island *longiflorus* (light purple). (E) A clear increase in the number of alleles that are shared between Island *longiflorus* and *parviflorus* but absent in *aridus* (green) was observed within this region. In contrast, no increase in the number of alleles that are shared between *parviflorus* and Island *longiflorus* but absent in mainland *longilforus* (yellow) was observed within this region, indicating unidirectional introgression from Island *longiflorus* into *parviflorus*.

Table S1. Sampling locations and number of samples sequenced from each location for this study. Four additional samples from ECSC were sequenced in Stankowski et al. 2019.

Population	number of samples	Latitude	Longitude
ECSC	3	34.018	-119.673
SC2	8	34.01915	-119.6804
SC3	7	34.01925	-119.68023
SCB	9	34.019267	-119.68493

Table S2. The t-ratio from a linear mixed-effects model testing if the mean admixture proportion (f_d) varied due to the divergence between the taxa used as P1 and P2 for the calculation of f_d. The name of the taxon used as P1, and the mean taxonomic divergence (da) between the taxa used as P1 and P2 are presented. Statistical significance is denoted as: *** for $p < 0.001$, ** for $p < 0.01$, and * for $p < 0.05$.

Factor Level	**Estimate**	***df***	***t*-ratio (Fd)**
aridus (0.00868) vs *aurantiacus* (0.00409)	0.02629	10587	4.621***
aridus (0.00868) vs *calycinus* (0.00236)	0.03681	10587	6.588***
aridus (0.00868) vs *longiflorus* (0.00219)	0.035	10587	6.269***
aridus (0.00868) vs OC (0.00333)	0.03209	10587	5.726***
aridus (0.00868) vs red (0.00328)	0.032	10587	5.709***
aridus (0.00868) vs yellow (0.00311)	0.03233	10587	5.765***
aurantiacus (0.00409) vs *calcyinus* (0.00236)	0.01052	10587	1.831
aurantiacus (0.00409) vs *longiflorus* (0.00219)	0.00871	10587	1.517
aurantiacus (0.00409) vs OC (0.00333)	0.0058	10587	1.007
aurantiacus (0.00409) vs red (0.00328)	0.00571	10587	0.991
aurantaicus (0.00409) vs yellow (0.00311)	0.006	10587	1.048
calycinus (0.00236) vs *longiflorus* (0.00219)	-0.00181	10587	-0.32
calycinus (0.00236) vs OC (0.00333)	-0.00472	10587	-0.833
calycinus (0.00236) vs red (0.00328)	-0.00481	10587	-0.849
calycinus (0.00236) vs yellow (0.00311)	-0.00448	10587	-0.79
longiflorus (0.00219) vs OC (0.00333)	-0.00291	10587	-0.514
longiflorus (0.00219) vs red (0.00328)	-0.003	10587	-0.53
longiflorus (0.00219) vs yellow (0.00311)	-0.00267	10587	-0.471
OC (0.00333) vs red (0.00328)	-0.00009	10587	-0.016
OC (0.00333) vs yellow (0.00311)	0.00024	10587	0.042
red (0.00328) vs yellow (0.00311)	0.00033	10587	0.058

Table S3. The t-ratio from a linear mixed-effects model testing if admixture proportion (f_d) varied due to the divergence between Island *longiflorus* and the taxon used as P3 for the calculation of f_d. The name of the taxon used as P3 and the mean taxonomic divergence between Island *longiflorus* and the taxon used as P3 for the calculation of f_d are presented. Statistical significance is denoted as: *** for p < 0.001, ** for p < 0.01, and * for p < 0.05.

Factor Level	**Estimate**	***df***	***t*-ratio (Fd)**
aurantiacus (0.00409) vs *calcyinus* (0.00236)	-0.01163	9044	-2.54
aurantiacus (0.00409) vs CI *longiflorus* (0.0)	-0.06997	9044	-16.035***
aurantiacus (0.00409) vs *longiflorus* (0.00219)	-0.01129	9044	-2.48
aurantiacus (0.00409) vs OC (0.00333)	-0.00909	9044	-1.942
aurantiacus (0.00409) vs red (0.00328)	-0.01068	9044	-2.248
aurantaicus (0.00409) vs yellow (0.00311)	-0.01018	9044	-2.13
calycinus (0.00236) vs CI *longiflorus* (0.0)	-0.05833	9044	-13.188***
calycinus (0.00236) vs *longiflorus* (0.00219)	0.00034	9044	0.074
calycinus (0.00236) vs OC (0.00333)	0.00254	9044	0.536
calycinus (0.00236) vs red (0.00328)	0.00096	9044	0.199
calycinus (0.00236) vs yellow (0.00311)	0.00145	9044	0.301
CI *longiflorus* (0.0) vs *longiflorus* (0.00219)	0.05867	9044	13.351***
CI *longiflorus* (0.0) vs OC (0.00333)	0.06087	9044	13.443***
CI *longiflorus* (0.0) vs red (0.00328)	0.05929	9044	12.897***
CI *longiflorus* (0.0) vs yellow (0.00311)	0.05978	9044	12.917***
longiflorus (0.00219) vs OC (0.00333)	0.0022	9044	0.467
longiflorus (0.00219) vs red (0.00328)	0.00062	9044	0.129
longiflorus (0.00219) vs yellow (0.00311)	0.00111	9044	0.231
OC (0.00333) vs red (0.00328)	-0.00158	9044	-0.323
OC (0.00333) vs yellow (0.00311)	-0.00109	9044	-0.221
red (0.00328) vs yellow (0.00311)	0.005	9044	0.099

Table S4. The t-ratio from a linear mixed-effects model testing if the mean admixture proportion for Test 1 f_d varies among recombination rate quantile bins. Quantile bins of recombination rate (in cM/Mb) are presented. Statistical significance is denoted as: *** for $p < 0.001$, ** for $p < 0.01$, and * for $p < 0.05$. Calculations of Test 1 f_d were performed using mainland *longiflorus* as P1, Island *longiflorus* as P2, and *parviflorus* as P3. *M. clevelandii* was used as the outgroup.

Factor Level	**Estimate**	***df***	***t*-ratio (Fd)**
[0,0.824] vs (0.824,1.51]	0.04604	1463	3.419**
[0,0.824] vs (1.51,2.33]	0.04239	1463	3.156*
[0,0.824] vs (2.33,3.66]	0.04099	1463	3.085*
[0,0.824] vs (3.66,10.9]	0.07756	1463	5.939***
(0.824,1.51] vs (1.51,2.33]	0.00365	1463	0.271
(0.824,1.51] vs (2.33,3.66]	0.00505	1463	0.380
(0.824,1.51] vs (3.66,10.9]	-0.03152	1463	-2.411
(1.51,2.33] vs (2.33,3.66]	0.00140	1463	0.106
(1.51,2.33] vs (3.66,10.9]	-0.03517	1463	-2.698
(2.33,3.66] vs (3.66,10.9]	-0.03657	1463	-2.838*

Table S5. The t-ratio from a linear mixed-effects model testing if the Test 2 admixture proportion (*fd*) varies among recombination rate quantile bins. Quantile bins of recombination rate (in cM/Mb) are presented. Statistical significance is denoted as: *** for $p < 0.001$, ** for $p < 0.01$, and * for $p < 0.05$. Calculations of Test 2 f_d were performed using *aridus* as P1, *parviflorus* as P2, and Island *longiflorus* as P3. *M. clevelandii* was used as the outgroup.

Factor Level	**Estimate**	***df***	***t*-ratio (Fd)**
[0,0.824] vs (0.824,1.51]	0.02829	1514	2.322
[0,0.824] vs (1.51,2.33]	0.02597	1514	2.114
[0,0.824] vs (2.33,3.66]	0.02086	1514	1.739
[0,0.824] vs (3.66,10.9]	0.05104	1514	4.258***
(0.824,1.51] vs (1.51,2.33]	0.00232	1514	0.191
(0.824,1.51] vs (2.33,3.66]	0.00744	1514	0.626
(0.824,1.51] vs (3.66,10.9]	-0.02275	1514	-1.918
(1.51,2.33] vs (2.33,3.66]	0.00511	1514	0.427
(1.51,2.33] vs (3.66,10.9]	-0.02507	1514	-2.095
(2.33,3.66] vs (3.66,10.9]	-0.03018	1514	-2.587

Table S6. The t-ratio from a linear mixed-effects model testing if the ancient admixture proportion (f_d) varies among recombination rate quantile bins. Quantile bins of recombination rate (in cM/Mb) are presented. Statistical significance is denoted as: *** for $p < 0.001$, ** for $p < 0.01$, and * for $p < 0.05$. Ancient f_d was calculated by taking the difference between f_d calculated from Test 2 and Test 1.

Factor Level	**Estimate**	***df***	***t*-ratio (Fd)**
[0,0.824] vs (0.824,1.51]	-0.00643	1049	-0.633
[0,0.824] vs (1.51,2.33]	-0.01527	1049	-1.506
[0,0.824] vs (2.33,3.66]	-0.01299	1049	-1.303
[0,0.824] vs (3.66,10.9]	-0.01479	1049	-1.433
(0.824,1.51] vs (1.51,2.33]	0.00884	1049	0.862
(0.824,1.51] vs (2.33,3.66]	0.00656	1049	0.650
(0.824,1.51] vs (3.66,10.9]	0.00836	1049	0.801
(1.51,2.33] vs (2.33,3.66]	-0.00229	1049	-0.227
(1.51,2.33] vs (3.66,10.9]	-0.00048	1049	-0.046
(2.33,3.66] vs (3.66,10.9]	0.00181	1049	0.176

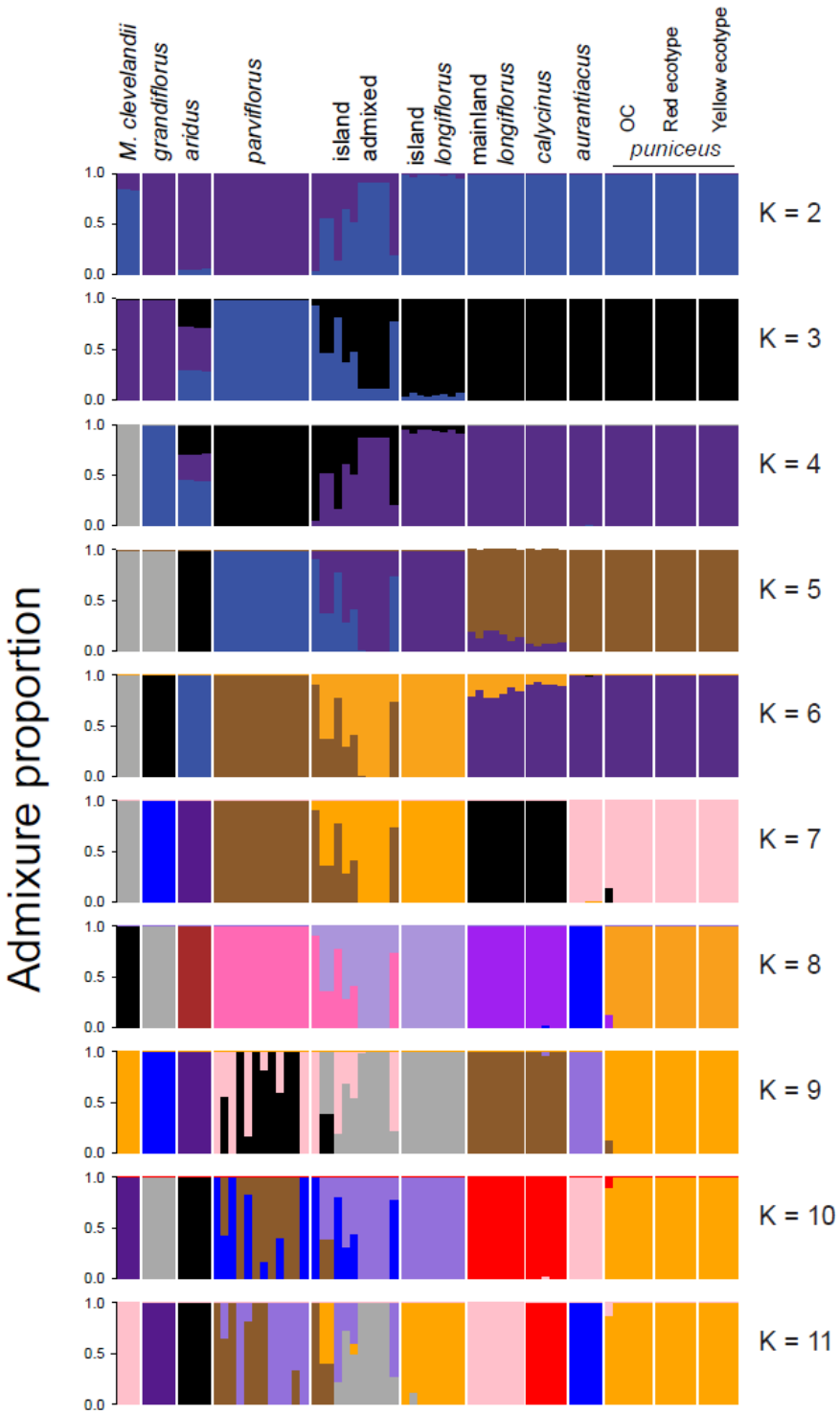

Figure S1. The ancestry proportions from *Admixture* at K = 2 to K=11 for samples from all the subspecies and their sister species, *M. clevelandii*.

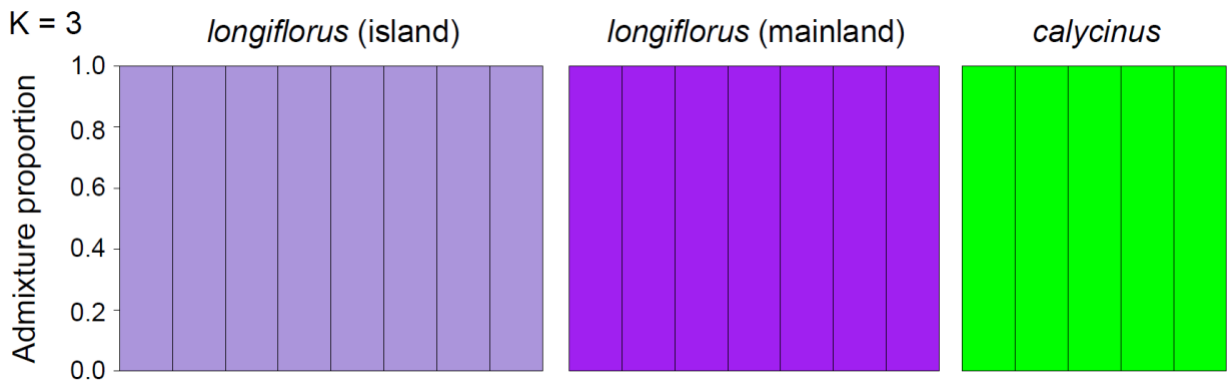

Figure S2. The ancestry proportions from *Admixture* at K = 3 from a run that included only the island *longiflorus*, mainland *longiflorus*, and *calycinus* samples.

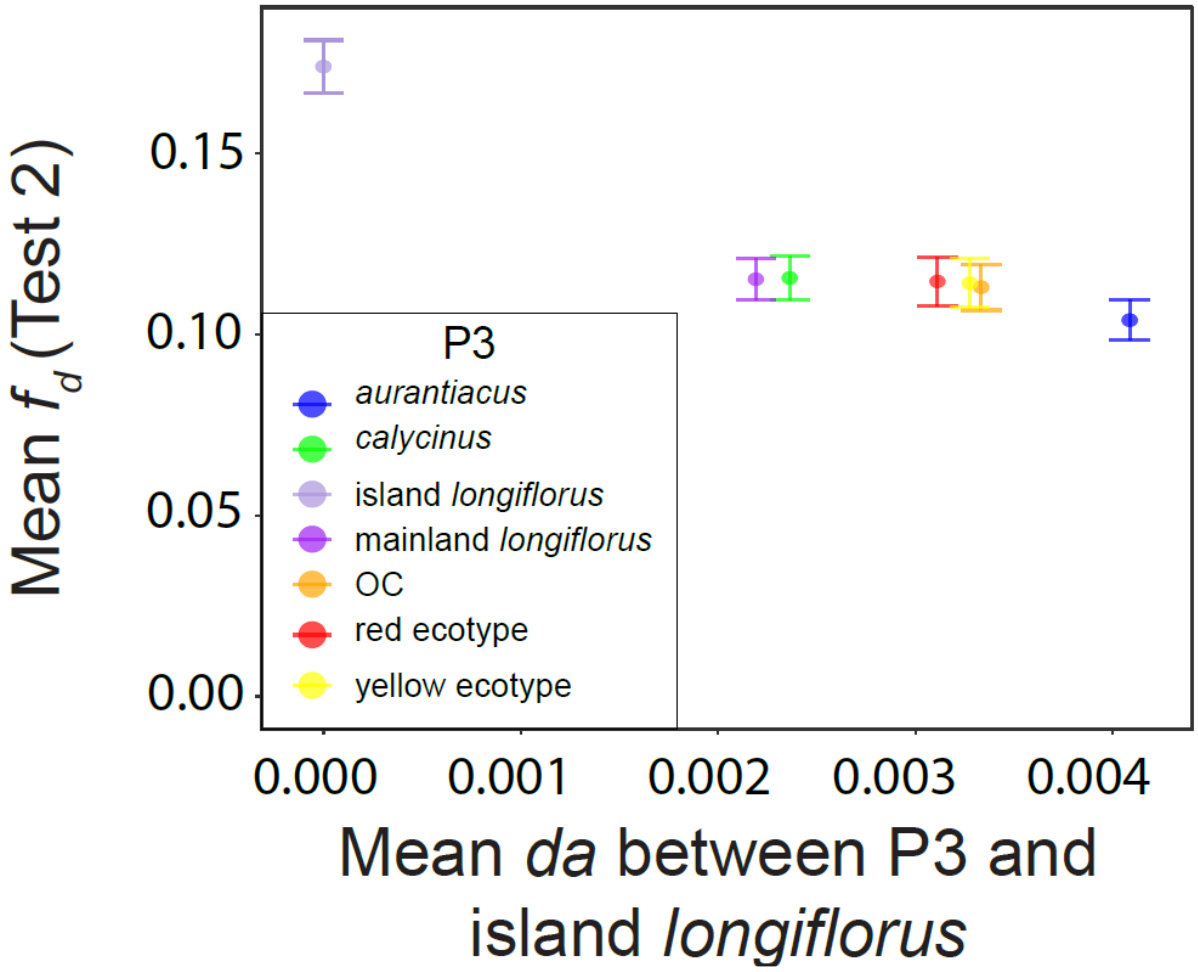

Figure S3. Mean and 95% confidence intervals of Test 2 f_d values calculated in 100kb windows are plotted against mean levels of sequence divergence (da) between Island *longiflorus* and the clade C or D taxa used as P3 for the calculation of f_d. Test 2 calculations of fd were performed using *aridus* as P1, *parviflorus* as P2, and the various clade C and D taxa as P3. Colors indicate the P3 taxon used to calculate f_d.

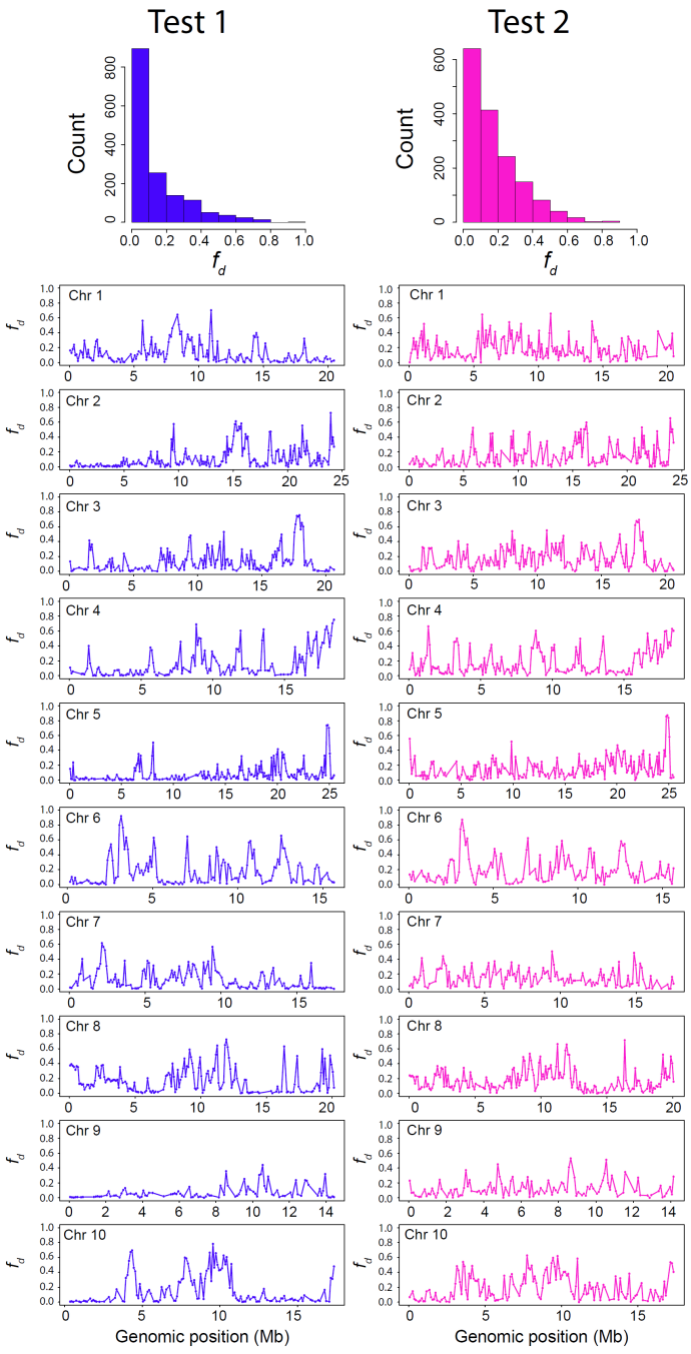

Figure S4. Histogram of the distribution of the raw Test 1 (blue) and Test 2 (pink) f_d values calculated in 100 kb windows. Genome wide variation of raw f_d values plotted across the 10 chromosomes of the *M. aurantiacus* genome. Test 1 f_d values are in blue and Test 2 f_d values are in pink. Test 1 calculations of f_d were performed using mainland *longiflorus* as P1, island *longiflorus* as P2, and *parviflorus* as P3. Test 2 calculations of f_d were performed using *aridus* as P1, *parviflorus* as P2, and island *longiflorus* as P3.

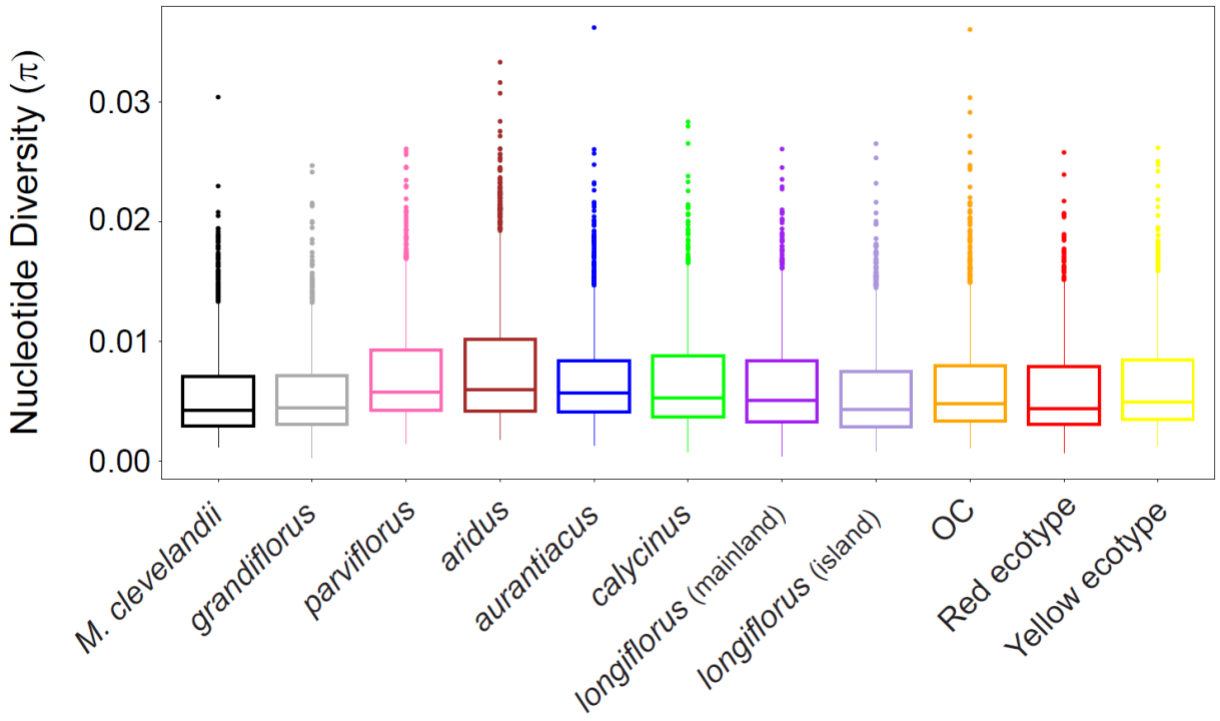

Figure S5. Boxplots of the distribution of nucleotide diversity (π) calculated in 100 kb windows for all the subspecies and their sister species, *M. clevelandii*.

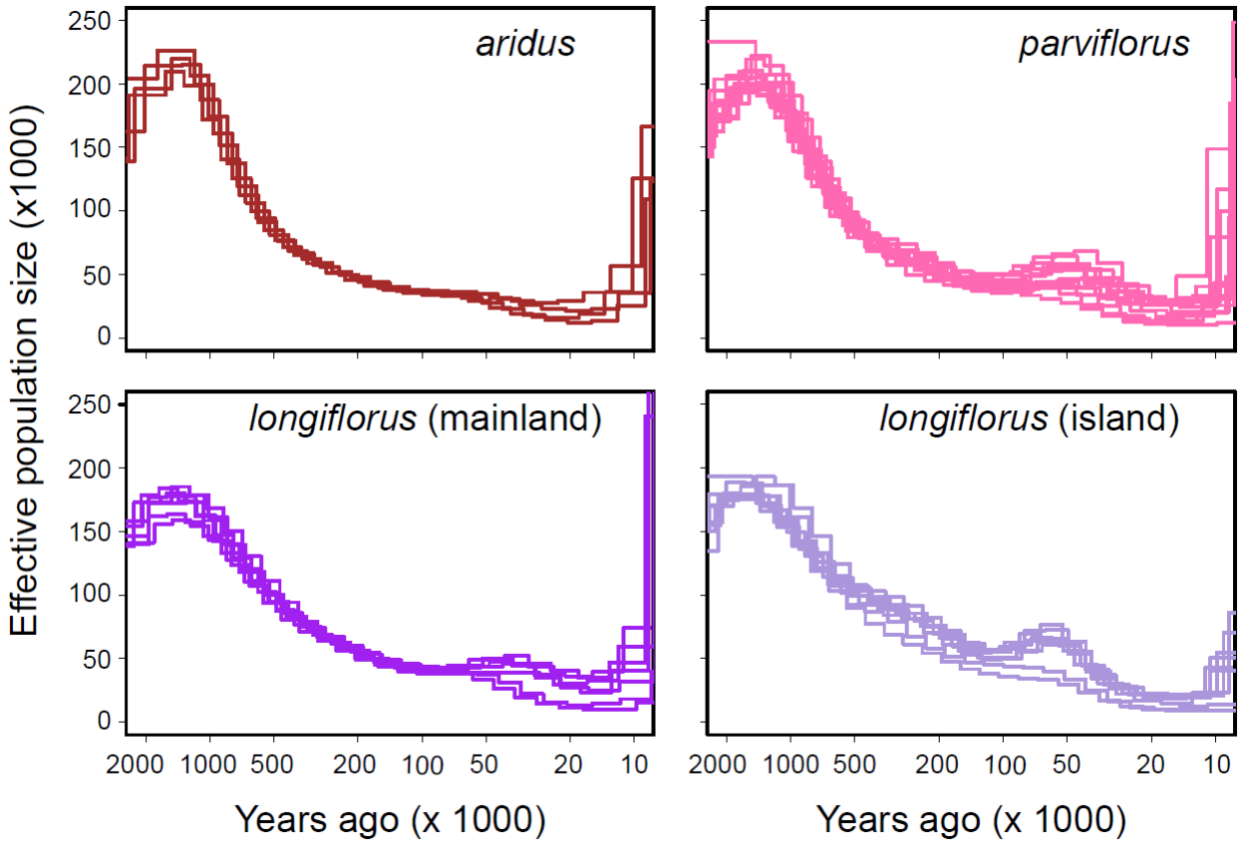

Figure S6. The estimated variation in effective population size over time from PSMC for each sequenced sample of *aridus*, *parviflorus*, mainland *longiflorus*, and island *longiflorus*.

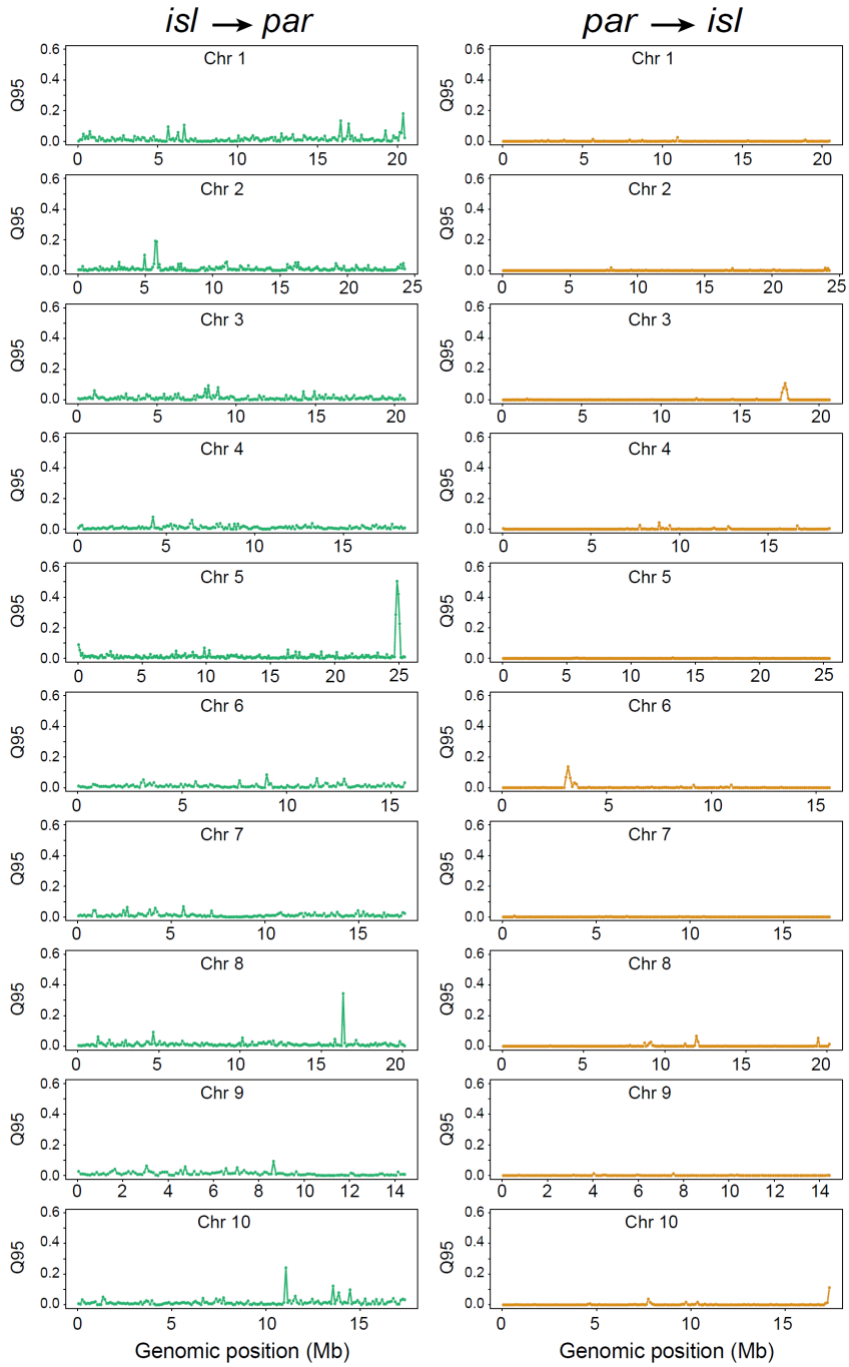

Figure S7. Genome wide variation in the Q95 statistic calculated in 100 kb windows plotted across the 10 chromosomes of the *M. aurantiacus* genome. Colors indicate the directionality of introgression, with gene flow from Island *longiflorus* into *parviflorus* in green, and gene flow from *parviflorus* into Island *longiflorus* in yellow.

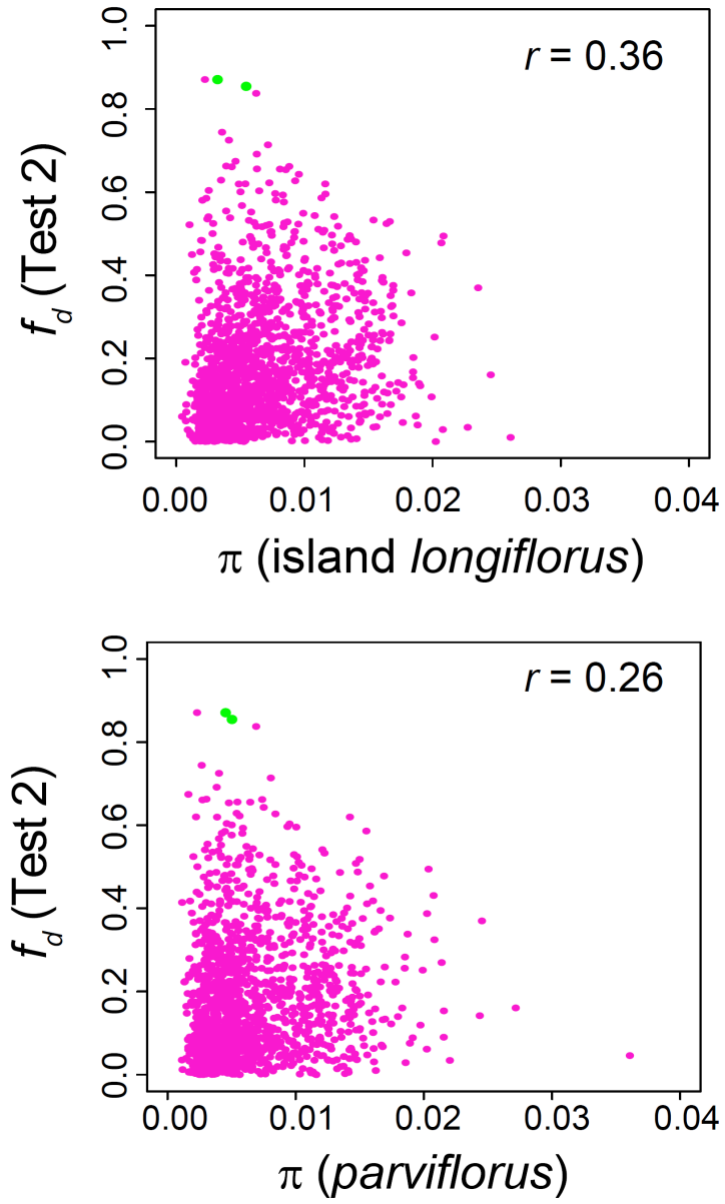

Figure S8. Scatterplots showing the relationship between Test 2 f_d and nucleotide diversity (π) for Island *longiflorus* (top) and *parviflorus* (bottom). The correlation coefficient between the statistics is presented in the upper right-hand corner of each plot. Test 2 calculations of fd were performed using *aridus* as P1, *parviflorus* as P2, and Island *longiflorus* as P3. Green points indicate the windows with the highest f_d and Q95 on chromosome 5.

3.7 References

1. Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, 19(9), 1655-1664.
2. Axelrod, D. I. (1965). Geologic history of the California insular flora. Pages 267-315 In: 1st Symposium on the Biology of the California Islands. National Park Service, 1965. 267-314.
3. Beeks, R. M. (1962). Variation and hybridization in southern California populations of *Diplacus*(Scrophulariaceae). Aliso 5: 83 – 122.
4. Brandvain, Y., Kenney, A. M., Flagel, L., Coop, G., & Sweigart, A. L. (2014). Speciation and introgression between *Mimulus nasutus* and *Mimulus guttatu*s. *PLoS Genetics*, **10**(6), e1004410. https://doi. org/10.1371/journal.pgen.1004410
5. Brauer, C. J., Sandoval-Castillo, J., Gates, K., Hammer, M. P., Unmack, P. J., Bernatchez, L., & Beheregaray, L. B. (2023). Natural hybridization reduces vulnerability to climate change. Nature Climate Change, 13(3), 282-289.
6. Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics*, *81*(5), 1084-1097.
7. California Native Plant Society (CNPS). 2023. Inventory of Rare and Endangered Plants (online edition, v8-01a). California Native Plant Society. Sacramento. Accessed 29 Nov 2023.
8. Carlquist, S. (1966). The biota of long-distance dispersal. I. Principles of dispersal and evolution. *The Quarterly Review of Biology*, 41(3), 247-270.

9. Chase, M. A., Stankowski, S., & Streisfeld, M. A. (2017). Genomewide variation provides insight into evolutionary relationships in a monkeyflower species complex (*Mimulus* sect. *Diplacus*). *American Journal of Botany*, **104**(10), 1510–1521. https://doi.org/10.3732/ajb.1700234
10. Coughlan, J. M., & Matute, D. R. (2020). The importance of intrinsic postzygotic barriers throughout the speciation process. *Philosophical Transactions of the Royal Society* B, 375(1806), 20190533.
11. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., & McVean, G. (2011). The variant call format and VCFtools. *Bioinformatics*, **27**(15), 2156-2158.
12. Duranton, M., & Pool, J. E. (2022). Interactions between natural selection and recombination shape the genomic landscape of introgression. *Molecular biology and evolution*, *39*(7), msac122.
13. Feng, X., Merilä, J., & Löytynoja, A. (2024). Secondary contact, introgressive hybridization and genome stabilization in sticklebacks. *Molecular Biology and Evolution*, msae031.
14. Gaio, D., Anantanawat, K., To, J., Liu, M., Monahan, L., & Darling, A. E. (2022). Hackflex: low-cost, high-throughput, Illumina Nextera Flex library construction. *Microbial Genomics*, *8*(1), 000744.
15. Gompert, Z., & Alex Buerkle, C. (2010). INTROGRESS: a software package for mapping components of isolation in hybrids. Molecular Ecology Resources, 10(2), 378-384.

16. 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.

17. Johnson, D. L. (1978). The Origin of Island Mammoths and the Quaternary Land Bridge of the Northern Channel Islands, California1. *Quaternary research*, *10*(2), 204-225.

18. Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493-496.

19. Liu, S., Zhang, L., Sang, Y., Lai, Q., Zhang, X., Jia, C., Long, Z., Wu, J., Ma, T., Mao, K., Street, N. R., Ingvarsson, P. K., Liu, J., & Wang, J. (2022). Demographic history and natural selection shape patterns of deleterious mutation load and barriers to introgression across *Populus* genome. *Molecular Biology and Evolution*, 39(2), msac008. https://doi.org/10.1093/molbev/msac008

20. Ma, Y., Wang, J.I., Hu, Q., Li, J., Sun, Y., Zhang, L., Abbott, R.J., Liu, J. & Mao, K. (2019). Ancient introgression drives adaptation to cooler and drier mountain habitats in a cypress species complex. *Communications Biology*, 2(1), 213.

21. Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., & Durbin, R. (2018). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature ecology & evolution*, *2*(12), 1940-1955.

22. Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite-Fast D-statistics and related admixture evidence from VCF files. *Molecular ecology resources*, *21*(2), 584-595.

23. Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, **20**(5), 229–237. https://doi.org/10.1016/j. tree.2005.02.010

24. Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., & 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

25. Martin, S. H., & Van Belleghem, S. M. (2017). Exploring evolutionary relationships across the genome using topology weighting. *Genetics*, 206(1), 429-438.

26. Martin, S. H., & Jiggins, C. D. (2017). Interpreting the genomic landscape of introgression. *Current opinion in genetics & development*, *47*, 69-74.

27. Martin, S. H., Davey, J. W., Salazar, C., & Jiggins, C. D. (2019). Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biology*, **17**(2), e2006288. https:// doi.org/10.1371/journal.pbio.2006288

28. Mead, A. (2023). *The genomic basis of adaptation to climate across oak (Quercus) species and populations in California*. University of California, Los Angeles.

29. Mead, A., Fitz-Gibbon, S., Knapp, J., & Sork, V. (2024). Comparison of conservation strategies for California Channel Island Oak (Quercus tomentella) using climate suitability predicted from genomic data. *bioRxiv*, 2024-05.

30. Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. Nature communications, 8(1), 14363.

31. Meier, J. I., Stelkens, R. B., Joyce, D. A., Mwaiko, S., Phiri, N., Schliewen, U. K., Selz, O. M., Wagner, C. E., Katongo, C., & Seehausen, O. (2019). The coincidence of ecological opportunity with hybridization explains rapid adaptive radiation in Lake Mweru cichlid fishes. *Nature communications*, *10*(1), 5391.

32. Meier, J. I., McGee, M. D., Marques, D. A., Mwaiko, S., Kishe, M., Wandera, S., Neumann, D., Mrosso, H., Chapman, L. J., Chapman, C.A., Kaufman, L., Taabu-Munyaho, A., Wagner, C. E., Bruggman, R., Excoffier, L., & Seehausen, O. (2023). Cycles of fusion and fission enabled rapid parallel adaptive radiations in African cichlids. *Science*, *381*(6665), eade2833.

33. Momigliano, P., Florin, A. B., & Merilä, J. (2021). Biases in demographic modeling affect our understanding of recent divergence. *Molecular biology and evolution*, *38*(7), 2967-2985.

34. Muhs, D. R., Simmons, K. R., Groves, L. T., McGeehin, J. P., Schumann, R. R., & Agenbroad, L. D. (2015). Late Quaternary sea-level history and the antiquity of mammoths (Mammuthus exilis and Mammuthus columbi), Channel Islands National Park, California, USA. *Quaternary Research*, *83*(3), 502-521.

35. Mychajliw, A. M., Rick, T. C., Dagtas, N. D., Erlandson, J. M., Culleton, B. J., Kennett, D. J., Buckley, M., & Hofman, C. A. (2020). Biogeographic problem-solving reveals the Late Pleistocene translocation of a short-faced bear to the California Channel Islands. *Scientific reports*, *10*(1), 15172.

36. Nelson, T. C., Stathos, A. M., Vanderpool, D. D., Finseth, F. R., Yuan, Y. W., & Fishman, L. (2021). Ancient and recent introgression shape the evolutionary history of pollinator adaptation and speciation in a model monkeyflower radiation (Mimulus section Erythranthe). PLoS Genetics, 17(2), e1009095.

37. Ortego, J., Gugger, P. F., & Sork, V. L. (2018). Genomic data reveal cryptic lineage diversification and introgression in Californian golden cup oaks (section Protobalanus). *New Phytologist*, *218*(2), 804-818.

38. Racimo, F., Marnetto, D., & Huerta-Sánchez, E. (2016) Signatures of archaic adaptive introgression in present-day human populations. *Mol Biol Evol*. 2016:34(2):296–317. https://doi.org/10.1093/molbev/msw216.

39. Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M.A.F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of evolutionary biology*, 30(8), 1450-1477.

40. Reatini, B., & Vision, T. J. (2023). The two faces of secondary contact on islands: Introgressive hybridization between endemics and reproductive interference between endemics and introduced species. *Journal of Biogeography*, 00, 1–16. https://doi.org/10.1111/jbi.14759

41. Roux, C., Fraisse, C., Romiguier, J., Anciaux, Y., Galtier, N., & Bierne, N. (2016). Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS biology*, *14*(12), e2000234.

42. Schoenherr, A. A., Feldmeth, C. R., & Emerson, M. J. (2003). *Natural history of the islands of California* (Vol. 61). Univ of California Press.

43. Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., Blazier, J. C., Sankararaman, S., Andolfatto, P., Rosenthal, G. G., & Przeworski, M. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, **360**(6389), 656–660. https://doi.org/10.1126/science.aar3684

44. Short, A. W., & Streisfeld, M. A. (2023). Ancient hybridization leads to the repeated evolution of red flowers across a monkeyflower radiation. *Evolution Letters*, *7*(5), 293-304.

45. Stankowski, S., Chase, M. A., Fuiten, A. M., Rodrigues, M. F., Ralph, P. L., & Streisfeld, M. A. (2019). Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers. *PLoS biology*, *17*(7), e3000391.

46. Stone, B. W., & Wessinger, C. A. (2024). Ecological diversification in an adaptive radiation of plants: the role of de novo mutation and introgression. *Molecular Biology and Evolution*, *41*(1), msae007.

47. Suarez-Gonzalez, A., Lexer, C., & Cronk, Q. C. (2018). Adaptive introgression: a plant perspective. *Biology letters*, *14*(3), 20170688.

48. Thorne, R. F. (1969). The California Islands. Annals of the Missouri Botanical Garden, 391-408.

49. Todesco, M., Pascual, M.A., Owens, G.L., Ostevik, K.L., Moyers, B.T., Hübner, S., Heredia, S.M., Hahn, M.A., Caseys, C., Bock, D.G., & Rieseberg, L.H. (2016). Hybridization and extinction. *Evolutionary applications*, 9(7), 892-908.

50. Tulig, M. (2000). Morphological variation in Mimulus section Diplacus (Scrophulariaceae). Ph.D. dissertation, California State Polytechnic University, Pomona, California, USA.

51. Tulig, M. C. and G. L. Nesom. (2012). Taxonomic overview of Diplacus sect. Diplacus (Phrymaceae). *Phytoneuron*, 45, 1–20.

52. Wells, H. (1980). A distance coefficient as a hybridization index: An example using *Mimulus longiflorus* and *M. flemingii* (Scrophulariaceae) from Santa Cruz Island, California. *Taxon,* 29, 53–65.

53. Zhang, D., Rheindt, F.E., She, H., Cheng, Y., Song, G., Jia, C., Qu, Y., Alström, P., & Lei, F. (2021a). Most genomic loci misrepresent the phylogeny of an avian radiation because of ancient gene flow. *Systematic Biology*, 70(5), 961-975.

54. Zhang, X., Witt, K.E., Bañuelos, M.M., Ko, A., Yuan, K., Xu, S., Nielsen, R., & Huerta-Sanchez, E. (2021b). The history and evolution of the Denisovan-EPAS1 haplotype in Tibetans. *Proceedings of the National Academy of Sciences*, *118*(22), p.e2020803118.

CHAPTER 4: A Few Barrier Loci Maintain Divergence Between Ecotypes of *Mimulus aurantiacus* subspecies *puniceus*

4.1 Introduction

The accumulation of reproductive barriers enables the maintenance of phenotypic and genetic differences between populations, which can eventually result in the formation of new species (Mayr 1942, Coyne & Orr 2004). Thus, determining how reproductive barriers accumulate through time is critical for understanding how variation within populations translates to divergence between species (Matute & Cooper 2021). Adaptation to divergent ecological conditions can act as a reproductive barrier if natural selection limits gene flow between populations (Nosil et al. 2005). In addition, genetic incompatibilities can restrict gene flow if hybrids have reduced fitness (Orr & Turelli 2001, Phadnis & Orr 2009). Divergent lineages that continue to hybridize in nature provide ideal opportunities to identify these barriers to gene flow before they become confounded with other species differences.

Gene flow between species (i.e., introgression) is now known to occur frequently between closely related species (Martin & Jiggins 2017). Even though introgression has served as an important source of genetic variation for adaptation (Suarez-Gonzalez et al. 2018), diversification (Marques et al. 2019), and speciation (Duranton et al. 2020), in most cases, genetic variation that is shared between species is expected to be deleterious. Thus, selection can result in a heterogeneous pattern of introgression across the genome (Turner et al. 2005, Ellegren et al. 2012). Introgression resistant loci are termed "barrier loci," and as speciation proceeds, they should accumulate across the genome, leading to more widespread differentiation (Wu 2001, Feder et al. 2012). Although barriers can be formed due to local adaptation and genetic incompatibilities, in the absence of gene flow, similar genomic patterns of variation can arise due

to within-species processes, such as selective sweeps, background selection, recombination rate variation, and drift (Cruickshank & Hahn 2014, Burri 2017). However, when divergence occurs in the face of gene flow, these within-species processes may be less relevant for divergence, such that genomic regions of locally elevated differentiation likely represent true barriers to gene flow (Matthey-Doret & Whitlock 2019). Hybridizing species pairs provide excellent systems to study the evolution of reproductive isolation and its genomic impacts, because they simultaneously reveal the contrasting effects of natural selection and gene flow on speciation.

Studies have attempted to use the ongoing gene flow between hybridizing species to identify evidence of selection against gene flow (for review see Harrison & Larson 2014, Ravinet et al. 2017). For example, allopatric and sympatric populations have been used to identify regions of elevated genetic differentiation and reduced introgression (Martin et al. 2013, Duranton et al. 2018). For species pairs that display geographic variation in hybridization, geographic clines in ancestry or allele frequency have been used to identify regions that are not exchanged across a hybrid zone (Stankowski et al. 2017, 2023; Westram et al. 2021). Bayesian genomic clines (BGC) is another approach that builds on the logic behind geographic clines by examining how site-specific ancestry changes with the genome-wide hybrid index (Gompert et al. 2011, 2012). In this way, BGC distributes hybrid samples based on how closely their genomes resemble one of the two parental species rather than relying on spatial variation (e.g. Guo et al. 2023). In this study, we identified barriers to gene flow between a pair of ecotypes in the *Mimulus aurantiacus* species complex that display evidence of hybridization where they come into contact but continue to remain differentiated over much of their geographic ranges.

The *M. aurantiacus* species complex is a radiation of seven closely related wood shrub species that are distributed throughout California and display extensive variation in floral traits and ecological habitat. However, despite these differences, many of these taxa come into contact in the wild and hybridize with one another. The majority of work in this complex has focused on the red- and yellow-flowered ecotypes of subspecies *puniceus*, which display a phenotypic transition in floral traits along a west to east gradient (Figure 1) (Streisfeld and Kohn 2005). Divergence between these taxa is maintained by differences in pollinator preference and ecological variation in habitat (Streisfeld and Kohn 2007; Sobel and Streisfeld 2015; Sobel et al. 2019). Differences in flower color are determined by allelic variation at the *MaMyb2* locus (Streisfeld et al. 2013). The *MaMyb2* allele that confers red flower color was introgressed deep in the history of this radiation, where it has since gone to fixation in the red ecotype due to selection, implying that this region likely acts as a barrier to gene flow (Stankowski and Streisfeld 2015, Short & Streisfeld 2023). In addition, a previous study that examined geographic clines in ancestry between these ecotypes revealed that other putative barriers to gene flow were distributed broadly across the genome (Stankowski et al. 2023).

However, this prior work focused only on broad-scale patterns of hybridization across the range of both ecotypes. In this chapter, we investigated the genomic location and extent of barriers to gene flow across a narrow hybrid zone transect that covers the full scope of the transition in ancestry between the ecotypes. This allowed us to more carefully explore a single location of contact between the ecotypes, so we could investigate the effects of gene flow and selection to the maintenance of differentiation. To identify genomic regions that show restricted introgression between the red and yellow ecotypes, we used Bayesian Genomic Clines (Gompert et al. 2012). Then, to determine whether these putative barriers also display reduced spatial

patterns of introgression across the hybrid zone, we characterized how allele frequencies and ancestry varied between populations on either end of the hybrid zone transect that resemble neighboring allopatric populations. Together, these analyses revealed evidence of a limited number of barriers that were distributed across the genome. These findings raise important questions about how these closely related ecotypes have remained distinct despite extensive allele sharing and gene flow.

4.2 Materials and Methods

4.2.1 Geographic variation in *MaMyb2* allele frequency

The primary phenotypic difference between the ecotypes is the change in flower color, which is determined by allelic variation at the *MaMyb2* locus (Streisfeld et al 2013). Red and yellow ecotypes are nearly fixed for a SNP in the second intron of the *MaMyb2* gene, with hybrids displaying a mixture of both alleles (Streisfeld et al. 2013). Previous work has focused on how *MaMyb2* allele frequency changes over broad geographic scales (Stankowski et al 2017). However, in this study, we explored how allele frequency at *MaMyb2* transitioned over a narrow geographic region through the center of the hybrid zone. We estimated the allele frequency of *MaMyb2* from six sampling locations across a transect through the hybrid zone (mean N = 19.5 per site; range: 17 – 53), as well as 10 additional populations across San Diego County that were combined with an existing dataset from 25 sites (Stankowski et al 2017, Table S1). Genotyping of the *MaMyb2*-M3 genetic marker was performed as described in Streisfeld et al. (2013). In total, the dataset contained genotypes from 801 plants across the county.

4.2.2 Whole genome sequencing and variant calling

In this chapter, we investigated how ancestry changed over geographic space within the hybrid zone by sequencing 170 individuals from 7 populations (JMC, SKT3, SKT1, SKT2, SKT4, SKT5, and SKT6) across a 16.65 km transect through the center of the hybrid zone. In addition, samples were collected from allopatric populations of the red (n = 14) and yellow (n = 13) ecotype to serve as unadmixed, reference populations. To provide the necessary sequence data for future studies on barriers to gene flow to the north of San Diego, samples were also collected from five sampling locations in Orange County (n =26), three that had yellow flowers and two that had red flowers. To examine the outcomes of admixture between San Diego and Orange County populations of *puniceus*, samples were collected from 8 populations from a suspected contact zone between these groups (n = 71). Leaf tissue was dried in silica in the field, and DNA was isolated using previously described CTAB methods (Sobel and Streisfeld 2015).

Illumina sequencing libraries were prepared according to Gaio et al (2022), with slight modifications. Bead-Linked Transposase from the Illumina Nextera XT Kit was used for initial tagmentation, generating insert sizes in the range of 400 – 1200 bp. Multiplexed libraries were sequenced on the Illumina Novaseq 6000 using paired-end 150 bp reads at the University of Oregon's Genomics Core Facility. New sequences of *puniceus* (Table S1) were combined with previously generated whole genome sequences from the subspecies of *M. aurantiacus* and their sister species *M. clevelandii* (Stankowski et al. 2019, Short & Streisfeld 2023; Short & Streisfeld 2024), resulting in a final dataset containing 331 individuals. Raw reads were filtered using fastp to remove reads with uncalled bases or poor-quality scores (Chen et al. 2018). The retained reads were then aligned to the reference assembly (Stankowski et al 2019) using BWA version 0.7.17 (Li & Durbin 2009). An average of 91.6% of reads aligned (range: 62.58% – 98.19%), and the average sequencing depth was 13.8× per individual (range: 6.5–30.2×). PCR duplicates were

marked using Picard (https://broadinstitute.github.io/picard/). Variant calling was performed following Stankowski et al. (2019). We then filtered the VCF file for biallelic SNPs with no missing data among the 331 samples using vcftools (Danecek et al. 2011). The final data set contained 8,064,402 SNPs across all 331 samples. Finally, we ran UnifiedGenotyper with the EMIT_ALL_CONFIDENT_SITES option to output all variant and invariant genotyped sites.

4.2.3 Estimation of evolutionary relationships, isolation by distance, and population structure

To estimate genetic patterns of similarity and divergence within and between populations of subspecies *puniceus*, we merged the whole genome dataset with another dataset consisting of RAD-sequences generated from 519 samples of *puniceus* collected across Orange County and San Diego. RAD-seq data from 294 samples from 25 populations were previously published (Stankowski et al 2017). In addition to these data, leaf tissue for RAD sequencing was collected from 20 locations in Orange and San Diego Counties (Table S1). DNA isolation and sequencing libraries were prepared using *PstI*, following the methods described in Etter et al. (2011), Sobel & Streisfeld (2015) and Stankowski et al. (2015). Single-end 100-bp sequencing was conducted on an Illumina HiSeq 2000 at the University of Oregon's Genomics Core Facility. RAD-seq reads were cleaned using the process_radtags script in Stacks v1.48 (Catchen et al 2013). Cleaned reads were aligned to the *Mimulus aurantiacus* genome using BWA (LI & Durbin 2019), and variants were called using Stacks version 1.48 (Catchen et al. 2013). The resulting VCF file was then combined with the VCF from the whole genome sequencing using vcftools (Danecek et al. 2011). We Only retained sites that were present in both the whole genome and RADseq VCF files, were present in 70% of all individuals, and had more than 50% of sites genotyped. This resulted in a final data set containing 139,959 SNPs across all 792 samples

(Table S1). We then used Admixture (Alexander et al. 2009) to estimate ancestry proportions from all samples, with the number of clusters (K) varying from 2 to 5.

We estimated the phylogenetic relationships among samples by generating a maximum likelihood consensus tree using IQ-TREE v1.6.12 (Nguyen et al. 2015). To do this, we concatenated SNPs from the whole genome sequencing dataset, which consisted of biallelic SNPs from 100 of the individuals that showed no evidence of admixture based on the Admixture analyses.

The role of geography in causing patterns of genetic differentiation is well established in this system (Stankowski et al. 2015). Indeed, previous work in this system found a strong signal of isolation by distance among 16 populations of both ecotypes and their hybrids sampled broadly across San Diego County (Stankowski et al 2015). To determine if genetic divergence between populations also scaled with geographic distance across the narrow hybrid zone transect sampled in this study, we calculated Fst for all 8,064,402 SNPs between each of the 7 populations from the hybrid zone transect using vcftools (Danecek et al. 2011) and regressed the mean Fst from each comparison against the pairwise geographic distance between populations.

4.2.4 Identifying patterns of introgression across the hybrid zone

To identify ancestry informative SNPs, we first identified fixed differences between unadmixed reference populations of the allopatric red and southern yellow ancestry groups from San Diego (see Results). We then used the program *bgc* to fit genomic clines for 170 admixed individuals from the 7 populations sampled along the hybrid zone transect, as they showed evidence of admixture and a weak signal of isolation by distance (see Results). *bgc* uses a Bayesian approach to identify SNPs whose probability of ancestry deviates from the genome wide hybrid index (Gompert & Buerkle 2011, 2012). This method estimates two parameters: one that reveals excess

ancestry from either group (alpha), and one that represents the rate of introgression relative to the gradient in genome wide hybrid index (beta). Sites that deviate from the neutral expectation are considered outliers for alpha or beta. Positive beta outliers indicate SNPs whose ancestry changes more gradually than the hybrid index, revealing variants with restricted introgression across the hybrid zone. We used ClineHelpR (Martin et al. 2021) to compile tables of the Bayesian Genomic Cline (BGC) results and identify alpha or beta outliers that exceeded the 97.5^{th} percentile of beta values.

Although BGC examines how ancestry changes with hybrid index, it does not consider how it varies across geographic space. Given the spatial variation in ancestry across the hybrid zone between the red and yellow ecotypes, we have an opportunity to corroborate the BGC outliers by identifying those sites that also show restricted introgression from one side of the hybrid zone to the other. We compared allele frequencies at two populations on the western side of the hybrid zone (SKT3 and JMC; both with median hybrid index greater than 0.75 for red ancestry) with a single population on the eastern side of the zone (SKT6; median hybrid index greater than 0.75 for yellow ancestry). We first calculated the allele frequency at each of the beta outlier SNPs for all 7 populations. We then fit linear models for each SNP, with allele frequency as the response variable and population as the predictor variable. We used the *emmeans* package in R (Lenth, 2019) to perform pairwise comparisons of the estimated marginal mean allele frequencies from the linear model, focusing on the contrasts between the populations on the western and eastern side of the hybrid zone and adjusting the p-value for multiple tests. To determine if these results differed between outlier and non-outlier SNPs, we repeated this analysis using all of the fixed SNPs.

To determine whether putative barrier SNPs were more genetically differentiated than non-barrier SNPs, we used vcftools to calculate Fst at each of the fixed SNPs between SKT3 and SKT6 (and between JMC and SKT6). We then calculated the Spearman's correlation between Fst and the beta score estimated by *bgc*.

4.2.5 Identifying additional barriers to gene flow

To estimate genome wide variation in gene flow between the red and yellow ecotypes of *puniceus*, we first phased the whole-genome VCF using Beagle (Browning, & Browning 2007), and then performed local ancestry inference using FLARE (Browning et al. 2023) based on the unadmixed red and southern yellow ecotype samples as references. We retained ancestry calls with a probability greater than 80%, resulting in a dataset of 353,165 SNPs. We calculated the difference in the frequency of yellow ancestry for each of the sites between JMC, SKT3, and SKT6, with the expectation being that a barrier to gene flow would show high yellow ancestry in SKT6 but low yellow ancestry in both JMC and SKT3. For the region that displayed the largest number of positive beta outliers and a corresponding large difference in ancestry between populations on either end of the zone, we explored how the haplotypes varied among individuals from the different populations across the hybrid zone.

4.3 Results

4.3.1 Population structure within *Mimulus aurantaicus* subspecies *puniceus*

There is a clear and rapid transition in *MaMyb2* allele frequencies between western and eastern populations in San Diego County, as the yellow allele is nearly fixed in the east but almost completely absent from the west (Figure 1). Across the narrow hybrid zone transect, the yellow allele remained nearly fixed on the eastern edge of the hybrid zone (frequency of 0.972 in

population SKT6), but it rapidly declined to a frequency of 0.333 on the western edge of the contact zone (population SKT3). This transition occurred over a distance of only 5.5 km, with both alleles present at roughly similar frequencies through the center of the hybrid zone (mean yellow allele frequency: 0.53). Moving 11.2 km farther to the west resulted in the almost complete loss of the yellow allele (frequency 0.04 in population JMC).

At K=4, *Admixture* assigned samples to four ancestry groups that varied according to geography (Figure 2). Samples of the red ecotype formed one group in western San Diego County, while samples to the north in Orange County formed a second group, consistent with previous analyses that identified a genetic break between the two regions (Chase et al. 2017; Short and Streisfeld 2023). While prior population genetic studies in San Diego revealed a single yellow ecotype that was common in eastern portions of the county, the more complete spatial sampling available in the current study identified a clear break in ancestry between northern and southern populations of the yellow ecotype (Figure 2). This is consistent with the presence of two reciprocally monophyletic clades consisting of samples of the yellow ecotype in the genome-wide phylogeny (Figure S1), which suggests that there is greater genetic diversity in the yellow ecotype than the red ecotype. This could be due to historical geographic isolation between the northern and southern populations of the yellow ecotype, admixture with a more diverged lineage, or adaptation to local environmental conditions.

In addition, there appears to be evidence of admixture along multiple geographic axes among these different ancestry groups. In particular, we found evidence of mixing between Orange and San Diego Counties, as well as between the red ecotype and both the northern and southern yellow ecotypes (Figure 2). Indeed, there was a transition in southern yellow ancestry across the narrow hybrid zone transect that we sampled, with nearly 100% yellow ancestry in

SKT6 on the eastern edge of the hybrid zone but a decline to 65% yellow ancestry in SKT3 on the western edge. Slightly farther to the west, we found that southern yellow ancestry dropped even further to 11% in the JMC population. The geographic distance between populations was only a weak predictor of mean Fst between populations ($R^2 = 0.069$; Figure 3). Previous work showed a much stronger relationship among populations sampled more broadly across San Diego County ($R^2 = 0.281$; Stankowski et al. 2015). This indicates that the transition in ancestry between these populations is not heavily influenced by their relative geographic distances. While future analyses should also focus on hybridization and barriers to gene flow among the northern ancestry groups, in this study we examined the contact zone between the red and southern-yellow ancestry groups.

4.3.2 Identification of barriers to gene flow

To explore patterns of introgression across a gradient in hybrid index scores, we fit genomic clines to 301 SNPs from 170 individuals across the 7 hybrid zone populations. The 301 sites were differentially fixed between the red and southern yellow ancestry groups and were found scattered across all 10 chromosomes (Figure S2). Similar to patterns of ancestry observed from Admixture analyses (Figure 2), the hybrid index estimated from *bgc* for these 301 SNPs showed a clear spatial gradient (Figure 4A-B). Populations on the western edge (JMC and SKT3) and eastern edge (SKT6) of the hybrid zone had substantially different hybrid indexes from each other and relative to the populations in the center of the zone.

Based on this variation in hybrid index, *bgc* estimated two parameters for each SNP, known as alpha and beta. The alpha parameter measures excess ancestry from one of the two parental species relative to the genome wide hybrid index (Figure 4C). Specifically, a negative alpha indicates a site with excess red ancestry relative to the overall hybrid index, whereas a

positive alpha value indicates an excess of southern yellow ancestry. We found a median alpha of 0.77, with values ranging from -7.29 and 5.07 (Figure S3, S4). Considerably more of the sites had negative alpha values than positive (215 vs 86) (Figure S3, S4). Of these, 71 were described as negative alpha outliers and 35 were positive alpha outliers, implying more biased introgression from the red ecotype into the southern yellow ecotype. The second parameter, beta, describes how rapidly a transition in ancestry at a particular site deviates from the genome-wide gradient in hybrid index. We found a wide range of beta values (-9.70 — 32.12), with a median of -2.86 (Figure S3, S5). Most sites (223/301) had negative beta, such that transitions in ancestry occurred more slowly than the genome wide average. By contrast, 78 sites displayed positive beta values (Figure S3, S5). Of these, 61 of the sites were considered positive beta outliers, implying that transitions from red to southern yellow ancestry occurred more rapidly than expected based on the distribution of hybrid index (Figure 4C). These 61 sites represent putative barrier loci, as introgression appears to be restricted relative to the gradient in hybrid index among hybrid individuals.

Genomic clines are based on the variation in hybrid index among hybrid individuals, rather than spatial differences in ancestry. The narrow hybrid zone we sampled varied both in hybrid index and along a west-east transect. Therefore, we determined whether the 61 putative barrier SNPs identified from *bgc* also differed spatially across the hybrid zone (Figure 5), which would add support to the conclusion that introgression across the hybrid zone was restricted for these SNPs. We found that 50 of the 61 SNPs displayed a significant difference in allele frequency between JMC and SKT6, 53 SNPs displayed a significant difference in allele frequency between SKT3 and SKT6, and 49 of the SNPs were found to differ in both tests (based on a Bonferroni corrected p-value = 0.0003; 0.05/183 tests). There were no significant

differences in allele frequency for any of the 61 outlier SNPs between JMC and SKT3. We repeated these analyses using the 240 non-outlier SNPs to determine whether any of them had significant differences in allele frequency between these populations. Despite a larger number of SNPs, there were fewer sites that showed differences in allele frequency between populations on either edge of the hybrid zone. Specifically, of the 240 SNPs, only 43 differed in allele frequency between JMC and SKT6, 51 varied between SKT3 and SKT6, and 28 differed between JMC and SKT3 (based on a Bonferroni corrected p-value = 0.00006; 0.05/720 tests). Consistent with these patterns, allele frequency differences between the populations were significantly greater among the positive beta outliers relative to the non-outliers (Figure S6) (JMC vs SKT6: mean of positive beta outliers = 0.8, mean of non-beta outliers = 0.36, t = 10.673, p-value < 2.2e-16; SKT3 vs SKT6: mean of positive beta outliers = 0.6, mean of non-beta outliers = 0.19, t = 12.921, p-value < 2.2e-16). Thus, many of the 61 outlier SNPs identified from *bgc* also varied spatially, suggesting that introgression was restricted across the hybrid zone. By contrast, despite the large differences in allele frequency between populations on either edge of the hybrid zone (spanning a distance of only 5-16 km), allele frequencies tended to fluctuate more substantially in the center of the hybrid zone, which spanned a distance of only 2 km (Figure 5).

In addition, we calculated Fst between JMC and SKT6 (and between SKT3 and JMC) at each of the 301 SNPs that were fixed between the red and yellow ecotypes. We generally identified higher Fst between JMC and SKT6 (median = 0.2942) than between SKT3 and SKT6 (median = 0.1908). The Fst values also tended to be greater for those SNPs identified as beta outliers (median Fst between JMC and SKT6 = 1.0, median Fst between SKT3 vs SKT6 = 0.6637) than those not identified as beta outliers (median Fst between JMC and SKT6 = 0.2595,

median Fst between SKT3 vs SKT6 = 0.1503). This leads to a strong correlation between Fst and beta (JMC-SKT6: $r = 0.70$; SKT3-SKT6: $r = 0.61$; Figure 6).

4.3.3 Extensive variation in genome wide levels of introgression

Using local ancestry inference, we generated red or southern yellow ancestry calls at 353,165 sites. We identified extensive genome-wide variation in the frequency of southern yellow ancestry between SKT6 on the eastern edge of the hybrid zone and JMC and SKT3 on the western side (Figure S7,S8). Although the frequency of southern yellow ancestry ranged between 0.0 to 0.92 in JMC, 0.0 to 0.97 in SKT3, and 0 to 1 in SKT6 (Figure S7), most of the ancestry informative sites had a very low frequency of southern yellow ancestry in JMC, with 75% of the sites displaying a frequency of southern yellow ancestry less than 0.1 (median = 0.08). Ancestry patterns in SKT3 were also shifted toward the red ecotype, with a median southern yellow ancestry of 0.14. By contrast, SKT6 displayed a more even distribution of southern yellow ancestry, with a median of 0.42.

To identify regions that showed restricted introgression across the hybrid zone, we calculated the difference in southern yellow ancestry at each ancestry informative site between SKT3 and SKT6, and between JMC and SKT6 (Figure 8). Of the 353,165 sites, 84% showed higher southern yellow ancestry frequency in SKT6 than SKT3 and 94% of sites had higher southern yellow ancestry in SKT6 compared to JMC. Of these, the median difference between SKT3 and SKT6 was 0.28 (range 0.004 – 0.976), and the median between JMC and SKT6 was 0.36 (range 0.028 – 1.00). In general, comparisons involving JMC showed greater differences in southern yellow ancestry with SKT6 than comparisons between SKT3 and SKT6. This is likely due to the greater geographic distance between the populations. Across the genome, there were large regions on multiple chromosomes that showed a high frequency of southern yellow

ancestry in SKT6 but a low frequency of yellow ancestry in both SKT3 and JMC (Figure S8). However, only three genomic regions that contained positive beta outliers from *bgc* also showed large differences in the frequency of southern yellow ancestry between populations on either side of the hybrid zone (Figure 7,8). Specifically, of the 61 positive beta outliers, 39 of them were found in these three regions on chromosome: 3 (4 SNPs spanning 21,698 bp), 4 (32 SNPs spanning11,832 bp) chromosome 7 (3 SNPs spanning 13,703 bp. Although there were many other regions that showed substantial differences in yellow ancestry across the genome, most of these did not contain any fixed SNPs used for the *bgc* analysis, indicating that there may be additional barriers that we have not detected.

To further explore how ancestry changes across the hybrid zone in the genomic region on chromosome 4 containing 32 of the 61 positive beta outliers, we examined ancestry tracts from phased haplotypes across each individual. Consistent with ancestry proportions across the genome, we found greater red ancestry at these sites in JMC and SKT3 as compared to SKT6 but more mixed ancestry in the populations from the center of the hybrid zone (Figure S9).

4.4 Discussion

4.4.1 Population structure within *Mimulus aurantiacus* subspecies *puniceus*

We identified evidence of four ancestry groups within subspecies *puniceus*. These groups correspond with current and previous phylogenetic analyses in this radiation (Chase et al. 2017, Stankowski et al. 2019, Short and Streisfeld 2023, Figure S1). However, we also identified evidence of population structure between northern and southern populations of the yellow ecotype, which was initially speculated by Stankowski et al (2015). Furthermore, we identified evidence of distinct zones of hybridization between the red ecotype and both the northern and southern populations of the yellow ecotype. Independent hybrid zones between taxa with similar

phenotypes have been reported previously in a number of systems (Mandeville et al. 2015, Schumer et al. 2018, Westram et al. 2020, Sianta et al. 2024). These studies have reported that the outcomes of this hybridization generally vary across each contact zone due to differences in the local environment and demographic history (Harrison & Larson 2014, Schumer et al. 2018, Sianta et al. 2024). In future work, we plan to compare how hybridization between these northern and southern ancestry groups could lead to discoveries about global and local barriers between these taxa.

We also identified evidence of hybridization between the red ecotype and Orange County *puniceus*. Unlike in San Diego County, differences in flower color between populations are not correlated with patterns of ancestry (Figure 2). Instead, the populations in Orange County all share similar ancestry but form a phenotypic hybrid zone that is not geographically structured, with some populations displaying red flowers and others consisting of yellow flowers (Beeks 1962). In this study, we sequenced plants from a contact zone between the red ecotype in San Diego and populations in Orange County, revealing a gradual transition in ancestry between these groups. In future work, we will examine genome wide and local levels of ancestry within these admixed populations to identify barriers to gene flow between them. The greater sequence divergence between red ecotype and Orange County populations provides an opportunity to examine if the same barrier loci accumulate and vary across independent contact zones.

4.4.2 Barriers to gene flow between red and southern yellow ancestry groups

Geography plays an important role in determining the extent of introgression across a hybrid zone (Sedghifar et al. 2015). In general, gene flow is expected to be more common between geographically proximate populations, but it will decline with increasing geographic distance between populations (Wright 1943). In this study, we sampled individuals from

populations across a narrow transect through the center of the hybrid zone. This allowed us to identify regions of the genome that showed restricted introgression across the hybrid zone. Previous work in this system examined how ancestry changed spatially across all of San Diego County (Stankowski et al 2023). While such a broad scale analysis may provide insight into how global barriers to introgression formed across the entirety of the hybrid zone, it does not take into account that there are likely multiple points of contact between these ecotypes. Indeed, we have shown here that there are multiple contact zones between the red ecotype and other lineages of subspecies *puniceus* (Figure 2). Thus, any reproductive barriers that exist might be the same or may differ among these contact points (Harrison & Larsen 2016; e.g. Sianta et al. 2024). By sampling densely across a small region, we can examine how ancestry changes over a single point of contact between these ecotypes. Furthermore, by sampling over such a small region, we were able to identify the impact that geography had on reducing levels of gene flow across this contact zone.

In this study, we identified very few SNPs that were differentially fixed between the red ecotype and southern populations of the yellow ecotype. Thus, there is extensive allele sharing between the groups, which may be the result of gene flow or shared ancestral polymorphism. Genome-wide patterns of genetic differentiation do not appear to be associated with geography through this narrow hybrid zone transect, implying that these populations are all exchanging genes (Figure 3). However, of the 301 fixed sites, we were able to demonstrate evidence of reduced gene flow at 61 of these variants. Many of these sites also showed spatial differences in allele frequency between populations on either edge of the hybrid zone, suggesting that introgression is likely restricted at these regions of the genome. Although we only found evidence of a few barriers, there is an increasing appreciation that a small number of loci can

contribute to the maintenance of species barriers (e.g. Wessinger et al. 2023). By contrast, previous analysis in this same system using samples from throughout San Diego County suggested that there may be many barriers across the genome (Stankowski et al. 2023). However, as noted above, this previous study examined spatial patterns in ancestry broadly across multiple contact points, potentially combining barriers that are unique to each contact point. Future work that sequences genomes from samples across additional contact points will be required to test this hypothesis.

Although speciation in the face of continuous gene flow may be unlikely, it is now recognized that divergence with gene flow following a period of geographic isolation is possible (Feder et al. 2012, Feder et al. 2013; Abbott et al. 2013, Harrison & Larson 2014). Previous demographic modelling between the red and yellow ecotypes using samples collected from across San Diego indicated that the ecotypes have come into contact after a period of geographic isolation (Stankowski et al 2023). After contact resumed, the model inferred that there has been ample migration between the ecotypes for the past 1,800 generations (roughly ~3,600 years) (Stankowski et al. 2023). Our BGC analysis confirms this extensive gene flow, as it showed widespread evidence of introgression. Interestingly, introgression appears to be occurring primarily from red into southern yellow, which may be related to the incomplete preference of hummingbirds to visit red flowers (Streisfeld and Kohn 2007). Despite evidence for widespread introgression, we were able to detect a small number of fixed differences between these taxa, and of those, a small number appeared to correspond to barriers. Interestingly, and consistent with previous findings (Short and Streisfeld 2023; Stankowski et al. 2023), the region with the most positive beta outliers (32 or 61 outlier SNPs) localized to a roughly 12 kb region on chromosome

4 that contains the *MaMyb2* gene responsible for the transition from yellow to red flowers (Streisfeld et al 2013).

As previously mentioned, the red allele at *MaMyb2* was introgressed from a more distantly related subspecies into the ancestor of the clade to which both the red and yellow ecotypes belong (Stankowski & Streisfeld 2015, Short & Streisfeld 2023). However, this allele later went to fixation in the red ecotype due to selection and appears to have been lost in the yellow ecotype. Thus, the large number of fixed differences between the red and yellow ecotype at this locus can be attributed to the divergent history of this locus followed by strong selection in the red ecotype, resulting in its fixation. The flower color differences due to allelic variation at *MaMyb2* appear to contribute to pollinator-mediated reproductive isolation between these ecotypes, with hummingbirds preferring red flowers and hawkmoths preferring yellow flowers (Sobel and Streisfeld 2015; Streisfeld and Kohn 2007). The results of the current study provide additional evidence that divergent selection at *MaMyb2* appears to lead to substantial reductions in gene flow between these ecotypes, but in this case, we have been able to document this change over a very small spatial scale (i.e., on the order of 5 km).

Despite this important barrier, pollinator isolation is not complete, resulting in the ongoing production of hybrids. Nevertheless, the ecotypes remain phenotypically distinct outside of the hybrid zone. Geography and divergent environmental adaptation also appear to contribute to the maintenance of species boundaries between these ecotypes (Sobel and Streisfeld 2015; Sobel et al. 2019). Specifically, it has been shown that the red and yellow ecotypes inhabit distinct ecological niches and that hybrid populations appear to be largely present in geographic regions where the ecological niches of these ecotypes overlap (Sobel & Streisfeld 2015). Furthermore, differences in drought tolerance have been observed between these ecotypes (Sobel

et al. 2019). However, the difference in habitat occurs over a more gradual transition than what is seen for floral traits (Stankowski et al 2015). It is possible that a combination of both divergent pollinator and environmental adaptation is contributing to the maintenance of divergence between these ecotypes, and that some of the regions of reduced introgression identified in our study may correspond to these ecological differences. However, it is also possible that hybrid incompatibilities may contribute to their divergence. Ongoing work in this system is in the process of scanning genomes for unusual patterns of linkage disequilibrium between chromosomes that could be caused by hybrid incompatibilities (Schumer and Brandvain 2016). It will be interesting to determine the extent of overlap in the loci identified by these two parallel analyses.

In addition, as mentioned above, examining the outcomes of hybridization between different contact points with the red ecotype will provide a rare opportunity to examine the repeatability of the outcomes of hybridization between these different ancestry groups. Given that all of these taxa are at the very early stages of divergence from one another, it will be interesting to see how gene flow and selection vary among independent contact zones.

4.4.3 Conclusion

In this study, we examined the outcomes of hybridization between a pair of hybridizing taxa that are at the very early stages of speciation. Given the recent divergence between these hybridizing taxa, we might expect to identify weaker and fewer reproductive barriers compared to a pair of taxa further along the speciation continuum (Coyne & Orr 1989, Moyle et al, 2004). The hybrid zone between the red and yellow ecotypes of *puniceus* has been studied extensively. Previous work has detected potential barrier loci that were widespread across the genome, but ancestry changes were examined over broad geographic spaces across the entire county (Stankowski et al

2023). In this study, we focused on the transition in ancestry across a narrow transect through center of the hybrid zone.

This study advances our understanding of how ancestry changes over a narrow geographic space within a well-studied hybrid zone. In addition, we were able to take advantage of a previously identified reproductive barrier whose genetic basis has been identified. Consistent with previous work on hybridization in this group, we identified evidence of extensive gene flow between the ecotypes occurring over a very small geographic region. However, we identified extensive variation in gene flow across the genome, with some regions showing clear evidence of reduced gene flow between them. The strongest signal of reduced gene flow was found in the region surrounding the *MaMyb2* locus. However, we also identified additional regions of potentially reduced gene flow, indicating additional barriers that will require further characterization. Ongoing work has the opportunity to identify loci involved in hybrid incompatibilities, as well as the potential to determine the repeatability of these loci across multiple contact zones. Together, these analyses suggest that there are relatively few barriers to gene flow between these ecotypes, but they are sufficient to maintain distinct phenotypes. These findings support a role for divergence with gene flow, where ecological and genetic differences likely accumulated in allopatry but now contribute to reproductive isolation upon secondary contact.

4.5 Figures and Tables

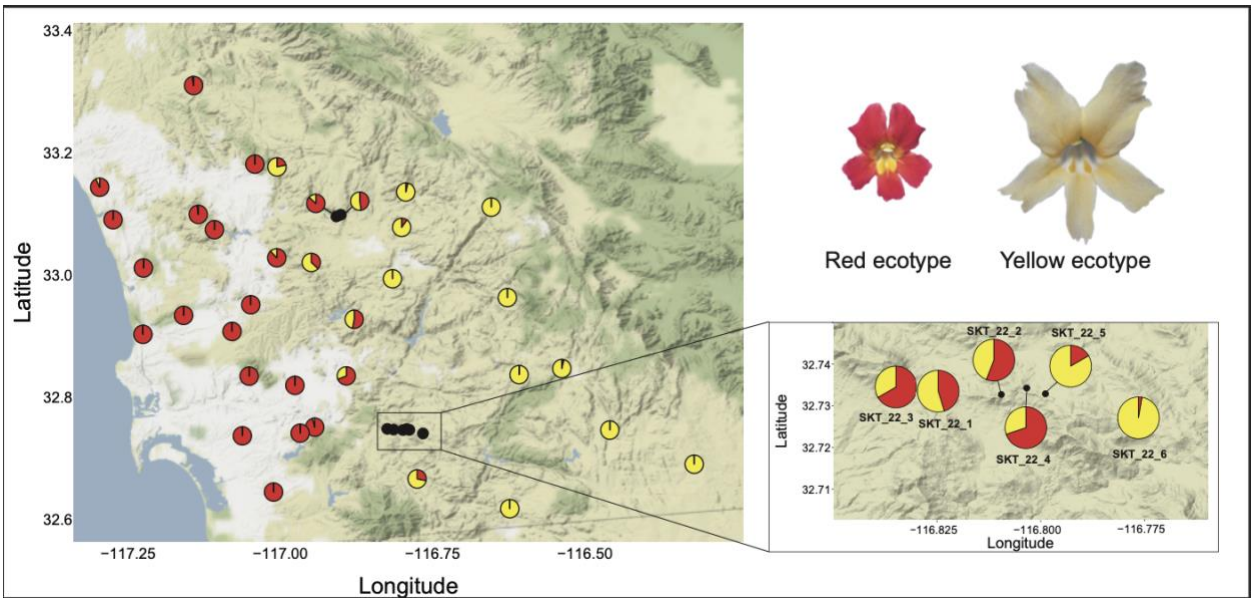

Figure 1. Map of MYB2 allele frequency in sampling locations from throughout San Diego County, California, USA

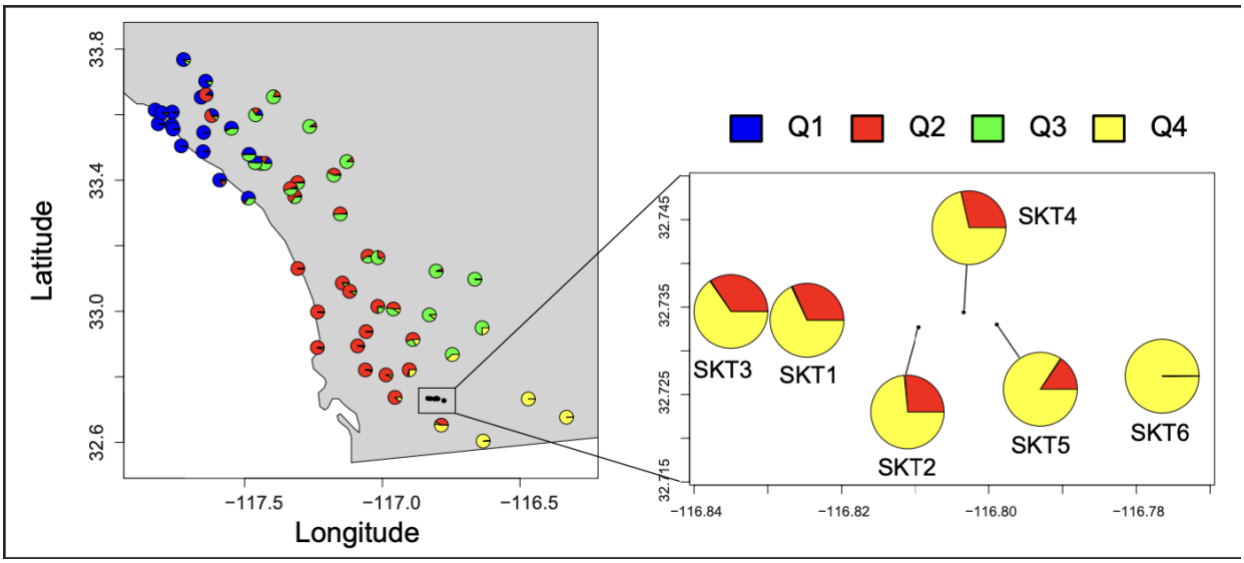

Figure 2. Map of ancestry scores from admixture at K=4 that displays the population structure that exists among populations of *puniceus* in southern California.

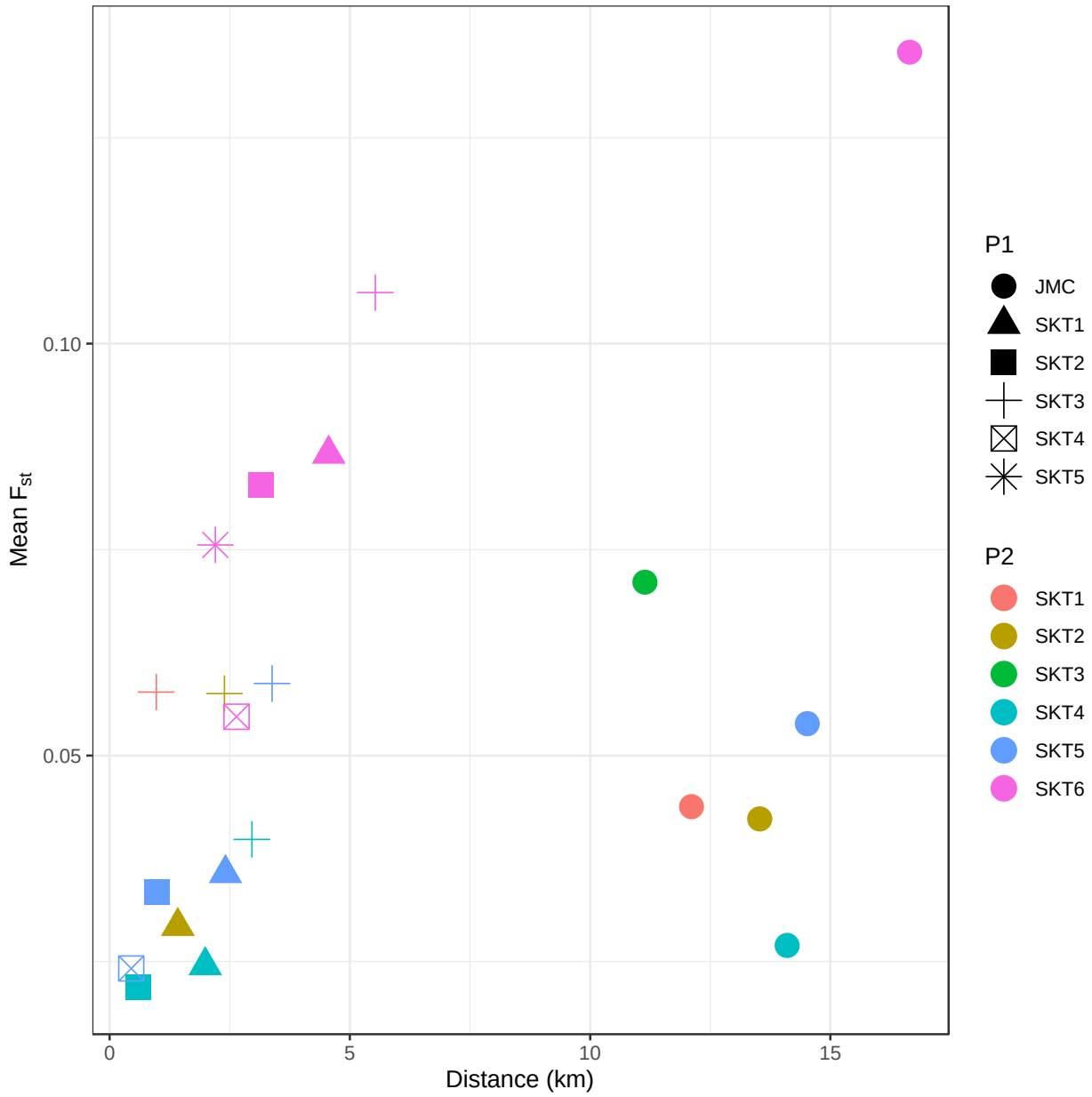

Figure 3. The relationship between geographic distance in kilometers and the mean Fst between populations from the SKT transect. The shape and color of the points indicates the populations being compared.

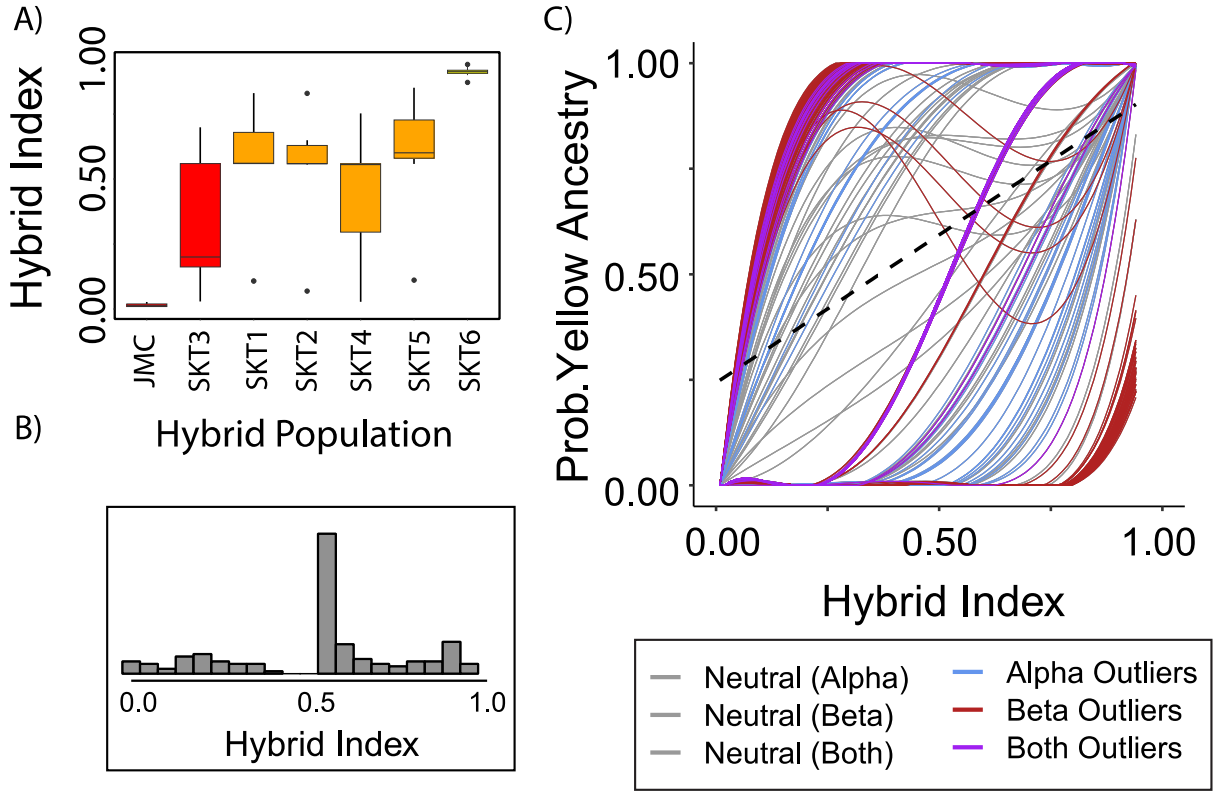

Figure 4. The hybrid index, and genomic clines estimated by Bayesian Genomic Clines (Gompert et al. 2011,2012) using the 301 fixed SNPs between the red and yellow ecotypes. (A) Boxplots of the hybrid index for samples from each of the populations from the SKT transect. The populations are arranged in a west to east orientation with the westernmost population on the left and the easternmost population on the right. Red shifted populations with a median hybrid index equal to or less than 0.25 are filled in red, intermediate populations with a median hybrid index between 0.75 and 0.25 are filled in orange, and yellow shifted populations with a median hybrid index equal to or greater than 0.75 are colored yellow. (B) Histogram of the hybrid index for the 170 samples included in the BGC analysis. (C) Plot of the genomic clines for the 301 fixed SNPs between the red and yellow ecotypes of *puniceus*. Each line shows the probability of yellow ancestry at each of the fixed SNPs for different values of the hybrid index.

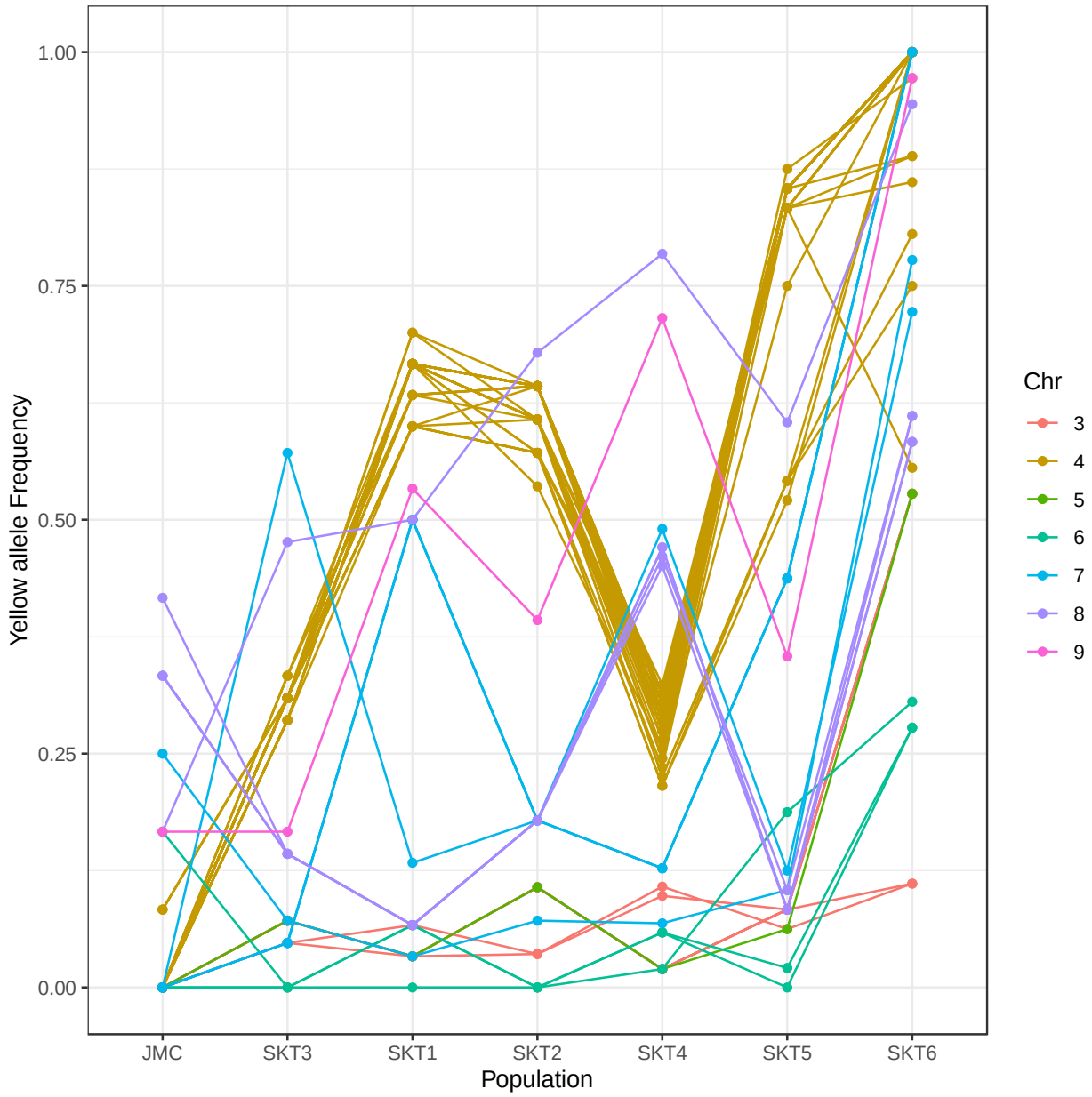

Figure 5. The yellow allele frequency for each of the populations from the SKT transect at the 61 outlier SNPs identified by BGC.

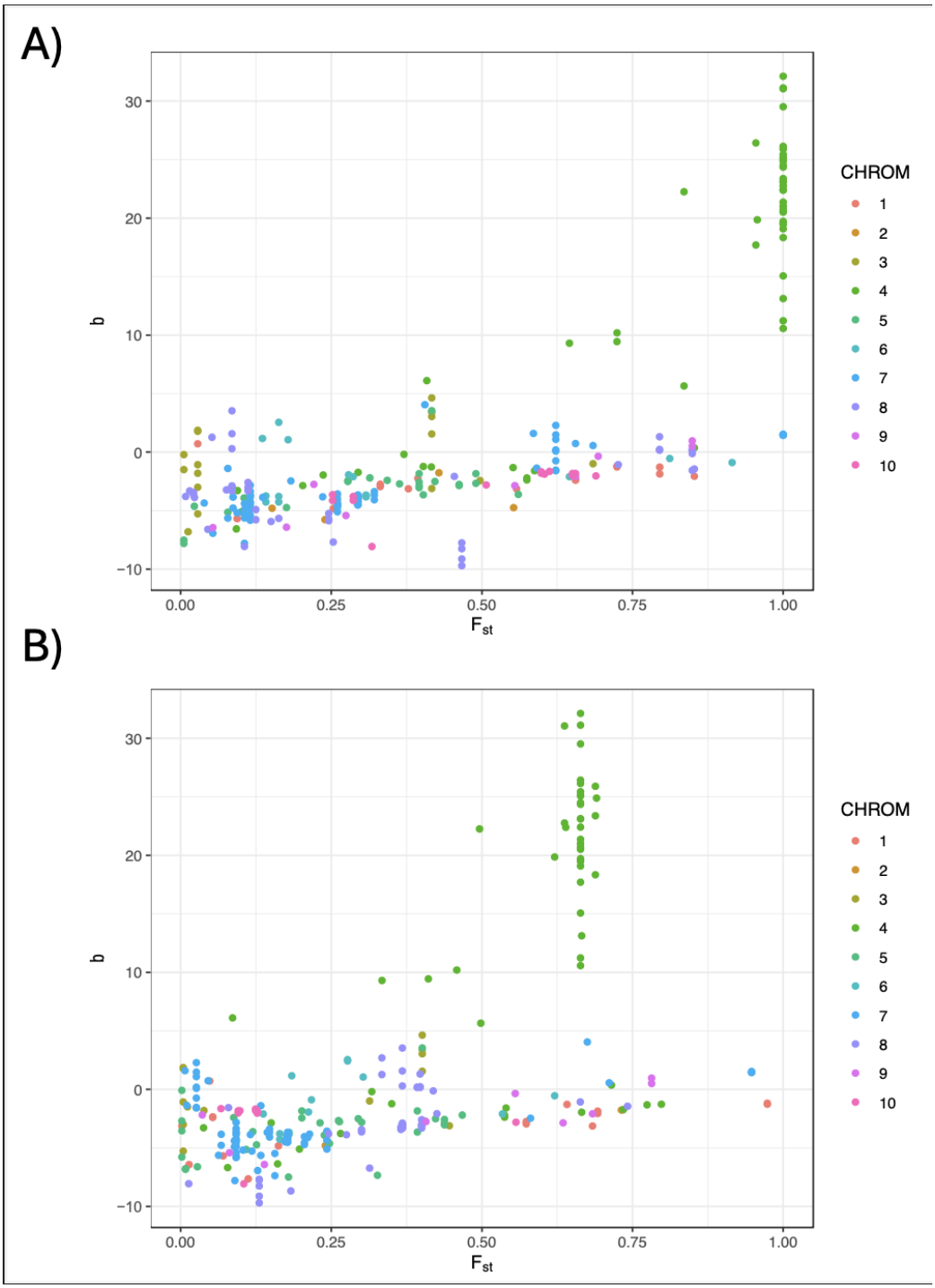

Figure 6: The relationship between the steepness of the genomic clines estimated by BGC's beta parameter and genetic differentiation estimated using Fst. Points are colored by the chromosome on which they are located. (A) Shows the relationship between beta and Fst between JMC and SKT6. (B) Shows the relationship between beta and Fst between SKT3 and SKT6.

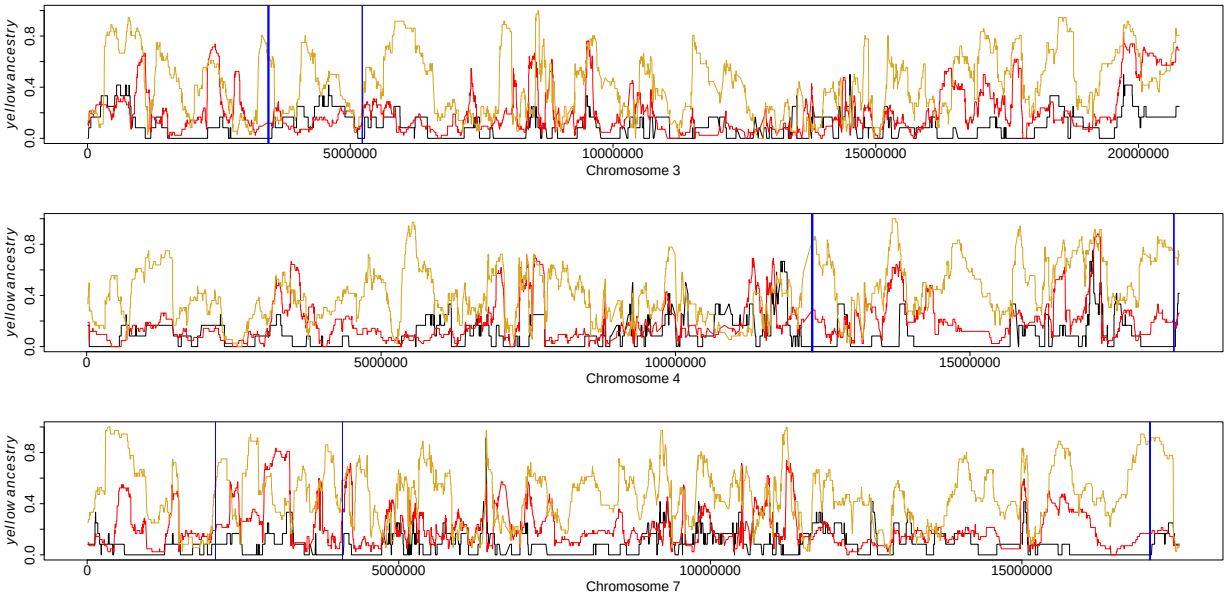

Figure 7. The frequency of yellow ancestry estimated on chromosome 3, 4, and 7. The colors of the lines correspond to the population in which the frequency of yellow ancestry was estimated with JMC in black, SKT3 in red, and SKT6 in yellow. The blue lines indicate the location of the beta outliers identified by BGC on these chromosomes.

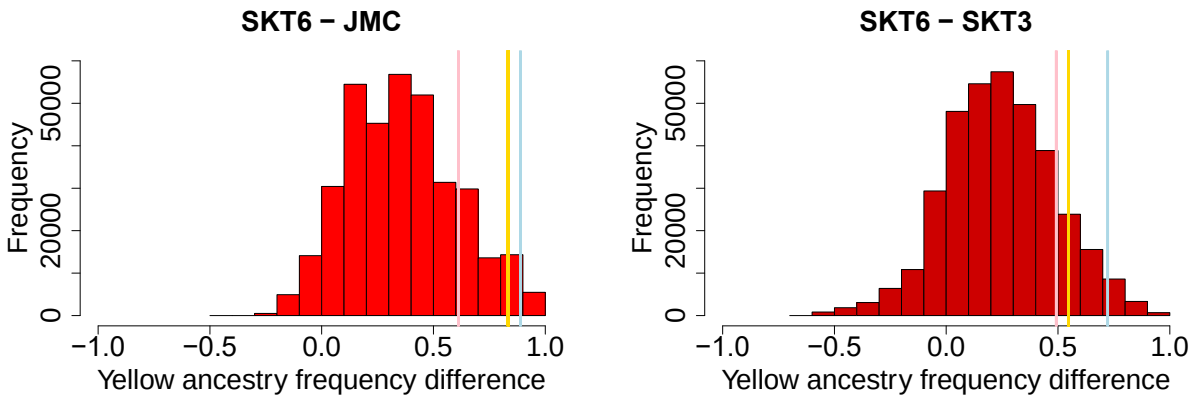

Figure 8. Histograms of the distribution of the difference in the yellow ancestry frequency between SKT6 and JMC (left) or SKT3 (right) at 353,000. The colored lines indicate the yellow allele frequency difference at putative barrier loci from three regions on chromosomes 3 (n=4), 4 (n=32) and 7 (n=3). With lines for the SNPs on chromosome 3 in pink, chromosome 4 in gold, and chromosome 7 in light blue.

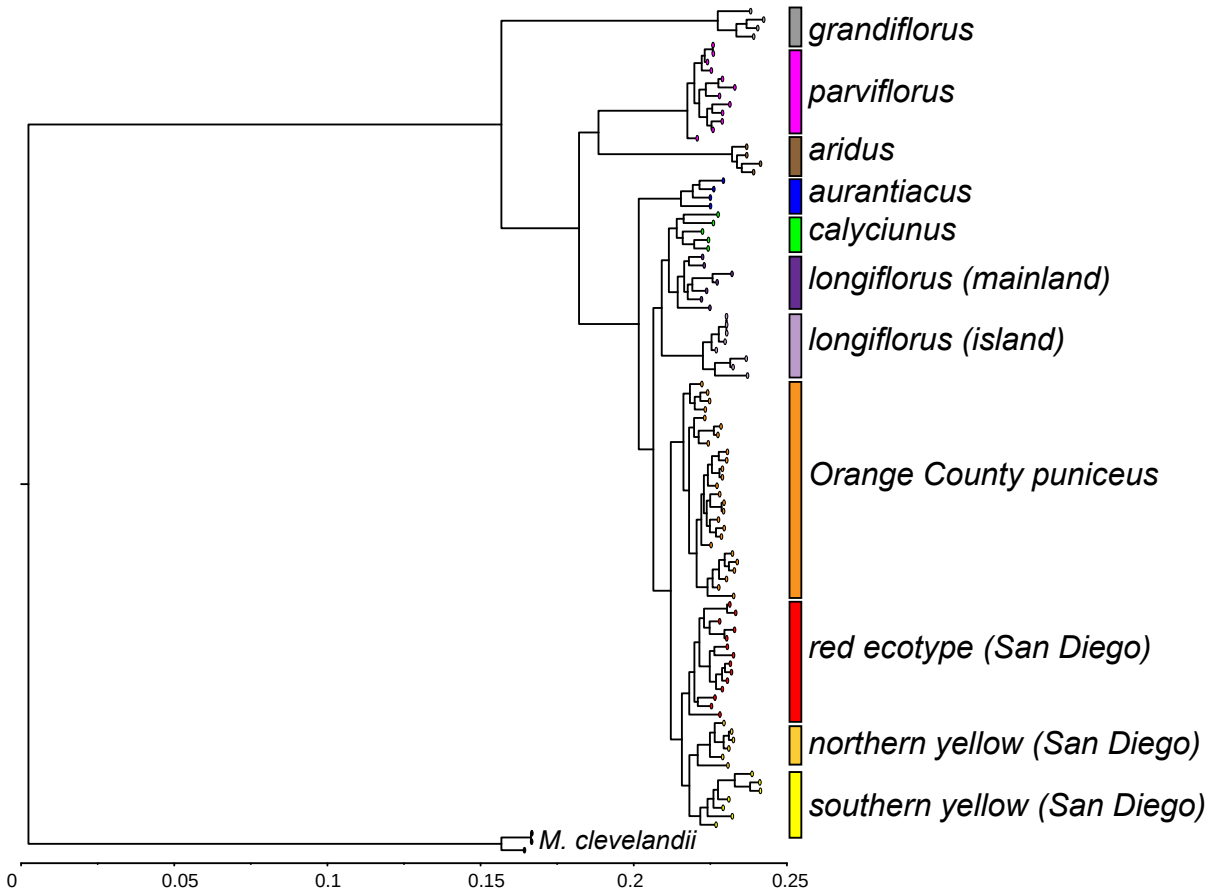

Figure S1. Phylogenetic tree showing the relationship of the taxa and the samples within taxa relative to one another. The colors of the circles correspond to the taxa to which that sample belongs with Mimulus *clevelandii* in black, *grandiflorus* in grey, *parviflorus* in pink, *aridus* in brown, *calycinus* in green, *longiflorus* in purple, Island *longiflorus* in light purple, orange county *puncieus* in orange, the red ecotype in red, the northern populations of the yellow ecotype in gold, and the southern populations of the yellow ecotype in yellow.

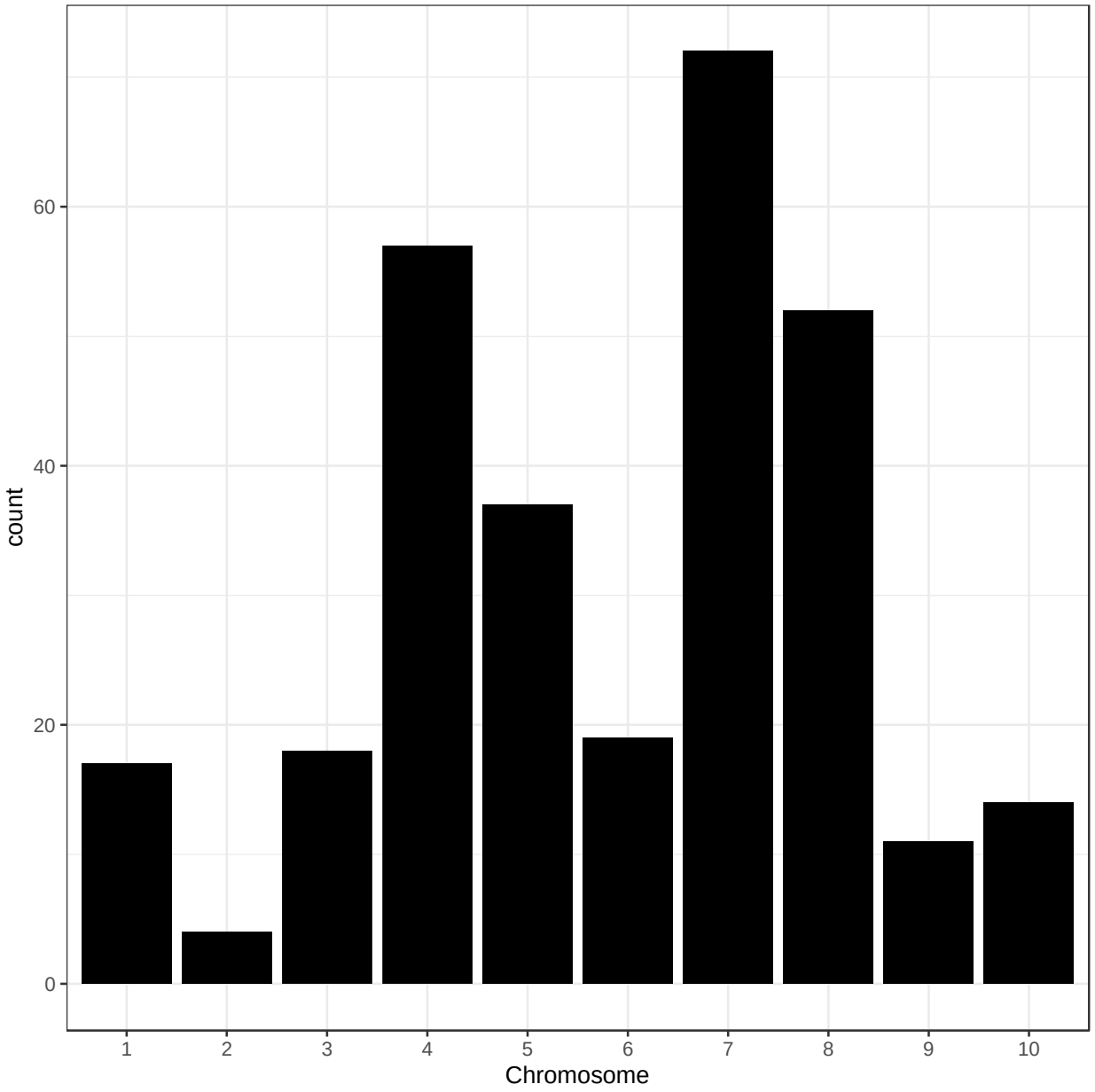

Figure S2. Barplot of the number of SNPs per chromosome that are fixed between the red and yellow ecotype.

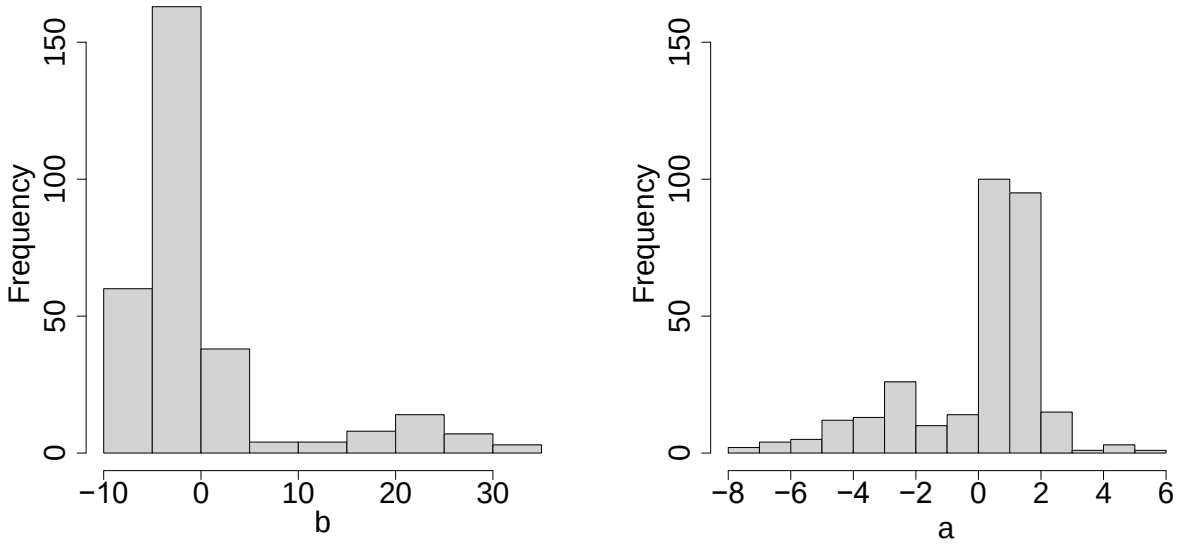

Figure S3. Distribution of the beta (left) and alpha (right) values from Bayesian Genomic Cline analysis using 301 fixed SNPs between the red and yellow ecotypes of *puniceus* from San Diego.

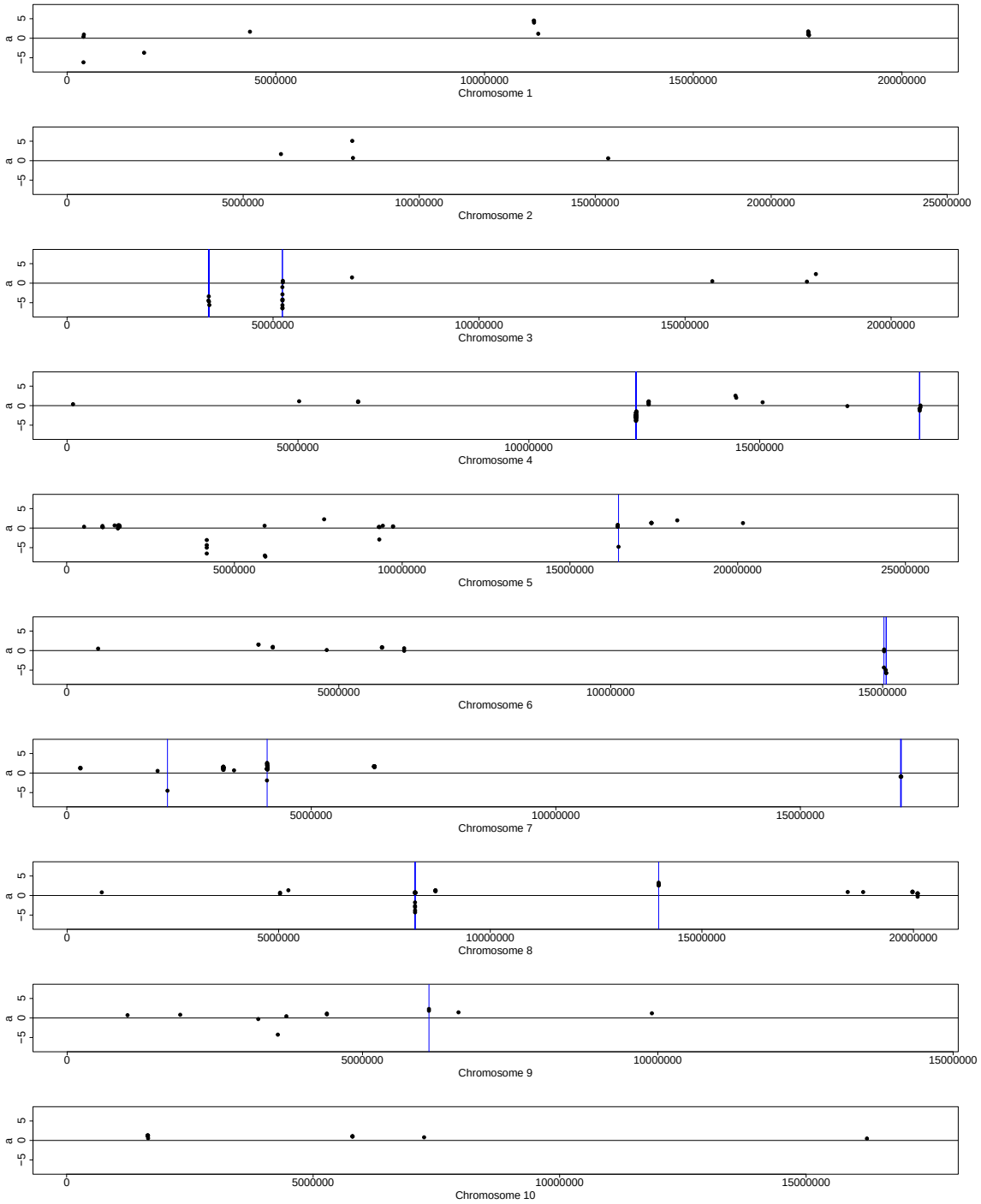

Figure S4. Genome wide variation in alpha (sites where there is an excess of ancestry from one of the two parental species as compared to the genome wide hybrid index). A negative alpha value indicates sites where there is excess red ecotype ancestry and a positive alpha indicates sites where there is excess yellow ecotype ancestry. The regions where positive beta outliers are present is indicated by blue lines.

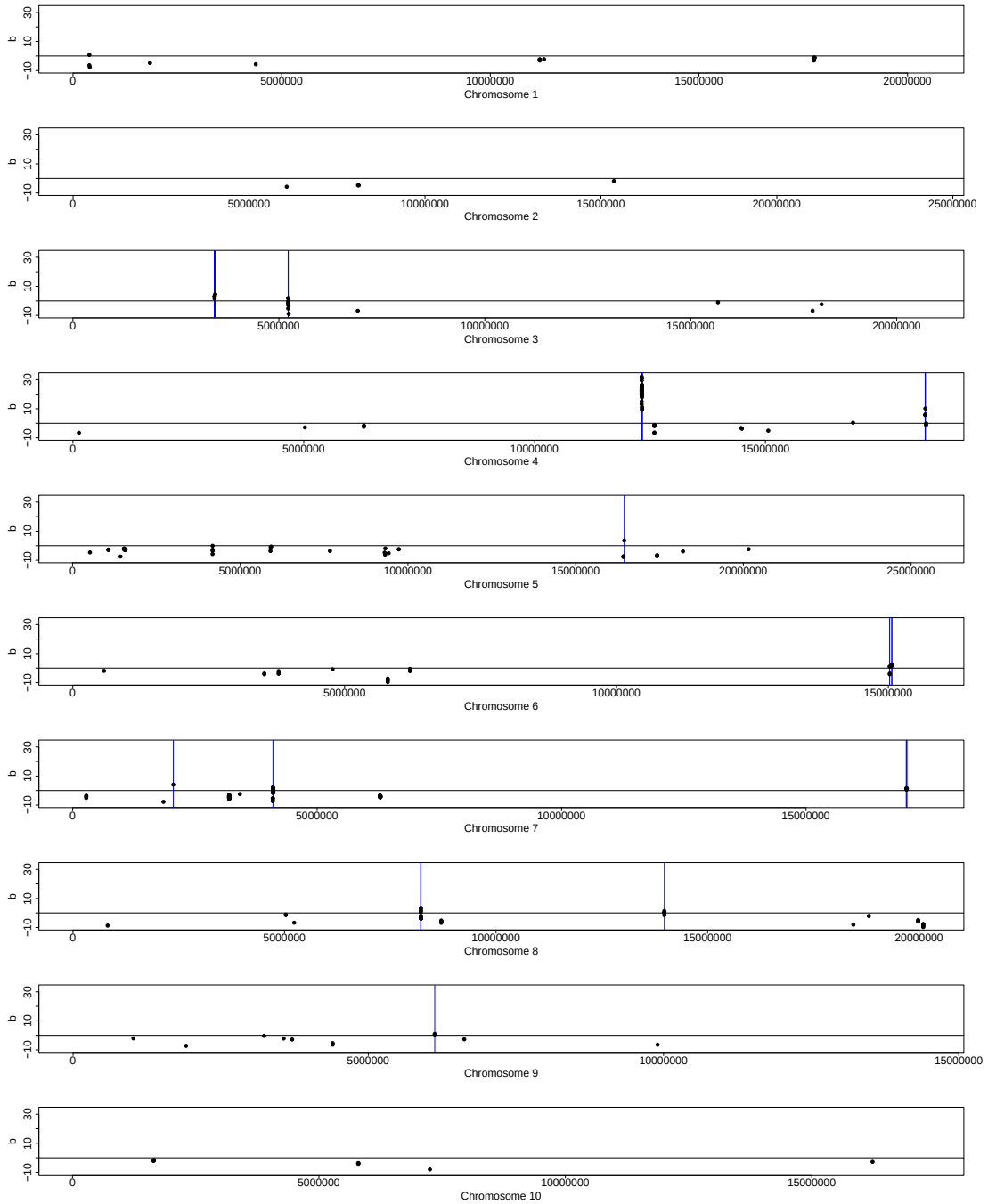

Figure S5. Genome wide variation in beta (sites where the transition in ancestry at a particular site deviates from the genome wide transition in hybrid index). A positive beta value indicates sites where transitions in ancestry occur more rapidly than genome wide transitions in ancestry potentially due to selection against gene flow between ancestry groups. A negative beta value indicates sites where transitions in ancestry occur more slowly than the genome wide average potentially due to adaptive introgression. The regions where positive beta outliers are present is indicated by blue lines.

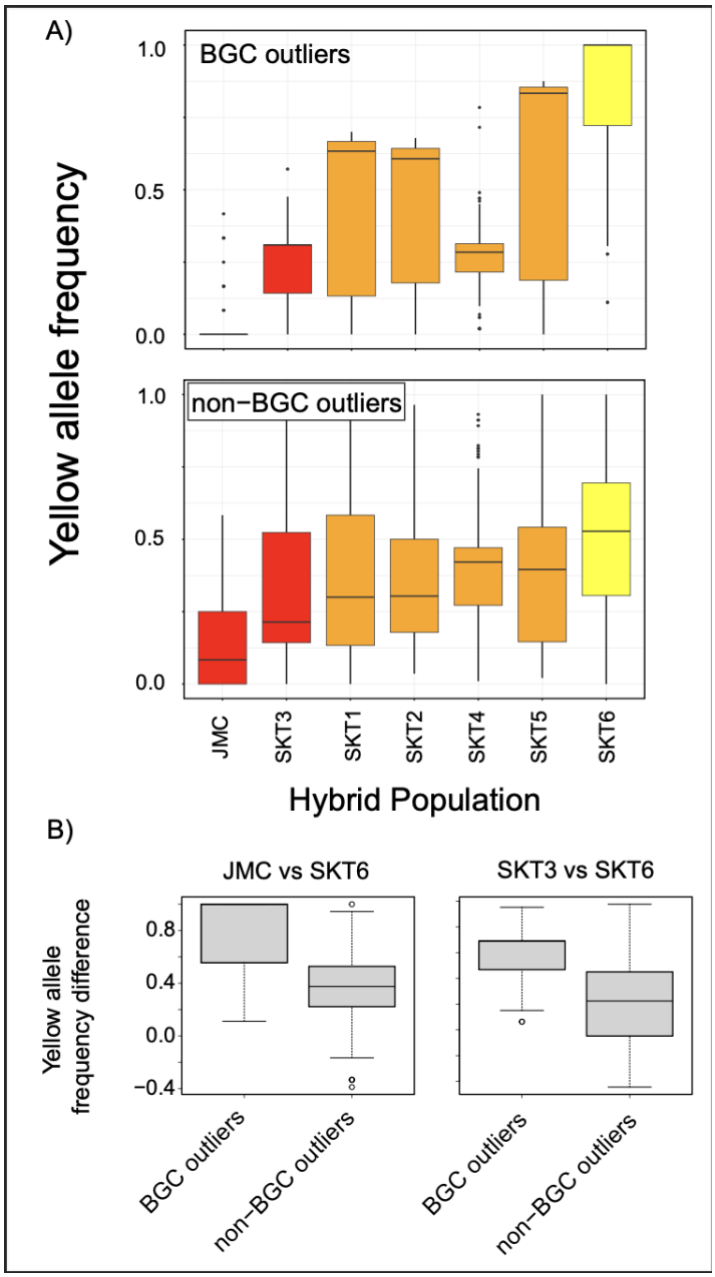

Figure S6. Boxplots of the yellow allele frequency and the yellow allele frequency difference for the 61 putative barrier loci identified by BGC and the 240 fixed SNPs that were not identified as barriers by BGC. (A) Boxplots of the yellow allele frequency for the putative barrier loci (top) and the SNPs that were not identified as barrier loci (bottom). Red shifted populations with a median hybrid index equal to or less than 0.25 are filled in red, intermediate populations with a median hybrid index between 0.75 and 0.25 are filled in orange, and yellow shifted populations with a median hybrid index equal to or greater than 0.75 are colored yellow. (B) Boxplot of the yellow allele frequency difference between SKT6 and JMC (left) and SKT6 and SKT3 (right) for the putative barrier loci and the 240 SNPs that were not identified as barriers.

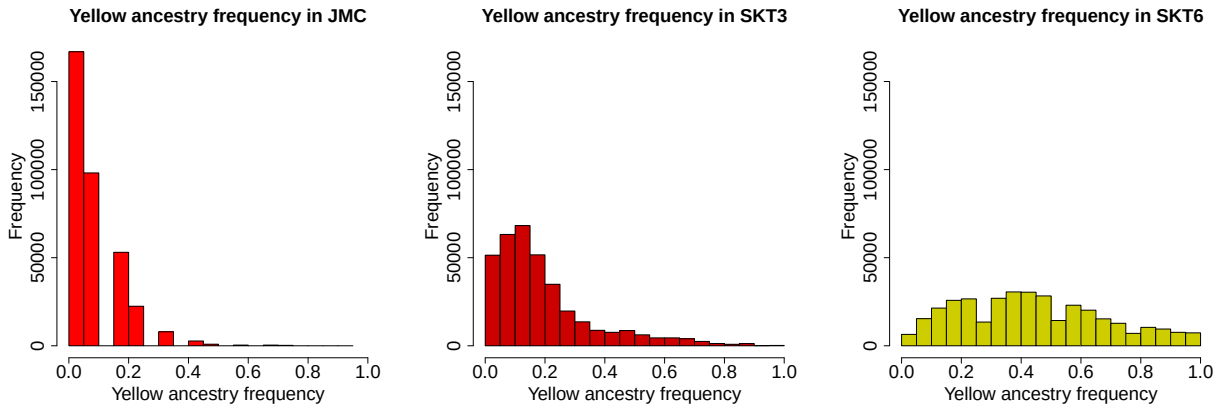

Figure S7. Histograms of the distribution of the frequency of yellow ancestry at 353,000 sites among samples from the JMC, SKT3, and SKT6 populations from the SKT hybrid zone transect.

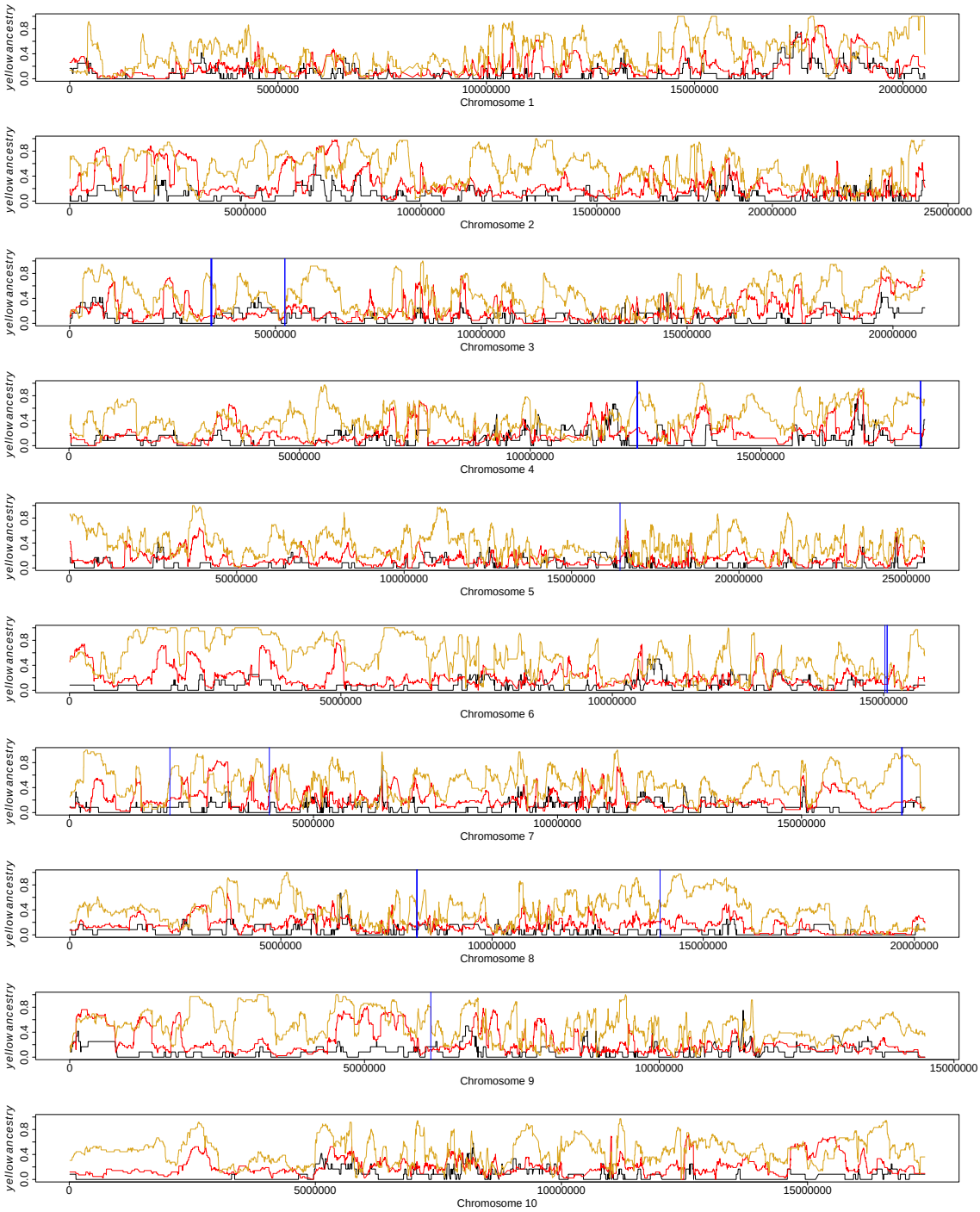

Figure S8. The frequency of yellow ancestry estimated at 353,000 sites across the genome. The colors of the lines correspond to the population in which the frequency of yellow ancestry was estimated with JMC in black, SKT3 in red, and SKT6 in yellow. The blue lines indicate the location of the 61 beta outliers identified by BGC.

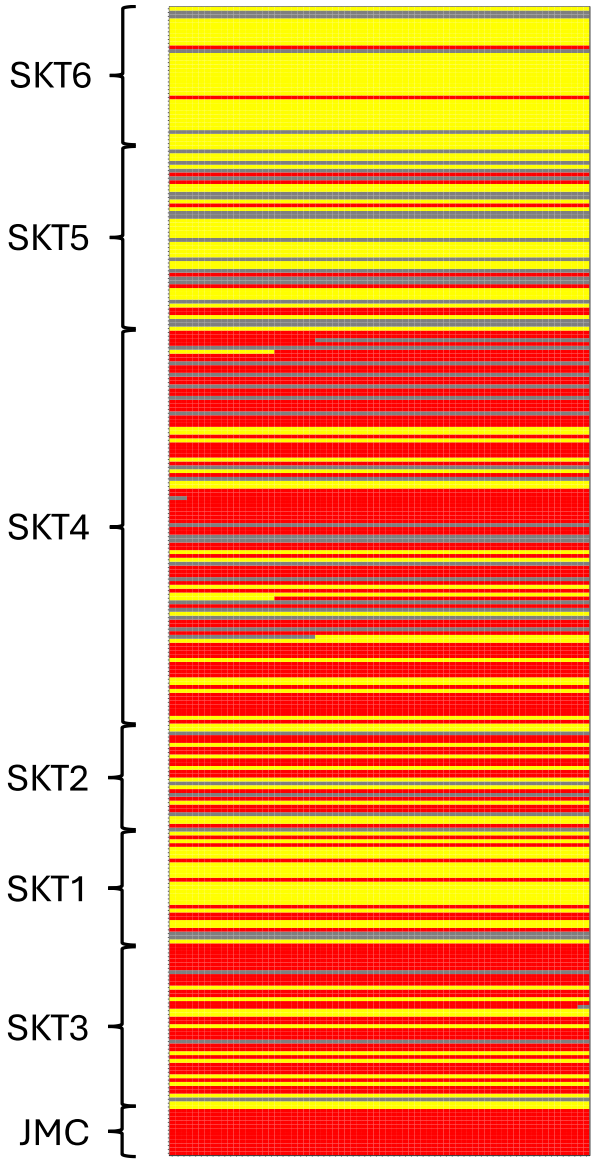

Figure S9. Local ancestry at the SNPs within the region containing the pileup of positive beta outliers from Chromosome 4 (Positions: 12317683 to 12329515) for all of the phased haplotypes from the 7 populations included in our Bayesian Genomic Cline analysis. The orientation of the populations from top to bottom corresponds to going from east to west across the hybrid zone. With the westernmost population JMC at the bottom and the easternmost population SKT6 at the top. Red squares indicate red ancestry and yellow squares indicate yellow ancestry at a particular position. Grey squares indicate SNPs where the ancestry probability was less than 0.80.

Table S1. Geographic coordinates, the number of samples whole genome or RAD-sequenced, and whether MYB2 genotyping was performed for populations of *Mimulus aurantiacus* subspecies *puniceus* collected from Orange County or San Diego County, California, USA.

Population	Latitude	Longitude	WGS samples	RAD-seq samples	MYB2 genotyped
395	-117.15295	33.2969	7	12	Yes
74_1	-117.460196	33.59937	10	0	Yes
74_2	-117.54821	33.55859	12	0	Yes
AND	-116.7461	32.86823	0	5	Yes
BC	-116.804683	33.12262	0	12	Yes
BCRD	-116.63795	32.94958	1	6	Yes
BS	-117.016433	33.0148	0	18	Yes
CKR	-117.264088	33.56396	0	11	Yes
CRS	-117.307167	33.13037	5	12	Yes
CRT	-117.590757	33.40024	0	12	Yes
DCR	-117.318133	33.3498	9	0	No
DLR	-117.052367	33.16818	0	9	Yes
DLT	-117.442252	33.45119	10	0	No
DLZ	-116.785967	32.6525	0	3	Yes
ECS	-117.426762	33.45156	12	0	No
ELF	-117.1453	33.08595	1	10	Yes
ELT	-117.089817	32.89422	0	12	Yes
FLP	-116.9867	32.80582	0	12	Yes
FLW	-117.649203	33.54575	0	12	Yes
HOT	-117.307743	33.39256	9	0	No
HRC	-117.486704	33.34514	8	0	No
IND	-117.333543	33.37451	9	0	No
INJ	-116.664317	33.09785	5	11	Yes
IRL	-117.72203	33.76821	4	0	No
JMC	-116.9541	32.73732	6	18	Yes
LB	-117.814	33.572	0	12	Yes
LFP	-117.657928	33.65318	4	0	No
LGC-1	-117.764286	33.5669	0	12	Yes
LGC-2	-117.763074	33.6078	0	12	Yes
LGH	-117.759565	33.55595	0	8	Yes
LH	-117.118767	33.06088	1	12	Yes
LKE	-117.395628	33.65434	0	12	Yes
LKW	-117.0161	33.16372	0	15	Yes
LO	-116.331233	32.6767	1	12	Yes
MT	-117.06175	32.82095	1	12	Yes
MTH	-117.825181	33.61412	6	4	Yes
MW	-116.959783	33.00718	0	17	Yes
OAK	-116.889317	32.91407	0	12	Yes
OSP	-117.035083	33.10197	0	10	No

OTR	-117.461976	33.45335	7	0	Yes
PCT	-116.469833	32.73258	5	7	Yes
PID	-117.730194	33.50474	0	12	Yes
PMD	-117.059133	32.93787	0	12	Yes
POTR	-116.633917	32.6038	1	11	Yes
RCR	-117.129235	33.45668	0	12	Yes
RG	-117.176344	33.41543	0	11	Yes
RPR	-117.802002	33.60575	6	12	Yes
SCRD	-117.641812	33.70225	0	12	Yes
SDP	-117.235383	32.9981	0	12	Yes
SKT1	32.7335	-116.8247	15	0	No
SKT2	32.7327	-116.8095	14	0	No
SKT3	32.7345	-116.835	21	0	No
SKT4	32.7344	-116.8034	51	0	No
SKT5	32.733	-116.7989	24	0	No
SKT6	32.7271	-116.7765	18	0	No
TOR	-117.640131	33.66101	0	12	Yes
UCSD	-117.236183	32.8894	5	13	Yes
VCR	-117.650862	33.48696	6	12	Yes
WCR	-116.8295	32.989	0	12	Yes
WM	-116.902283	32.82133	0	17	Yes
YNK	-117.484619	33.47892	7	0	No

4.6 References

1. Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C.A., Buggs, R., Butlin, R.K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S.H., Hermansen, J.S., Hewitt, G., Hudson, A.G., Jiggins, C., Jones, J,m Keller, B., … & Zinner, D. (2013). Hybridization and speciation. *Journal of evolutionary biology*, *26*(2), 229-246.
2. Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, *19*(9), 1655-1664.
3. Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics*, *81*(5), 1084-1097.
4. Browning, S. R., Waples, R. K., & Browning, B. L. (2023). Fast, accurate local ancestry inference with FLARE. *The American Journal of Human Genetics*, *110*(2), 326-335.
5. Burri, R. (2017). Linked selection, demography and the evolution of correlated genomic landscapes in birds and beyond. Molecular ecology, 26(15), 3853-3856.
6. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular ecology*, *22*(11), 3124-3140.
7. Coyne, J. A., & Orr, H. A. (1989). Patterns of speciation in *Drosophila*. Evolution, 43(2), 362-381.
8. Coyne, J. A., & Orr, H. A. (2004). Speciation (Vol. 37). Sunderland, MA: Sinauer Associates.

9. 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.
10. Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156-2158.
11. Duranton, M., Allal, F., Fraïsse, C., Bierne, N., Bonhomme, F., & Gagnaire, P. A. (2018). The origin and remolding of genomic islands of differentiation in the European sea bass. *Nature communications*, *9*(1), 2518.
12. Duranton, M., Allal, F., Valière, S., Bouchez, O., Bonhomme, F., & Gagnaire, P. A. (2020). The contribution of ancient admixture to reproductive isolation between European sea bass lineages. Evolution letters, 4(3), 226-242.
13. Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A. & Uebbing, S. (2012). The genomic landscape of species divergence in Ficedula flycatchers. Nature, 491(7426), 756-760.
14. Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. Trends in genetics, 28(7), 342-350.
15. Feder, J. L., Flaxman, S. M., Egan, S. P., Comeault, A. A., & Nosil, P. (2013). Geographic mode of speciation and genomic divergence. *Annual Review of Ecology, Evolution, and Systematics*, *44*(1), 73-97.
16. Gompert, Z., & Buerkle, C. A. (2011). Bayesian estimation of genomic clines. *Molecular Ecology*, *20*(10), 2111-2127.

17. Gompert, Z., & Buerkle, C. A. (2012). bgc: Software for Bayesian estimation of genomic clines. *Molecular Ecology Resources*, *12*(6), 1168-1176

18. Guo, J. F., Zhao, W., Andersson, B., Mao, J. F., & Wang, X. R. (2023). Genomic clines across the species boundary between a hybrid pine and its progenitor in the eastern Tibetan Plateau. *Plant Communications*, *4*(4).

19. Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, *105*(S1), 795-809.

20. Mandeville, E. G., Parchman, T. L., McDonald, D. B., & Buerkle, C. A. (2015). Highly variable reproductive isolation among pairs of *Catostomus* species. *Molecular ecology*, *24*(8), 1856-1872.

21. Matute, D. R., & Cooper, B. S. (2021). Comparative studies on speciation: 30 years since Coyne and Orr. Evolution.

22. Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., & 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

23. Martin, S. H., & Jiggins, C. D. (2017). Interpreting the genomic landscape of introgression. Current Opinion in Genetics & Development, 47, 69-74.

24. Martin, B. T., Chafin, T. K., Douglas, M. R., & Douglas, M. E. (2021). ClineHelpR: an R package for genomic cline outlier detection and visualization. *BMC bioinformatics*, *22*, 1-10

25. Marques, D.A., Meier, J.I. and Seehausen, O. (2019). A combinatorial view on speciation and adaptive radiation. Trends in ecology & evolution, 34(6), pp.531-544.

26. Matthey-Doret, R., & Whitlock, M. C. (2019). Background selection and FST: consequences for detecting local adaptation. Molecular ecology, 28(17), 3902-3914.

27. Mayr, E. (1942). Systematics and the origin of species, from the viewpoint of a zoologist. Harvard University Press.

28. Moyle, L. C., Olson, M. S., & Tiffin, P. (2004). Patterns of reproductive isolation in three angiosperm genera. Evolution, 58(6), 1195-1208.

29. Nosil, P., Vines, T. H., & Funk, D. J. (2005). Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution, 59 (4), 705-719.

30. Orr, H. A., & Turelli, M. (2001). The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. Evolution, 55 (6), 1085-1094.

31. Phadnis, N., & Orr, H. A. (2009). A single gene causes both male sterility and segregation distortion in Drosophila hybrids. Science, 323 (5912), 376-379.

32. Ravinet, M., Faria, R., Butlin, R.K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M.A.F., Mehlig, B. and Westram, A.M. (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of evolutionary biology*, 30(8), 1450-1477.

33. Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., Blazier, J. C., Sankararaman, S., Andolfatto, P., Rosenthal, G. G., & Przeworski, M. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, **360**(6389), 656–660. https://doi.org/10.1126/science.aar3684

34. Sedghifar, A., Brandvain, Y., Ralph, P., & Coop, G. (2015). The spatial mixing of genomes in secondary contact zones. Genetics, 201(1), 243-261.

35. Sedghifar, A., Brandvain, Y., & Ralph, P. (2016). Beyond clines: lineages and haplotype blocks in hybrid zones. Molecular ecology, 25(11), 2559-2576.

36. Sianta, S. A., Moeller, D. A., & Brandvain, Y. (2024). The extent of introgression between incipient Clarkia species is determined by temporal environmental variation and mating system. *Proceedings of the National Academy of Sciences*, *121*(12), e2316008121.

37. Stankowski, S., Sobel, J. M., & Streisfeld, M. A. (2017). Geographic cline analysis as a tool for studying genome-wide variation: A case study of pollinator-mediated divergence in a monkeyflower. *Molecular ecology*, *26*(1), 107-122.

38. Stankowski, S., Chase, M. A., McIntosh, H., & Streisfeld, M. A. (2023). Integrating top-down and bottom-up approaches to understand the genetic architecture of speciation across a monkeyflower hybrid zone. *Molecular Ecology*, *32*(8), 2041-2054.

39. Suarez-Gonzalez, A., Lexer, C. and Cronk, Q.C. (2018). Adaptive introgression: a plant perspective. Biology letters, 14(3), p.20170688.

40. Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in Anopheles gambiae. *PLoS biology*, 3(9).

41. Westram, A. M., Faria, R., Johannesson, K., & Butlin, R. (2021). Using replicate hybrid zones to understand the genomic basis of adaptive divergence. *Molecular ecology*, *30*(15), 3797-3814.

42. Wright, S. (1943). Isolation by distance. *Genetics*, *28*(2), 114

43. Wu, C. I. (2001). The genic view of the process of speciation. *Journal of evolutionary biology*, 14(6), 851-865.

CHAPTER 5: CONCLUSION

In this dissertation, I describe my work investigating the outcomes of natural hybridization between taxa at various stages of divergence across a diverse evolutionary radiation. Examination of the outcomes of hybridization in other well studied radiations, such as Darwin's finches (Grant & Grant 1994, 2008, 2016; Grant et al. 2004; Lamichhaney et al. 2015), African cichlids (Meier et al. 2017, 2019, 2023), *Heliconius* butterflies (Martin et al. 2013, Edelman et al. 2019, Kozak et al. 2021), and ancient hominids (Green et al. 2010, Huerta-Sánchez et al. 2014, Gokcumen 2020, Zhang et al. 2021), has revealed that hybridization occurs repeatedly throughout the history of rapidly diverging species, and in some cases, it can contribute to adaptation and diversification. In addition, new work in emerging systems, such as Alpine whitefish (De-Kayne et al. 2022), and *Penstemon* (Stone & Wessinger 2024), is helping to reveal the role of hybridization vs *de novo* mutations in the repeated evolution of adaptive traits and complex phenotypes. Furthermore, examining the outcomes of hybridization in natural hybrid zones between recently diverged species has revealed how divergent selection on a few differentiated regions can contribute to the maintenance of species differences (Martin et al. 2013, Wessinger et al. 2023).

In the second and third chapters of this thesis, I used the diversity of taxa present in the *Mimulus aurantiacus* radiation to estimate the history of hybridization among taxa. Using these estimates, I then determined how introgressive hybridization has contributed to the repeated evolution of an adaptive trait (Chapter 2), and how the fitness effects of introgression varied with divergence time (Chapter 3). In the fourth chapter, I studied a natural hybrid zone between the very closely related red and yellow flowered ecotypes of subspecies *puniceus*. By examining a very narrow transect across the center of this hybrid zone, I revealed regions of the genome that

appeared to act as barriers to gene flow. Given that these taxa are at a very early stage in the process of speciation, these loci may have driven the initial periods of divergence.

Together, the results of this thesis demonstrate that hybridization can be common throughout the speciation process, but that its impacts can vary at different stages of divergence. In some cases, hybridization will be deleterious, whereas it can also introduce beneficial alleles into populations, where it can fuel the diversification of entire radiations. Continuing to characterize the functions and identities of the barriers that block gene flow at the earliest stages of divergence will yield important information on the mechanisms that initiate speciation.

5.1 References

1. Grant, P. R., & Grant, B. R. (1994). Phenotypic and genetic effects of hybridization in Darwin's finches. *Evolution*, *48*(2), 297-316.
2. Grant, P. R., Grant, B. R., Markert, J. A., Keller, L. F., & Petren, K. (2004). Convergent evolution of Darwin's finches caused by introgressive hybridization and selection. *Evolution*, *58*(7), 1588-1599.
3. Grant, B. R., & Grant, P. R. (2008). Fission and fusion of Darwin's finches populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1505), 2821-2829.
4. Grant, P. R., & Grant, B. R. (2016). Introgressive hybridization and natural selection in Darwin's finches. *Biological Journal of the Linnean Society*, *117*(4), 812-822.
5. Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., ... & Andersson, L. (2015). Evolution of Darwin’s finches and their beaks revealed by genome sequencing. *Nature*, *518*(7539), 371-375.

6. Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature communications*, *8*(1), 14363.
7. Meier, J. I., Stelkens, R. B., Joyce, D. A., Mwaiko, S., Phiri, N., Schliewen, U. K., ... & Seehausen, O. (2019). The coincidence of ecological opportunity with hybridization explains rapid adaptive radiation in Lake Mweru cichlid fishes. *Nature communications*, *10*(1), 5391.
8. Meier, J. I., McGee, M. D., Marques, D. A., Mwaiko, S., Kishe, M., Wandera, S., ... & Seehausen, O. (2023). Cycles of fusion and fission enabled rapid parallel adaptive radiations in African cichlids. *Science*, *381*(6665), eade2833.
9. 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.
10. Edelman, N. B., Frandsen, P. B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R. B., ... & Mallet, J. (2019). Genomic architecture and introgression shape a butterfly radiation. *Science*, *366*(6465), 594-599
11. Kozak, K. M., Joron, M., McMillan, W. O., & Jiggins, C. D. (2021). Rampant genome-wide admixture across the Heliconius radiation. *Genome Biology and Evolution*, *13*(7), evab099.
12. Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., ... & Pääbo, S. (2010). A draft sequence of the Neandertal genome. *science*, *328*(5979), 710-722.

13. Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B. M., Vinckenbosch, N., ... & Nielsen, R. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, *512*(7513), 194-197.
14. Gokcumen, O. (2020). Archaic hominin introgression into modern human genomes. *American journal of physical anthropology*, *171*, 60-73.
15. Zhang, X., Witt, K. E., Bañuelos, M. M., Ko, A., Yuan, K., Xu, S., ... & Huerta-Sanchez, E. (2021). The history and evolution of the Denisovan-EPAS1 haplotype in Tibetans. *Proceedings of the National Academy of Sciences*, *118*(22), e2020803118.
16. De-Kayne, R., Selz, O. M., Marques, D. A., Frei, D., Seehausen, O., & Feulner, P. G. (2022). Genomic architecture of adaptive radiation and hybridization in Alpine whitefish. *Nature Communications*, *13*(1), 4479.
17. Stone, B. W., & Wessinger, C. A. (2024). Ecological diversification in an adaptive radiation of plants: the role of de novo mutation and introgression. *Molecular Biology and Evolution*, *41*(1), msae007.
18. Wessinger, C. A., Katzer, A. M., Hime, P. M., Rausher, M. D., Kelly, J. K., & Hileman, L. C. (2023). A few essential genetic loci distinguish Penstemon species with flowers adapted to pollination by bees or hummingbirds. *PLoS biology*, *21*(9), e3002294.