

Ancestry and Ecology Shape the Genomic Landscape of  
Divergence in *Mimulus aurantiacus*

by

CONNOR LANE

A THESIS

Presented to the Department of Biology  
and the Robert D. Clark Honors College  
in partial fulfillment of the requirements for the degree of  
Bachelor of Science

June 2020

## An Abstract of the Thesis of

Connor Lane for the degree of Bachelor of Science  
in the Department of Biology to be taken <Graduation Month and Year>

Title: Ancestry and Ecology but not Geography Explain Genome Correlations  
Indicative of Divergence in *Mimulus aurantiacus*

Approved: Matthew Streisfeld, Ph.D.  
Primary Thesis Advisor

One of the main goals of evolutionary biology is to understand the speciation process that creates the vast diversity of life we observe on earth. In the case of speciation with gene flow, it is important to understand how populations become isolated from each other to allow genetic differences to accumulate. This isolation can occur due to distance (IBD) and adaptation to the local ecology (IBA). While both have been characterized in different systems, it is unknown how these processes affect certain genome correlations involving  $F_{ST}$  that are indicative of linked selection and genome divergence. Using reduced representation sequencing (RAD-seq), we first characterized population structure in our study system *Mimulus aurantiacus ssp. puniceus*, where IBD and local adaptation have been well-characterized. We found that *puniceus* displays more levels of genetic subdivision than related subspecies, and we were able to characterize two distinct lineages of *puniceus* that display near-identical distributions of floral phenotypes. Our population genomics analysis then revealed that IBA was a better predictor of genome correlations than IBD in a subset of populations that occurred across a considerable climatic gradient. However, the best predictor of the genome correlations was differences in ancestry between the two discrete lineages characterized in our population structure analysis. These results show that ancestry differences can explain patterns of genomic divergence without clear signs of adaptation, but that divergent selection is likely important for explaining patterns of genomic divergence during early stages of speciation.

## Acknowledgements

I would like to thank Matt Streisfeld for guiding me through the research and writing of the thesis process. His contributions and mentorship helped me grow as a student and scientist, and his consideration for my personal well-being helped me when I was working through roadblocks. Along those lines, I'd like to thank Sean Stankowski for being on my committee and providing mentorship when I first became involved in research. I always pictured science as being stiff and formal, so it was good for my anxiety to work with someone whose personable, relaxed, and supportive personality filled the little office we would meet in. I'd also like to thank Samantha Hopkins for serving on my committee and talking with me about how I can be a scientist that is also involved in the liberal arts principles that the CHC stands for. Additionally, my thanks go out to Sean Stankowski and Keith Karoly for sequencing genomes of the *calycinus*, *longiflorus*, and northern *puniceus* individuals used in this thesis, and for Sean Stankowski gathering floral trait data for the northmost *puniceus* populations. This thesis wouldn't have been possible without their contributions. This work also wouldn't have been possible without the funding of UO CURE and the NSF, and I am grateful for their support. Finally, I'd like to thank my parents for loving me from the day I was born, and for accepting me as the person I am.

## Table of Contents

Introduction	1
Methods	6
Results	16
Discussion	28
Conclusion	34
Works Cited	35

## List of Figures

**Figure 1.** Location and subspecies of all populations sampled in southern California. Colors of the points indicate subspecies sampled (orange for *punicieus*, dark blue for *longiflorus*, and light blue for *calycinus*). Letter codes indicate site name.

Figure 2. Admixture scores and a PCA of bioclimatic variables. (a) Stacked Admixture barplots for all individuals with  $K = 2-5$ . (b) PC1 vs. PC2 of the same genetic dataset. Individuals with an ancestry score of at least 0.8 for an admixture group at  $K = 5$  are colored according to that admixture group, as in Fig. 2a. (c) Pie plots colored by admixture scores for  $K = 5$  from Figure 2a, averaged for each population and plotted by latitude and longitude. The background interpolation represents PC1 scores for 23 bioclimatic variables from ClimateWNA (shown in the lower left corner; full descriptions the variables in Table 2)

**Figure 3.** A summary of floral traits within *punicieus*. (a) PC1 vs. PC2 for all families measured, colored by population allele frequency at the Myb-2 locus linked to flower color. (b) Average PC1 of each population plotted against latitude and colored by Myb-2 allele frequency, with error bars representing standard error of the mean. (c) Loadings of all floral traits measured (Anthocyanin concentration, stigma exertion, pedicel length, corolla length, and tube width).

**Figure 4.** Scatterplots and linear regressions of  $F_{ST}$  and the Pearson's  $r$  between  $F_{ST}$  and other genome statistics plotted across geographic and ecological distance for all pairs of populations. The blue dots, blue trendlines, and blue  $r^2$  values represent the SD subset, and the red points represent, trendlines, and  $r^2$  values represent comparisons that had an ancestry score difference of at least 0.8 for the orange-colored admixture group at  $K = 3$  (Fig. 2). Black trendlines and  $r^2$  values represent the full dataset. All  $r^2$  values are derived from Pearson's  $r$  between response variables and measures of distance.

**Figure 5.** Partial Mantel correlation coefficients for  $F_{ST}$  and correlations involving  $F_{ST}$  over geographic distance (blue points), differences in the orange-colored admixture group from Fig. 2 (yellow-orange points) and ecological distance (red points), for the full dataset (a) and the SD subset (b). Data markers point in the direction of the expected correlation for increasingly isolated lineages, as in Stankowski et al. (2019). Asterisks indicate that the  $r$ -value is significant ( $p < 0.05$ ).

## **List of Tables**

Table 1. Summary of sampling and locations for all populations

Table 2. All 23 measured bioclimatic variables from ClimateWNA

## Introduction

Determining the mechanisms by which populations become genetically isolated is crucial for understanding how speciation produces the vast diversity we observe on earth. In particular, it is now common to find examples where speciation can occur in the face of gene flow. The application of genomics to the field has introduced the possibility of identifying the regions of the genome responsible for this divergence. A common outcome of these studies is that genomic differentiation is often heterogeneous, leading to a pattern of “islands of divergence,” where certain areas of the genome are highly differentiated, whereas the rest of the genome is not. The conventional wisdom is that the islands represent the regions of the genome responsible for divergence, while the rest of the genome is homogenized due to gene flow (Feder, Egan, & Nosil, 2012; Feder & Nosil, 2010; Nosil, Egan, & Funk, 2008; Renaut et al., 2013; Turner, Hahn, & Nuzhdin, 2005). However, key to understanding the mechanisms contributing to this heterogeneity is the idea of linked selection, which occurs when positive or negative selection indirectly impacts patterns of genetic diversity in linked genetic regions. While one cause of “islands” is that they are due to divergent natural selection reducing local gene flow due to hitchhiking, it is also possible that selection against areas of the genome nearby deleterious alleles (a process known as background selection) generates heterogeneity in genetic variation across the genome. Therefore, determining the role of local adaptation in divergence becomes critical.

An under-addressed topic within the study of speciation genomics is how to measure the effects of linked selection and determine where these islands of divergence

are located. A recent review shows that most speciation genomics studies use the statistic  $F_{ST}$  as a measure of genetic differentiation (Wolf & Ellegren, 2017).  $F_{ST}$  is inherently tied to levels of genetic diversity ( $\pi$ ) (Charlesworth, 2013), so areas of low diversity in the genome may have high  $F_{ST}$  even under neutral models. This is especially relevant given that areas of low recombination tend to have low genetic diversity and therefore elevated  $F_{ST}$  (Noor & Bennett, 2009), potentially explaining why some studies find peaks of  $F_{ST}$  near centromeres where recombination is limited (Turner et al., 2005) or across the genome in areas with low recombination (Renaut et al., 2013). When studies examine variation in  $F_{ST}$  without considering genome architecture, it can be difficult to make definitive conclusions about adaptation.

Because genome architecture is heterogeneous, we can use correlations between  $F_{ST}$  and other genome statistics to inform us about linked selection. For example, linked selection affects wider genomic regions when recombination rates are low and acts more often in gene-rich regions, leading to locally reduced genetic diversity (Payseur & Nachman, 2002). Therefore, the heterogeneous effects of linked selection within populations result in the frequent positive correlation between  $\pi$  and recombination rate. Between populations, varying strengths of linked selection will result in a negative correlation between  $F_{ST}$  and  $\pi$ , and  $F_{ST}$  and recombination rate, but a positive correlation of  $F_{ST}$  with gene density. Moreover,  $F_{ST}$  is expected to be stochastic near the beginning of the speciation process, leading to a weak correlation with  $\pi$ , recombination rate, and gene density (Burri et al., 2015). However, as divergence time increases, the effects of linked selection become more pronounced, and the strength of these correlations is expected to increase (Burri, 2017). This phenomenon has been seen in both bird and



plant systems (Burri et al., 2015; Stankowski et al., 2019). However, it has been argued that at least in some cases, the strengthening of these correlations may be due to recurrent background selection rather than repeated positive selection (Burri, 2017; Burri et al., 2015; Vijay et al., 2016), though this limitation has been overcome in the past through use of simulations showing that background selection alone is unlikely to be responsible for observed patterns of heterogeneous genome divergence in *M. aurantiacus* (Stankowski et al., 2019).

While we often consider genetic differences accumulating over time, it is also important to examine the role of geography and ecology in shaping variation in these genomic landscapes in order to properly characterize speciation with gene flow. Isolation by distance (IBD) is caused when geographically restricted gene flow leads to genetic differentiation due to the effects of genetic drift, and has been observed on both small and large scales (Peterson & Denno, 1998; Rose, Paynter, & Hare, 2006). On the other hand, isolation by adaptation (IBA) is the concept that differences in local ecology can restrict gene flow due to selection acting on maladapted immigrant individuals (Nosil et al., 2008). A review analyzing 70 studies found that IBD and IBA were prevalent across different systems, and several systems showed both processes acting in the same system (Sexton, Hangartner, & Hoffmann, 2014). These studies mostly used average  $F_{ST}$  as their measure of genetic differentiation (Sexton et al., 2014) and showed that geographic space and ecological differences may be analogous to differences in divergence time. However, because correlations involving  $F_{ST}$  can tell us about the localized effects of linked selection (Stankowski et al., 2019), the covariation of these correlations with geographic distance and/or ecological distance also can inform us

about the effects of ecologically-based divergent selection driving heterogeneity in patterns of genomic variation. In this study, we will determine if variation in geography and the environment can inform us about the importance of positive selection in shaping genomic patterns. If correlations involving  $F_{ST}$  strengthen over ecology rather than geographic space, it suggests that adaptation to the local environment is responsible for keeping populations isolated.

The *Mimulus aurantiacus* (bush monkeyflower) species complex is excellent for tackling problems about the roles of history, geography, and ecology in driving divergence. *M. aurantiacus* consists of eight phenotypically-differentiated taxa distributed across southern California that diverged in a recent radiation (Chase, Stankowski, & Streisfeld, 2017). The plants occur across a range of elevations, temperatures, and moisture levels (Thompson, 2005), making them ideal for studies investigating local adaptation. As previously mentioned, genome correlations involving  $F_{ST}$  increase with time among subspecies at different stages of divergence (Stankowski et al., 2019), and IBD has been demonstrated in subspecies *puniceus* in San Diego county (Stankowski, Sobel, & Streisfeld, 2015). Additionally, the presence of red- and yellow-flowered ecotypes within *puniceus* is an established system for studying local adaptation due to pollinator selection (Sobel & Streisfeld, 2015; Stankowski & Streisfeld, 2015; M. A. Streisfeld & Kohn, 2007; Matthew A. Streisfeld & Kohn, 2005). In addition, differences in the environments occupied by populations of *puniceus* may contribute to ecophysiological trait adaptations, further suggesting local adaptation to the abiotic environment (Sobel, Stankowski, & Streisfeld, 2019). These factors make

*M. aurantiacus* an ideal system for addressing the effects of ecologically-based divergent selection on patterns of genomic variation.

The first part of our study is a survey of population structure and trait evolution in southern California populations of *M. aurantiacus*, which was designed to reveal historical patterns of divergence and admixture across geographic and taxonomic levels. This survey is similar to one performed by Chase et al. (2017), but instead of sampling broadly across the range of the entire species complex, we concentrated on a diverse group of three subspecies. Our primary focus was sampling subspecies *puniceus* across its geographic range, but we also included populations from the more divergent subspecies *calycinus* and *longiflorus*. Specifically, we sampled *puniceus* populations to the north of San Diego, which were regarded previously as belonging to a different “population series” based on floral phenotypes (Beeks, 1962). Our population genomic survey allowed us to test whether these northern populations represented a distinct lineage from the southern San Diego populations. We additionally surveyed floral phenotypes from northern and southern plants grown in a common greenhouse to see how traits varied across the range.

From there, we tested the role of local adaptation in driving genomic variation in subspecies *puniceus*. Under a model of local adaptation, we predict that ecological variation will explain patterns of genomic divergence better than geography. We find that northern *puniceus* populations are highly differentiated from San Diego populations, potentially representing two distinct lineages with little gene exchange. Thus, we expect comparisons between north and south to exhibit signatures of linked selection greater than would be predicted by their geographic distance alone

(Stankowski et al., 2019). The results of these two studies provide novel insight into the role of geography and ecology in shaping patterns of genomic variation during the speciation process.

## Methods

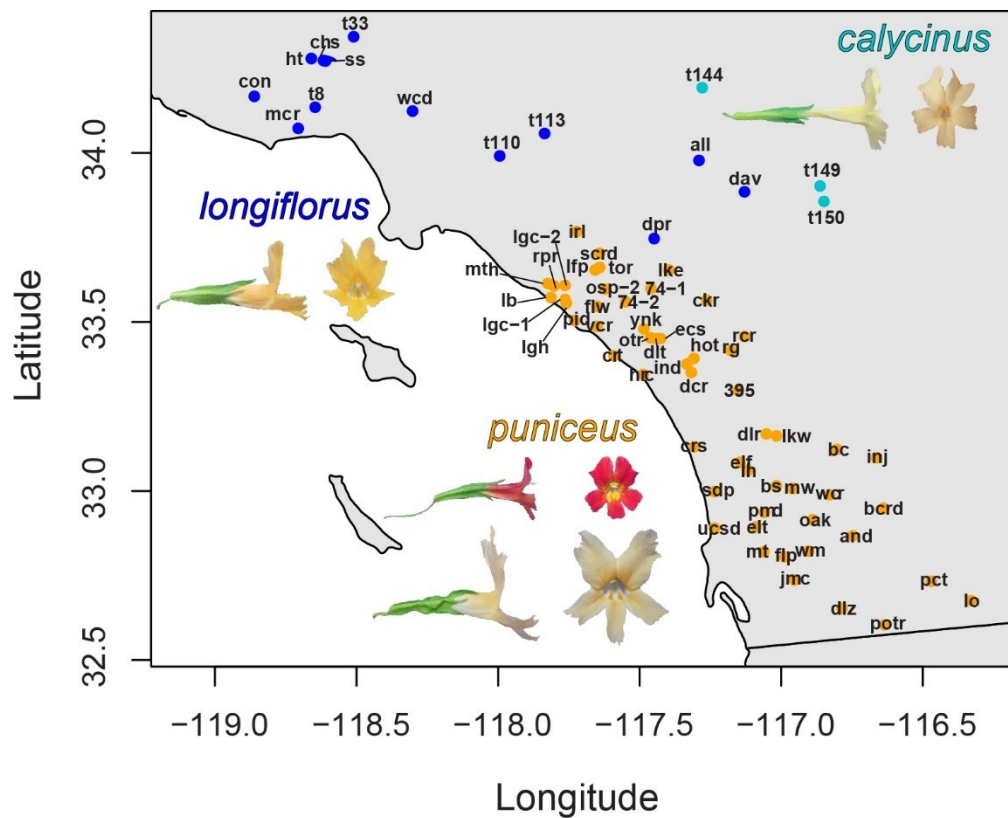
### *Sampling and Sequencing*

Individuals were collected from 71 populations comprising three subspecies of *M. aurantiacus* across southern California. *Mimulus aurantiacus* ssp. *puniceus*, the main subspecies of study, comprised 55 of the sampled populations. In addition, we included samples from three populations of ssp. *calycinus* and 13 populations from ssp. *longiflorus* to examine patterns of admixture across southern California. Details on the location of populations are in Fig. 1. Sample sizes, and the specific analyses performed with each are in Table 1.

From these populations, we generated new sequence data for 352 plants (227 *puniceus*, 25 *calycinus*, and 100 *longiflorus*) from 36 populations (20 *puniceus*, 3 *calycinus*, and 13 *longiflorus*) located to the north of San Diego County. In addition, we included 263 *puniceus* plants from 25 populations from San Diego County that were included from previously published work (Stankowski, Sobel, & Streisfeld, 2017), for a total dataset of 615 individuals from 61 populations (range 3 to 17 individuals per population). Genomic DNA was isolated from leaf tissue samples stored at -80C. We sequenced restriction-site associated DNA tags (RAD-seq) from each sample using single-end 100 bp Illumina HiSeq sequencing. Library construction and sequencing procedures followed previous analyses (Stankowski et al., 2015, 2017).

In addition, we phenotyped floral traits from the northern range of *puniceus* by growing field-collected seeds from 188 maternal plants representing 18 populations. These data were combined with previously published information on the same floral

traits from 136 individuals representing 16 populations from the southern range of *puniceus* (Stankowski et al., 2015). This resulted in a total dataset of 324 individuals from 34 populations, ranging from 7 to 13 individuals per population. Full details on sampling can be found in Table 1. Finally, we generated genotypes from 1000 *puniceus* plants from a single SNP within the *MaMyb2* gene that is responsible for flower color differences (Matthew A. Streisfeld, Young, & Sobel, 2013) to determine how floral traits varied based on allelic variation at this flower color locus.



**Figure 1.** Location and subspecies of all populations sampled in southern California. Colors of the points indicate subspecies sampled (orange for *puniceus*, dark blue for *longiflorus*, and light blue for *calycinus*). Letter codes indicate site name.

Table 1. Summary of sampling and locations for all populations

population	taxon	correlation datasets	individuals genotyped	families phenotyped	MaMyb2-sequenced individuals	latitude	longitude
t144	<i>calycenus</i>	-	9	-	-	34.19	-117.28
t149	<i>calycenus</i>	-	7	-	-	33.90	-116.86
t150	<i>calycenus</i>	-	9	-	-	33.86	-116.85
t33	<i>longiflorus</i>	-	6	-	-	34.34	-118.51
ht	<i>longiflorus</i>	-	7	-	-	34.28	-118.66
chs	<i>longiflorus</i>	-	7	-	-	34.27	-118.62
ss	<i>longiflorus</i>	-	9	-	-	34.27	-118.61
con	<i>longiflorus</i>	-	11	-	-	34.17	-118.86
t8	<i>longiflorus</i>	-	8	-	-	34.13	-118.65
wcd	<i>longiflorus</i>	-	8	-	-	34.12	-118.30
mcr	<i>longiflorus</i>	-	6	-	-	34.07	-118.71
t113	<i>longiflorus</i>	-	7	-	-	34.06	-117.84
t110	<i>longiflorus</i>	-	4	-	-	33.99	-117.99
all	<i>longiflorus</i>	-	8	-	-	33.98	-117.29
dav	<i>longiflorus</i>	-	9	-	-	33.88	-117.13
dpr	<i>longiflorus</i>	-	10	-	-	33.75	-117.45
irl	<i>puniceus</i>	-	-	11	21	33.77	-117.72
scrd	<i>puniceus</i>	full	12	12	20	33.70	-117.64
tor	<i>puniceus</i>	full	12	8	19	33.66	-117.64
lke	<i>puniceus</i>	full	12	-	20	33.65	-117.40
lfp	<i>puniceus</i>	-	-	12	17	33.65	-117.66
mth	<i>puniceus</i>	full	4	10	20	33.61	-117.83
lgc-2	<i>puniceus</i>	full	12	12	20	33.61	-117.76
rpr	<i>puniceus</i>	full	12	-	20	33.61	-117.80
74-1	<i>puniceus</i>	full	11	-	17	33.60	-117.46
osp-2	<i>puniceus</i>	full	12	12	20	33.60	-117.62
lb	<i>puniceus</i>	full	12	-	20	33.57	-117.81
lgc-1	<i>puniceus</i>	full	12	12	20	33.57	-117.76
ckr	<i>puniceus</i>	full	12	-	20	33.56	-117.26
74-2	<i>puniceus</i>	full	12	-	21	33.56	-117.55
lgh	<i>puniceus</i>	full	8	11	20	33.56	-117.76
flw	<i>puniceus</i>	full	12	-	20	33.55	-117.65
pid	<i>puniceus</i>	full	12	-	20	33.50	-117.73
vcr	<i>puniceus</i>	full	12	12	20	33.49	-117.65
ynk	<i>puniceus</i>	-	-	8	11	33.48	-117.48
rcr	<i>puniceus</i>	full	12	-	20	33.46	-117.13

otr	<i>puniceus</i>	-	-	9	15	33.45	-117.46
ecs	<i>puniceus</i>	-	-	13	19	33.45	-117.43
dlt	<i>puniceus</i>	-	-	8	9	33.45	-117.44
rg	<i>puniceus</i>	full	12	-	20	33.42	-117.18
crt	<i>puniceus</i>	full	12	-	21	33.40	-117.59
hot	<i>puniceus</i>	-	-	11	14	33.39	-117.31
ind	<i>puniceus</i>	-	-	10	11	33.37	-117.33
dcr	<i>puniceus</i>	-	-	10	12	33.35	-117.32
hrc	<i>puniceus</i>	-	-	7	11	33.35	-117.49
395	<i>puniceus</i>	full	12	-	20	33.30	-117.15
dlr	<i>puniceus</i>	full, sd subset	8	10	16	33.17	-117.05
lkw	<i>puniceus</i>	full, sd subset	13	9	24	33.16	-117.02
crs	<i>puniceus</i>	full, sd subset	9	-	15	33.13	-117.31
bc	<i>puniceus</i>	full, sd subset	12	7	24	33.12	-116.80
inj	<i>puniceus</i>	full, sd subset	11	8	16	33.10	-116.66
elf	<i>puniceus</i>	full, sd subset	8	9	24	33.09	-117.15
lh	<i>puniceus</i>	full, sd subset	11	11	24	33.06	-117.12
bs	<i>puniceus</i>	full, sd subset	15	-	24	33.01	-117.02
mw	<i>puniceus</i>	full, sd subset	16	6	23	33.01	-116.96
sdp	<i>puniceus</i>	full, sd subset	12	-	16	33.00	-117.24
wcr	<i>puniceus</i>	full, sd subset	12	-	12	32.99	-116.83
bcrd	<i>puniceus</i>	full, sd subset	4	7	8	32.95	-116.64
pmd	<i>puniceus</i>	full, sd subset	12	-	16	32.94	-117.06
oak	<i>puniceus</i>	full, sd subset	10	8	16	32.91	-116.89
elt	<i>puniceus</i>	full, sd subset	11	-	16	32.89	-117.09
ucsd	<i>puniceus</i>	full, sd subset	11	9	32	32.89	-117.24
and	<i>puniceus</i>	full, sd subset	6	-	-	32.87	-116.75
wm	<i>puniceus</i>	full, sd subset	17	9	24	32.82	-116.90
mt	<i>puniceus</i>	full, sd subset	10	9	16	32.82	-117.06
flp	<i>puniceus</i>	full, sd subset	9	-	16	32.81	-116.99
jmc	<i>puniceus</i>	full, sd subset	17	13	24	32.74	-116.95
pct	<i>puniceus</i>	full, sd subset	5	8	21	32.73	-116.47
lo	<i>puniceus</i>	full, sd subset	12	7	23	32.68	-116.33
dlz	<i>puniceus</i>	full, sd subset	3	-	9	32.65	-116.79
potr	<i>puniceus</i>	subset	9	6	23	32.60	-116.63



### *Population Structure*

To detect patterns of genomic ancestry and levels of admixture among the sequenced samples, we ran the program ADMIXTURE (Alexander, Novembre, & Lange, 2009). Raw RAD-seq reads were processed in a similar manner to previously published work (Chase et al., 2017; Stankowski et al., 2019) using STACKS version 1.44 (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). For this analysis, reads were filtered with the process\_radtags script from STACKS, aligned to the *M. aurantiacus* v.1 assembly (Stankowski et al., 2019) using bowtie2 (Langmead & Salzberg, 2013), and then run through the refmap.pl script from STACKS to call SNPs. To filter for errors and linkage disequilibrium, we then further filtered SNPs using a minor allele frequency cutoff of 0.01 (across the entire dataset), we only included loci present in at least 80% of samples, and we limited the output to one random SNP per RAD tag, resulting in 42,386 SNPs. To further reduce the extent of linkage disequilibrium among tightly linked sites, we used the command indep-pairwise 50 10 0.1 in PLINK v1.90 to filter overlapping windows of 50 variant SNPs (step size = 10 SNPs) down to  $r = 0.1$  (Purcell & Chang, 2015). This resulted in a final dataset of 15,558 SNPs.

We ran ADMIXTURE v1.3 with 2 – 5 clusters (K), and 20 independent runs for each value of K. The runs were aggregated with CLUMPP (Jakobsson & Rosenberg, 2007), using the greedy method, weighting populations by number of individuals using the G' similarity matrix, and using random input orders with 1000 repeats (M = 2, W = 1, S = 2, GREEDY\_OPTION = 2, REPEATS = 1000). Additionally, we ran a principal component analysis (PCA) using PLINK on the same SNP data, writing all

615 principal components to accurately calculate variance explained from the positive eigenvalues of the covariance matrix.

### *Floral Trait Analysis*

To understand how floral traits were distributed geographically across the range of *puniceus*, we grew plants from seeds collected from 324 maternal families from 34 populations (minimum 6 families per population, maximum 12), 16 of which were previously published (Sobel et al., 2019; Stankowski & Streisfeld, 2015). Seedlings were initially grown in plug trays in a growth chamber for two weeks. Up to three individuals from each family were repotted and grown in a common greenhouse environment. From two flowers per individual within 48 hours of anthesis, we measured four floral traits known to differ significantly between the red and yellow ecotypes of *puniceus* in San Diego County (Stankowski et al., 2015): corolla length (CL), pedicel length (PL), stigma exertion (SE), tube width (TW), and the concentration of the red anthocyanin pigment, as described previously (Stankowski et al., 2015). A principal component analysis (PCA) was performed to summarize the multivariate relationship among the traits. To test whether the floral traits differed between northern and southern regions, the PC trait scores in each population were compared to latitude. To test whether multi-trait adaptation associated with flower color was consistent across different geographic regions, we compared PC scores of these traits to the *MaMyb2* allele frequency for each population.

## *Bioclimatic Data and Visualization*

In addition to differences in floral traits likely associated with visitation by different pollinators (Streisfeld and Kohn 2007; Sobel and Streisfeld 2015), we investigated the potential role of local adaptation to the abiotic environment by examining variation in climate among sites. Previous work demonstrated physiological traits associated with climate variation between western and eastern populations of *puniceus* in San Diego. For all populations that contained sequence data, we downloaded data on 23 bioclimatic variables for the years 1981-2010 from ClimateWNA (Wang, Hamann, Spittlehouse, & Murdock, 2012). A description of all variables can be found in Table 2. We then summarized correlations among variables using PCA and extracted the first principal component from these data, which explained 63.7% of the variation in climate. To visualize spatial variation in climate across the region, we interpolated variation in climate across the sampled region from normalized PC1 data using kriging in the R package “kriging.” To generate a metric for the ecological difference in climate between sampled sites, we calculated the difference between mean PC1 scores between pairs of populations, which we called “ecological distance.”

Table 2. All 23 measured bioclimatic variables from ClimateWNA

Variable	abbreviation	calculated
mean annual temperature (°C)	MAT	directly
mean warmest month temperature (°C)	MWMT	directly
mean coldest month temperature (°C)	MCMT	directly
temperature difference between MWMT and MCMT, or continentality (°C)	TD	directly
mean annual precipitation (mm)	MAP	directly

May to September precipitation (mm)	MSP	directly
annual heat-moisture index (MAT+10)/(MAP/1000)	AHM	directly
summer heat-moisture index ((MWMT)/(MSP/1000)	SHM	directly
degree-days below 0°C, chilling degree-days	DD<0	derived
degree-days above 5°C, growing degree-days	DD>5	derived
degree-days below 18°C, heating degree-days	DD<18	derived
degree-days above 18°C, cooling degree-days	DD>18	derived
the number of frost-free days	NFFD	derived
frost-free period	FFP	derived
the day of the year on which FFP begins	bFFP	derived
the day of the year on which FFP ends	eFFP	derived
precipitation as snow (mm) between August in previous year and July in current year	PAS	derived
extreme minimum temperature over 30 years	EXT	derived
extreme maximum temperature over 30 years	EXT	derived
Hargreaves reference evaporation (mm)	Eref	derived
Hargreaves climatic moisture deficit (mm)	CMD	derived
mean annual solar radiation (MJ m <sup>-2</sup> d <sup>-1</sup> )	MAR	derived
mean annual relative humidity (%)	RH	derived

---

Notes: information obtained from

<http://www.climatewna.com/help/climateBC/help.htm>

### *Population Genomic Analyses*

To examine how genome-wide patterns of differentiation varied across geographic space and ecology, we re-called variants in STACKS using only individuals from subspecies *puniceus*. Beginning with processed, aligned sequences, we re-ran the ref-map.pl script to call SNPs. We then filtered the data to include loci present in at least 70% of samples. We did not include minor allele frequency cutoffs or LD pruning in this dataset to allow rare alleles to be included in population genetic estimates of diversity and differentiation. This resulted in a final dataset of 436,935 SNPs. We calculated  $\pi$  and  $F_{ST}$  within and between all sequenced populations in 500 kb non-overlapping windows using the script popgenwindows.py (downloaded from [https://github.com/simonhmartin/genomics\\_general](https://github.com/simonhmartin/genomics_general)). Estimates of  $\pi$  were adjusted to

account for invariant sites by multiplying our measure of  $\pi$  by the proportion of variant to total sites in each window. In addition, we used previously published data on recombination rate and gene count in the same 500 kb windows, which were derived from a genetic map and genome annotation from subspecies *punicieus* (Stankowski et al., 2019).

To investigate how genomic variation was impacted by space and ecology, we examined the relationship between average  $F_{ST}$  among genomic windows and distance metrics. In addition, we estimated the pairwise population-specific effects of linked selection by calculating the correlation coefficient across windows (Pearson's  $r$ ) between  $F_{ST}$  and average  $\pi$ ,  $F_{ST}$  and gene count, and  $F_{ST}$  and recombination rate. We created a subset of the full dataset consisting of southern populations of *punicieus*, which we referred to as the "San Diego subset" or "SD subset" (Table 1), as previous work in these populations revealed that local adaptation to the abiotic environment likely contributed to ecotypic differentiation (Sobel et al. 2019). For both the full dataset and the SD subset, we began by evaluating the relationships between these correlations and geographic and ecological distance between all pairs of populations using linear regression. However, because geographic distance tends to be correlated with ecological distance in San Diego, we then used partial Mantel tests using the R package *ncf* v1.2-9 (Bjornstad & Bjornstad, 2020) to compare the independent effects of geographic or ecological distance on these correlations. To evaluate the statistical significance of the relationship, we used a permutation test with 1000 random permutations of the data. Moreover, due to extensive differences in ancestry between northern and southern populations of *punicieus* (see results), we created a distance matrix based on differences

in ancestry scores from the Admixture analysis. Specifically, “ancestry distance” between all pairs of populations was calculated by taking the difference in mean Q score between populations from the cluster that primarily characterized the northern *punicieus* populations (fig. 2a,b). Differences in ancestry and geography tend to be correlated for populations separated in allopatry. Therefore, we ran partial Mantel tests to examine the independent effects of ancestry and geography on correlations involving  $F_{ST}$ . Because IBD was previously characterized in southern *punicieus* and IBA was likely to occur due to known ecogeographic isolation (Sobel and Streisfeld 2015), we expected both geography and ecology to explain variation in  $F_{ST}$  and correlations involving  $F_{ST}$  in the SD subset. Additionally, under the hypothesis that divergence time strengthens genome correlations (Stankowski et al., 2019), we expected that differences in ancestry would explain variation in genome correlations for the full dataset where there are extensive differences in ancestry between north and south (see results).

## Results

### *Multiple levels of genetic subdivision due to geographical, ecological, and ancestry differences within puniceus*

Our SNP dataset reveals large amounts of population stratification present in *puniceus* relative to *longiflorus* and *calycinus*. We expected the ADMIXTURE dataset to segregate individuals by subspecies, but we were surprised to find that *puniceus* individuals do not all group together at  $K = 2$  or  $K = 3$ . Instead, we find that a number of geographically clustered northern *puniceus* populations separate from the rest of their subspecies for all values of  $K$  (Fig. 2a,c). Indeed, a PCA on the same dataset reveals that the individuals cluster into three essentially discrete groups along the first two principal components that correspond to *longiflorus/calycinus*, northern *puniceus*, and southern *puniceus*. At  $K = 4$  and  $K = 5$ , we see additional subdivision within *puniceus* that corresponds to eastern and western populations in San Diego (Fig. 2a).

Within *puniceus*, separation due to geography at  $K = 2$  points to extensive differences in ancestry between northern and southern groups (Fig. 2a,c). This abrupt change in ancestry score occurs over a relatively short geographic distance and persists at higher values of  $K$ , suggesting that northern and southern *puniceus* populations have extreme differences in ancestry. The possibility that there are two distinct lineages is relevant for the population genomic analyses below, as the model of IBD assumes a single lineage that becomes increasingly differentiated through space. Therefore, differences in history due to unique ancestry must be accounted for when considering the effects of geography and ecology on patterns of genomic variation.

Within southern *puniceus*, patterns of ancestry and admixture appear related to geographic and ecological differences. At  $K = 4$ , southern *puniceus* separates into eastern and western groups, consistent with differences in climates (Fig. 2a,c). This trend supports previous results that demonstrated IBD and adaptation to climate in these same populations (Sobel et al., 2019; Stankowski et al., 2015). Unlike the separation of northern *puniceus* from the rest of its subspecies, the genetic separation between southeastern and southwestern groups occurs over a gradient with extensive admixture across many individuals (Fig. 2a,c). This pattern is consistent with IBD and ongoing hybridization where the ranges of the red and yellow ecotypes overlap (Stankowski et al., 2015, 2017). Additionally, the western and eastern groups of southern *puniceus* group close to each other along the first two principal components (Fig. 2b). In addition to southern *puniceus* grouping together at  $K = 3$ , this suggests differences in ancestry within the southern populations are much less substantial relative to comparisons between the regions.

Our analysis also finds evidence of admixture among populations of *puniceus* and between *puniceus* and *longiflorus/calycinus*. In addition to the well-studied hybridization between the red and yellow ecotypes of *puniceus* in San Diego (Stankowski et al., 2015; Stankowski & Streisfeld, 2015), we find evidence of admixture between the northern and southern ancestry groups of *puniceus*. While this could be due to secondary contact between two previously isolated populations, further work will be needed to distinguish this hypothesis from the maintenance of shared ancestral polymorphism among the populations. Moreover, there is evidence for admixture between northern *puniceus* populations and *longiflorus/calycinus* in areas



where geographic overlap is evident between these subspecies. Finally, *longiflorus* and *calycinus* curiously failed to become differentiated from each other, even at  $K = 5$  (Fig. 2a), suggesting that these two subspecies are extremely closely related. Rather,  $K=5$  revealed a small group of yellow ecotype individuals with unique ancestry from extreme southern San Diego County. These findings are consistent with previous results (Stankowski et al 2015) and further highlight the many levels of population stratification found within *puniceus*.

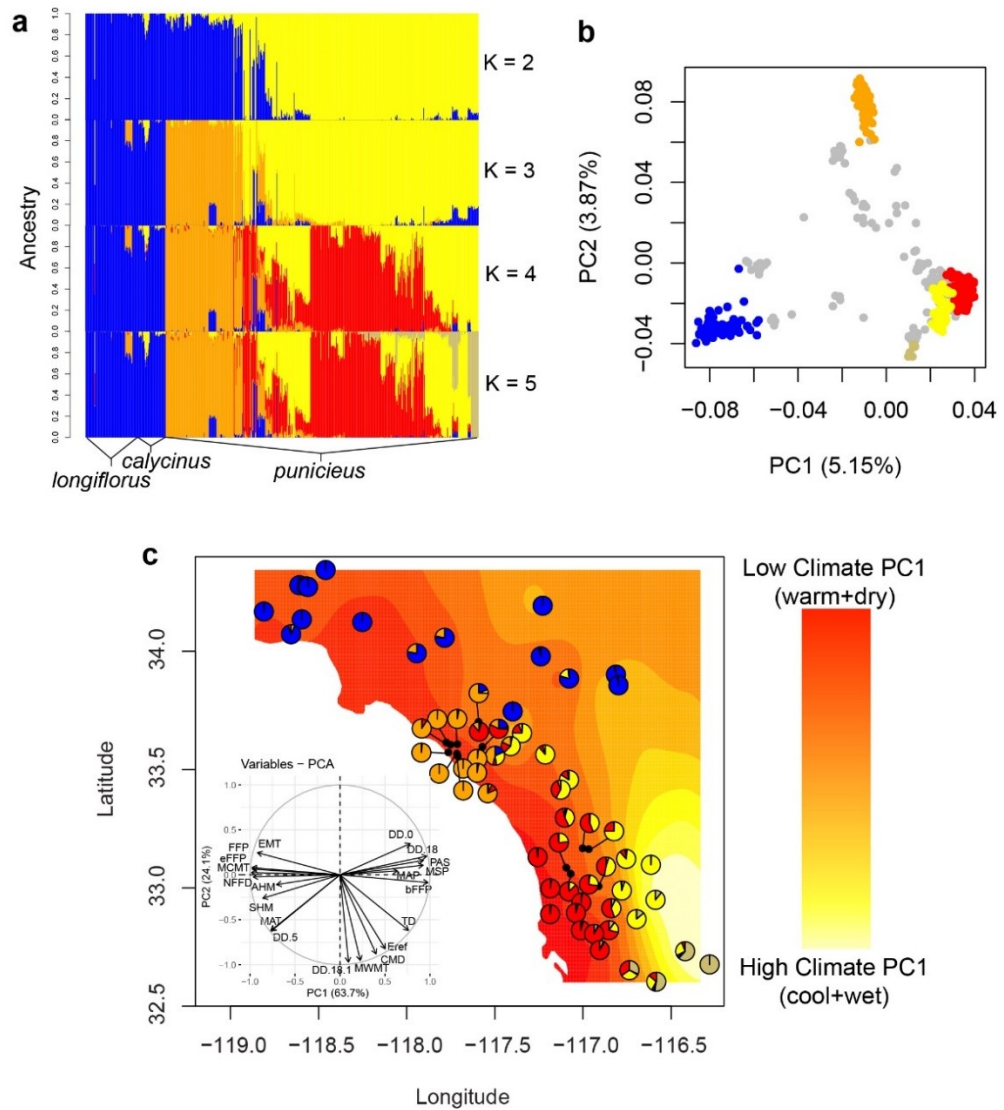


Figure 2. Admixture scores and a PCA of bioclimatic variables. (a) Stacked Admixture barplots for all individuals with K = 2-5. (b) PC1 vs. PC2 of the same genetic dataset. Individuals with an ancestry score of at least 0.8 for an admixture group at K = 5 are colored according to that admixture group, as in Fig. 2a. (c) Pie plots colored by admixture scores for K = 5 from Figure 2a, averaged for each population and plotted by latitude and longitude. The background interpolation represents PC1 scores for 23 bioclimatic variables from ClimateWNA (shown in the lower left corner; full descriptions the variables in Table 2)

*Trait distributions are uniform across northern and southern populations of puniceus*

Given the distinct patterns of ancestry between northern and southern populations of subspecies *puniceus*, we compared floral trait distributions between these areas. In particular, we wished to see if the genetic differentiation between northern and southern populations of *puniceus* was accompanied by divergence in floral features (Fig. 2). As noted previously, floral traits are highly differentiated between the ecotypes in San Diego and contribute to pollinator isolation (Stankowski et al 2015; Sobel and Streisfeld 2015). Despite differences in ancestry, multivariate floral phenotypes were highly similar between southern and northern regions of the range (linear regression of average trait PC1 against collection latitude; multiple  $r^2 = 0.0071$ ,  $p = 0.63$ ).

Additionally, we found no correlation between mean trait PC1 and the mean ancestry score (difference in the orange bars at  $K=3$  in figure 2a, largely representing for northern *puniceus* individuals; linear regression; multiple  $r^2 = 0.0310$ ,  $p = 0.40$ ). Instead, we observed that floral traits grouped by flower color genotype when all families were aggregated (Fig. 3a; linear model of population trait PC1 means against MaMyb2 allele frequency; multiple  $r^2 = 0.811$ ,  $p < 0.001$ ). Moreover, PC1 explained 65.6% of the variation in these traits, and clearly separated red and yellow ecotype populations. Indeed, red ecotype populations are characterized by higher concentrations of anthocyanin pigment, larger stigma exertion and longer pedicels, but shorter and narrower corollas. These findings suggest that substantial floral trait variation exists across the range of *puniceus*, but this variation remains similar between the genetically differentiated northern and southern regions.

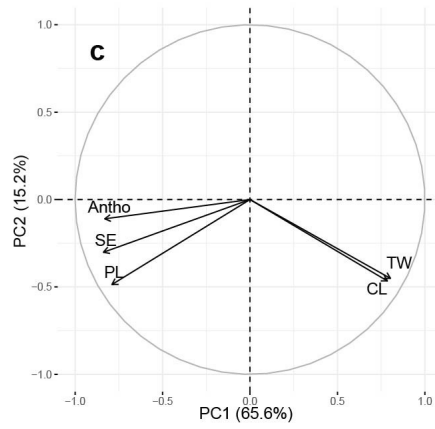
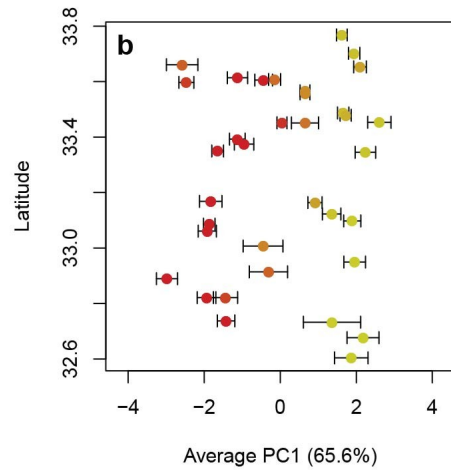
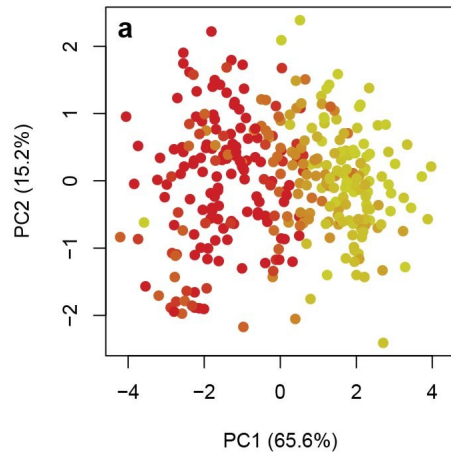


Figure 3. A summary of floral traits within *M. puniceus*. (a) PC1 vs. PC2 for all families measured, colored by population allele frequency at the Myb-2 locus linked to flower color. (b) Average PC1 of each population plotted against latitude and colored by Myb-2 allele frequency, with error bars representing standard error of the mean. (c) Loadings of all floral traits measured (Anthocyanin concentration, stigma exertion, pedicel length, corolla length, and tube width).

*Widespread selection is associated with ancestry and ecology, but not geography*

To assess the role of local adaptation in shaping the genomic landscape of differentiation, we investigated how genomic correlations involving  $F_{ST}$  varied due to geography and ecology. If the strength of genome correlations is associated with variation in geographic but not ecological differences, this suggests that divergent natural selection and adaptation may not be the primary driver of divergence. By contrast, if the correlations are more strongly predicted by ecological differences between populations, this suggests that local adaptation is likely important in explaining genomic patterns. However, it is possible that correlations may form between previously separated lineages that have nothing to do with local adaptation. Therefore, we marked comparisons where ancestry estimates differed between populations by at least 0.8 (orange bars at  $K=3$  in Fig. 2). If this ancestry-differentiated subset represented comparisons between two separate lineages, we would expect  $F_{ST}$  to be high and correlations involving  $F_{ST}$  to be strong regardless of geographic or ecological distances. Additionally, we focused separately on the southern *puniceus* populations (in San Diego

County; SD), where there exists substantial variation in climate along a west to east gradient (Fig 2c), and IBD has been well characterized (Stankowski et al., 2015). Therefore, the potential exists that IBA might be prominent within the SD region.

IBD and IBA both appear to affect patterns of genetic differentiation across subspecies *puniceus* (Fig. 4 a,b). Within the SD subset, the correlation between  $F_{ST}$  and ecological distance ( $r^2 = 0.5287$ ) is stronger than the correlation with geography ( $r^2 = 0.2767$ ), but for the full dataset across all *puniceus* populations, geographic distance is a better predictor of  $F_{ST}$  than ecological distance ( $F_{ST}$ -geographic-distance;  $r^2 = 0.2434$ ;  $F_{ST}$ -ecological distance,  $r^2 = 0.1815$ ). In comparisons between populations that differ substantially in admixture scores, geographic distance is not a good predictor for  $F_{ST}$  ( $r^2 = 0.0470$ ). However, these comparisons show signs that ecological differences might play a role in driving genetic differentiation ( $F_{ST}$ -ecological distance  $r^2 = 0.2418$ ).

Across the full dataset, genome-wide correlations between  $F_{ST}$  and measures of genetic diversity and genome architecture appear to strengthen due to geographic distance. The  $F_{ST}$ - $\pi$  and  $F_{ST}$ -recombination correlations grow increasingly negative at greater geographic distances between populations, and the  $F_{ST}$ -gene-count correlation grows increasingly positive (Fig. 4 c,e,g, black lines), which is expected as divergence increases (Stankowski et al., 2019). Considered separately, these results imply that neutral processes and background selection may be important for shaping the genomic landscape. However, geographic distance is a relatively poor predictor of  $F_{ST}$  and correlations involving  $F_{ST}$  when looking within the ancestry-differentiated subset where only northern and southern populations are compared, and these correlations are in the opposite direction based on predictions due to linked selection (Fig. 4 a,c,e,g red lines).

Therefore, comparisons between northern and southern populations may result in strong trends regardless of geographic distance. Furthermore, we observed stronger correlations when considering ecological distance rather than geographic distance for the ancestry-differentiated subset (Fig. 4 b,d red lines), but the correlations again are in the opposite direction expected under a model of linked selection (Fig. 4 f,h red lines).

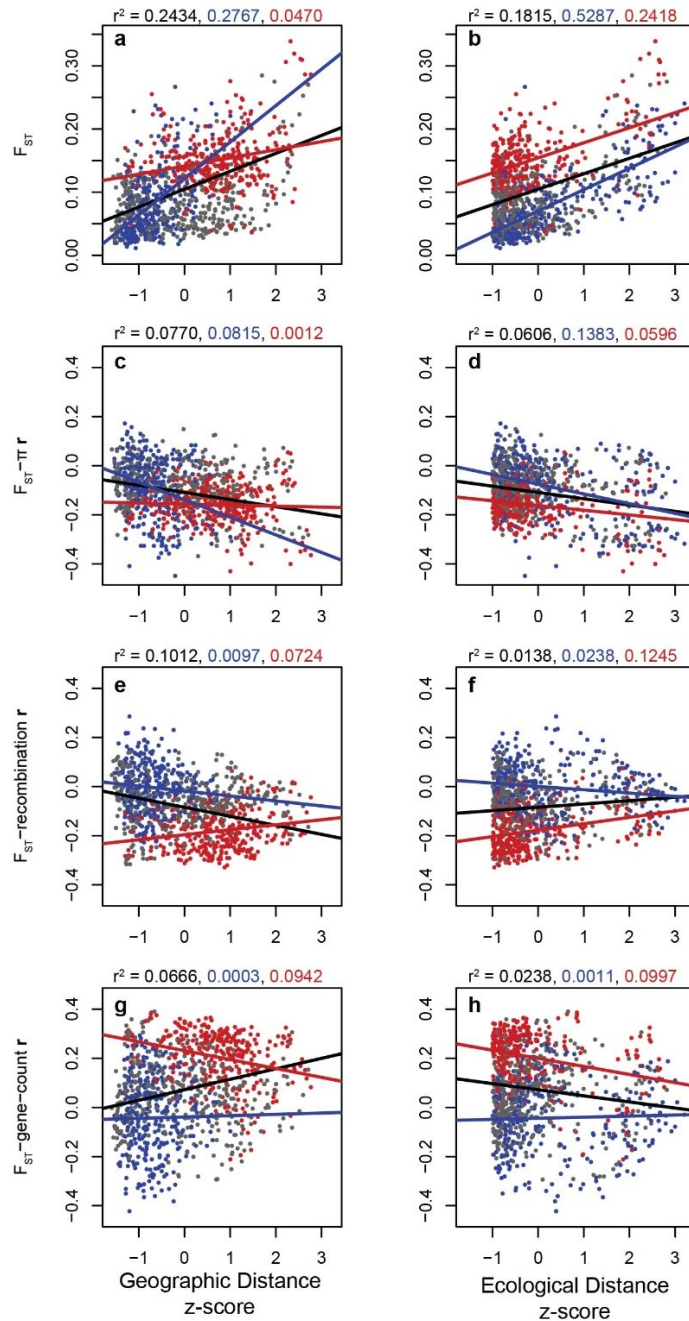


Figure 4. Scatterplots and linear regressions of  $F_{ST}$  and the Pearson's  $r$  between  $F_{ST}$  and other genome statistics plotted across geographic and ecological distance for all pairs of populations. The blue dots, blue trendlines, and blue  $r^2$  values represent the SD subset, and the red points, red trendlines, and red  $r^2$  values represent comparisons that had an ancestry score difference of at least 0.8 for the orange-colored admixture group at  $K = 3$  (Fig. 2). Black trendlines and  $r^2$  values represent the full dataset. All  $r^2$  values are derived from Pearson's  $r$  between response variables and measures of distance.



Within the SD subset, the  $F_{ST}-\pi$  correlation varies more with ecological distance ( $r^2 = 0.1383$ ) than geographic distance ( $r^2 = 0.0815$ ), and the relationship between the  $F_{ST}-\pi$  correlation and ecology is stronger within the SD subset than in the full dataset or the admixture-differentiated subset (Fig. 4d). By contrast, the  $F_{ST}$ -recombination and  $F_{ST}$ -gene-count correlations do not covary substantially with either geographic or ecological distance for the SD subset (Fig. 4 e-h, blue lines). These results suggest that among SD populations, ecology plays a larger role in shaping the genome than geography.

Because geography covaries with both ecology and ancestry (Full dataset: geography-ecology  $r = 0.2659$ , geography-ancestry  $r = 0.5374$ ; SD subset: geography-ecology  $r = 0.6042$ ; all  $r$  values from Mantel tests), we used partial Mantel tests to elucidate the relative effects of each distance measure. When geography and ecology are compared in the full dataset, both are good predictors of average  $F_{ST}$ , but only geography is a good predictor of the strength of the correlations involving  $F_{ST}$  (Fig 5a). These results line up with what we observed in Fig. 4, but they do not account for the effects of ancestry, which also covaries with geographic distance. Therefore, we examined the predictive power of both ancestry and geography and found that admixture was a better predictor than geography for all variables measured (Fig. 5b). Worth noting is that the effects of ecology are not accounted for in this comparison, so we cannot discount that the correlations arising from this analysis are due to ecology rather than geography. However, as ecological distance shows little covariance with ancestry for the full dataset (Mantel  $r = -0.0983$ ,  $p = 0.0859$ ), trends that occur across ancestry differences are unlikely to be confounded by ecology. Finally, we compared the

independent effects of geography and ecology in the SD subset, where admixture differences at  $K = 3$  are negligible and environmental differences are the most extreme (Fig. 2c). We found that both geography and ecology are significant predictors of average  $F_{ST}$ . In addition, ecology (but not geography) is a significant predictor of the correlation between  $F_{ST}$  and  $\pi$ , but neither is a significant predictor of the correlation between  $F_{ST}$  and gene-count or recombination (Fig. 5c). In sum, these data show that despite ecology, geography, and ancestry being good predictors of average  $F_{ST}$  between populations, genome correlations involving  $F_{ST}$  consistently strengthen only due to differences in ancestry. However, in the SD subset, ecology explains more variation in the genome wide correlation between  $F_{ST}$  and  $\pi$  than geography, suggesting that local adaptation may help shape patterns of heterogeneous genomic divergence in these populations that are early in the speciation process.

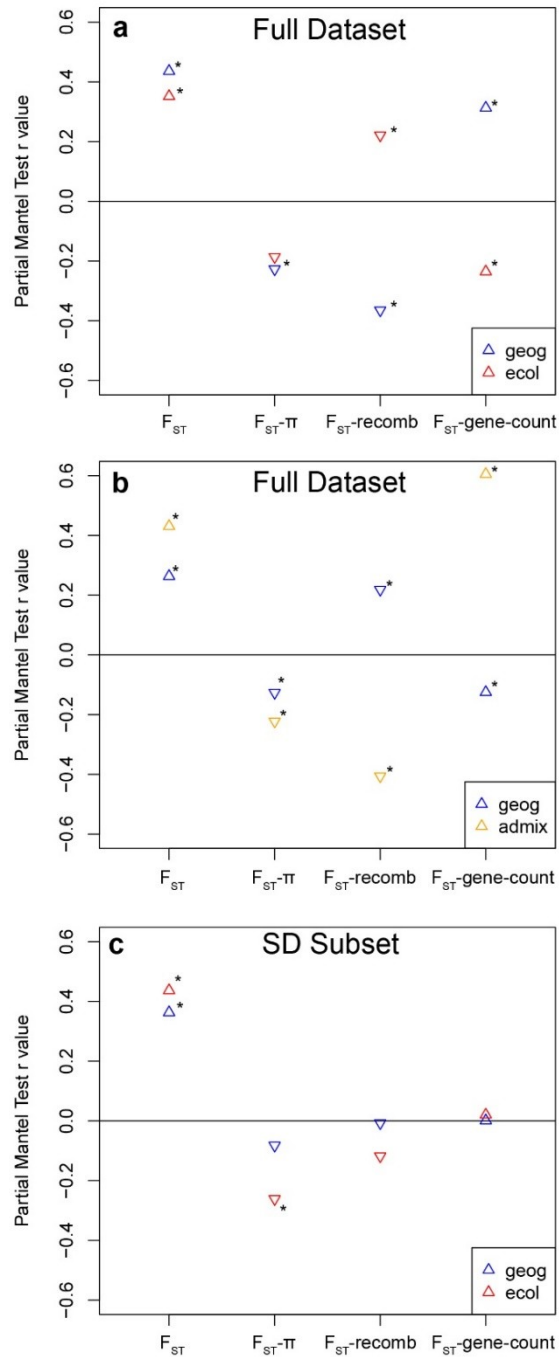


Figure 5. Partial Mantel correlation coefficients for  $F_{ST}$  and correlations involving  $F_{ST}$  over geographic distance (blue points), differences in the orange-colored admixture group from Fig. 2 (yellow-orange points) and ecological distance (red points), for the full dataset (a) and the SD subset (b). Data markers point in the direction of the expected correlation for increasingly isolated lineages, as in Stankowski et al. (2019). Asterisks indicate that the

## Discussion

*Population structure allows us to characterize a genetically distinct lineage of puniceus*

Our population structure analysis identified two separate lineages of *puniceus* corresponding to northern and southern areas (Fig. 2). Climate does not appear to vary considerably between these regions, suggesting that IBA is unlikely to explain this pattern. Additionally, the sharp transition between the groups argues against IBD (Fig. 2c), which contrasts with the gradual pattern of transition in southern, San Diego populations (Stankowski et al, 2015). Therefore, these findings suggest that northern populations represent a separate lineage of *puniceus* that was historically isolated from the southern populations.

By contrast, the ancestry-differentiated subset shows signs of IBD and IBA, which appears to challenge the notion that northern and southern *puniceus* were allopatrically isolated. However, differences in ancestry rather than geography or ecology are the best predictor of these trends. If northern and southern populations were genetically distinct, we expect  $F_{ST}$  and the strength of genome correlations to be consistently high and unrelated to geographic and ecological distances. Instead, geography and ecology predicted some of the variation in  $F_{ST}$ , with the correlation between ecology and  $F_{ST}$  being stronger (Fig. 4 a,b). However, these results align with a previous phylogeny that indicated northern populations were more closely related to southwestern than southeastern populations (Chase et al., 2017). We therefore expect the northern populations to be the most differentiated from populations in the southeast, and these comparisons have slightly higher geographic distance and much higher

ecological distance than north-southwest comparisons (Fig. 2c). Therefore, the patterns of  $F_{ST}$  we observe in our ancestry-differentiated subset most likely can be attributed to differences in evolutionary history that resulted in two separate lineages of *puniceus*.

While northern *puniceus* was originally described as part of a separate population series (the “Laguna series”) on the basis of perceived differences in floral traits (Beeks, 1962), our common garden experiment in the greenhouse failed to detect heritable trait differences between the regions. Indeed, we found that the distribution of floral traits in northern *puniceus* populations largely overlapped with southern populations (Fig. 3b), but that ancestry was not a significant predictor of trait PC1 scores. Interestingly, northern populations maintain differences in floral phenotypes despite close geographic proximity to each other, while in San Diego, floral trait differences are structured east to west (Stankowski et al., 2015). It is currently unclear why this is the case. Subspecies of *M. aurantiacus* are often categorized by floral phenotype (Chase et al., 2017), but as we can see here, phenotypically indistinguishable populations are in fact genetically distinct.

Despite ancestry being the best predictor of differences in  $F_{ST}$  and correlations involving  $F_{ST}$  between northern and southern *puniceus*, IBA and IBD appear to contribute to variation in southern *puniceus*. At  $K = 4$ , southern *puniceus* splits along a geographic and ecological gradient from west to east (Fig. 4a,c). IBD has been characterized previously in southern *puniceus* (Stankowski et al., 2015), showing a gradual change in genetic differentiation over geographic distance. In addition, our kriging analysis revealed that climate also changed over a similar gradient from east to west (Fig. 4c). This is consistent with previous work showing that ecophysiological

traits exhibited a gradual clinal transition along the same geographic axis (Sobel et al., 2019). Therefore, the east-west gradient of ancestry scores in southern *puniceus* populations could be caused by ecological differences, as well as, or instead of, geographic distance. The separation within eastern San Diego *puniceus* populations at  $K = 5$  also is consistent with previous results (Stankowski et al., 2015), suggesting some isolation among yellow ecotype populations, perhaps because suitable habitat is more patchy in eastern San Diego or pollen and seed dispersal may be more limited. Further work will be necessary to evaluate these hypotheses.

In addition to the previously well-characterized hybridization between *puniceus* ecotypes in San Diego (Sobel & Streisfeld, 2015; Stankowski et al., 2015, 2017; Streisfeld & Kohn, 2007; Streisfeld & Kohn, 2005), we find evidence of admixture between southern and northern *puniceus* populations, as well as between *longiflorus/calycinus* and northern *puniceus* (Fig. 4c). In both cases, the extent of admixture appears to vary across geographic gradients, patterns that are characteristic of gene flow following secondary contact between lineages (Barton, Gale, & Harrison, 1993; Endler, 1977). These results highlight the importance of genetic admixture and differences in ancestry when considering continuously distributed populations of separate lineages.

#### *Adaptation and ancestry differences predict heterogenous genome divergence*

In San Diego populations of *puniceus*, there is considerable evidence of speciation occurring between red and yellow-flowered ecotypes along a geographic and ecological gradient (Chase et al., 2017; Sobel et al., 2019; Stankowski et al., 2017). Our population genomic analyses find that local adaptation contributes to shaping the

heterogeneous genomic landscape in the SD subset. Variation in both geography and ecology explain a significant amount of variation in  $F_{ST}$ , confirming the effects of both IBD and IBA. In addition, ecological distance was a significant predictor of the strength of the correlation between  $F_{ST}$  and  $\pi$ , implying the effects of heterogeneous linked selection in driving this pattern (Fig. 5c). However, there were no effects of ecological or geographic distance on the strength of the  $F_{ST}$ -recombination or  $F_{ST}$ -gene count correlations that are hallmark predictors of linked selection (Burri et al., 2015). It is worth noting that our measure of climate differences only represents a single axis of ecological variation that may be important for adaptation. If we understood more about various aspects of the biotic and abiotic environments, such as soil type and organisms that consume these plants, we might be able to find stronger signals of linked selection. Nevertheless, it appears that differences in climate are associated with a stronger association between  $F_{ST}$  and  $\pi$  than geography, which is consistent with the effects of divergent natural selection shaping genomic divergence between these emerging species.

By contrast, when we examine the full dataset, it becomes clear that ancestry differences are a strong predictor of variation in genome correlations. While both ancestry and geography are significant predictors of variation in average  $F_{ST}$ , only ancestry is a good predictor of the  $F_{ST}$ -recombination-rate and  $F_{ST}$ - $\pi$  correlations in the direction we expect due to linked selection (Fig. 5b). Although geography is a significant predictor of the  $F_{ST}$ - $\pi$  correlation in the full dataset, this conclusion must be tempered because geography covaries with both ecology and ancestry, which cannot both be accounted for in a partial Mantel test. We do not have the same concerns about

ancestry differences covarying with the climate variables measured, but there may be covariation between ancestry and unmeasured sources of ecological variation.

Therefore, we cannot rule out the possibility of IBA differentiating southern and northern populations.

Our results support the role of adaptive evolution facilitating genetic divergence through linked selection in the southern *puniceus* lineage, but there is little support for a role of divergent selection in the north. While the relationship between the strength of the correlations and ancestry suggest a possible role for background selection in shaping heterogeneous patterns of genomic differentiation, we also cannot rule out the possibility of local adaptation due to unmeasured selective pressures. Moreover, floral traits are differentiated similarly in both northern and southern populations, but we did not measure ecological features typically associated with divergence in floral traits, such as pollinator visitation. That being said, our results suggest the possibility that divergent selection due to variation in climate is capable of driving genomic patterns of differentiation in San Diego, but because climate differences are less substantial among populations to the north, we are unable to detect a signal of local adaptation.

In sum, these findings indicate the importance of having a strong understanding of a system's ecology when investigating genomic trends arising from adaptation.

Because similar genomic patterns can arise from both positive and background selection, it is important to understand the mechanism by which populations become partially or fully isolated from one another. An increasing number of studies is testing for IBA (Sexton et al., 2014), but unlike IBD, it is not always clear how to measure ecological distance. Understanding the important ecological factors within a system of



study is vital for interpreting genetic divergence. It is also important to consider many factors of the genome and what they can tell us about divergence and linked selection. Genome-wide  $F_{ST}$  is limited in what it can tell us about important aspects of the genome, such as linked selection. Ancestry, geography, and ecology were all significant predictors of  $F_{ST}$  (Fig. 5), but they varied dramatically in their ability to explain correlations associated with the presence of linked selection and genomic divergence. Sampling a variety of genome statistics and understanding what each one tells us is vital for establishing a holistic view of speciation genomics.

## Conclusion

In summary, we found that local adaptation can shape heterogeneity in the genomic landscape during the early phases of divergence with gene flow. However, this pattern was present across only a portion of the geographic range of subspecies *puniceus*. Our survey identified two lineages of *puniceus*, and a population genomic analysis determined that this difference in ancestry between populations is a strong predictor of genomic correlations that form due to heterogeneous linked selection. On the other hand, within a subset of *puniceus* where populations are distributed across an ecological gradient, adaptation to the local ecology appears to contribute to heterogeneous genomic patterns. Characterizing a system's ecology and the many features of the genome are both crucial for identifying how the speciation process affects the genome, which gets us closer to understanding the vast diversity in life on earth that surrounds us every day.

## Works Cited

- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664.
- Barton, N. H., Gale, K. S., & Harrison, R. G. (1993). Hybrid zones and the evolutionary process. *Genetic Analysis of Hybrid Zones*, pp. 13–42. Oxford University Press Oxford.
- Beeks, R. M. (1962). Variation and hybridization in southern California populations of *Diplacus* (Scrophulariaceae). *Aliso: A Journal of Systematic and Evolutionary Botany*, 5(2), 83–122.
- Bjornstad, O. N., & Bjornstad, M. O. N. (2020). Package ‘ncf.’
- Burri, R. (2017). Interpreting differentiation landscapes in the light of long-term linked selection. *Evolution Letters*, 1(3), 118–131. doi: 10.1002/evl3.14
- Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., ... 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. doi: 10.1101/gr.196485.115
- 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.
- Charlesworth, B. (2013). Background selection 20 years on. *Journal of Heredity*, 104(2), 161–171. doi: 10.1093/jhered/ess136
- 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. doi: 10.3732/ajb.1700234
- 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. doi: 10.1111/mec.12796
- Endler, J. A. (1977). *Geographic variation, speciation, and clines*. Princeton University Press.
- Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, 28(7), 342–350. doi: 10.1016/j.tig.2012.03.009

- Feder, J. L., & Nosil, P. (2010). The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution; International Journal of Organic Evolution*, 64(6), 1729–1747. doi: 10.1111/j.1558-5646.2009.00943.x
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23(14), 1801–1806.
- Langmead, B., & Salzberg, S. L. (2013). Langmead. 2013. Bowtie2. *Nature Methods*, 9, 357–359.
- Noor, M. A. F., & Bennett, S. M. (2009). Islands of speciation or mirages in the desert Examining the role of restricted recombination in maintaining species. *Heredity*, 103(6), 439–444. doi: 10.1038/hdy.2009.151
- Nosil, P., Egan, S. P., & Funk, D. J. (2008). Heterogeneous genomic differentiation between walking-stick ecotypes: “Isolation by adaptation” and multiple roles for divergent selection. *Evolution*, 62(2), 316–336. doi: 10.1111/j.1558-5646.2007.00299.x
- Payseur, B. A., & Nachman, M. W. (2002). Gene Density and Human Nucleotide Polymorphism. *Molecular Biology and Evolution*, 19(3), 336–340. doi: 10.1093/oxfordjournals.molbev.a004086
- Peterson, M. A., & Denno, R. F. (1998). The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects. *American Naturalist*, 152(3), 428–446. doi: 10.1086/286180
- Purcell, S., & Chang, C. (2015). PLINK 1.9. URL [https://www. Cog-Genomics. Org/Plink2](https://www.cog-genomics.org/Plink2).
- Renaut, S., Grassa, C. J., Yeaman, S., Moyers, B. T., Lai, Z., Kane, N. C., ... Rieseberg, L. H. (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications*, 4(May). doi: 10.1038/ncomms2833
- Rose, C. G., Paynter, K. T., & Hare, M. P. (2006). Isolation by distance in the eastern oyster, *Crassostrea virginica*, in Chesapeake Bay. *Journal of Heredity*, 97(2), 158–170. doi: 10.1093/jhered/esj019
- Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution*, 68(1), 1–15. doi: 10.1111/evo.12258
- 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. doi: 10.1111/jeb.13442

- Sobel, J. M., & Streisfeld, M. A. (2015). Strong premating reproductive isolation drives incipient speciation in *Mimulus aurantiacus*. *Evolution*, 69(2), 447–461. doi: 10.1111/evo.12589
- 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. In *PLoS Biology* (Vol. 17). doi: 10.1371/journal.pbio.3000391
- Stankowski, S., Sobel, J. M., & Streisfeld, M. A. (2015). The geography of divergence with gene flow facilitates multitrait adaptation and the evolution of pollinator isolation in *Mimulus aurantiacus*. *Evolution*, 69(12), 3054–3068. doi: 10.1111/evo.12807
- 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. doi: 10.1111/mec.13645
- 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). doi: 10.1098/rspb.2015.1666
- 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. doi: 10.1111/j.1420-9101.2006.01216.x
- Streisfeld, Matthew A., & Kohn, J. R. (2005). Contrasting Patterns of Floral and Molecular Variation Across a Cline in *Mimulus aurantiacus*. *Evolution*, 59(12), 2548. doi: 10.1554/05-514.1
- Streisfeld, Matthew 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). doi: 10.1371/journal.pgen.1003385
- Thompson, D. M. (2005). Systematics of *Mimulus* Subgenus *Schizoplacus* (Scrophulariaceae). *Systematic Botany Monographs*, 75, 1–213. Retrieved from <http://www.jstor.org/stable/25027942>
- Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, 3(9), 1572–1578. doi: 10.1371/journal.pbio.0030285

- Vijay, N., Bossu, C. M., Poelstra, J. W., Weissensteiner, M. H., Suh, A., Kryukov, A. P., & Wolf, J. B. W. (2016). Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature Communications*, 7, 1–10. doi: 10.1038/ncomms13195
- Wang, T., Hamann, A., Spittlehouse, D. L., & Murdock, T. Q. (2012). ClimateWNA-high-resolution spatial climate data for western North America. *Journal of Applied Meteorology and Climatology*, 51(1), 16–29. doi: 10.1175/JAMC-D-11-043.1
- Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, 18(2), 87–100. doi: 10.1038/nrg.2016.133