PHYLOGENETICS AND THE GENOMIC CONSEQUENCES OF DIVERGENCE IN A RADATION OF MONKEYFLOWERS

by

MADELINE A. CHASE

A THESIS

Presented to the Department of Biology
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Master of Science

March 2018

THESIS APPROVAL PAGE

Student: Madeline A. Chase

Title: Phylogenetics and the Genomic Consequences of Divergence in a Radiation of Monkeyflowers

This thesis has been accepted and approved in partial fulfillment of the requirements for the Master of Science degree in the Department of Biology by:

Bill Cresko	Chairperson
Peter Ralph	Co-Chair
Matt Streisfeld	Advisor

and

Sara D. Hodges	Interim Vice Provost and Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded March 2018

THESIS ABSTRACT

Madeline A. Chase

Master of Science

Department of Biology

March 2018

Title: Phylogenetics and the Genomic Consequences of Divergence in a Radiation of Monkeyflowers

Understanding the forces that drive divergence, and the genomic consequences of this process, is a central goal in evolutionary biology. Evolutionary radiations provide excellent opportunities to study speciation, as taxa span a continuum of divergence. However, inherent features of radiations resulting from rapid diversification create challenges for inferring evolutionary history. In this thesis, I address outstanding questions relating to the process of divergence in a diverse radiation of monkeyflower, the *Mimulus aurantiacus* species complex. I first use reduced representation sequencing to infer evolutionary relationships, examine patterns of phenotypic evolution among taxa, and assess previous taxonomic treatments. I then employ whole genome sequencing of samples from across the radiation to examine phylogenetic discordance and to understand what forces shape patterns of genome-wide variation among taxa. This work furthers our understanding of factors that drive diversification and the genomic consequences of divergence.

This thesis includes previously published and unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Madeline A Chase

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene

DEGREES AWARDED:

Master of Science, Biology, 2018, University of Oregon
Bachelor of Arts, Biology, 2015, University of Oregon

AREAS OF SPECIAL INTEREST:

Ecology and Evolutionary Biology

PROFESSIONAL EXPERIENCE:

Graduate Teaching Fellow, University of Oregon, September 2015 – March 2018

PUBLICATIONS:

Chase, M.A., Stankowski, S. and 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), pp.1510-1521.

ACKNOWLEDGMENTS

I would like to sincerely express my appreciation to Matt Streisfeld and Sean Stankowski for their assistance in the preparation of this work and support throughout the process. Many thanks to Bill Cresko and Peter Ralph for their thoughts and feedback. I would also like to acknowledge Allison Fuiten for performing the genome annotation used in Chapter 3.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I
INTRODUCTION

Understanding how organisms diverge and diversify during the speciation process is a central goal in evolutionary biology. The remarkable power of speciation is apparent when we consider evolutionary radiations – groups of taxa that display rapid bursts of diversification from a single common ancestor and that are accompanied by extensive phenotypic or ecological divergence (Schluter 2000). Because taxa in radiations span a continuum of divergence, they offer excellent opportunities to dissect the evolutionary processes driving diversification and the genomic consequences of lineage splitting.

Radiations can provide amazing insights into the power of natural selection and adaptation, however, inferring the evolutionary history of divergence remains a considerable challenge. For example, incomplete lineage sorting and ongoing gene flow across porous species boundaries (i.e., introgression) complicate the reconstruction of evolutionary relationships but are natural outcomes associated with the rapid timing between speciation events that defines radiations (Heliconius Genome Consortium 2012; Lamichhaney et al. 2015; Mallet et al. 2016; Pease et al. 2016; Wallbank et al. 2016). As a consequence, relatedness among taxa varies across the genome, which leads to genomic regions with diverse histories that do not match the species branching order (Fontaine et al. 2012; Gante et al. 2016; Richards and Martin 2017). The complex evolutionary history, combined with the common features of phenotypic convergence and divergence (Schluter 2000; Berner and Salzburger 2015), means that phylogenetic relatedness may be of little predictive power for determining patterns of phenotypic evolution. Thus, taxonomic treatments of radiations may be incomplete if they rely solely on phenotypic information to assign classifications, and efforts to understand the history of phenotypic evolution may be misled if interpreted only in the light of a species tree (Hahn and Nakhleh 2016). Therefore, genome-wide analyses are necessary to infer evolutionary relationships when these conflicts are common.

Until recently, DNA sequences from only a handful of genes were available to infer relationships among taxa. However, advances in sequencing technology over the past decade now provide access to data from across entire genomes, even for non-model organisms. This increase in data availability has allowed for unprecedented resolution of

evolutionary relationships across many radiations (Emerson et al. 2010; Heliconius Genome Consortium 2012; Wagner et al. 2013; Eaton and Ree 2013; McCluskey and Postlethwait 2015; Pease et al. 2016; Wessinger et al. 2016). In addition, whole genome sequence data now allow us to address many outstanding questions related to the process of divergence, including: how does lineage sorting vary across the genome? What processes are responsible for shaping patterns of genetic variation across lineages? And, how do levels of differentiation across the genome evolve?

One of the major open areas in the field of evolutionary genomics is to explain the processes that are responsible for generating patterns of genome-wide differentiation between taxa. A frequent observation is that genetic differentiation does not evolve in a uniform fashion across the genome (Turner et al. 2005; Hohenlohe et al. 2010; Ellegren et al. 2012; Martin et al. 2013; Soria-Carrasco 2014; Poelstra et a. 2014; Renaut et al. 2013; Lamichhaney et al. 2015; Malinsky et al. 2015; Vijay et al. 2016). Instead, a heterogeneous pattern of peaks and valleys is almost always observed between taxa. Peaks of elevated differentiation were originally referred to as "islands of speciation," as it was assumed that they harbored loci directly involved in reproductive isolation, while the rest of the genome was homogenized through gene flow (Wu et al. 2001; Turner et al. 2005). The simplicity of this interpretation, the ease of testing it with genome scans, and the fact that islands are almost always found, have resulted in numerous conclusions that speciation occurred in the face of gene flow (Turner et al. 2005; Nosil 2008; Ellegren et al. 2012; Nadeau et al. 2012; Poelstra et al. 2014).

However, despite the excitement that it might be possible to identify the genomic regions controlling reproductive isolation, recent work has demonstrated that heterogeneity in measures of differentiation and divergence had little or nothing to do with the process of speciation *per se*. Rather, they noted that peaks of differentiation were the inevitable outcome of variation in intrinsic features of the genome that shaped patterns of nucleotide diversity across the genome (Cruickshank and Hahn 2014; Burri et al. 2015; Van Doren et al. 2017; Vijay et al. 2017). Support for this interpretation comes from the observation that genome-wide patterns of differentiation, divergence, and diversity (hereafter referred to as genomic landscapes) were strongly correlated among different taxonomic comparisons. As these species tend to live in different habitats and/or

have been diverged for long periods of time (millions of years in some cases), correlated peaks across multiple comparisons are unlikely to have evolved via repeated instances of divergent selection. Thus, other explanations for these correlations have been proposed, including the long-term action of purifying selection against deleterious mutations that removed linked neutral variation (i.e. linked background selection). This model has been supported further by the fact that levels of differentiation, divergence, and diversity were correlated not only with each other, but also with variation in intrinsic features of the genome that impact the extent of background selection, such as gene density and recombination rate. Although background selection has been shown to shape patterns of diversity and differentiation in many taxa, the available data are taxonomically limited, making it unclear whether this phenomenon is common across diverse forms of life.

In this thesis, I examine the evolutionary genomic processes that have contributed to divergence within a recent radiation in a group of wildflowers commonly known as the bush monkeyflowers (*Mimulus aurantiacus* species complex). The radiation lies within the genus *Mimulus*, which itself represents a radiation across western North America, with around 100 species that vary in life history strategies, mating systems, and habitat types (Wu et al. 2008; Twyford et al. 2015). Specifically, the section Diplacus that contains the bush monkeyflowers is a group of phenotypically diverse perennial shrubs distributed across semi-arid chaparral habitats, deserts, higher elevations, and islands in California. Due to extensive hybridization in regions where taxonomic boundaries overlap, there has been considerable controversy among taxonomists on how to delineate this diversity, with some treatments describing as many as thirteen species but others that recognized only two species. The radiation spans a broad range of divergence times and thus provides an excellent opportunity to study the impact of varying evolutionary forces at different stages of speciation.

In Chapter 2, I use reduced representation sequencing approaches to determine evolutionary relationships among taxa, examine patterns of phenotypic evolution, and discuss issues surrounding the group's taxonomy. The work in this chapter has been previously published with the co-authors Sean Stankowski and Matthew Streisfeld. In Chapter 3, I then employ whole genome sequencing from samples across the radiation to investigate the factors that shape patterns of genetic variation and the evolution of

genomic landscapes of differentiation. This chapter includes unpublished work with the co-authors Sean Stankowski, Allison Fuiten, and Matthew Streisfeld. Ultimately, this work furthers our understanding of the evolutionary processes that are responsible for driving divergence in a diverse radiation, and reveals the evolution of the genomic landscape primarily occurs due to factors unrelated to the processes of adaptation and speciation.

CHAPTER II

GENOMEWIDE VARIATION PROVIDES INSIGHT INTO EVOLUTIONARY RELATIONSHIPS IN A MONKEYFLOWER SPECIES COMPLEX (*MIMULUS* SECT. *DIPLACUS*)

From Chase, M.A., Stankowski, S. and 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), pp.1510-1521.

INTRODUCTION

Evolutionary radiations provide excellent opportunities to study the processes that drive phenotypic divergence and speciation. However, because rapid divergence is a hallmark of radiations, past efforts to infer evolutionary relationships among their taxa often have been unsuccessful due to the low levels of sequence variation contained in one or a few genes (Qiu et al. 1999; Wolfe et al. 2006; Jarvis et al. 2014). Advances in sequencing technology have overcome this limitation by greatly expanding the amount of the genome that can be queried. For example, reduced representation techniques, like restriction site associated DNA sequencing (RADseq), can be used to obtain data from thousands of genomic regions that can be combined in a single analysis (Miller et al. 2007; Baird et al. 2008). This approach has allowed relationships to be resolved in some radiations for the first time (Emerson et al. 2010; Heliconius Genome Consortium 2012; Wagner et al. 2013; Eaton & Ree 2013; McCluskey and Postlethwait 2015; Wessinger et al. 2016; Pease et al. 2016).

Although genome-wide phylogenies provide an excellent framework for understanding the history of radiations, a single bifurcating topology also may obscure important details about the divergence process (Hahn & Nakhleh 2015, Mallet et al.

2016). This is because the genomes of recently-radiated taxa often are complex genealogical mosaics that have been shaped by a range of processes, including incomplete lineage sorting and introgressive hybridization (Mallet et al. 2016; Pease et al. 2016). Although these processes were once considered to generate noise that prevented the resolution of taxon-level relationships, recent studies have shown that they often are important sources of adaptive alleles that can drive speciation (Heliconius Genome Consortium 2012; Lamichhaney et al. 2015; Pease et al. 2016; Wallbank et al. 2016). Thus, a holistic understanding of relationships within radiations requires integrated approaches using both tree based and non-tree based analyses that can reveal both patterns of divergence and sources of shared variation among taxa.

These new insights from genomic data also have sparked discussion about the nature of species, which has important implications for how we choose to delineate them. Although reproductive isolation always has been the cornerstone of the biological species concept (BSC) (Mayr 1995), we now know that speciation is a continuous process, and radiations will contain taxa at different stages of divergence. During this process, reproductive barriers can remain highly porous for long periods of time (Rieseberg et al. 1999; Turner et al. 2005; Harrison and Larson 2014). Indeed, the genic view of speciation suggests that divergence occurs heterogeneously across the genome (Wu 2001), such that the loci that underlie isolating traits become differentiated before the rest of the genome (Turner et al. 2005; Ellegren et al. 2012; Malinsky et al. 2015; Vijay et al. 2016). Therefore, even taxa at intermediate levels of divergence typically continue to share alleles, leading to a "gray zone" where species concepts fail to reflect the realities of biological diversity (Mallet et al. 2016; Roux et al. 2016). These issues are paramount to the way that we consider patterns of taxonomic diversity in radiations, and they indicate the need for a more fluid, modern view of speciation that takes into account the continuous and multifaceted nature of the process.

In this study, we combine genomic and morphological data to shed light on evolutionary relationships within a recent radiation of monkeyflowers. *Mimulus* section *Diplacus* (Phrymaceae) is a monophyletic group of perennial shrubs distributed mainly in California (Beardsley et al. 2004). The phenotypically and ecologically diverse group (Fig 2.1A) consists of at most thirteen previously described taxa that are interfertile and

continue to hybridize in narrow areas where their geographic ranges overlap. Although evolutionary studies have focused primarily on divergence between two parapatrically distributed taxa in San Diego county (Streisfeld and Kohn 2005; Streisfeld and Kohn 2007; Sobel and Streisfeld 2015; Stankowski et al. 2015; Stankowski et al. 2017), little is known about the evolutionary history of divergence across the rest of the radiation. One reason for this is that the relationships among taxa remain unclear, as phylogenetic analyses have been limited to a handful of genes and included only some of the taxa (Beardsley et al. 2004; Stankowski and Streisfeld 2015).

In addition to an incomplete understanding of evolutionary relationships, taxonomists have struggled to describe the extensive phenotypic diversity within *Diplacus*. As a consequence, there have been 12 different taxonomic revisions over the past century (Fig 2.1B; Grant 1924; Munz 1935, 1959, 1973; McMinn 1951; Pennell 1951; Beeks 1962; Thompson 1993, 2005, 2012; Tulig 2000; Tulig and Nesom 2012). As few as two and as many as thirteen species have been described, and many of the treatments also recognize additional subspecies or varieties. For example, the two most recent taxonomies were both published in 2012, but they differ dramatically in how they delimit the taxa. Thompson (2012) recognized two species, one of which included six varieties. By contrast, Tulig and Nesom (2012) split this same variation into thirteen species, three of which were reported to be of hybrid origin. While much of the disagreement about the number and status of species results from the absence of intrinsic barriers to gene flow and the natural hybridization that occurs across their ranges (McMinn 1951; Beeks 1962; Streisfeld and Kohn 2005), these taxonomic conclusions were based entirely on phenotypic data. Therefore, integrating genomic data with this phenotypic information will allow for an explicit evaluation of these taxonomic hypotheses.

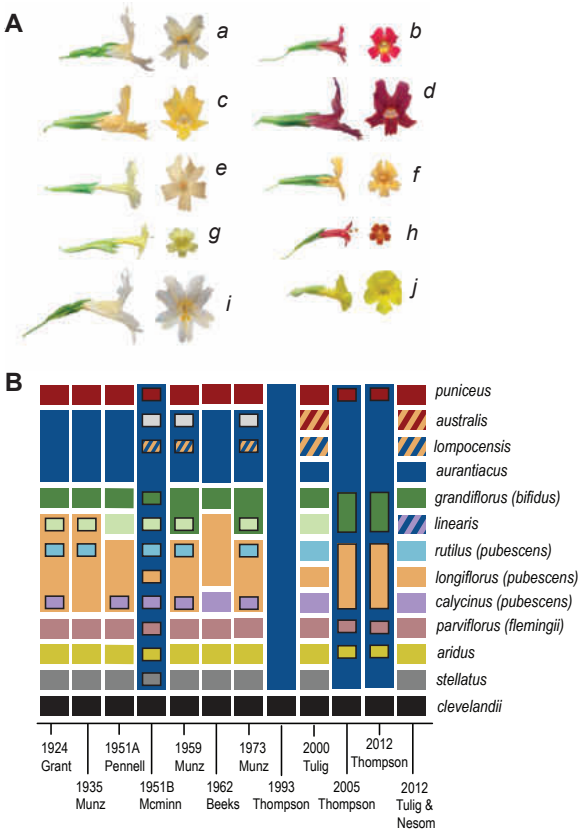


Figure 2.1. Morphological diversity and taxonomic history within *Mimulus* sect. *Diplacus*. (A) Representative photos in front and side view of some of the floral diversity present in *Mimulus* sect. *Diplacus.* Taxa included here are (a) *M. australis*, (b) *M. puniceus*, (c) *M. longiflorus*, (d) *M. rutilus*, (e) *M. calycinus*, (f) *M. aurantiacus*, (g) *M. aridus*, (h) *M. parviflorus*, (i) *M. grandiflorus,* and (j) *M. clevelandii*. (B) A summary of the 12 taxonomic revisions that have been published over the past century, beginning with Grant (1924). Across the different treatments, species status is represented by colored rectangles, and subspecies or variety status is represented by smaller rectangles with black outlines that occur within the colored rectangle for a species. The color of the box is associated with the name given by Tulig and Nesom (2012), presented to the right of the figure. Other names previously used to define taxa are included in parentheses. The location of the taxon names lines up with their treatment in each taxonomy. Hatched boxes indicate that a taxon is described as a hybrid species, with the color of the two lines representing the proposed progenitor species.

In this study, we use a combination of phylogenetic and population genomic approaches to elucidate the evolutionary history and patterns of shared variation among taxa in section *Diplacus*. In addition, we combine phylogenomic and morphological data from a nearly complete sampling of taxa to explore patterns of phenotypic evolution across the group. In doing so, we provide a critical assessment of previously published

taxonomic hypotheses in the light of new genomic analyses. This work will inform conservation and management practices, and it provides a framework for future taxonomic treatments of this group. Finally, this work creates new opportunities for comparative evolutionary, ecological, and genomic studies of the history of divergence in this species complex.

METHODS

Study system

Members of *Mimulus* section *Diplacus* are perennial shrubs that vary most notably in floral characteristics (Fig. 2.1A). They occur throughout semi-arid regions of California, including most coastal sage scrub and inland chaparral communities, as well as some mountain peaks and deserts (Beeks 1962). Hummingbirds and insects are their primary pollinators (Grant 1994), and their preferences have been suggested to play an important role in the divergence of some taxa (Grant 1993; Streisfeld and Kohn 2007). Intrinsic crossing barriers appear to be absent among all taxa, with the exception that crosses involving *M. clevelandii* Brandegee frequently were unsuccessful (McMinn 1951). This suggests that significant reproductive isolation exists between *M. clevelandii* and other members of the group. Consistent with this observation, all previous taxonomies recognize *M. clevelandii* as a separate species.

By contrast, there has been little consensus about the ranks of other taxa (Fig 2.1B). With the exception of Thompson (1993; 2005; 2012), who treated most taxa as varieties of the species *M. aurantiacus* Curtis, all other treatments consistently recognized six species (*M. aridus* Abrams*, M. parviflorus* Greene, *M. puniceus* Nutt.*, M. longiflorus* Nutt.*, M. grandiflorus* Groenland*,* and *M. aurantiacus*; Fig 2.1B). Although *M. stellatus* Kellogg is also treated consistently as a species, it has not been collected since 1940 (McMinn 1951), and most taxonomists make no mention of it other than noting it was recognized as a species by Grant (1924). As a consequence, we do not consider *M. stellatus* further in this study. The remaining taxa have been more controversial. For example, *M. calycinus* Eastw. and *M. rutilus* A.L. Grant have been described either as separate species (McMinn 1951; Beeks 1962; Tulig 2000; Tulig and Nesom 2012) or as subspecies of *M. longiflorus* (Grant 1924; Munz 1935; Pennell 1951;

Munz 1959; Munz 1973). In addition, *M. linearis* Benth. has been described as a distinct species (Pennell 1951; McMinn 1951; Tulig 2000), a subspecies of *M. longiflorus* (Grant 1924; Munz 1935), a subspecies of *M. grandiflorus* (Munz 1959; Munz 1973), and as a species of hybrid origin between *M. aurantiacus* and *M. calycinus* (Tulig and Nesom 2012). Different treatments also have recognized *M. lompocensis* McMinn as both a subspecies of *M. aurantiacus* (Munz 1959; Munz 1973) and as a species of hybrid origin between *M. aurantiacus* and *M. longiflorus* (McMinn 1951; Tulig 2000; Tulig and Nesom 2012). Finally, *M. australis* McMinn ex Munz has been described as its own species (McMinn 1951), a subspecies of *M. aurantiacus* (Munz 1959; Munz 1973), and most recently, being of hybrid origin between *M. puniceus* and *M. longiflorus* (Tulig 2000; Tulig and Nesom 2012). Each of these taxa has also at some point been considered a synonym of another, less controversial taxon (Fig 2.1B). Due to this extreme confusion over naming conventions, we choose to be as inclusive as possible with the taxonomy by addressing every previously described taxon without regard to species concepts. Therefore, unless otherwise noted, we refer to each taxon using its specific binomial epithet, according to Tulig and Nesom's (2012) treatment.

Taxonomic and population sampling

A recently published analysis of phylogenetic relationships included the eight most widely distributed taxa but avoided some of the more controversial groups (i.e., *M. lompocensis, M. linearis, M. rutilus*) (Stankowski and Streisfeld 2015). Additionally, some taxa were sampled across a limited portion of their geographic range (i.e. *M. puniceus* and *M. australis*). We include samples of these taxa here to provide a more complete examination of the group. Thus, our analyses included individuals from 12 taxa from section *Diplacus* and one outgroup species (*M. kelloggii* Curran, which is sister to *Diplacus* in section *Oenoe*; Beardsley et al. 2004).

Leaf tissue was collected either from the field or from field-collected seeds grown in the University of Oregon greenhouses. For ingroup taxa, samples included between one and fourteen individuals across the taxon's geographic range, totaling 73 individuals (Fig 2.2; Appendix A). One individual from the outgroup species *M. kelloggii* also was included. Samples were identified according to Tulig and Nesom (2012). Sixty-one of the

seventy-three ingroup individuals were sequenced previously (Stankowski and Streisfeld 2015); the new individuals included here are two *M. lompocensis*, four *M. rutilus*, one *M. parviflorus*, one *M. aridus*, and two additional *M. puniceus* and *M. australis* from the northern portion of their range.

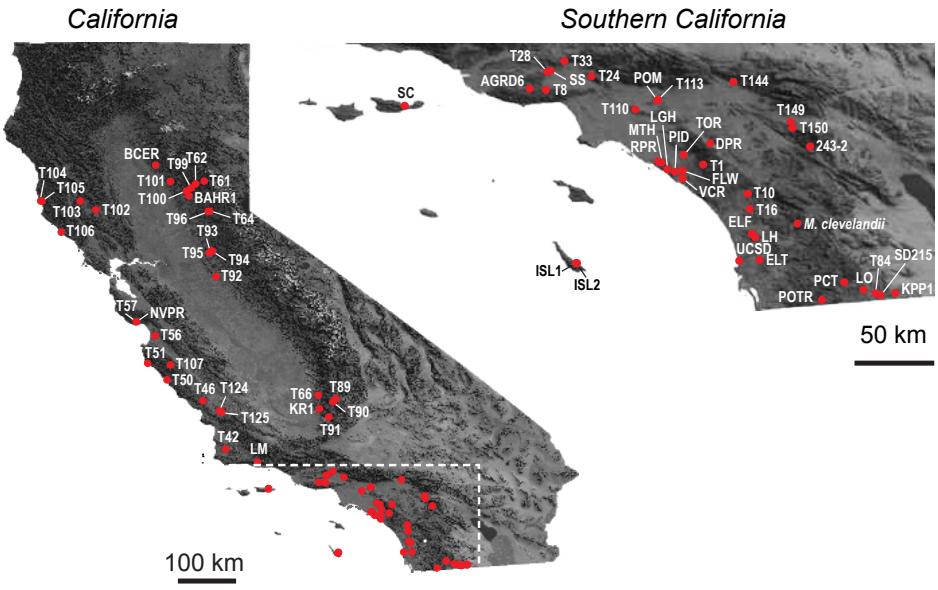


Figure 2.2. Geographic range and distribution of sampled individuals. Red dots represent the sampling locations used in this study. Population codes that begin with the letter "T" followed by a number indicate populations that were sampled previously for floral trait data by Tulig (2000). The region depicted by the inset in southern California is shown by the dashed line.

Forty-five of our samples come from locations that were previously visited by Tulig (2000) in a study of floral trait variation. Therefore, morphological data are available for these populations (described below). Although hybridization is known to occur between some taxa, we avoided sampling from zones of contact, because we did not want our analysis of broadly distributed taxa to be impacted by dynamics in narrow hybrid zones (except in the case of *M. rutilus*, which only occurs within populations described as *M. longiflorus*).

Analyses of evolutionary relationships

To generate genome-wide data to infer the evolutionary history of section *Diplacus*, we used Illumina sequencing of restriction site associated DNA tags (RADtags). DNA was isolated using a modified CTAB extraction described by Sobel and Streisfeld (2015), or ZYMO Plant/Seed DNA miniprep kits. RAD libraries were then prepared using the *PstI* restriction enzyme, followed by single-end 100-bp Illumina Hiseq 2000 sequencing, according to methods described previously (Etter et al. 2011; Sobel and Streisfeld 2015). We used the *process_radtags* module of the *Stacks* v. 1.35 package (Catchen et al. 2013) to remove reads with low quality or uncalled bases. Errors in the barcode and restriction site sequences were corrected prior to downstream analysis. Reads were aligned to an initial draft reference assembly from *M. puniceus* (described in Stankowski et al. 2017) using the *very-sensitive* settings of *bowtie2* (Langmead and Salzberg 2012). Loci were constructed with the *ref_map.pl* script in *Stacks* v. 1.35 (Catchen et al. 2011; Catchen et. al. 2013). Single nucleotide polymorphisms (SNPs) were identified for our phylogenetic analysis using the *populations* module in *Stacks* v. 1.35, requiring that SNPs were present in at least 90% of the individuals and had a minimum minor allele frequency of 0.02 to exclude any SNPs found in a single heterozygote.

To infer relationships among samples, we used a maximum-likelihood method, implemented in RAxML v. 8.2.3 (Stamatakis 2014). For each sample, we generated an alignment of all 24,699 polymorphic 95-bp RAD-tags, which included invariant sites (specified using the *--phylip_var_all flag* in *populations*). Methods of phylogenetic reconstruction were developed for the use of sequence data that include invariant sites; therefore, using whole RAD-tags is more appropriate than including only polymorphic sites (Stamatakis 2014). *RAxML* was run using the GTR+GAMMA model of nucleotide substitution. Support for each node was obtained by running 100 bootstrap replicates. Previous analyses using Bayesian, distance, and coalescent-based approaches yielded qualitatively similar results (Stankowski and Streisfeld 2015).

Closely related populations often share high levels of sequence variation as a result of both incomplete lineage sorting (ILS) and ongoing hybridization (Lamichhaney et al. 2015; Mallet et al. 2016; Hahn and Nakhleh 2016; Pease et al. 2016). Therefore,

forcing individuals to conform to a bifurcating tree may obscure the complex evolutionary history of a group (Huson and Bryant 2006). We therefore also constructed a split network using the program *SplitsTree* v4 (Huson and Bryant 2006). This method allows us to visualize more complex signals in the data by adding splits that are not permitted in a bifurcating tree.

We then used the Bayesian clustering algorithm implemented in *Structure 2.3.4* (Pritchard et al. 2000) as an alternative method for inferring patterns of ancestry within *Diplacus*. Unlike phylogenetic methods, *Structure* analysis reveals shared variation among inferred genetic groups, which could result from admixture or ancestral polymorphism. Because the phylogenetic analysis revealed four major clades, we conducted six replicate runs at $K = 4$, assuming the admixture model, correlated allele frequencies, 50,000 iterations of burn-in and 200,000 iterations of sampling. Additional runs were added with subsets of the individuals to address hypotheses that emerged from the phylogenetic analysis (see below). *M. clevelandii* and *M. kelloggii* were not included in these analyses. Due to computational limitations, we used a reduced dataset of 6095 SNPs, generated by including one SNP per RAD-tag and a minimum minor allele frequency of 0.15. Results from each run were evaluated using Structure Harvester (Earl 2012), and multiple runs were summarized in CLUMPP (Jakobssen and Rosenberg 2007).

Tests for introgressive hybridization

Of the four primary clades identified in the phylogenetic analysis, one clade (Clade D) was especially diverse and contained up to six described taxa from southern California. Although three sub-clades are evident in the phylogenetic analysis, *Structure* revealed substantial levels of shared variation among the sub-clades. To investigate whether this shared variation reflects ancestral polymorphism or recent gene flow, we calculated Patterson's D statistic (Green et al. 2010). Patterson's D is calculated using four taxa with the relationship $(((P_1, P_2), P_3), O)$ and provides a test for introgression between the donor population, P_3, and either of the two ingroup taxa, P_1 and P_2 (Green et al. 2010). The statistic is calculated as the ratio of SNPs that fits an ABBA pattern to the number of SNPs that fits a BABA pattern across the four taxa, where A is the ancestral

allele and B is the derived allele. Under random sorting of ancestral variation, the number of SNPs fitting both patterns is expected to be equal; however, an excess of either pattern indicates introgression has occurred between the donor taxon and one of the ingroup taxa (Green et al. 2010). The taxa used in this dataset include *M. longiflorus* (P1), *M. calycinus* (P2), *M. australis* and *M. puniceus* (P3), and *M. grandiflorus* (O). *M. australis* and *M. puniceus* were combined to form P3, because the two formed a single phylogenetic group (see results). Two *M. calycinus* individuals that grouped in Clade C and showed high levels of admixture (see results) were not included in this analysis. SNPs included in this dataset were required to be present in 90% of the individuals included, and to have a minimum minor allele frequency of 0.02. To reduce the effects of linkage, we included only a single SNP per RADtag. We did not require SNPs to be fixed within a taxon, and 16,920 polymorphic sites were included in the analysis. To assess if D was significantly different from 0, we followed the approach of Eaton and Ree (2013) to calculate a *p*-value from the *Z*-score obtained from 1000 bootstrap replicates of the test statistic.

In addition to exploring the origins of shared variation revealed by our analyses, we tested Tulig's (2000) hypothesis that *M. australis* is of hybrid origin between *M. puniceus* and *M. longiflorus*. Tulig (2000) used multivariate analysis of floral traits (see below) to identify two groups that differed in flower size. The small-flowered group included *M. puniceus, M. aurantiacus*, and *M. parviflorus*, while the large-flowered group included *M. grandiflorus, M. longiflorus,* and *M. calycinus*. Populations of *M. australis* were intermediate in size and overlapped with populations of *M. lompocensis*, which McMinn (1951) previously suggested was of hybrid origin. Tulig (2000) took these patterns to be evidence that *M. australis* also is of hybrid origin. Tulig and Nesom (2012) further speculated that, based on their geographic range and phenotypic similarity, *M. longiflorus* and *M. puniceus* were likely to be the progenitors of modern day *M. australis*.

We used the F3 test (Reich et al. 2009) to ask whether there was genomic evidence that *M. australis* arose through hybridization between *M. longiflorus* and *M. puniceus*. The F3 test compares three populations, X, Y, and W, and evaluates whether Y is of mixed ancestry between X and W. The test is calculated by measuring the allele

frequency difference between Y and either X or W, and taking the product of the two values. In this case, we combined all *M. australis* individuals as population Y, and *M. puniceus* and *M. longiflorus* were populations X and W, respectively. A negative value would support the hybrid origin of *M. australis* and a non-negative value would refute it. The F3 statistic was calculated from the dataset used to calculate Patterson's D, with 3204 sites polymorphic among these three taxa. We applied the same bootstrap approach as described above to determine the significance of the observed F3 statistic. The F3 and Patterson's D statistics were calculated using ADMIXTOOLS (Patterson et al. 2012).

Analyses of floral trait data

Previous studies used morphological characteristics — mainly floral traits — to delimit taxa in *Diplacus* (Pennell 1951; McMinn 1951; Beeks 1962; Munz 1973; Tulig and Nesom 2012; Thompson 2012). However, given that 12 revisions have been published over the past century, a critical assessment of the taxonomic utility of floral traits is warranted. Indeed, the two most recent treatments differ considerably in how taxa are delimited (Thompson 2012; Tulig and Nesom 2012; Fig. 2.1B). Therefore, we used an existing morphometric dataset that consisted of 18 floral traits that were measured on 1-30 plants from 45 of our collection sites (mean = 6 plants per site; SD = 4.17; Tulig 2000; trait descriptions provided in Appendix B) to ask how well each of these treatments performed at delineating taxa. We performed separate discriminant function analyses (DFA) with either Tulig and Nesom's (2012) taxonomy or Thompson's (2012) taxonomy as the grouping variable. If morphological characteristics alone can be used to delineate taxa, we would expect that one of the treatments would assign individuals to taxa more reliably than the other.

In addition, by combining floral trait data with phylogenetic and population genomic analysis, we now have the capacity to test whether trait variation can be used to reconstruct an accurate picture of evolutionary relationships. Specifically, if traits have strong phylogenetic signal, individuals within the same clade should be more phenotypically similar than individuals in different clades. However, this relationship may be obscured by the effects of convergent and divergent phenotypic evolution, which are common during radiations (Berner and Salzburger 2015). For example, we would

expect convergent evolution to result in phenotypic overlap among taxa from different clades, while divergence would cause pronounced phenotypic differences among taxa within clades. Therefore, to examine how trait variation is partitioned within and among clades, we summarized the multivariate trait data using a principal components analysis (PCA) and mapped the four phylogenetic clades in the bivariate space of the first two principal components. This approach separates samples based on the two largest sources of phenotypic variation across the entire dataset. Therefore, if these trait data reflect the evolutionary history of divergence, the first two principal components should correspond to the deepest evolutionary divisions in the group. However, convergent and divergent phenotypic evolution would prevent the accurate reconstruction of evolutionary history from these traits.

Finally, although it may be possible that the primary sources of floral trait variation fail to reflect the history of this group, there may be more subtle trait variation that does carry a phylogenetic signal. To test for such traits, we used DFA with phylogenetic clade as the grouping variable to examine how often individuals were assigned to clade using the 18 floral traits. If this analysis reliably assigns individuals to clade, then we can identify the traits that vary in accordance with the main evolutionary history of the group. All analyses of floral trait data were performed in *R* (R Core Team 2015).

RESULTS

Evolutionary relationships within section *Diplacus*

After quality filtering and aligning raw reads to the *M. puniceus* reference genome, an average of 69.8% of reads mapped uniquely when excluding *M. kelloggii*. The high percentage of reads aligning across taxa reflects the recent history of the group. In contrast, only 37.61% of the *M. kelloggii* reads aligned uniquely, as it is more distantly related to *M. puniceus*. The final dataset for phylogenetic analysis included 24,699 loci (RAD-tags), totaling 2,346,405 bp, with 68,889 variable sites. Of these loci, 38.3% were missing from *M. kelloggii.* After further filtering for *Structure* analyses, we retained 6,095 of the most informative SNPs.

Phylogenetic analysis revealed four highly supported (100% bootstrap support) clades (A-D; Fig. 2.3A), consistent with those previously identified in Stankowski and Streisfeld (2015). Clade A consists of all samples of *M. grandiflorus* and *M. linearis*, which are reciprocally monophyletic. Clade B includes all samples of *M. aridus* and *M. parviflorus*, which are reciprocally monophyletic. Clade C contains *M. aurantiacus* as a monophyletic group, but also includes one *M. lompocensis* individual and two individuals that phenotypically resemble *M. calycinus*. Clade D contains the highest number of described taxa, including *M. australis, M. puniceus, M. rutilus, M. longiflorus*, the remaining *M. calycinus* samples, and one *M. lompocensis* individual.

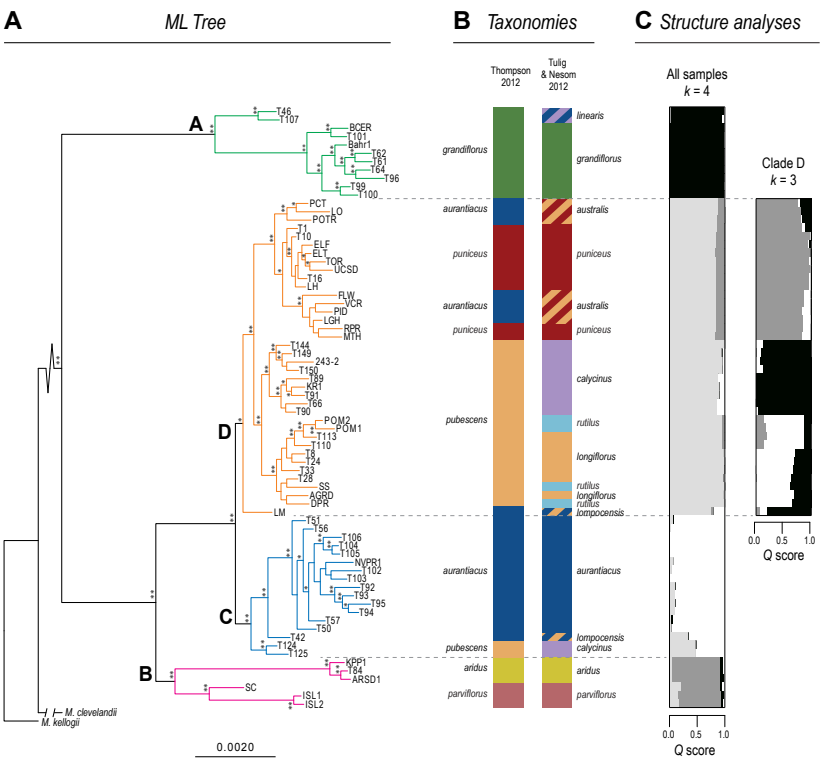


Figure 2.3. Evolutionary relationships within *Mimulus* sect. *Diplacus*. (A) Maximum-likelihood tree illustrating the relationships among samples. Nodes with 100% bootstrap support are represented with two stars, and nodes with >90% bootstrap support are represented with one star. We identified four highly supported clades (A-D), which are labeled on the tree, with branches colored according to clade. (B) Taxonomic identity of samples that is based on the two most recent taxonomies. Colors in the rectangles correspond with Figure 1. Taxon names colored in red represent populations that contain red-flowered individuals. (C) Structure analyses that show ancestry scores (*Q*) for all samples at *K* = 4 (left), and only for Clade D at *K* = 3 (right). Individuals line up with tips of the phylogeny. Dashed lines separate the four clades.

By including *M. kelloggii* as an outgroup, we were able to test the phylogenetic position of *M. clevelandii*. Although *M. clevelandii* frequently has been described as a separate species (Fig. 2.1B), and was used as the outgroup in a past analysis (Stankowski and Streisfeld 2015), the only previous molecular phylogenetic analysis revealed that *M. clevelandii* grouped within the rest of the taxa (Beardsley et al. 2004). However, by rooting with *M. kelloggii*, we confirm that *M. clevelandii* indeed is sister to the remaining taxa. In addition to the phylogenetic tree, we illustrated relationships through a split network (Appendix C). This analysis highlights the deep division that separates Clades A and B from Clades C and D, and it reveals the complex nature of the relationships among taxa, especially within the rapidly radiating Clades C and D.

Prevalence of shared variation between and within clades

As an alternative to phylogenetic analysis, we used *Structure* to infer patterns of admixture among the individuals. The analysis at $K = 4$ revealed clusters of individuals that largely agreed with the clades recovered in the phylogenetic analysis. However, it also revealed shared variation among Clades B, C, and D that was not apparent from the bifurcating tree (Fig. 2.3C). For example, we detected extensive admixture in the two *M. lompocensis* individuals and the two *M. calycinus* individuals that group within Clade C. The intermediate *Structure* scores of these *M. calycinus* individuals suggest that they are hybrids, so they were excluded from other analyses. In addition, consistent with previous evidence of introgressive hybridization (Stankowski and Streisfeld 2015), *M. puniceus* and *M. australis* from Clade D show some mixed ancestry with individuals in Clade B.

We performed an additional *Structure* analysis to test for divergence and admixture in Clade D. The *Structure* analysis at $K = 3$ recapitulated the three highly supported sub-clades from the phylogeny (one that includes both *M. puniceus* and *M. australis*, one that only includes *M. calycinus*, and one that includes *M. longiflorus* and *M. rutilus*). However, it also revealed extensive shared variation across all of Clade D (Fig. 2.3C). For example, southern populations of *M. calycinus* share some ancestry with individuals of *M. longiflorus*, and the four southern populations of *M. longiflorus* and *M. rutilus* share variation with individuals of *M. puniceus* and *M. australis.*

This shared variation may be due to retained ancestral polymorphisms or recent gene flow. In order to determine whether gene flow can explain some amount of shared variation among taxa, we performed an ABBA-BABA test within clade D. Because *M. puniceus* and *M. australis* are indistinguishable based on *Structure* analyses, we treated them as one taxon for this test. Moreover, since this test requires sister ingroup taxa, we used *M. calycinus* and *M. longiflorus* as the ingroups. Patterson's D for the test was 0.1672 (Table 2.1), which reflects a 16% excess of BABA sites over ABBA sites and suggests likely introgression between the *M. australis* and *M. puniceus* lineage and the *M. longiflorus* lineage. Bootstrap analysis revealed that the value of Patterson's D was highly significant (Table 1, $p < 0.00001$).

Table 2.1: Results from Patterson's D and F3 tests.

Test	Comparison	Test Statistic	Z-score	P-value	BABA Sites	ABBA Sites	Total Sites
D	P1: *M. longiflorus* P2: *M. calycinus* P3: *M. aus/pun* O: *M. grandiflorus*	0.1672	7.578	< 0.00001	196	140	16920
F3	X: *M. puniceus* Y: *M. australis* W: *M. longiflorus*	0.0244	6.135	< 0.00001	N/A	N/A	3204

To address Tulig and Nesom's (2012) claim that *M. australis* is a hybrid species, we also tested whether *M. australis* individuals are significantly admixed between *M. puniceus* and *M. longiflorus*. The result from the F3 test, designed to measure whether population Y is admixed between populations X and W, was 0.024408, which is a significantly positive value (Table 2.1, $p < 0.00001$). This result suggests that *M. australis* is not the product of hybridization between *M. puniceus* and *M. longiflorus*, as proposed by Tulig and Nesom (2012). Although it is possible that *M. australis* arose due to hybridization between other taxa, neither the split network nor the *Structure* analysis provides substantial evidence for admixture that would support this conclusion.

Patterns of phenotypic variation

Both discriminant function analyses reliably assigned individuals to the set of taxa described by each taxonomic treatment. Tulig and Nesom's (2012) treatment and Thompson's (2012) treatment each correctly assigned 97.06% of the individuals into taxa (Fig 2.4A-B). Based on Tulig and Nesom's (2012) treatment, the traits that loaded most highly on discriminant function 1 (DF1) were pedicel width (PDWD), corolla height (CRHT), and the length of the short filament (FSLN; Table 2.2). By contrast, width of the throat opening (THRO), calyx height (CAHT), and length of the lower central petal lobe (LLCL) loaded most strongly on discriminant function 2 (DF2). Based on Thompson's (2012) treatment, the traits that explained most of the variation on both DF1 and DF2 were pedicel width (PDWD), width of the throat opening (THRO), and length of the lower central petal lobe (LLCL). Thus, these traits appear to be most important in separating taxa. Notably, both taxonomies performed equally well at separating taxa based on this floral trait variation. Consequently, given the substantial differences between the taxonomic treatments, this analysis provides little guidance as to which taxonomy more accurately describes the diversity.

Table 2.2: Loadings for the first two discriminant function axes using the taxonomy of Tulig and Nesom (2012) or Thompson (2012) as the grouping variable. Descriptions of trait name abbreviations can be found in Appendix B.

	Tulig and Nesom (2012)		**Thompson (2012)**	
	DF1	DF2	DF1	DF2
CRLN	-0.05332177	0.05981519	-0.098544301	0.049198411
CULN	0.11349505	-0.027188252	-0.064752232	-0.178416174
CLLN	-0.11456936	-0.16278508	0.012548931	-0.061930311
BLLN	-0.09347819	0.196527236	-0.042122811	0.017305656
UCOS	0.08753006	-0.074278674	-0.107440042	-0.083783893
INFL	-0.07794273	-0.104483587	-0.133694509	-0.01175704
UCIS	-0.0581908	-0.067652583	-0.036843668	0.004904862
WLCL	-0.08329381	-0.044438801	0.000161317	-0.026132869
LLCL	0.09515379	0.383080129	0.461992819	0.447229467
THRO	-0.26309451	0.646242479	-0.384743181	0.683290511
CRHT	-0.56351185	0.213665212	0.145023746	-0.096769739
CAHT	0.09380583	-0.598988778	-0.229591581	-0.239920631
PDLN	0.07142001	0.040301165	0.100065766	0.149577015
PDWD	-1.461039	0.01673164	-0.44245656	-2.063862501
CTLN	-0.27308312	-0.329314029	-0.209082137	-0.092024316
FLLN	0.3154864	-0.000242331	0.333242237	-0.148756483
FSLN	0.35834455	0.092175164	0.131968689	0.153662132
STLN	-0.02811806	-0.00905903	0.080392938	0.116118715

To test whether floral trait variation can reconstruct evolutionary history, we performed a PCA with individuals colored by phylogenetic clade and performed a DFA with phylogenetic clade as the grouping variable. The first two principal components explained 84.9% of the variation among the 18 floral traits (PC1: 60.4%; PC2: 15.4%). However, rather than revealing a series of discrete groups, the samples were distributed along a continuum of phenotypic variation. In addition, there was almost no discrete clustering of samples from the same clade in PC space. Rather, individuals from different clades broadly overlapped one another. The only exception was Clade B, which was distinct from the other three clades and formed two clusters corresponding to *M. parviflorus* and *M. aridus*. This analysis indicates that the largest sources of phenotypic variation present in the dataset do not separate the samples into distinct groups that correspond to the deep evolutionary divisions revealed by the phylogenetic and population genomic analyses. Alternatively, substantial phenotypic overlap exists among the individuals from Clades A and D, and among individuals from Clades C and D, indicating convergence on similar phenotypes across clades. Moreover, the distinctness of the two taxa in Clade B, and the well-studied differences in floral traits between the closely related *M. puniceus* and *M. australis* (Streisfeld and Kohn 2005; Stankowski et al. 2015), reveal a complex history of phenotypic evolution in this group that involves both convergent evolution between clades and divergent evolution within clades.

Although the most conspicuous traits do not carry a phylogenetic signal, more subtle characters might distinguish the major clades from one another. To test for such traits, we conducted a discriminant function analysis using phylogenetic clade as the grouping variable. In contrast to the PCA, the individuals within each clade were largely separated from each other across discriminant space and were correctly assigned to clade 94.12% of the time (Fig 2.4D). Clade B is once again distinct from all other groups, but it no longer forms two separate clusters. Clades A and D are differentiated more clearly in discriminant space than in PC space, with only minor overlap between them. The greatest overlap occurred between Clades C and D, but there was less overlap evident than in the PCA. Pedicel width (PDWD) and the width of the throat opening (THRO) loaded most

heavily on these canonical axes (Table 2.3), suggesting that these traits carried the strongest phylogenetic signal based on the clustering of clades.

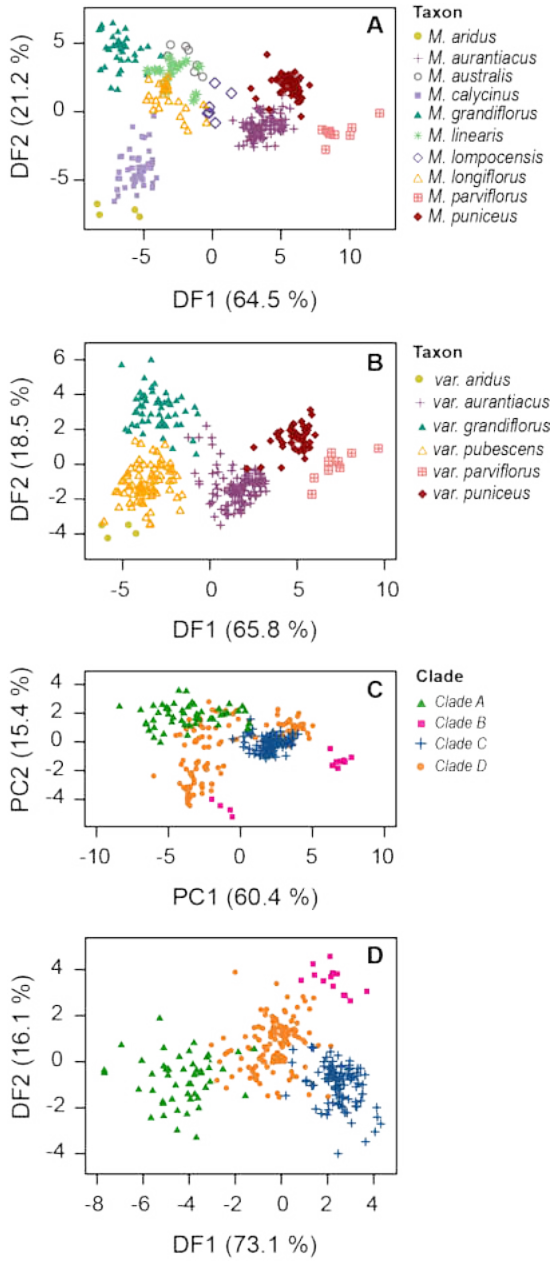


Figure 2.4. Phenotypic variation within section *Diplacus*. (A-B) Discriminant function analyses on 18 floral traits, based on individuals from the 45 populations that were previously phenotyped by Tulig (2000). Individuals within each plot are grouped according to the taxonomy of (A) Tulig and Nesom (2012) or (B) Thompson (2012). (C) Principal component analysis on floral traits, with individuals colored according to phylogenetic clade, as identified in Figure 3. (D) Discriminant function analysis on floral traits using phylogenetic clade as the grouping variable. Individuals are colored by clade. Based on these 18 traits, individuals were correctly assigned to Clade A 94% of the time, Clade B 100% of the time, Clade C 93.3% of the time, and Clade D 90.8% of the time.

Table 2.3: Loadings for the first two principal components and first two discriminant functions using phylogenetic clade as the grouping variable. Descriptions of trait name abbreviations can be found in Appendix B.

Trait	PC1	PC2	Clade DF1	Clade DF2
CRLN	0.87484	0.38394	-0.007534267	-0.23149174
CULN	0.64359	0.63356	0.133768577	0.03993585
CLLN	0.67048	0.68768	0.082474462	-0.03988584
BLLN	0.91132	-0.32478	-0.049643515	-0.0069482
UCOS	0.90306	-0.31785	0.112837875	-0.21298351
INFL	0.87773	-0.36545	0.065978965	-0.42316548
UCIS	0.83717	-0.42061	-0.129622483	0.32149581
WLCL	0.75442	-0.09692	0.050158241	-0.07255096
LLCL	0.85984	-0.38537	-0.412127261	0.40931716
THRO	0.87132	-0.34805	-0.791635928	0.09593173
CRHT	0.83665	-0.22986	0.026632214	-0.18093956
CAHT	0.71113	0.39742	0.203823105	-0.16578586
PDLN	-0.48729	-0.07971	-0.082918263	-0.07242255
PDWD	0.84532	0.0959	1.750621312	1.4659284
CTN	0.77697	0.53037	-0.019122676	0.20307343
FLLN	-0.67735	0.16934	0.121460739	0.44573151
FSLN	-0.74221	0.11235	0.083141571	-0.39595641
STLN	0.52842	0.6334	-0.132464288	0.16109519

DISCUSSION

Evolutionary relationships in radiations can be complex, and their resolution often requires detailed sampling and integrated analyses. By combining phylogenetic, population genomic, and phenotypic analysis, we show that these monkeyflowers exhibit the hallmarks of a rapid radiation, including a range of diverse taxa at different stages of divergence, extensive shared variation across the group, and evidence for divergent and convergent phenotypic evolution. Our results also have taxonomic implications, and we discuss how they might inform a future revision.

Patterns of divergence and shared variation across the radiation

Our analysis of genomic data provides some of the first insight into the evolutionary history of this diverse group of monkeyflowers. Specifically, we show that the taxa are closely related but are at different stages of divergence, which creates exciting opportunities for comparative studies across the speciation continuum. For example, in the early stages of speciation, the genomes of taxa are thought to be largely

undifferentiated as a result of their very recent history (Rundle and Nosil 2005; Nosil 2012). This is the case between *M. puniceus* and *M. australis* from clade D, which have divergent floral phenotypes as a consequence of pollinator-mediated selection (Streisfeld and Kohn 2007; Handlemann and Kohn 2014; Sobel and Streisfeld 2015), but do not form separate monophyletic groups in our phylogenetic analysis. In contrast, another pair of ecologically divergent taxa from Clade D, *M. calycinus* and *M. longiflorus*, form shallow monophyletic sister clades, suggesting they are at an intermediate stage of speciation (Beeks 1962; Grant 1993). A much more distantly related pair of taxa, *M. parviflorus* from Clade B and *M. longiflorus* from Clade D, are able to co-occur in sympatry on Santa Cruz Island off the coast of California despite hybridization between them (Wells 1980; M. Chase, personal observation). Future comparative, ecological, and genomic studies in these and other taxa will examine how the factors that generate and maintain diversity change with progress toward speciation.

While our phylogenetic analysis provides insight into the patterns of divergence between taxa, our population genomic analyses reveal a complex pattern of shared variation among taxa. Although incomplete lineage sorting probably accounts for most of the shared variation within and between clades, our analyses indicate that some is due to introgressive hybridization. Hybridization is a relatively common phenomenon in radiations, and in some cases, can be so extensive that relationships cannot be illustrated accurately with a tree (Malinsky et al. 2017). In *Mimulus* section *Diplacus*, many studies have noted hybridization between taxa in areas where their ranges overlap. Although this mixing has been a major cause of taxonomic conflict in this group (McMinn 1951; Beeks 1962; Thompson 1993, 2005, 2012; Tulig 2000), our data indicate that hybridization does not have a major effect on the core structure of clades and taxa. Rather, hybridization probably occurs in areas that coincide with transitions between different environments. This is consistent with the observation that floral trait differences between the taxa are stable over large geographic areas, and that hybrid zones are narrow in comparison.

Evidence for divergent and convergent phenotypic evolution

By examining floral trait variation in combination with phylogenetic analyses, we show striking phenotypic similarity between comparatively distantly related taxa, and

remarkable dissimilarity between very closely related taxa. This pattern of convergent and divergent evolution has been observed in many rapidly diverging groups (Muschick et al. 2012; Heliconius Genome Consortium 2012; Mahler et al. 2013) and is common in adaptive radiations (Schluter 2000; Berner and Salzburger 2015).

Multiple processes can cause the patterns of phenotypic evolution we observe. Divergent phenotypic evolution is thought to occur most commonly when populations adapt to contrasting environments, which can cause ecological isolating barriers to evolve (Rundle and Nosil 2005; Nosil 2012). Previous work indicates that this is the case between *M. puniceus* and *M. australis*, which show low levels of genomic differentiation despite selection on flower color and other floral traits (Streisfeld and Kohn 2005; Handelman and Kohn 2014; Stankowski et al. 2015). However, in other cases, sister taxa are geographically isolated from each other. Thus, phenotypic divergence may be the result of neutral processes rather than adaptation (Schluter 2009). For instance, *M. aridus* and *M. parviflorus* have entirely distinct ranges, but they differ in flower color (Fig. 1) and are completely separated from each other in the DFA of floral traits (Fig. 2.4). Given their allopatric distributions, drift or selection may have played a role in their phenotypic divergence. Thus, further study is required to determine the evolutionary forces responsible for phenotypic divergence across different taxa.

Convergent phenotypic evolution also may arise through various processes. Shared features among clades could result from independent origins of a trait through new mutations, the sharing of ancestral polymorphisms, or through introgressive hybridization. The latter two possibilities may be especially common in systems marked by rapid diversification (Hahn and Nahkleh 2016), and they signal the need for caution when interpreting phenotypic evolution in the context of a phylogeny. Although our phylogeny reflects the demographic history of divergence, there are likely regions of the genome with discordant evolutionary histories, some of which may underlie adaptive traits. Indeed, previous work in this system has provided evidence that a mutation causing red flowers was shared between Clade B and *M. puniceus* through historical introgression (Stankowski and Streisfeld 2015). Future analyses, aided by an improved, chromosome-level genome assembly, will allow us to reveal the underlying genomic features responsible for the patterns of divergent and convergent phenotypic evolution we

observe. In addition, these data will provide new opportunities to document the evolutionary history of ecologically important phenotypic transitions associated with adaptive divergence in a recent radiation.

Taxonomic implications and recommendations for formal revision

Although our primary focus was to infer the evolutionary relationships in this group, this work has important taxonomic implications and highlights the need for a new revision of section *Diplacus*. Over the last century, all 12 of the published treatments disagree to some extent on the appropriate number and rank of the taxa described. This is most apparent in the two most recent treatments that were published in 2012 (Tulig and Nesom 2012; Thompson 2012), as our results indicate that neither one is better at describing the floral trait variation analyzed here. This uncertainty is especially problematic for managers and conservationists, as well as for evolutionary biologists, who are left without a clear conceptual framework for how to appropriately refer to the diversity in the group. While we do not provide a formal revision here, we present recommendations for future changes that are based on the integration of genomic and phenotypic analyses that emerge from this study.

Five of the taxa that we examined have faced frequent revision, including *M. calycinus, M. rutilus, M. linearis, M. lompocensis,* and *M. australis.* Both *M. calycinus and M. rutilus* have been described previously as subspecies of *M. longiflorus,* and *M. calycinus* recently has been grouped together with *M. longiflorus* to form *M. aurantiacus* var. *pubescens* (Thompson 2005, 2012). Although the genomic and phenotypic data clearly separate *M. calycinus* from *M. longiflorus*, *M. rutilus* is not genetically distinct from *M. longiflorus*, even though they differ considerably in flower color (Fig. 2.1). Thus, based on these results, we would recommend that *M. calycinus* be treated as a distinct entity. However, given that red-flowered *M. rutilus* is found growing only within otherwise yellow-flowered populations of *M. longiflorus,* the genomic data suggest that *M. rutilus* should be recognized more appropriately as a simple flower color polymorphism that is restricted to a few geographic areas.

M. linearis has had many proposed evolutionary histories, including being a subspecies of either *M. longiflorus* or *M. grandiflorus*, as well as being a species of

hybrid origin between *M. aurantiacus* and *M. calycinus*. Our data reveal that even though *M. linearis* and *M. grandiflorus* are geographically distinct from each other, they emerge as sister taxa in the phylogeny and the split network, and there is little shared variation between *M. linearis* and taxa from other clades. Therefore, it remains unclear whether a future taxonomic revision should consider *M. linearis* to be its own entity or a form of *M. grandiflorus*, as proposed previously (Munz 1959; Munz 1973; Thompson 2005; Thompson 2012).

M. lompocensis has been described as a hybrid species between *M. aurantiacus* and *M. longiflorus* by several authors (McMinn 1959; Tulig 2000; Tulig and Nesom 2012). The two individuals included in this study grouped in different clades in the tree and split network (Clades C and D), and they showed high levels of admixture in the *Structure* analysis. While these results are consistent with a history of hybridization, it will be necessary to determine if *M. lompocensis* is ecologically distinct from its presumed progenitors (Gross and Rieseberg 2004) prior to concluding that this admixture reflects a stable taxon of hybrid origin (as in Tulig and Nesom 2012) rather than a product of recent natural hybridization.

Finally, *M. australis* has been described as a subspecies of *M. aurantiacus*, its own species, a species of hybrid origin, or in some treatments, *M. australis* has not been described at all (Grant 1924; Munz 1935; Pennell 1951; Beeks 1962; Thompson 1993; Thompson 2005; Thompson 2012). Based on the genomic data analyzed in the current study, *M. australis* is not distinguishable from *M. puniceus*, and the two are interdigitated in the phylogeny. In addition, populations described as *M. australis* show no evidence of being hybrids between *M. puniceus* and *M. longiflorus* (Table 2.1), as proposed previously by Tulig and Nesom (2012). Nevertheless, partial reproductive isolation has evolved between western red-flowered populations and eastern yellow-flowered populations (Sobel and Streisfeld 2015). Moreover, multiple floral and vegetative traits are differentiated along this same geographical transition (Stankowski et al. 2015; Sobel et al. *in revision*), indicating an early stage of ecological divergence between the taxa. Therefore, based on these data, we would not recommend that *M. puniceus* and *M. australis* be defined as distinct entities. However, even though no previous description of the red-flowered *M. puniceus* exists that also would include the yellow-flowered *M.*

australis, we suggest that future revisions incorporate these genomic and ecological patterns into a description that recognizes this divergence in the form of 'ecotypes' of the consistently recognized *M. puniceus*.

In addition to delimiting taxa, a new treatment also must consider the appropriate taxonomic rank for each entity. The difficulty of assigning ranks at or below the species level for this group has been recognized for a long time, as demonstrated by McMinn (1951) who wrote:

> ". . . since complete agreement has not been reached by botanists as to the status of species, subspecies, and varieties, I have chosen to treat all these field entities (taxa) simply as binomials. Inasmuch as binomials to most botanists indicate species, I have endeavored not to use the word species when writing of these various entities. I must point out, however, that if sterility and geographical distribution tests were the main criteria applied in delimiting species and subspecies, then the field entities . . . probably would be classified as two taxonomic species, eleven subspecies, and numerous hybrids."

Although McMinn (1951) ends by considering the BSC as one way to delimit taxa, this statement foreshadowed the need for integrative taxonomic approaches that recognize the different stages of divergence present among taxa in radiations. In most of the previous treatments of this group, the rank employed appears arbitrary and often was not justified by the authors. However, given the interfertility, natural hybridization, and shared genomic variation present among taxa, we support the view by McMinn (1951), and more recently by Thompson (2012), who treated the taxa (with the exception of *M. clevelandii*) as intraspecific subspecies or varieties of *M. aurantiacus*. This view, which acknowledges the reproductive continuity and close relationships among these taxa, emphasizes our need to understand how and why so much diversity arose and has been maintained within this group.

With a thorough understanding of the evolutionary relationships among taxa in this radiation, we are able to examine the genomic consequences of divergence. In the following chapter, I apply the framework that is developed here to understand the process of lineage formation in this group in greater detail and to explore how variation is partitioned across the genome.

CHAPTER III

MONKEYFLOWER GENOMES REVEAL THE TEMPORAL DYNAMICS OF LINKED BACKGROUND SELECTION ON PATTERNS OF GENETIC VARIATION

This chapter includes unpublished work with other co-authors. The genome assembly and genetic map were performed by Matthew Streisfeld and Sean Stankowski. The genome annotation was performed by Allison Fuiten. I performed all other analyses.

INTRODUCTION

Understanding the processes that partition genetic variation among populations, lineages, and species is essential for understanding the origins of biodiversity (Lewontin 1974; Seehausen et al. 2014; Wolf and Ellegren 2017). Until recently, efforts to study the evolutionary processes that have shaped patterns of diversity have been hampered by our inability to detect and adequately measure genetic variation (Lewontin 1974). Next-generation sequencing has provided novel opportunities to characterize patterns of genetic variation across the genomes of many non-model organisms. However, interpreting patterns of genome-wide variation remains a formidable challenge, because many evolutionary processes can leave similar genetic signatures (Wolf and Ellegren 2017; Ravinet et al. 2017).

The challenge of inferring process from pattern is illustrated by recent efforts to interpret patterns of genome-wide differentiation in light of the speciation process. For example, many studies have revealed a heterogeneous pattern of genome-wide differentiation between taxa (Turner et al. 2005; Hohenlohe et al. 2010; Ellegren et al. 2012; Martin et al. 2013; Renaut et al. 2013; Soria-Carrasco 2014; Poelstra et a. 2014; Lamichhaney et al. 2015; Malinsky et al. 2015; Vijay et al. 2015) A pattern of peaks and valleys of differentiation (F_{ST}), commonly referred to as the 'genomic landscape,' was originally interpreted as the consequence of local adaptation, with peaks of differentiation harboring loci that have become differentiated by selection in the face of gene flow (Wu 2001). However, in contrast to this view, recent studies have shown that heterogeneous genomic landscapes can arise due to unrelated processes (Cruickshank and Hahn 2014;

Burri et al. 2015; Burri. 2017a; Ellegren and Wolf 2017; Ravinet et al. 2017; Vijay et al. 2017; Wolf and Ellegren 2017).

Long-term linked selection can have a profound impact on patterns of genome-wide variation (Hahn 2008; Burri 2017a; Ellegren and Wolf 2017; Ravinet et al. 2017). Linked selection describes the combined effects of purifying selection against deleterious mutations and positive selection on globally beneficial alleles, which causes reductions in diversity (π) and effective population size (N_e) and leads to increased genetic differentiation (F_{ST}) and faster rates of lineage sorting (Charlesworth et al. 1993; Cutter and Payseur 2013). Several recent studies, primarily based on different bird genomes, have found empirical support for linked selection as a primary driver of genomic landscapes of genetic differentiation, indicating that it can be a major force shaping patterns of genome-wide variation (Poelstra et al. 2014; Burri et al. 2015; Burri 2017a; Burri 2017b; Ellegren and Wolf 2017; Van Doren et al. 2017; Vijay et al. 2017). However, empirical studies in other taxonomic groups are needed to refine our understanding of the process and to determine its general importance (Ellegren and Wolf 2017).

In this study, we sequenced the genomes of multiple taxa across a recent and rapid radiation to examine how genomic variation has been shaped and structured through time. The bush monkeyflower radiation consists of perennial shrubs from seven subspecies (one with two ecotypes) of *Mimulus aurantiacus* that are distributed primarily in California (Chase et al. 2017). Together with their sister species *M. clevelandii,* multiple pairs of taxa exist at different stages of divergence, ranging from locally adapted ecotypes to partially isolated subspecies (McMinn 1951; Sobel and Streisfeld 2015). The plants inhabit a range of environments, including temperate coastal regions, mountain ranges, semi-arid habitats, and offshore islands (Thompson 2012). Most of the taxa are geographically isolated from one another, though some have parapatric distributions and hybridize in narrow regions where their distributions overlap (Streisfeld and Kohn 2005; Streisfeld and Kohn 2007; Thompson 2012; Streisfeld et al. 2013; Sobel and Streisfeld 2015; Stankowski et al. 2015; Stankowski and Streisfeld 2015; Stankowski et al. 2017). However, recent studies have confirmed the monophyly of the radiation and established the basic relationships among its taxa (Stankowski and Streisfeld 2015; Chase et al.

2017). The presence of many closely related taxa at varying stages of differentiation make this a powerful system for studying how genome-wide variation has evolved over the course of divergence.

Aided by the first chromosome-level genome assembly for *M. aurantiacus,* we used whole-genome sequencing to study the evolutionary processes that have shaped variation across this recent and rapid radiation. After inferring relationships among the taxa, we reveal heterogeneous patterns of lineage sorting, diversity, and differentiation that have been shaped by linked background selection (BGS). The pattern of linked BGS is conserved across the radiation and is determined by the genomic distribution of functional elements and the local recombination rate. By examining taxa along a continuum of divergence, we show that lineage-specific BGS has been able to maintain an ancestral diversity landscape, which has resulted in the buildup of a heterogeneous differentiation landscape over time.

RESULTS AND DISCUSSION

A chromosome-level genome assembly, map and annotation for the bush monkeyflower

To facilitate genomic analyses in this group, we constructed the first chromosome-level reference genome for the bush monkeyflower using a combination of long-read Single Molecular Real Time (SMRT) sequencing (PacBio), overlapping and mate-pair short-reads (Illumina), and a high-density genetic map (7,589 segregating markers across 10 linkage groups). Contig building and multiple rounds of scaffolding yielded 1,547 scaffolds with an N50 size of 1,578 kb and a total length of 207 Mbp. The high-density map allowed us to anchor and orient 94% of the assembled genome onto the 10 linkage groups, which is the number of chromosomes inferred from karyotypic analyses in all subspecies of *M. aurantiacus* and *M. clevelandii* (Vickery 1995). Analysis of assembly completeness based on conserved gene space revealed that 95% of 956 universal single copy orthologous genes were completely (931) or partially (891) assembled (Simão et al. 2015). Subsequent annotation yielded 23,018 predicted genes. Comparison of the synonymous substitution rate (K_s) with the only other available

monkeyflower genome (*M. guttatus*) revealed an average divergence time between species of 20.5 million years (SD 8.4 million years).

Variation in the extent of lineage sorting across the genome

As a first step toward understanding the processes driving divergence and lineage formation across this radiation, we inferred evolutionary relationships among the nine bush monkeyflower taxa. Rapid diversification is a hallmark of evolutionary radiations and can result in complex patterns of phylogenetic discordance among genomic regions due primarily to incomplete lineage sorting (ILS) (Jarvis et al. 2014; Pease et al. 2016). To infer relationships and quantify discordance, we sequenced 37 whole genomes from the seven subspecies and two ecotypes of *Mimulus aurantiacus* (n = 4-5 taxon) and the sister species *M. clevelandii* (n = 3) (average sequencing depth of 21x). Close sequence similarity allowed us to align reads from all samples to the reference assembly with high confidence (average 91.7% reads aligned; Appendix D). After mapping and variant calling, we identified 13.2 million SNPs that we use in subsequent analyses.

We inferred evolutionary relationships among these nine taxa using maximum-likelihood (ML) phylogenetic analysis (implemented in RAxML; Stamatakis 2014) and two different datasets: whole-genome concatenation and 500 kb non-overlapping windows. The resulting tree topologies from the whole-genome dataset (Fig. 3.1) confirmed the same relationships among taxa as previous analyses based on reduced representation sequencing and five different methods of phylogenetic reconstruction (Stankowski and Streisfeld 2015; Chase et al. 2017). Specifically, all seven of the subspecies formed monophyletic groups with 100% bootstrap support. However, within subspecies *puniceus*, relationships are more complex, as the red ecotype formed a monophyletic sub-clade within the paraphyletic yellow ecotype, a result that reflects the recent origin of red flowers from a yellow-flowered ancestor (Stankowski and Streisfeld 2015).

Although the whole genome phylogeny provides a well-supported summary of the evolutionary relationships among the taxa, our window-based analysis revealed extensive phylogenetic discordance across local genomic regions (Fig 3.1). Even at the relatively large scale of 500 kb regions, only 22 (6%) of the 387 window-based trees showed the

same taxon branching order as the whole-genome tree. Nevertheless, four of the subspecies formed separate monophyletic groups across nearly all windows (*grandiflorus* 99.7%, *aridus* 99.7%, *parviflorus* 94.1%, and *aurantiacus* 88.1%). Subspecies *puniceus* also was monophyletic in the majority of trees (72.6%), but individuals from the red and yellow ecotypes only formed monophyletic groups in 27% and 9% of the trees, respectively. Similarly, the closely related subspecies *longiflorus* (45.7%) and *calycinus* (24.1%) were monophyletic in fewer trees than the other subspecies.

While some of the phylogenetic discordance that we observed could be generated by gene flow after divergence (e.g., Lamichhaney et al. 2015; Stankowski and Streisfeld 2015; Pease et al. 2016; Richards and Martin 2017), our data indicate that the majority of the discordance is due to incomplete lineage sorting (ILS). Specifically, the level of discordance at nodes that define taxa and clades is strongly predicted by internode length ($r = 0.76$; Appendix E). Thus, the strongest signal of discordance occurs in areas of the tree where multiple divergence events have occurred within a short period of time. Thus, as predicted by theory (Maddison 1997) and as shown in other diverse radiations (Suh et al. 2015; Pease et al. 2016), we conclude that ILS is the major source of phylogenetic conflict across the bush monkeyflower radiation.

To assess the genomic distribution of tree discordance, we estimated the correlation (Pearson's *r*) between the distance matrix for each window-based tree and the genome-wide tree. Plotting this tree concordance score across the 10 linkage groups revealed a striking pattern of lineage sorting across the genome (Figure 3.2a). Rather than being randomly distributed, trees with similar concordance scores tend to cluster together, spanning large sections of all ten chromosomes. This pattern indicates that local rates of lineage sorting are determined by differences in the nature and or strength of the evolutionary processes that shape genetic variation across the genome.

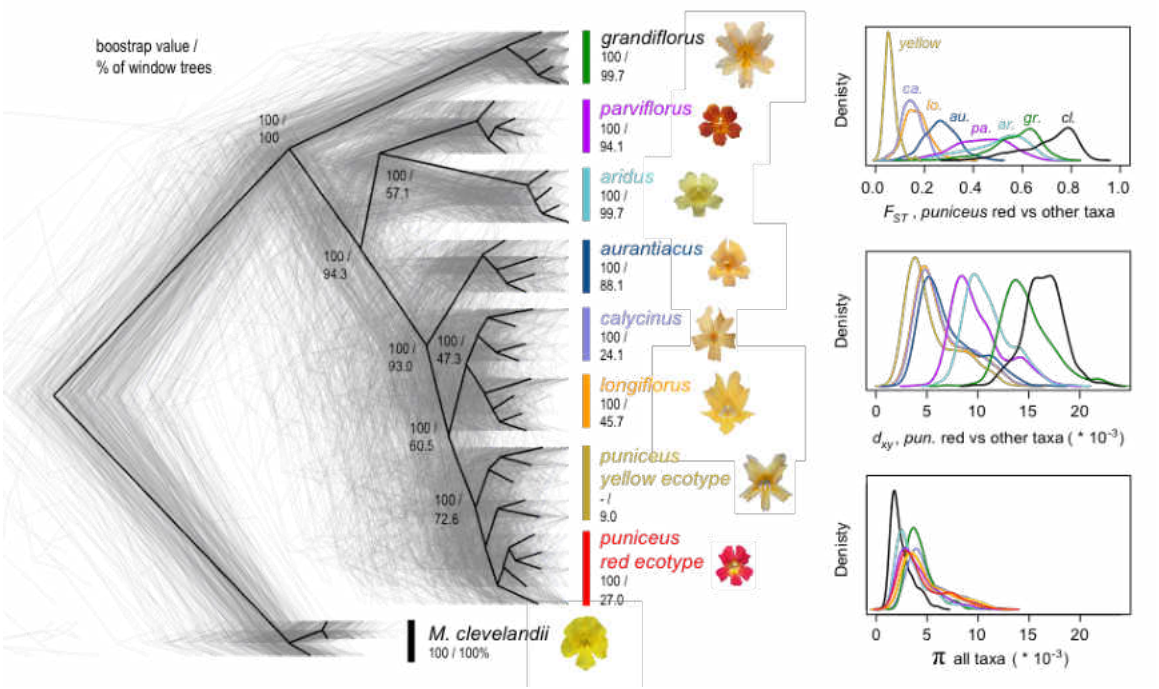
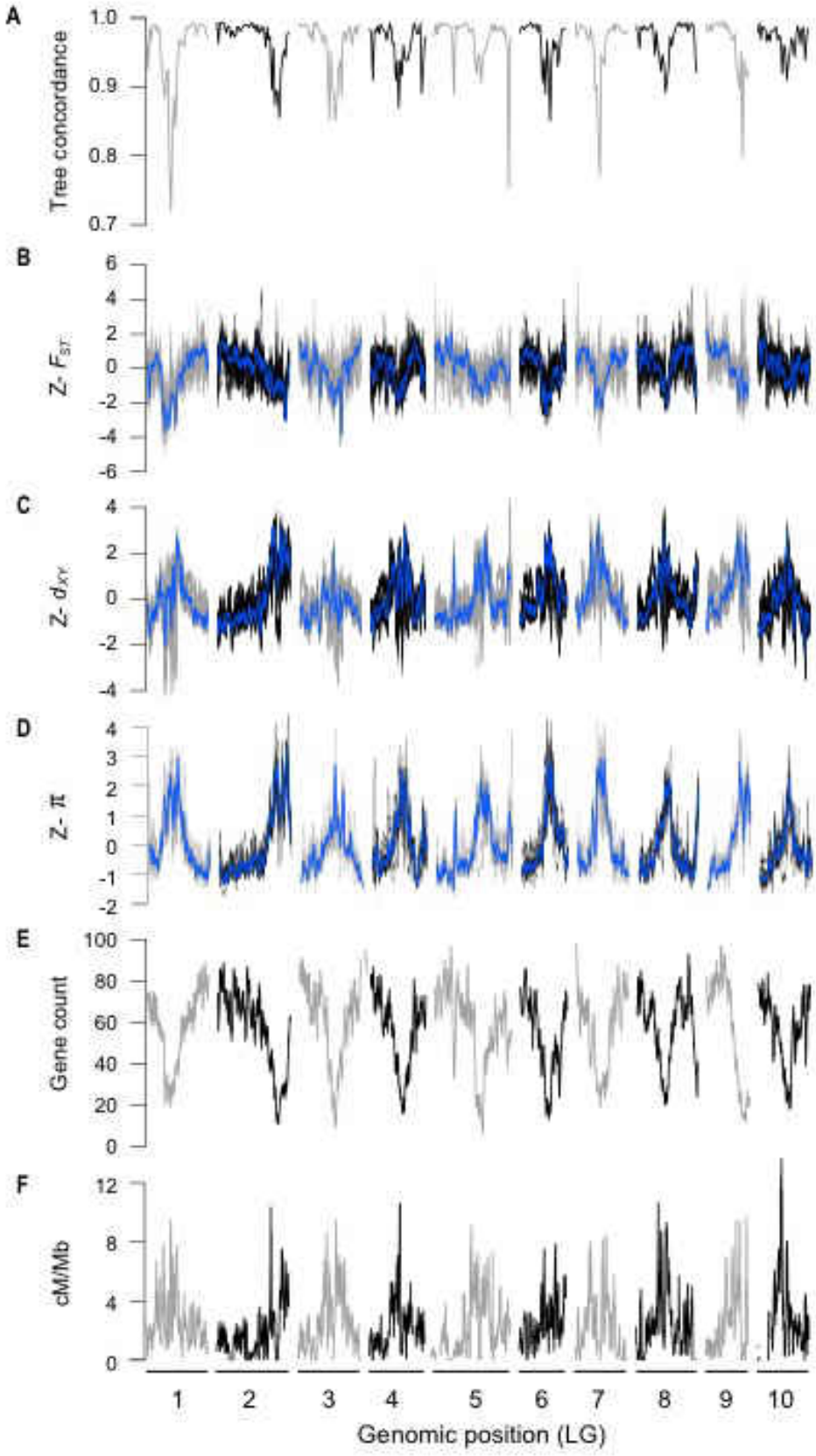


Figure 3.1. Evolutionary relationships among subspecies and ecotypes of the bush monkeyflower. The black topology is the ML tree constructed from a concatenated alignment across the whole genome and is rooted using *M. clevelandii*. The 387 gray trees show the ML topologies for 500 kb genomic windows. The first number along each node indicates the percent bootstrap support of that node in the whole genome tree, and the second number describes the percent of the window-based trees that match that node in the whole genome tree. The first number below each taxon name reflects the bootstrap support from the whole genome tree of the node that defines the multiple samples from that taxon as a monophyletic group. The second number describes the percent of window-based trees that demonstrate the taxon as a monophyletic group. b) Levels of genomic differentiation (F_{ST}), divergence (d_{xy}), and levels of diversity (π) within and among bush monkeyflower taxa based on the same 500 kb windows. For simplicity, F_{ST} and d_{xy} data are shown only for comparisons between the red ecotype of subspecies *puniceus* and all other taxa, while π is calculated for all nine taxa.

Figure 3.2. (Next page) Common differentiation and diversity landscapes mirror variation in the local properties of the genome. a) Tree concordance in 500kb genomic windows across the 10 bush monkeyflower chromosomes. b – d) z-transformed F_{ST}, d_{xy}, and π in overlapping 500 kb windows (step size = 50 kb). The alternating gray and black lines are the z-transformed scores for each of the 36 pairwise comparisons (b,c) or nine taxa (d), and the blue line is the *z*-transformed score of the first principal component (PC1), which explains 66%, 70%, and 85% of the variation within and among taxa in, F_{ST}, d_{xy} and π, respectively. e and f) Gene count and recombination rate (cM/Mb) in 500 kb overlapping genomic windows.


A
Tree concordance
1.0
0.9
0.8
0.7
B
Z- F_{ST}
6
4
2
0
-2
-4
-6
C
Z- d_{XY}
4
2
0
-2
-4
D
Z- π
4
3
2
1
0
-1
-2
E
Gene count
100
80
60
40
20
0
F
cM/Mb
12
8
4
0
1
2
3
4
5
6
7
8
9
10
Genomic position (LG)

A common pattern of linked background selection shapes genome-wide variation and drives lineage sorting

To gain insight into the evolutionary processes that have shaped patterns of genome-wide variation, we used summary statistics in the same 500 kb windows to quantify patterns of differentiation (F_{ST}), divergence (d_{xy}), and diversity (π) among and within these taxa (Fig. 3.1). The range of mean estimates of F_{ST} across all 36 pairwise comparisons among the nine taxa highlights the continuous nature of divergence across the group (Fig. 3.1b; Appendix F), with estimates spanning from 0.06 (red vs. yellow ecotypes of *puniceus*) to more than 0.70 for comparisons with the sister species *M. clevelandii.* More striking, the broad distributions of the window-based estimates reveal extreme heterogeneity in the levels of all three statistics among genomic regions. For F_{ST}, the average range of window-based estimates within the 36 pairwise comparisons (0.46) is roughly equal to the mean difference in F_{ST} between taxon pairs (0.47). Distributions of absolute divergence (d_{xy}) show a similar pattern, with mean values ranging from 0.54% (red vs. yellow ecotypes) to 1.6% (yellow ecotype vs. *M. clevelandii*), and the average range of window based estimates (1.30%) exceeding the average difference among taxa (1.14%). Remarkable variation in π was also observed among windows, ranging from 0.09% to 1.26%, even though mean estimates were very similar among the ingroup taxa (0.37% to 0.53%).

As with tree concordance, variation in these summary statistics was non-randomly distributed across the genome, as genome scans revealed rugged differentiation (F_{ST}), divergence (d_{xy}), and diversity (π) landscapes (Fig. 3.2b-d). Strikingly, the shapes of these landscapes were highly correlated among all comparisons. The first axis of a principal components analysis (PC1) across all 36 taxonomic comparisons explained a remarkable 65.9% of the variation in $F_{ST,}$ with loadings ranging from 0.37 to 0.92 (Fig. 3.2b). Patterns of divergence (d_{xy}) and diversity (π) were also highly correlated among comparisons, with PC1 explaining 69.5% and 84.7% of the variation among window-based estimates, respectively (Fig 3.2c-d).

The presence of a correlated genomic landscape is remarkable and is unlikely to reflect evolutionary processes that are unique to each taxon (Burri 2017a; Burri 2017b; Ellegren and Wolf 2017). Rather, it suggests that a common mechanism is responsible for

shaping patterns of variation. Recent studies have observed correlated genomic landscapes among related taxa, concluding that they were generated by a common pattern of linked background selection (BGS) across the genome (Poelstra et al. 2014; Burri et al. 2015; Van Dorren et al. 2017; Vijay et al. 2017). Under this model, regions of the genome with higher BGS are predicted to show reduced diversity, because purifying selection removes variation at linked neutral sites (Charlesworth et al. 1993; Charlesworth 2012). When the pattern of genome-wide BGS is shared across the genomes of taxa, a common genomic landscape is predicted to evolve (Burri 2017a; Burri 2017b). In agreement with this prediction, we observe a strong negative correlation between PC1 F_{ST} and PC1 π (r = -0.84), indicating that differentiated regions tend to show lower diversity. These same regions of low diversity also tend to show lower divergence (r = 0.84) and higher levels of lineage sorting (r = -0.69), both of which are predicted under BGS because recurrent bouts of selection cause long-term reductions in N_e (Cruikshank and Hahn 2014).

Another possible selective explanation for these correlated landscapes is recurrent positive selection within the same genomic regions in all nine taxa (Smith and Haigh 1974). Although positive selection may play some role in shaping differentiation landscapes, BGS is expected to have a greater impact on differentiation and diversity, as most new mutations are more likely to be deleterious than beneficial (Eyre-Walker and Keightley 2007). In addition to being more frequent, the targets for deleterious mutations are widespread across the genome, whereas not all functional elements are necessarily expected to be targets of positive selection. Consistent with a widespread role of BGS in our system, the majority of genomic regions show high levels of differentiation, while regions of exceptional differentiation appear as troughs of low F_{ST} (Fig 3.2a). This indicates that some regions of the genome experience less BGS than the average level.

Heterogeneous BGS is determined by the distribution of conserved genomic features

Given the evidence for a prominent role of BGS in shaping patterns of differentiation and lineage sorting in this system, we next identified factors that affect variation in its intensity. In models of BGS, the frequency of purifying selection is predicted to be higher in regions of the genome that are enriched for functional elements,

as they provide rich targets for deleterious mutations (Charlesworth et al. 1993; Charlesworth 2012). Recombination rate variation is also expected to modify the local impact of purifying selection. In regions of high recombination, linked neutral variants are more likely to escape their association with deleterious mutations, so they are less likely to experience reductions in diversity (Charlesworth et al. 1993; Charlesworth 2012). Conversely, the physical impact of BGS is magnified where the local recombination rate is low due to tighter linkage among sites. While gene density and recombination rate may act independently to shape patterns of diversity, these genomic properties are expected to interact, further increasing the impact of BGS (Charlesworth et al. 1993; Charlesworth 2012; Burri 2017a; Burri 2017b).

To test for an association between nucleotide diversity and these intrinsic genomic features, we used our annotated genome and genetic map to estimate the number of protein coding genes and the recombination rate (cM/Mb) in each non-overlapping 500 kb window across the genome (Fig. 3.2e-f). We then estimated the correlation between these variables and the corresponding estimates of diversity in each window. The results support the prediction that gene rich regions experience a higher rate of background selection, as we observed a strong negative correlation between gene count and PC1 π (r = -0.84; Fig. 3.3). Similarly, we observed a strong, positive correlation between PC1 π and recombination rate, which is consistent with its role in moderating the impact of linked selection (r = 0.44; Fig. 3.3). However, we did not observe a significant interactive effect of gene count and recombination rate on levels of nucleotide diversity (p = 0.057). This may be because genomic features are also correlated (r = -0.40; Fig. 3.3), making it difficult to tease apart their relative impacts.

Despite only having a direct estimate of gene density and recombination rate variation from one subspecies (*puniceus*), a common diversity landscape implies that the genomic distribution of these features is conserved across the radiation. This is not surprising given the recent shared history of the group, and the fact that all subspecies of *M. aurantiacus* and the outgroup, *M. clevelandii*, have the same number of chromosomes (n = 10). Further, the inclusion of the outgroup reveals that all taxa inherited a similar diversity landscape that was pre-shaped by historical BGS. However, the maintenance of

this landscape by lineage specific BGS has resulted in the build-up of a heterogeneous differentiation landscape over time.

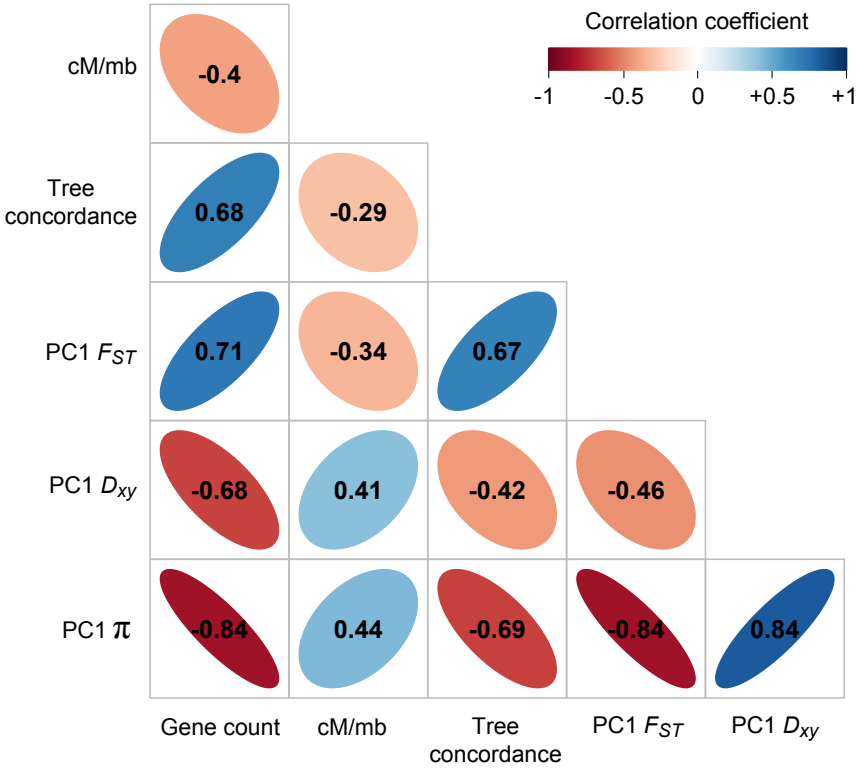


Figure 3.3. Correlations reveal heterogeneity in linked selection across the genome. A matrix of pairwise correlation coefficients is shown that relates the first principal component (PC1) of F_{ST}, d_{xy}, and π to tree concordance, gene density, and recombination rate. The heat map indicates the strength of the correlation and its sign, while the ellipse represents 95% of the bivariate normal density in the data. All correlations are statistically significant at $p < 0.001$. The strength and direction of all of the correlations support a model of heterogeneous linked selection across the genomes of the taxa in the bush monkeyflower radiation.

The dynamic build-up of the differentiation landscape by BGS

The temporal buildup of differentiation by linked BGS begins when a population first splits (assuming a simple model of allopatric divergence). At this moment, estimates of diversity (π) and divergence (d_{xy}) should be equal, so there should be no heterogeneous pattern of genome-wide differentiation (F_{ST}) and no relationship between π and F_{ST}. As time progresses and new mutations accumulate within each subpopulation, d_{xy} becomes greater than π. New mutations with deleterious effects will be purged from the population along with linked neutral variants, thus maintaining the shape of the diversity landscape

over time. Soon after the split, when mutational input is low, elevated differentiation should only partially coincide with genomic regions that have experienced a high rate of historical BGS. This is because ancestral variation is shared between taxa in areas of the genome that have not yet experienced lineage specific BGS. However, as more deleterious mutations arise, the negative relationship between diversity and differentiation should become stronger until they mirror one another. Similarly, the replacement of ancestral variants by BGS should gradually increase the extent of lineage sorting in these same regions.

Given that the different pairs of monkeyflower taxa all shared a common ancestor at different points in the past, we were able to examine temporal changes in the strengths of the various relationships, as predicted by the BGS model. As expected for pairs of taxa at the earliest stages of divergence, the relationship between π and d_{xy} is almost perfect ($r \sim 1$), but the correlation decays to 0 as lineage-specific mutations accumulate over time (Fig. 3.4). Consistent with maintenance of the diversity landscape by BGS in the face of new mutations, the correlations between π and gene density and recombination rate remain constant as the divergence time between taxon pairs increases (Fig. 3.4). However, the dynamic buildup of heterogeneous differentiation can be observed as the relationships between F_{ST} and levels of diversity, gene count, and recombination rate all become stronger with greater divergence between taxa (Fig. 3.4).

CONCLUSIONS

We have demonstrated that a shared diversity landscape across all taxa leads to the rapid evolution of a common, heterogeneous genome-wide pattern of differentiation. Although adaptive evolution is considered a hallmark of radiations and is almost certainly responsible for the remarkable phenotypic diversity in this group (Streisfeld and Kohn 2005; Stankowski et al. 2015; Stankowski et al. 2017), we show that broad scale patterns of genome-wide variation are driven primarily by non-adaptive processes. Specifically, background selection against deleterious alleles has had a major impact on patterns of genomic variation and had led to variation in rates of lineage sorting. Even after lineages diverge, the pattern of background selection across the genome is maintained in all taxa, resulting in correlated patterns of differentiation across taxon pairs.

An important implication of this work is that patterns of genetic diversity are shaped largely by variation in natural selection across the genome rather than neutral processes. This finding adds to the growing body of work that reveals the widespread role that selection plays in molding genetic variation in nature (Begun et al. 2007; Hahn 2008) and provides insight into the temporal dynamics driving the buildup of differentiation due to BGS. Thus, as more genomes become available for more distantly related *Mimulus* species, it will be interesting to determine whether these processes have been operating to shape genomic variation at even deeper timescales.

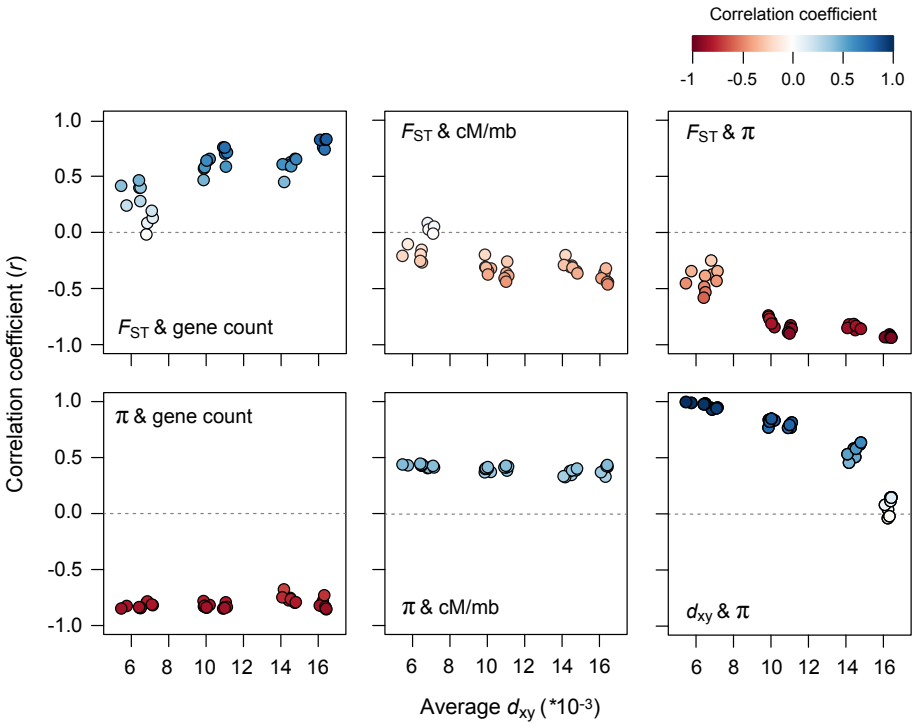


Figure 3.4. The dynamic build-up of linked selection through time across the radiation. The correlation coefficients across genomic windows for each of the 36 pairwise taxonomic comparisons are plotted against levels of sequence divergence between taxa (average dxy among windows). Sequence divergence is used here as a proxy for divergence time. The top row shows the dynamic relationship between F_{ST} and gene count, recombination rate, and π, while the bottom row examines how correlations between π and gene count, recombination rate, and d_{xy} change with divergence time. The heat map describes the strength of the correlation, and the dotted line is positioned at zero, indicating no relationship between the variables.

MATERIALS AND METHODS

Genome assembly

We used a combination of short-read Illumina and long-read Single Molecule, Real Time (SMRT) sequencing to assemble the genome of a single individual from the red ecotype of *M. aurantiacus* subspecies *puniceus* (population UCSD). Genomic DNA was isolated from leaf tissue using either ZR plant/seed DNA miniprep kits (Zymo Research) or GeneJet Plant Genomic DNA purification kits (Thermo Fisher). Illumina libraries were generated following the *Allpaths-LG* assembly pipeline (Gnerre et al. 2011), which included a single fragment library with average 180 bp insert size and three mate pair libraries (average insert sizes: 3.5-5 kb, 5-7 kb, and 7-13 kb). Libraries were sequenced on the Illumina HiSeq 2500 using paired-end 100 bp reads. An initial scaffold-level assembly was performed with *Allpaths-LG* using default parameters and the *haploidify* function enabled. This assembly yielded 11,123 contigs (N50 = 40.5 kb) and 2,299 scaffolds (N50 = 1,310 kb), for a total assembly size of 193.3 Mbp. Long-read sequencing was performed from the same individual using 12 SMRT cells sequenced on the Pacific Biosystems RS II machine at Duke University. We obtained a total of 6.4 Gbp of sequence, which corresponds to ~21 × coverage of the genome. The PacBio reads were used to re-scaffold the *Allpaths-LG* scaffolds using *Opera-LG* (Gao et al. 2016). This reduced the number of scaffolds to 1,547 (N50 = 1,578 kb).

Finally, we gap filled the assembly using the PacBio data and the program *PBJelly* (English et al. 2012). Resulting scaffolds were assembled into pseudomolecules using *Chromonomer* (Amores et al. 2014), according to the online manual. This software anchored and oriented scaffolds based on the order of markers in a high-density linkage map (see below) and made corrections to scaffolds when differences occurred between the genetic and physical positions of markers in the map. A final round of gap filling with *PBJelly* was performed to fill any gaps that were created by broken scaffolds in *Chromonomer*. To assess the completeness of the gene space in the assembly, we used both the BUSCO and CEGMA pipelines to estimate the proportion of 956 single copy plant genes (BUSCO) or 248 core eukaryotic genes (CEGMA) that were completely or partially assembled (Parra et al. 2007; Simao et al. 2015). The proportion of these genes

present in an assembly has been shown to be correlated with the total proportion of assembled gene space, and thus serves as a good predictor of assembly completeness.

Construction of high-density linkage map

We generated an outbred F_2 mapping population by crossing two F_1 individuals, each the product of crosses between different greenhouse-raised plants collected from one red ecotype and one yellow ecotype population (populations UCSD and LO, respectively; Table S1). We then used restriction-site associated DNA sequencing (RADseq) to genotype F_1 and F_2 individuals. DNA was extracted from leaf material using Zymo ZR plant/seed DNA miniprep kits, and RAD library preparation followed the protocol outlined in Sobel and Streisfeld (2015). Libraries were sequenced on the Illumina HiSeq 2000 platform using single-end 100 bp reads at the Genomics Core Facility, University of Oregon.

Reads were filtered based on quality, and errors in the barcode sequence or RAD site were corrected using the *process_radtags* script in *Stacks v. 1.35* (Catchen et al. 2013). Loci were created using the *denovo_map.pl* function of *Stacks*, with three identical raw reads required to create a stack, two mismatches allowed between loci for an individual, and two mismatches allowed when processing the catalog. Single nucleotide polymorphisms (SNPs) were determined and genotypes called using a maximum-likelihood (ML) statistical model implemented in *Stacks* and a stringent χ^2 significance level of 0.01 to distinguish between heterozygotes and homozygotes (Hohenlohe et al. 2010, 2012; Catchen et al. 2011). We then used the *genotypes* program implemented in *Stacks* to identify a set of 9,029 mappable markers. We specified a 'CP' cross design (F_1 individuals coded as the parents), requiring that a marker was present in at least 85% of progeny at a minimum depth of 12 reads per individual, and we allowed automated corrections to be made to the data.

Linkage map construction was performed using *Lep-MAP2* (Rastas et al. 2015). The data were filtered using the *Filtering* module to include only individuals with less than 15% missing data and excluded markers that showed evidence for extreme segregation distortion (χ^2 test, $P < 0.01$). To assign markers to linkage groups, we used the *SeparateChromosomes* module with a logarithm of odds score limit of 20 and no

minimum size for linkage groups (LG). This assigned 7,217 markers to 10 linkage groups, which matches the number of chromosomes in *M. aurantiacus*. The *JoinSingles* module was executed again with a LOD limit of 10 to join an additional 877 ungrouped markers to the 10 previously formed LGs. Fifty-seven singles that were not joined at this stage were discarded from the dataset. Initial marker orders were determined using sex-averaged and sex-specific recombination rates using the *OrderMarkers* module. For each LG, we conducted 10 independent runs using the Kosambi mapping function (*useKosambi=1*), with the dataset split into seven pseudofamilies to take advantage of parallel processing. When multiple markers had identical genotypes, only the duplicate marker with the least missing data was used in marker ordering. We retained the marker order from the run with the best likelihood. After removing markers with an error rate > 0.05, the ML order was re-evaluated using the *evaluateOrder* flag. The map contained 8,094 informative loci from 269 F_2 individuals, with an average of 3.5% ± SD 3.86 missing data per individual.

After the integration of our assembly and genetic map using the *Chromonomer* software (Amores et al. 2014), we made corrections to our map order based on the physical position of markers within assembled scaffolds. Using the output of *Chromonomer*, we identified markers that were out of order in the map compared to their local assembly order and aligned these markers to the assembly from *Chromonomer* using *Bowtie2* v. 2.2.5 (Langmead and Salzberg 2012) with the *very_sensitive* settings to obtain their physical order. We then re-estimated the map using the *evaluateOrder* flag in *Lep-MAP2* as described above, but with the marker order constrained to the physical order (*improveOrder=0*) and with all duplicate markers included in the analysis (*removeDuplicates=0*). After initial map construction, we removed 17 markers with an estimated error rate greater than 5% and estimated the map using the same settings. The final map contained 7,589 markers across the 10 linkage groups.

Genome annotation

Prior to genome annotation, the assembly was soft-masked for repetitive elements and areas of low complexity with *RepeatMasker* (RepeatMasker Open-4.0) using a custom *Mimulus aurantiacus* library created by *RepeatModeler* (RepeatModeler Open-1.0),

Repbase repeat libraries (Jurka et al. 2005), and a list of known transposable elements provided by *MAKER* (Holt and Yandall 2011). In total, 30.99% of the genome assembly was masked by *RepeatMasker*. Repetitive elements were annotated with *RepeatModeler*. Hidden Markov Models for gene prediction were generated by *SNAP* (Korf 2004) and *Augustus* (Stanke and Waack 2003) and were trained iteratively to the assembly using *MAKER*, as described by Cantarel et al. (2008). Training was performed on the 14.5 Mbp sequence from LG9. Evidence used by *MAKER* for annotation included protein sequences from *Arabidopsis thaliana*, *Oryza sativa*, *Solanum lycopersicum*, *Solanum tuberosum*, *Daucus carota*, *Vitis vinifera*b (all downloaded from EnsemblPlants on 9 August 2016), *Salvia miltiorrhiza* (downloaded from Herbal Medicine Omics Database on 9 August 2016), *Mimulus guttatus v. 2* (downloaded from JGI Genome Portal on 9 August 2016), and all Uniprot/swissprot proteins (downloaded on 18 August 2016) (Kersey et al. 2015; Goodstein et al. 2011; Nordberg et al. 2014; Herbal Medicine Omics Database; Uniprot). We filtered the annotations with *MAKER* to include: 1) only evidence-based information that also contained assembled protein support, and 2) those *ab initio* gene predictions that did not overlap with the evidence-based annotations and that contained protein family domains, as detected with InterProScan (Quevillon et al 2005).

Estimation of divergence time with *Mimulus guttatus*

To estimate divergence times between *M. aurantiacus* and the congeric *M. guttatus*, we compiled the distribution of synonymous substitution rates (K_s) from orthologous gene pairs between the species. We used the program *Inparanoid* v. 4.1 (O'Brien et al. 2005) to identify orthologous gene pairs from the proteomes of each species. Ortholog pairs were filtered initially to include only those genes that mapped simultaneously to one of the 10 *M. aurantiacus* and one of the 14 *M. guttatus* linkage groups, which generated 8,580 orthologous gene pairs. Protein sequences then were aligned using *Clustal W*, and only those sequences containing predicted start and stop codons and with alignment scores > 30 were included in the final dataset, resulting in 4,449 orthologus gene pairs. Protein sequences were back translated using the coding sequences associated with each protein and the *reverse.align* function in the R package *seqinr*, and K_s values were

calculated for each gene alignment using the *kaks* function (Charif and Lobry 2007). Divergence time (T) for each gene was estimated according to the equation: $T = K_s/2\mu$, where μ is the per site mutation rate per generation (Kimura 1980). We used a mutation rate of 1.5×10^{-8}, as has been used previously for other *Mimulus* species (Brandvain et al. 2014).

Genome re-sequencing and variant calling

We collected leaf tissue from from four to five individuals from the eight taxa within the *Mimulus aurantiacus* species complex. In addition, we collected leaf tissue from three individuals of *M. clevelandii*, which is sister to *M. aurantiacus* (Appendix D). We extracted DNA from dried tissue using the Zymo Plant/Seed MiniPrep DNA kit following the manufacturer's instructions. We prepared sequencing libraries using the Kapa Biosystems HyperPrep kit, and libraries with an insert size between 400-600 bp were sequenced on the Illumina HiSeq 4000 using paired-end 150 bp reads at the Genomics Core Facility, University of Oregon.

We filtered raw reads using the *process_shortreads* script in *Stacks* v1.46 to remove reads with uncalled bases or poor quality scores. We then aligned the retained reads to the reference assembly using the BWA-MEM algorithm in *BWA* v0.7.15 (Li and Durbin 2009). An average of 91.7% of reads aligned (range: 82.6-96.0%), and the average sequencing depth was 21x (range: 15.16x – 30.86x). We then marked PCR duplicates using Picard. We performed an initial run of variant calling using the UnifiedGenotyper tool in GATK v3.8 (McKenna et al. 2010) and filtered the data to remove variants with a mapping quality < 50, a quality depth < 4 and a Fisher Strand score > 50. We then used these variants to perform base quality score recalibration for each individual, before performing another run of the UnifiedGenotyper to call final variants. After the second run of variant calling, we removed variants with a mapping quality < 40, a quality depth < 2, and a Fisher Strand score > 60. The final VCF file contained 13,233,829 SNPs across the nine taxa. Finally, we ran UnifiedGenotyper with the EMIT_ALL_SITES option to output all variant and invariant genotyped sites.

Phylogenetic analyses

We reconstructed the evolutionary relationships among the nine taxa using RAxML v8 (Stamatakis 2014) by concatenating variant sites across the genome to obtain the topology for the whole genome tree. To investigate patterns of phylogenetic discordance across the genome, we also built trees from windows across the genome. We first phased SNPs using BEAGLE v4.1 (Browning and Browning 2007), using a window size of 100,000 bp and an overlap of 10,000 bp. We incorporated information on recombination rate from the genetic map and did not impute missing genotypes. After phasing, we used MVFtools (Pease and Rosenzweig 2016; [https://www.github.com/jbpease/mvftools]) to run RAxML from 500 kb non-overlapping windows, with the *M. clevelandii* populations set as the outgroups. We then visualized the window trees in DensiTree v2.2.5 (Bouckaert 2010).

To assess concordance between the trees from different windows and the species tree, we converted trees to distance matrices using the *Ape* package in R (Paradis et al. 2004). We then calculated the Pearson's correlation coefficient between the distance matrix from each window and the species tree, with a stronger correlation indicating higher concordance with the species tree.

Population genomic analyses

To examine how genome-wide patterns of diversity, differentiation, and divergence varied among taxa, we calculated within-taxon nucleotide diversity (π), between-taxon relative differentiation (F_{ST}), and between-taxon absolute divergence (d_{XY}) across non-overlapping 500kb windows using Python scripts downloaded from https://github.com/simonhmartin/genomics_general. We calculated measures of differentiation and divergence across all 36 pairwise comparisons among the nine taxa, and diversity was calculated within each of the nine taxa. We filtered the data separately for each comparison so that each site was genotyped in at least three individuals for each comparison within the *M. aurantiacus* complex, or at least two individuals for each comparison to *M. clevelandii.* To incorporate invariant sites into measures of diversity and divergence, we divided the number of pairwise differences (within and between taxa,

respectively) by the total number of genotyped sites (variant and invariant) within a window.

We summarized the variation in each statistic across comparisons using a principal components analysis (PCA), with taxon or taxon pair as the variables. Thus, across each window, the first principal components of π, F_{ST}, and d_{XY} provide multivariate measures that explain the greatest covariance in the data. To examine the relationships among diversity, differentiation, and divergence, we estimated the pairwise correlation in PC1 across genomic windows among all three statistics. Further, we estimated correlations among these three statistics and tree concordance, gene density, and recombination rate. Recombination rate was estimated by comparing the genetic and physical distance (in cM/Mbp) between all pairs of adjacent markers on each LG from the genetic linkage map described above. We removed the top 5% of recombination rates, as these represented unrealistically high values of recombination. A minimum of three estimates were required to obtain an average recombination rate estimate within each window. Gene density was estimated as the number of predicted genes in each window, as determined from the annotation described above.

To determine how the correlations among the statistics (diversity, differentiation, divergence, recombination rate, gene count, tree concordance) change through time, we examined the correlation coefficient among all pairs of statistics individually for each of the 36 comparisons. As a measure of overall divergence time between all pairs of taxa, we estimated the percent sequence divergence from the average value of d_{XY} among each 500 kb window. Because diversity is measured within taxa rather than between them, we calculated the mean value of π between each pair of taxa.

APPENDIX A

SUPPLEMENTAL TABLE 1

Supplemental Table 1: GPS coordinates for sampling locations with taxon names given according to Tulig and Nesom (2012).

POP	LAT	LONG	Taxon	Morphological data (Y/N)
ELF	33.08595	-117.1453	*M. puniceus*	N
ELT	32.89421667	-117.0898167	*M. puniceus*	N
LH	33.06088333	-117.1187667	*M. puniceus*	N
T1	33.590639	-117.498333	*M. puniceus*	Y
T10	33.378139	-117.174333	*M. puniceus*	Y
T16	33.266861	-117.159389	*M. puniceus*	Y
UCSD	32.8894	-117.2361833	*M. puniceus*	N
ISL1	32.87361	-118.42611	*M. parviflorus*	Y
ISL2	32.87278	-118.41917	*M. parviflorus*	Y
SC	34.018	-119.673	*M. parviflorus*	N
M. clevelandii	33.158858	-116.812194	*M. clevelandii*	N
AGRD6	34.142931	-118.7647917	*M. longiflorus*	N
T110	33.99175	-117.993139	*M. longiflorus*	Y
T113	34.057194	-117.835083	*M. longiflorus*	Y
T24	34.233389	-118.3145	*M. longiflorus*	Y
T28	34.268083	-118.634667	*M. longiflorus*	Y
T33	34.34375	-118.509944	*M. longiflorus*	Y
T8	34.134722	-118.645222	*M. longiflorus*	Y
243-2	33.721253	-116.720761	*M. calycinus*	N
KR1	35.467298	-118.754533	*M. calycinus*	N
T124	35.427583	-120.547194	*M. calycinus*	N
T125	35.401139	-120.507139	*M. calycinus*	N
T144	34.192861	-117.278361	*M. calycinus*	Y
T149	33.901389	-116.862444	*M. calycinus*	Y
T150	33.856389	-116.848056	*M. calycinus*	Y
T66	35.707722	-118.775722	*M. calycinus*	Y
T89	35.649222	-118.458056	*M. calycinus*	Y
T90	35.591833	-118.505194	*M. calycinus*	Y
T91	35.317194	-118.587139	*M. calycinus*	Y
BAHR1	39.295667	-121.098114	*M. grandiflorus*	N
BCER	39.847733	-121.696267	*M. grandiflorus*	N
T100	39.379639	-121.146111	*M. grandiflorus*	Y

T101	39.553556	-121.430056	*M. grandiflorus*	Y
T61	39.559	-120.824278	*M. grandiflorus*	Y
T62	39.511806	-120.977639	*M. grandiflorus*	Y
T64	39.018194	-120.723889	*M. grandiflorus*	Y
T96	39.012194	-120.755194	*M. grandiflorus*	Y
T99	39.437556	-121.059861	*M. grandiflorus*	Y
LO	32.6767	-116.3312333	*M. australis*	N
PCT	32.73258333	-116.4698333	*M. australis*	Y
POTR	32.6038	-116.6339167	*M. australis*	Y
NVPR1	37.032444	-122.044972	*M. aurantiacus*	N
T102	39.042389	-122.772722	*M. aurantiacus*	Y
T103	39.202944	-123.051833	*M. aurantiacus*	Y
T104	39.204528	-123.764583	*M. aurantiacus*	Y
T105	39.197306	-123.746194	*M. aurantiacus*	Y
T106	38.642639	-123.394361	*M. aurantiacus*	Y
T50	35.986528	-121.492778	*M. aurantiacus*	Y
T51	36.284778	-121.839694	*M. aurantiacus*	Y
T56	36.781444	-121.703194	*M. aurantiacus*	Y
T57	37.032444	-122.044972	*M. aurantiacus*	Y
T92	37.845889	-120.610972	*M. aurantiacus*	Y
T93	38.296194	-120.675611	*M. aurantiacus*	Y
T94	38.302333	-120.692944	*M. aurantiacus*	Y
T95	38.266556	-120.735306	*M. aurantiacus*	Y
SD215	32.64056	-116.207712	*M. aridus*	N
KPP1	32.653778	-116.100556	*M. aridus*	Y
T84	32.652639	-116.244889	*M. aridus*	Y
DPR	33.745925	-117.448484	*M. rutilus*	N
VCR	33.487141	-117.648984	*M. australis*	N
RPR	33.605751	-117.802002	*M. puniceus*	N
MTH	33.614124	-117.825181	*M. puniceus*	N
TOR	33.661005	-117.640131	*M. puniceus*	N
T107	36.255806	-121.4315	*M. linearis*	Y
T46	35.65	-120.807444	*M. linearis*	Y
FLW	33.545745	-117.649203	*M. australis*	N
LGH	33.555947	-117.75956	*M. australis*	N
T42	34.73792	-120.4384	*M. lompocensis*	Y
SS	34.27222	-118.61	*M. rutilus*	N
LM	34.50498	-119.866	*M. lompocensis*	N

POM1	34.06039	-117.8289	*M. rutilus*	N
POM2	34.06039	-117.8289	*M. rutilus*	N
PID	33.50474	-117.73	*M. australis*	N
M. kelloggii	38.7717	-120.4450	*M. kelloggii*	N

APPENDIX B

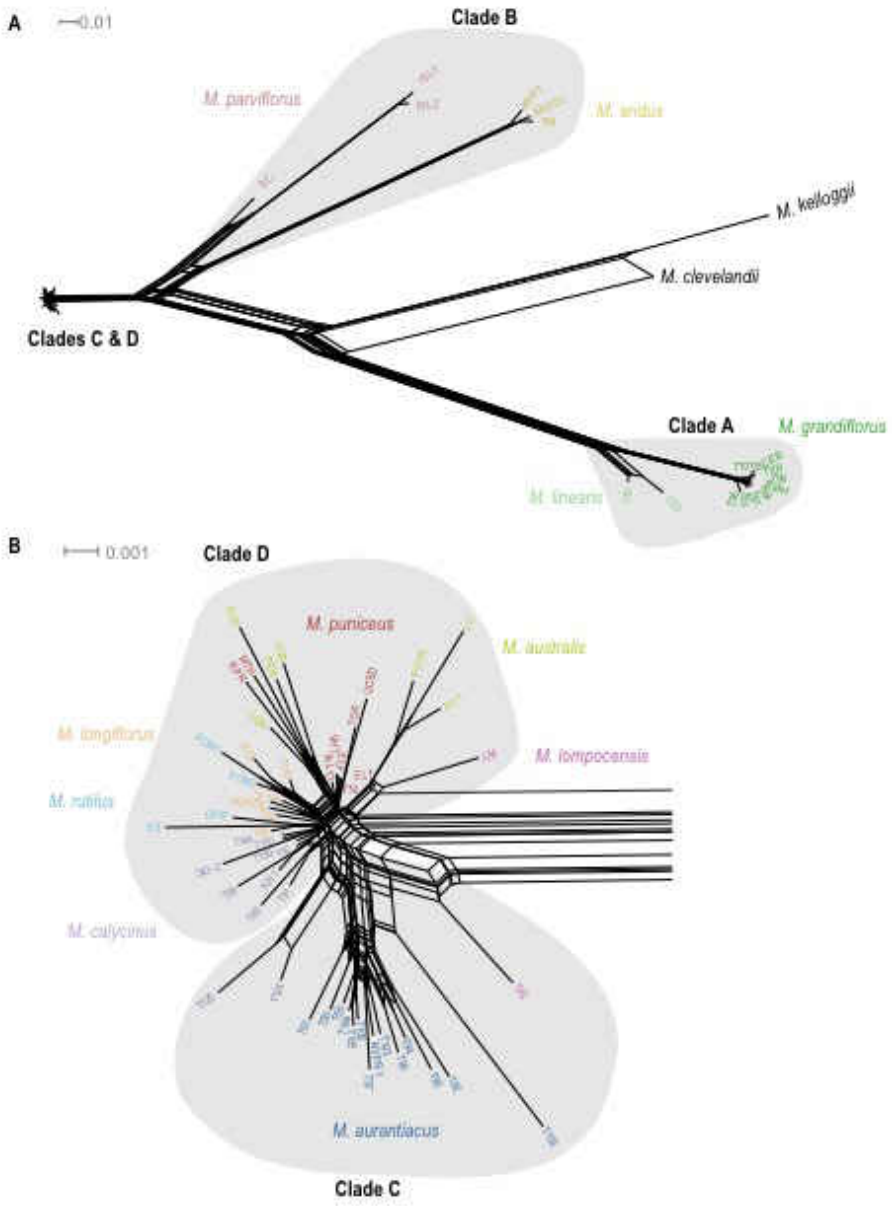
SUPPLEMENTAL TABLE 2

Trait abbreviations from Tulig (2000) used in principal components and discriminant functions analysis (Fig. 4).

Abbreviation	Trait measured
CRLN	Corolla length
BLLN	Length across bottom lobes
UCIS	Apex of upper corolla lobe to inner sinus
UCOS	Apex of upper corolla lobe to outer sinus
INFL	Height of inflection
WLCL	Width of lower center lobe
LLCL	Length of lower center lobe
THRO	Opening of throat
CRHT	Corolla height
CTLN	Corolla tube length
STLN	Style length
FLLN	Length of long filament
FSLN	Length of short filament
CULN	Calyx length to upper lobe
CLLN	Calyx length to lower sinus
CAHT	Calyx height
PDLN	Pedicel length
PDWD	Pedicel width

APPENDIX C

SUPPLEMENTAL FIGURE 1



Supplemental figure 1 (Next page). Split network representation of evolutionary relationships. A split network was calculated in SplitsTree under default settings, based on a dataset of 19, 294 SNPs. We present both (A) the whole network is presented with all taxa and (B) a closer view of Clades C and D. Sample names are given at the tips, and the color of the sample name corresponds with the color of the taxon name, based on the taxonomy of Tulig and Nesom (2012). Shading demonstrates the phylogenetic clade that samples belong to.

APPENDIX D

SUPPLEMENTAL TABLE 3

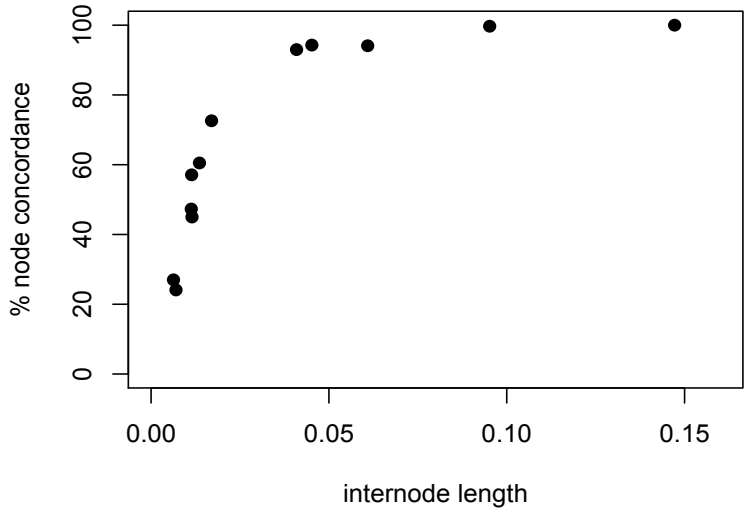
Supplemental Table 3. Summary of the the 37 whole genome samples, including their taxon identity, sampling location, percent read alignment, and total sequencing depth.

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	ssp. *longiflorus*	33.7459	-117.4485	96.0	26.88
SS	ssp. *longiflorus*	34.2722	-118.6100	94.2	30.86
T33	ssp. *longiflorus*	34.3438	-118.5099	94.6	18.87
T8	ssp. *longiflorus*	34.1347	-118.6452	82.6	25.11
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
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

APPENDIX E

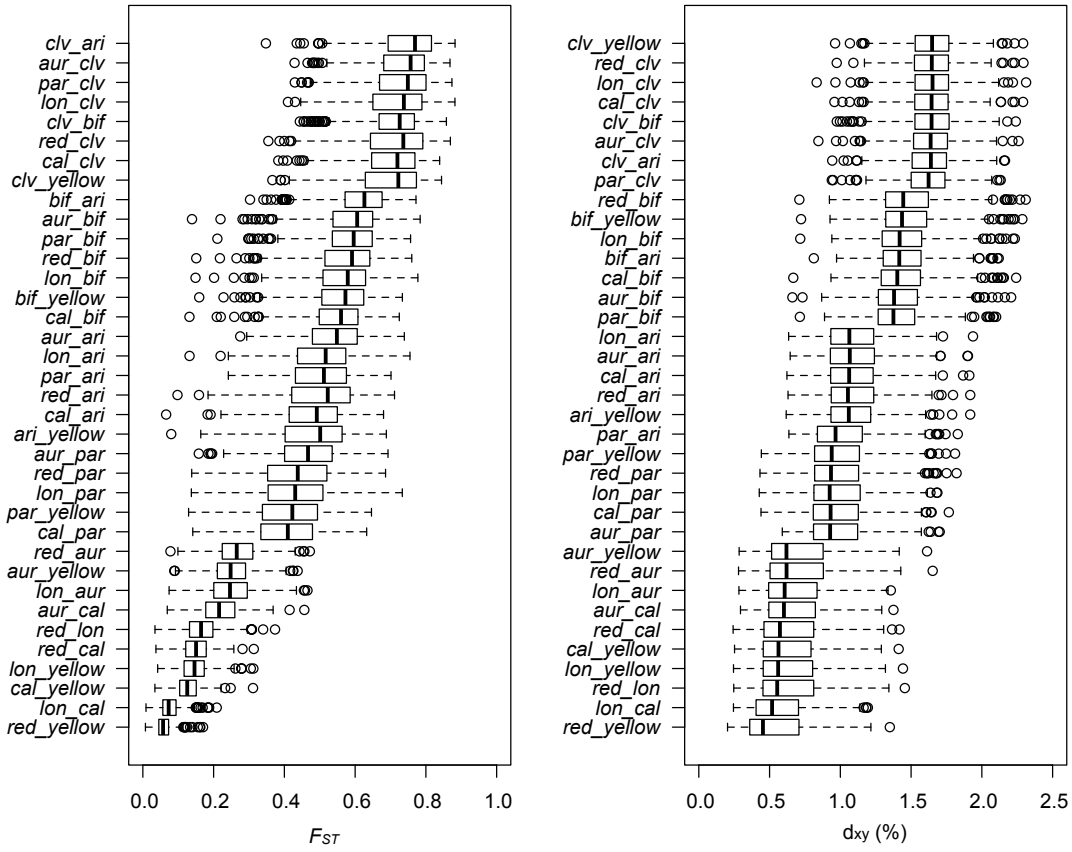
SUPPLEMENTAL FIGURE 2



Supplemental figure 2: Phylogenetic discordance is primarily the result of ILS. % Node concordance is calculated for each node in the whole-genome concatenated tree as the percent of 500kb window trees in which that node is present. Internode length is calculated from the distance matrix of the whole-genome concatenated tree.

APPENDIX F

SUPPLEMENTAL FIGURE 3



Supplemental figure 3. Box plots for each of the 36 pairwise taxonomic comparisons reveal differences in mean F_{ST} and d_{xy}, and show extensive variance among genomic windows within each comparison. Vertical black lines indicate the median, boxes represent the lower and upper quartiles, and whiskers extend to 1.5 times the interquartile range.

REFERENCES CITED

CHAPTER I

Berner, D. and Salzburger, W., 2015. The genomics of organismal diversification illuminated by adaptive radiations. *Trends in Genetics*, *31*(9), pp.491-499.

Cruickshank, T.E. and Hahn, M.W., 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular ecology*, *23*(13), pp.3133-3157.

Eaton, D.A. and Ree, R.H., 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology*, *62*(5), pp.689-706.

Ellegren, H., Smeds, L., Burri, R., Olason, P.I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A. and Uebbing, S., 2012. The genomic landscape of species divergence in Ficedula flycatchers. *Nature*, *491*(7426), pp.756-760.

Emerson, K.J., Merz, C.R., Catchen, J.M., Hohenlohe, P.A., Cresko, W.A., Bradshaw, W.E. and Holzapfel, C.M., 2010. Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the national academy of sciences*, *107*(37), pp.16196-16200.

Fontaine, M.C., Pease, J.B., Steele, A., Waterhouse, R.M., Neafsey, D.E., Sharakhov, I.V., Jiang, X., Hall, A.B., Catteruccia, F., Kakani, E. and Mitchell, S.N., 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, *347*(6217), p.1258524.

Gante, H.F., Matschiner, M., Malmstrøm, M., Jakobsen, K.S., Jentoft, S. and Salzburger, W., 2016. Genomics of speciation and introgression in Princess cichlid fishes from Lake Tanganyika. *Molecular ecology*, *25*(24), pp.6143-6161.

Hahn, M.W. and Nakhleh, L., 2016. Irrational exuberance for resolved species trees. *Evolution*, *70*(1), pp.7-17.

Heliconius Genome Consortium, 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, *487*(7405), pp.94-98.

Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A. and Cresko, W.A., 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS genetics*, *6*(2), p.e1000862.

Lamichhaney, S., Berglund, J., Almén, M.S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C.J., Wang, C., Zamani, N. and Grant, B.R., 2015. Evolution of Darwin/'s finches and their beaks revealed by genome sequencing. *Nature*, *518*(7539), pp.371-375.

Malinsky, M., Challis, R.J., Tyers, A.M., Schiffels, S., Terai, Y., Ngatunga, B.P., Miska, E.A., Durbin, R., Genner, M.J. and Turner, G.F., 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*, *350*(6267), pp.1493-1498.

Mallet, J., Besansky, N. and Hahn, M.W., 2016. How reticulated are species?. *BioEssays*, *38*(2), pp.140-149.

Martin, S.H., Dasmahapatra, K.K., Nadeau, N.J., Salazar, C., Walters, J.R., Simpson, F., Blaxter, M., Manica, A., Mallet, J. and Jiggins, C.D., 2013. Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome research*, *23*(11), pp.1817-1828.

McCluskey, B.M. and Postlethwait, J.H., 2015. Phylogeny of zebrafish, a "model species," within Danio, a "model genus". *Molecular biology and evolution*, *32*(3), pp.635-652.

Nadeau, N.J., Whibley, A., Jones, R.T., Davey, J.W., Dasmahapatra, K.K., Baxter, S.W., Quail, M.A., Joron, M., Blaxter, M.L., Mallet, J. and Jiggins, C.D., 2012. Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. *Phil. Trans. R. Soc. B*, *367*(1587), pp.343-353.

Pease, J.B., Haak, D.C., Hahn, M.W. and Moyle, L.C., 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biol*, *14*(2), p.e1002379.

Poelstra, J.W., Vijay, N., Bossu, C.M., Lantz, H., Ryll, B., Müller, I., Baglione, V., Unneberg, P., Wikelski, M., Grabherr, M.G. and Wolf, J.B., 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*, *344*(6190), pp.1410-1414.

Renaut, S., Grassa, C.J., Yeaman, S., Moyers, B.T., Lai, Z., Kane, N.C., Bowers, J.E., Burke, J.M. and Rieseberg, L.H., 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature communications*, *4*, p.1827.

Richards, E.J. and Martin, C.H., 2017. Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes. *PLoS genetics*, *13*(8), p.e1006919.

Schluter, D., 2000. *The ecology of adaptive radiation*. OUP Oxford.

Soria-Carrasco, V., Gompert, Z., Comeault, A.A., Farkas, T.E., Parchman, T.L., Johnston, J.S., Buerkle, C.A., Feder, J.L., Bast, J., Schwander, T. and Egan, S.P., 2014. Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, *344*(6185), pp.738-742.

Turner, T.L., Hahn, M.W. and Nuzhdin, S.V., 2005. Genomic islands of speciation in Anopheles gambiae. *PLoS Biol*, *3*(9), p.e285.

Twyford, A.D., Streisfeld, M.A., Lowry, D.B. and Friedman, J., 2015. Genomic studies on the nature of species: adaptation and speciation in Mimulus. *Molecular ecology*, *24*(11), pp.2601-2609.

Van Doren, B.M., Campagna, L., Helm, B., Illera, J.C., Lovette, I.J. and Liedvogel, M., 2017. Correlated patterns of genetic diversity and differentiation across an avian family. *Molecular ecology*, *26*(15), pp.3982-3997.

Vijay, N., Bossu, C.M., Poelstra, J.W., Weissensteiner, M.H., Suh, A., Kryukov, A.P. and Wolf, J.B., 2016. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature Communications*, *7*.

Vijay, N., Weissensteiner, M., Burri, R., Kawakami, T., Ellegren, H. and Wolf, J.B., 2017. Genome-wide patterns of variation in genetic diversity are shared among populations, species and higher order taxa. *Molecular ecology*.

Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., Sivasundar, A. and Seehausen, O., 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular ecology*, *22*(3), pp.787-798.

Wallbank, R.W., Baxter, S.W., Pardo-Diaz, C., Hanly, J.J., Martin, S.H., Mallet, J., Dasmahapatra, K.K., Salazar, C., Joron, M., Nadeau, N. and McMillan, W.O., 2016. Evolutionary novelty in a butterfly wing pattern through enhancer shuffling. *PLoS Biol*, *14*(1), p.e1002353.

Wu, C.I., 2001. The genic view of the process of speciation. *Journal of Evolutionary Biology*, *14*(6), pp.851-865.

Wu, C.A., Lowry, D.B., Cooley, A.M., Wright, K.M., Lee, Y.W. and Willis, J.H., 2008. Mimulus is an emerging model system for the integration of ecological and genomic studies. *Heredity*, *100*(2), p.220.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. and Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one*, *3*(10), p.e3376.

Beardsley, P.M., Schoenig, S.E., Whittall, J.B. and Olmstead, R.G., 2004. Patterns of evolution in western north american mimulus (Phrymaceae). *American Journal of Botany*, *91*(3), pp.474-489.

Berner, D. and Salzburger, W., 2015. The genomics of organismal diversification illuminated by adaptive radiations. *Trends in Genetics*, *31*(9), pp.491-499.

Beeks, R. M. 1962. Variation and hybridization in southern California populations of *Diplacus* (Scrophulariaceae). *El Aliso* 5:83-122.

Catchen, J.M., Amores, A., Hohenlohe, P., Cresko, W. and Postlethwait, J.H., 2011. Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics,* 1:171-182.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: an analysis tool set for population genomics. *Molecular ecology* 22:3124-3140.

Earl, D.A., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources*, *4*(2), pp.359-361.

Eaton, D.A. and Ree, R.H., 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology*, *62*(5), pp.689-706.

Ellegren, H., Smeds, L., Burri, R., Olason, P.I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A. and Uebbing, S., 2012. The genomic landscape of species divergence in Ficedula flycatchers. *Nature*, *491*(7426), pp.756-760.

Etter, P.D., Bassham, S., Hohenlohe, P.A., Johnson, E.A. and Cresko, W.A. 2011. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. In Molecular methods for evolutionary genetics (pp. 157-178). Humana Press.

Emerson, K.J., Merz, C.R., Catchen, J.M., Hohenlohe, P.A., Cresko, W.A., Bradshaw, W.E. and Holzapfel, C.M., 2010. Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the national academy of sciences*, *107*(37), pp.16196-16200.

Fontaine, M.C., Pease, J.B., Steele, A., Waterhouse, R.M., Neafsey, D.E., Sharakhov, I.V., Jiang, X., Hall, A.B., Catteruccia, F., Kakani, E. and Mitchell, S.N., 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, *347*(6217), p.1258524.

Grant, A.L., 1924. A monograph of the genus Mimulus. *Annals of the Missouri Botanical Garden*, *11*(2/3), pp.99-388.

Grant, V. 1993a. Effect of hybridization and selection on floral isolation. *Proceedings of the National Academy of Sciences of the United States of America* 90:990-993.

Grant, V. 1993b. Origin of floral isolation between ornithophilous and sphingophilous plant species. *Proceedings of the National Academy of Sciences of the United States of America* 90:7729-7733.

Grant, V. 1994. Modes and origins of mechanical and ethological isolation in angiosperms. *Proceedings of the National Academy of Sciences of the United States of America* 91:3-10.

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.Y. and Hansen, N.F., 2010. A draft sequence of the Neandertal genome. *science*, *328*(5979), pp.710-722.

Gross, B.L. and Rieseberg, L.H., 2004. The ecological genetics of homoploid hybrid speciation. *Journal of heredity*, *96*(3), pp.241-252.

Hahn, M.W. and Nakhleh, L., 2016. Irrational exuberance for resolved species trees. *Evolution*, *70*(1), pp.7-17.

Handelman, C. and Kohn, J.R., 2014. Hummingbird color preference within a natural hybrid population of Mimulus aurantiacus (Phrymaceae). *Plant Species Biology*, *29*(1), pp.65-72.

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

Heliconius Genome Consortium, 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, *487*(7405), pp.94-98.

Huson, D.H. and Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution*, *23*(2), pp.254-267.

Jakobsson, M. and 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), pp.1801-1806.

Jarvis, E.D., Mirarab, S., Aberer, A.J., Li, B., Houde, P., Li, C., Ho, S.Y., Faircloth, B.C., Nabholz, B., Howard, J.T. and Suh, A., 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, *346*(6215), pp.1320-1331.

Keller, I., Wagner, C.E., Greuter, L., Mwaiko, S., Selz, O.M., Sivasundar, A., Wittwer, S. and Seehausen, O., 2013. Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology*, *22*(11), pp.2848-2863.

Lamichhaney, S., Berglund, J., Almén, M.S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C.J., Wang, C., Zamani, N. and Grant, B.R., 2015. Evolution of Darwin/'s finches and their beaks revealed by genome sequencing. *Nature*, *518*(7539), pp.371-375.

Langmead, B. and Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods*, *9*(4), p.357.

Maddison, W.P., 1997. Gene trees in species trees. *Systematic biology*, *46*(3), pp.523-536.

Mahler, D.L., Ingram, T., Revell, L.J. and Losos, J.B., 2013. Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science*, *341*(6143), pp.292-295.

Malinsky, M., Challis, R.J., Tyers, A.M., Schiffels, S., Terai, Y., Ngatunga, B.P., Miska, E.A., Durbin, R., Genner, M.J. and Turner, G.F., 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*, *350*(6267), pp.1493-1498.

Malinsky, Milan; Svardal, H., Tyers, A.M., Miska, E. A., Genner, M. J., Turner, G.F., and Durbin, R., 2017. Whole genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *BioRxiv 143859.*

Mallet, J., Besansky, N. and Hahn, M.W., 2016. How reticulated are species?. *BioEssays*, *38*(2), pp.140-149.

Mayr, E., 1995. Species, classification, and evolution. *Biodiversity and Evolution. National Science Museum Foundation, Tokyo*, pp.3-122.

McCluskey, B.M. and Postlethwait, J.H., 2015. Phylogeny of zebrafish, a “model species,” within Danio, a “model genus”. *Molecular biology and evolution*, *32*(3), pp.635-652.

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

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A. and Johnson, E.A., 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome research*, *17*(2), pp.240-248.

Munz, P.A., 1935. Manual of southern California botany.

Munz PA, Keck DD. 1959 A California flora. Berkeley, CA: University of California Press.

Munz PA, Keck DD. 1973 A California flora and supplement. Berkeley, CA: University of California Press.

Muschick, M., Indermaur, A. and Salzburger, W., 2012. Convergent evolution within an adaptive radiation of cichlid fishes. *Current biology*, *22*(24), pp.2362-2368.

Nosil, P. 2012. *Ecological speciation*. Oxford University Press.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T. and Reich, D., 2012. Ancient admixture in human history. *Genetics*, *192*(3), pp.1065-1093.

Pease, J.B., Haak, D.C., Hahn, M.W. and Moyle, L.C., 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biol*, *14*(2), p.e1002379.

Pennell, F.W., 1951. Mimulus. *Illustrated flora of the Pacific states*, *3*, pp.688-731.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.

Yin-Long, Q., Lee, J., Bernasconi-Quadroni, F. and Soltis, D.E., 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature*, *402*(6760), p.404.

Reich, D., Thangaraj, K., Patterson, N., Price, A.L. and Singh, L., 2009. Reconstructing Indian population history. *Nature*, *461*(7263), pp.489-494.

Rieseberg, L.H., Whitton, J. and Gardner, K., 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, *152*(2), pp.713-727.

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

Rundle, H. D., and P. Nosil, P. 2005. Ecological speciation. *Ecology letters* 8:336-352.

Schluter, D., 2000. *The ecology of adaptive radiation*. OUP Oxford.

Schluter, D., 2009. Evidence for ecological speciation and its alternative. *Science*, *323*(5915), pp.737-741.

Sobel, J.M., and M. A. Streisfeld. 2015. Strong premating isolation exclusively drives insipient speciation in *Mimulus aurantiacus. Evolution* 69:447-461.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), pp.1312-1313.

Stankowski, S., and M.A. Streisfeld. 2015. Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. In *Proc. R. Soc. B*: 282: p. 20151666.

Stankowski, S., J. M Sobel and M. A. Streisfeld. 2015. The geography of divergence with gene flow facilitates multitrait adaptation and the evolution of pollinator isolation in Mimulus aurantiacus. *Evolution* 69:3054-3068.

Stankowski, S., J. M Sobel and M. A. Streisfeld. 2017. Geographic cline analysis as a tool for studying genome-wide variation: a case study of pollinator-mediated divergence in a monkeyflower.

Streisfeld, M. A., and J. R. Kohn. 2005. Contrasting patterns of floral and molecular variation across a cline in *Mimulus aurantiacus*. *Evolution* 59:2548-2559.

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

Streisfeld, M. A., W. N. Young, and J. M. Sobel. 2013. Divergent selection drives genetic differentiation in an R2R3-Myb transcription factor that contributes to incipient speciation in *Mimulus aurantiacus*. *PLoS genetics* 9:e1003385.

Thompson DM. 1993 Mimulus. In The Jepson manual; higher plants of California (ed. JC Hickman), pp. 1037–1047. Los Angeles, CA: University of California Press.

Thompson, D.M. 2005. Systematics of Mimulus subgenus Schizoplacus (Scrophulariaceae). Syst. Bot. Monogr. 75: 1–213.

Thompson, D.M. 2012. Mimulus. Pp. 988–998, in B.G. Baldwin, D.H. Goldman, D.J. Keil, R.Patterson, and T.J. Rosatti (eds.), The Jepson Manual: Vascular Plants of California (ed. 2). Univ. of California Press, Berkeley.

Tulig, M. 2000. Morphological variation in Mimulus section Diplacus (Scrophulariaceae). Doctoral dissertation, California State Polytechnic University.

Turner, T.L., Hahn, M.W. and Nuzhdin, S.V., 2005. Genomic islands of speciation in Anopheles gambiae. *PLoS Biol*, *3*(9), p.e285.

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

Vijay, N., Bossu, C.M., Poelstra, J.W., Weissensteiner, M.H., Suh, A., Kryukov, A.P. and Wolf, J.B., 2016. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature Communications*, *7*.

Wallbank, R.W., Baxter, S.W., Pardo-Diaz, C., Hanly, J.J., Martin, S.H., Mallet, J., Dasmahapatra, K.K., Salazar, C., Joron, M., Nadeau, N. and McMillan, W.O., 2016. Evolutionary novelty in a butterfly wing pattern through enhancer shuffling. *PLoS Biol*, *14*(1), p.e1002353.

Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., Sivasundar, A. and Seehausen, O., 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular ecology*, *22*(3), pp.787-798.

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*, pp.53-65.

Wessinger, C.A., Freeman, C.C., Mort, M.E., Rausher, M.D. and Hileman, L.C., 2016. Multiplexed shotgun genotyping resolves species relationships within the North American genus Penstemon. *American journal of botany*, *103*(5), pp.912-922.

Wolfe, A.D., Randle, C.P., Datwyler, S.L., Morawetz, J.J., Arguedas, N. and Diaz, J., 2006. Phylogeny, taxonomic affinities, and biogeography of Penstemon (Plantaginaceae) based on ITS and cpDNA sequence data. *American Journal of Botany*, *93*(11), pp.1699-1713.

Wu, C.I., 2001. The genic view of the process of speciation. *Journal of Evolutionary Biology*, *14*(6), pp.851-865.

CHAPTER III

Amores, A., Catchen, J., Nanda, I., Warren, W., Walter, R., Schartl, M. and Postlethwait, J.H., 2014. A RAD-tag genetic map for the platyfish (Xiphophorus maculatus) reveals mechanisms of karyotype evolution among teleost fish. *Genetics*, *197*(2), pp.625-641.

Begun, D.J., Holloway, A.K., Stevens, K., Hillier, L.W., Poh, Y.P., Hahn, M.W., Nista, P.M., Jones, C.D., Kern, A.D., Dewey, C.N. and Pachter, L., 2007. Population genomics: whole-genome analysis of polymorphism and divergence in Drosophila simulans. *PLoS biology*, *5*(11), p.e310.

Bouckaert, R.R., 2010. DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics*, *26*(10), pp.1372-1373.

Brandvain, Y., Kenney, A.M., Flagel, L., Coop, G. and Sweigart, A.L., 2014. Speciation and introgression between Mimulus nasutus and Mimulus guttatus. *PLoS Genetics*, *10*(6), p.e1004410.

Browning, S.R. and 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), pp.1084-1097.

Burri, R., Nater, A., Kawakami, T., Mugal, C.F., Olason, P.I., Smeds, L., Suh, A., Dutoit, L., Bureš, S., Garamszegi, L.Z. and Hogner, S., 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), pp.1656-1665.

Burri, R., 2017a. Linked selection, demography and the evolution of correlated genomic landscapes in birds and beyond. *Molecular ecology*, *26*(15), pp.3853-3856.

Burri, R., 2017b. Interpreting differentiation landscapes in the light of long-term linked selection. *Evolution Letters*, *1*(3), pp.118-131.

Cantarel, B.L., Korf, I., Robb, S.M., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A.S. and Yandell, M., 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome research*, *18*(1), pp.188-196.

Catchen, J.M., Amores, A., Hohenlohe, P., Cresko, W. and Postlethwait, J.H., 2011. Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics,* 1:171-182.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: an analysis tool set for population genomics. *Molecular ecology* 22:3124-3140.

Charif, D. and Lobry, J.R., 2007. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In *Structural approaches to sequence evolution* (pp. 207-232). Springer, Berlin, Heidelberg.

Charlesworth, B., Morgan, M.T. and Charlesworth, D., 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics*, *134*(4), pp.1289-1303.

Charlesworth, B., 2012. The effects of deleterious mutations on evolution at linked sites. *Genetics*, *190*(1), pp.5-22.

Cruickshank, T.E. and Hahn, M.W., 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular ecology*, *23*(13), pp.3133-3157.

Chase, M.A., Stankowski, S. and 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), pp.1510-1521.

Cutter, A.D. and Payseur, B.A., 2013. Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature Reviews Genetics*, *14*(4), p.262.

Ellegren, H., Smeds, L., Burri, R., Olason, P.I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A. and Uebbing, S., 2012. The genomic landscape of species divergence in Ficedula flycatchers. *Nature*, *491*(7426), pp.756-760.

Ellegren, H. and Wolf, J.B., 2017. Parallelism in genomic landscapes of differentiation, conserved genomic features and the role of linked selection. *Journal of evolutionary biology*, *30*(8), pp.1516-1518.

English, A.C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D.M., Reid, J.G., Worley, K.C. and Gibbs, R.A., 2012. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PloS one*, *7*(11), p.e47768.

Eyre-Walker, A. and Keightley, P.D., 2007. The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, *8*(8), p.610.

Gao, S., Bertrand, D., Chia, B.K. and Nagarajan, N., 2016. OPERA-LG: efficient and exact scaffolding of large, repeat-rich eukaryotic genomes with performance guarantees. *Genome biology*, *17*(1), p.102.

Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., Sharpe, T., Hall, G., Shea, T.P., Sykes, S. and Berlin, A.M., 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences*, *108*(4), pp.1513-1518.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N. and Rokhsar, D.S., 2011. Phytozome: a

comparative platform for green plant genomics. *Nucleic acids research*, *40*(D1), pp.D1178-D1186.

Hahn, M.W., 2008. Toward a selection theory of molecular evolution. *Evolution*, *62*(2), pp.255-265.

Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A. and Cresko, W.A., 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS genetics*, *6*(2), p.e1000862.

Hohenlohe, P.A., Bassham, S., Currey, M. and Cresko, W.A., 2012. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Phil. Trans. R. Soc. B*, *367*(1587), pp.395-408.

Holt, C. and Yandell, M., 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC bioinformatics*, *12*(1), p.491.

Jarvis, E.D., Mirarab, S., Aberer, A.J., Li, B., Houde, P., Li, C., Ho, S.Y., Faircloth, B.C., Nabholz, B., Howard, J.T. and Suh, A., 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, *346*(6215), pp.1320-1331.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O. and Walichiewicz, J., 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and genome research*, *110*(1-4), pp.462-467.

Kersey, P.J., Allen, J.E., Armean, I., Boddu, S., Bolt, B.J., Carvalho-Silva, D., Christensen, M., Davis, P., Falin, L.J., Grabmueller, C. and Humphrey, J., 2015. Ensembl Genomes 2016: more genomes, more complexity. *Nucleic acids research*, *44*(D1), pp.D574-D580.

Korf, I., 2004. Gene finding in novel genomes. *BMC bioinformatics*, *5*(1), p.59.

Lamichhaney, S., Berglund, J., Almén, M.S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C.J., Wang, C., Zamani, N. and Grant, B.R., 2015. Evolution of Darwin/'s finches and their beaks revealed by genome sequencing. *Nature*, *518*(7539), pp.371-375.

Langmead, B. and Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods*, *9*(4), p.357.

Lewontin, R.C., 1974. *The genetic basis of evolutionary change* (Vol. 560). New York: Columbia University Press.

Li, H. and Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), pp.1754-1760.

Malinsky, M., Challis, R.J., Tyers, A.M., Schiffels, S., Terai, Y., Ngatunga, B.P., Miska, E.A., Durbin, R., Genner, M.J. and Turner, G.F., 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*, *350*(6267), pp.1493-1498.

Martin, S.H., Dasmahapatra, K.K., Nadeau, N.J., Salazar, C., Walters, J.R., Simpson, F., Blaxter, M., Manica, A., Mallet, J. and Jiggins, C.D., 2013. Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome research*, *23*(11), pp.1817-1828.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. and DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, *20*(9), pp.1297-1303.

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

Nordberg, H., Cantor, M., Dusheyko, S., Hua, S., Poliakov, A., Shabalov, I., Smirnova, T., Grigoriev, I.V. and Dubchak, I., 2013. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic acids research*, *42*(D1), pp.D26-D31.

O'brien, K.P., Remm, M. and Sonnhammer, E.L., 2005. Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic acids research*, *33*(suppl_1), pp.D476-D480.

Paradis, E., Claude, J. and Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), pp.289-290.

Parra, G., Bradnam, K. and Korf, I., 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, *23*(9), pp.1061-1067.

Pease, J.B., Haak, D.C., Hahn, M.W. and Moyle, L.C., 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biol*, *14*(2), p.e1002379.

Pease, J. and Rosenzweig, B., 2015. Encoding data using biological principles: the Multisample Variant Format for phylogenomics and population genomics. *IEEE/ACM transactions on computational biology and bioinformatics*.

Poelstra, J.W., Vijay, N., Bossu, C.M., Lantz, H., Ryll, B., Müller, I., Baglione, V., Unneberg, P., Wikelski, M., Grabherr, M.G. and Wolf, J.B., 2014. The genomic

landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*, *344*(6190), pp.1410-1414.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R. and Lopez, R., 2005. InterProScan: protein domains identifier. *Nucleic acids research*, *33*(suppl_2), pp.W116-W120.

Rastas, P., Calboli, F.C., Guo, B., Shikano, T. and Merilä, J., 2015. Construction of ultradense linkage maps with Lep-MAP2: stickleback F 2 recombinant crosses as an example. *Genome biology and evolution*, *8*(1), pp.78-93.

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), pp.1450-1477.

Renaut, S., Grassa, C.J., Yeaman, S., Moyers, B.T., Lai, Z., Kane, N.C., Bowers, J.E., Burke, J.M. and Rieseberg, L.H., 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature communications*, *4*, p.1827.

Richards, E.J. and Martin, C.H., 2017. Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes. *PLoS genetics*, *13*(8), p.e1006919.

Seehausen, O., Butlin, R.K., Keller, I., Wagner, C.E., Boughman, J.W., Hohenlohe, P.A., Peichel, C.L., Saetre, G.P., Bank, C., Brännström, Å. and Brelsford, A., 2014. Genomics and the origin of species. *Nature Reviews Genetics*, *15*(3), p.176.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), pp.3210-3212.

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

Sobel, J.M., and M. A. Streisfeld. 2015. Strong premating isolation exclusively drives insipient speciation in *Mimulus aurantiacus. Evolution* 69:447-461.

Soria-Carrasco, V., Gompert, Z., Comeault, A.A., Farkas, T.E., Parchman, T.L., Johnston, J.S., Buerkle, C.A., Feder, J.L., Bast, J., Schwander, T. and Egan, S.P., 2014. Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, *344*(6185), pp.738-742.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), pp.1312-1313.

Stanke, M. and Waack, S., 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*, *19*(suppl_2), pp.ii215-ii225.

Stankowski, S., and M.A. Streisfeld. 2015. Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. In *Proc. R. Soc. B*: 282: p. 20151666.

Stankowski, S., J. M Sobel and M. A. Streisfeld. 2015. The geography of divergence with gene flow facilitates multitrait adaptation and the evolution of pollinator isolation in Mimulus aurantiacus. *Evolution* 69:3054-3068.

Stankowski, S., J. M Sobel and M. A. Streisfeld. 2017. Geographic cline analysis as a tool for studying genome-wide variation: a case study of pollinator-mediated divergence in a monkeyflower.

Streisfeld, M. A., and J. R. Kohn. 2005. Contrasting patterns of floral and molecular variation across a cline in *Mimulus aurantiacus*. *Evolution* 59:2548-2559.

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

Streisfeld, M. A., W. N. Young, and J. M. Sobel. 2013. Divergent selection drives genetic differentiation in an R2R3-Myb transcription factor that contributes to incipient speciation in *Mimulus aurantiacus*. *PLoS genetics* 9:e1003385.

Suh, A., Smeds, L. and Ellegren, H., 2015. The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLoS biology*, *13*(8), p.e1002224.

Thompson, D.M. 2012. Mimulus. Pp. 988–998, in B.G. Baldwin, D.H. Goldman, D.J. Keil, R.Patterson, and T.J. Rosatti (eds.), The Jepson Manual: Vascular Plants of California (ed. 2). Univ. of California Press, Berkeley.

Turner, T.L., Hahn, M.W. and Nuzhdin, S.V., 2005. Genomic islands of speciation in Anopheles gambiae. *PLoS Biol*, *3*(9), p.e285.

Van Doren, B.M., Campagna, L., Helm, B., Illera, J.C., Lovette, I.J. and Liedvogel, M., 2017. Correlated patterns of genetic diversity and differentiation across an avian family. *Molecular ecology*, *26*(15), pp.3982-3997.

Vickery, R.K., 1995. Speciation by aneuploidy and polyploidy in Mimulus (Scrophulariaceae). *The Great Basin Naturalist*, *55*(2), pp.174-176.

Vijay, N., Bossu, C.M., Poelstra, J.W., Weissensteiner, M.H., Suh, A., Kryukov, A.P. and Wolf, J.B., 2016. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature Communications*, *7*.

Vijay, N., Weissensteiner, M., Burri, R., Kawakami, T., Ellegren, H. and Wolf, J.B., 2017. Genome-wide patterns of variation in genetic diversity are shared among populations, species and higher order taxa. *Molecular ecology*.

Wolf, J.B. and Ellegren, H., 2017. Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, *18*(2), p.87.

Wu, C.I., 2001. The genic view of the process of speciation. *Journal of Evolutionary Biology*, *14*(6), pp.851-865.