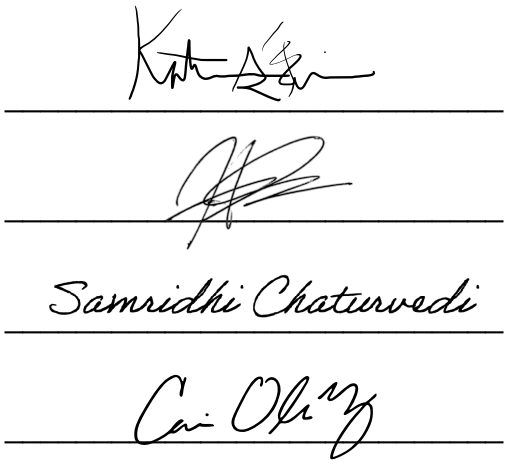
Genomic Insights into Species Divergence: Unraveling Genetic Architecture Amid Gene Flow in the *Mimulus guttatus* Complex

A Master's Thesis
Submitted on the Second day of May, 2024
to the Department of Ecology & Evolutionary Biology
Tulane University
In Fulfillment of the Requirements
For the degree of
Masters of Science
By

Charlotte M. Hankin



1 | Introduction

Understanding how adaptation may drive force for the process of speciation is crucial to advancing our knowledge of how evolutionary processes shape biodiversity. The formation of new species occurs by the accumulation of reproductive isolating barriers, which prevent gene flow and introgression (Coyne & Orr, 2004; Gould et al., 2017; Lowry et al., 2008; Sobel & Chen 2014). Historically, speciation was believed to occur due to geographical isolation (allopatric speciation) facilitating rapid genetic divergence through selection and drift, though recent work has pointed to speciation in absence of geographic barriers (sympatric speciation) as as also contributing to diversification (Kautt et al., 2020). Sympatric speciation is a surprising scenario as neighboring populations experience ongoing gene flow which is known to be a homogenizing force (Pinho & Hey, 2010). As genetic material is exchanged between populations, recombination events will break up linkage groups and make the populations more genetically similar to one another. However, heavily linked loci under selection can create islands of differentiation that essentially do not experience gene flow in the same way that neutral loci do (Pinho & Hey, 2010; Wu 2001).

Disruptive selection can facilitate the speciation process in the face of gene flow when micro-geographic differences in habitat are selecting for traits at different optima in the fitness landscape (Pinho & Hey, 2010). This divergent adaptation sometimes generates barriers to gene flow, allowing for both adaptive and neutral divergence to occur within the genomes (Hendrick, 2016; Levin, 2009). Habitat differentiation may

lead to increased divergence between populations (Gould et al., 2017; Hendrick, 2016) however the genetic architecture of speciation that occurs in the face of gene flow is still not well understood (Emelianov et al., 2004; Kautt et al., 2020; Kliman et al. 2000; Machado & Hey, 2003; Navarro & Barton, 2003; Rieseberg et al. 1999; Wu, 2001). Selective pressures often drive the evolution of a complex array of phenotypic changes allowing for local adaptation (Gould et al., 2017), but whether this process involves many different loci across the genome, or few genomic islands of differentiation, remains unknown.

Theory states that under prolonged bouts of adaptation in the face of gene flow the genetic architecture of locally adapted traits will gradually move from many small effect alleles to fewer large effect alleles or linkage groups of smaller effect alleles that together act as a large pleiotropic allele (Yeaman & Whitlock, 2011).). Large effect loci have larger selection coefficients and therefore stronger selection at these loci is more likely to maintain divergence in the face of gene flow. When locally adaptive loci are physically linked, the chances of recombination breaking up positive linkage disequilibrium generated by selection is decreased and the adaptive traits are inherited together, avoiding maladaptive intermediate genotypes (Lenormand and Otto 2000; Maynard Smith, 1977; Pylkov et al., 1998; Yeaman & Whitlock, 2011).

This theory has support from many genomic studies (Emelianov et al., 2004; Seehausen, O. et al., 2014) that have found that differentiation commonly takes the form of “barrier loci” which are loci of large effect restricted to few genomic regions (Kautt et al., 2020; Malinsky et al., 2015; Marques et al., 2016; Poelstra et al., 2014; Riesch et al., 2017; Westram et al., 2018). These expected “barrier loci” are

characterized as clustered genomic distributions of between-habitat genetic differences (Emelianov et al., 2004). They are known as islands of differentiation (Pinho & Hey, 2010; Wu, 2001) as the regions of linkage disequilibrium resists the homogenizing effects of gene flow by prevents recombination between locally adaptive loci (Gavrilets, 2017; Maynard Smith, 1966; Pinho & Hey, 2010; Servedio et al., 2011; Smadja & Butlin, 2011; Wu, 2001). The barrier loci theory has recently been demonstrated in the genomes of *Helianthus* (Rieseberg et al. 1999), *Drosophila* (Kliman et al. 2000; Machado & Hey, 2003; Wu, 2001) larch budmoth (Emelianov et al., 2004) and between *Homo and Pan* (Navarro & Barton, 2003).

In contrast to this “barrier loci” theory, recent work has pointed to the reverse genetic architecture as a possible mode of speciation with gene flow. Instead of simple architectures made up of few loci of large-effect resisting gene flow through linkage disequilibrium, the formation of ecotypes can also form through changes in allele frequencies at many loci across the genome underlying multiple adaptations working in a concerted fashion to cause reproductive isolation (Coyne & Orr 2004; Gould et al., 2017; Nosil, 2012; Schluter, 2009). Support for this multi-locus theory (Barton, 1983; Kautt et al., 2020) as a mode of divergence with gene flow is strengthened by genomic simulations (Feder et al., 2014; Flaxman et al., 2014), showing that the synergistic effects of many alleles with weak selection coefficients are able to promote the rapid accumulation of genome-wide differentiation (Barton & Bengtsson, 1986; Kautt et al., 2020; Nosil et al., 2017; Riesch et al., 2017). While many studies have found evidence of barrier loci (Emelianov et al., 2004; Kliman et al. 2000; Machado & Hey, 2003; Navarro & Barton, 2003; Rieseberg et al. 1999; Wu, 2001), the multi-locus theory has

recently been demonstrated as the mode of speciation with gene flow within a species complex of neotropical cichlid fishes (Kautt et al., 2020). This leads to the question: *Is divergence in the face of gene flow more often driven by many small-effect loci working in a concerted fashion, or few large-effect barrier loci?*

Candidate regions of the genome contributing to ecotype differentiation and reproductive barriers in the face of gene flow can be detected through the genomic analysis of ecologically differentiated incipient species that are not yet completely reproductively isolated (Emelianov et al., 2004). Divergence mapping using whole-genome sequencing data from pooled populations (Poolseq) can identify loci involved in adaptive trait divergence in sympatry (Gould et al., 2017; Schlöttereret et al., 2014). While taking a PoolSeq approach is not ideal for estimating parameters of a single population (Gould et al., 2017; Lynch et al. 2014) it is an effective method of identifying candidate genomic regions of differentiation between populations (Cheng et al., 2012; Fabian et al., 2012; Flagel et al., 2014a,b; Gould et al., 2017; Kapun et al., 2016; Kofler et al., 2012; Magwene et al., 2011; Raineri et al., 2012; Turner et al., 2010).

The *Mimulus guttatus* species complex is an excellent system to study how genetic architecture evolves when populations in sympatry experience disruptive selection toward different trait optima (Ferris, 2016; Yeaman & Otto, 2011). This species complex is a group of closely related interfertile wildflowers characterized by ongoing introgression and is a model system for studying microgeographic adaptation in sympatry (Brandvain et al., 2014; Wu et al., 2008). Ecotypes are differentially adapted to a variety of edaphic environments throughout western North America (Ferris et al., 2016; Ferris & Willis, 2018; Wu et al., 2007). Additionally, genomic analyses of trait

differentiation are easily conducted as a wealth of information about genomic structure is publicly available, including a fully sequenced and annotated genome of *M. guttatus* (Ferris et al., 2015; Wu et al., 2007; www.phytozome.org). Due to its high levels of genetic diversity and wide geographic range the ancestor of the species complex is expected to be *M. guttatus*-like (Ferris et al., 2015; Modliszewski & Willis, 2012; Sweigart & Willis, 2003). From this common ancestor, many ecotypes of *Mimulus* have since evolved, often through colonizing harsh environments. While *M. guttatus* perennially occurs in moist streams and seeps, many of its sympatric close relatives have populated habitats characterized by their low soil moisture (Ferris et al., 2015; Ferris & Willis, 2018; Wu et al., 2007). One such ecotype, *M. laciniatus*, has adapted to granite outcrops with shallow rocky soils with low water retention, extreme temperature fluctuations, and high light intensity (Ferris et al. 2014; Ferris & Willis, 2018; Tataru et al., 2023). *Mimulus laciniatus* has adaptively diverged from its close relative *M. guttatus* despite the two frequently being found within meters of each other (Ferris et al. 2016; Ferris & Willis, 2018; Wu et al. 2007). Divergent traits that are adaptive in *M. laciniatus's* rocky outcrops are (1) earlier flowering time than sympatric *M. guttatus* (Ferris et al., 2016; Ferris & Willis, 2018; Friedman & Willis, 2013), (2) a self-fertilizing mating system (Fenster & Ritland, 1994; Ferris et al., 2014; Ferris & Willis, 2018), (3) small stature (Ferris & Willis, 2018) and (4) a highly lobed leaf morphology (Ferris et al., 2015; Ferris & Willis, 2018). These traits are in contrast to the late-flowering, predominantly outcrossing, larger and round-leaved *M. guttatus* (Wu et al., 2007; Ferris et al., 2016; Ferris & Willis, 2018). Despite these major ecotypic differences in traits that commonly lead to reproductive isolation such as temporal isolation through flowering time and

mating system isolation, there is evidence of ongoing gene flow through the existence of stable hybrid zones between the two species (Vickery 1968; D. Tataru unpublished data).

To understand how adaptive divergence was able to occur despite the homogenizing effects of gene flow we used a pooled whole-genome resequencing approach to locate the genomic regions under disruptive selection responsible for ecological adaptation between *M. laciniatus* and *M. guttatus*, we used a pooled whole-genome resequencing approach. A previous quantitative trait loci (QTL) mapping experiment by Ferris et al (2017) identified regions underlying divergence in flowering time, mating system, and leaf shape between the two *Mimulus* species at a sympatric site near Shaver Lake, California. This Shaver Lake population also contains both phenotypic and genomic evidence of ongoing gene flow between species (Ferris et al 2017; D. Tataru unpublished data). All five traits relating to morphological and life history adaptations to the contrasting microhabitats were shown to have a genetically simple basis (Ferris et al. 2017). Four to five QTLs of large-to-moderate effect were identified for each trait with most of the variance seemingly explained by a major effect pleiotropic locus on chromosome 8 (LG8b; Ferris et al., 2017). The fact that few loci of large effect underlie adaptive trait divergence in this sympatric population of *M. laciniatus* and *M. guttatus* lends support to the “barrier loci” theory of speciation with gene flow. However, with a moderate sample size of 424 F2s, it is possible that some loci of small effect were not captured by this study (Beavis, 1998; Xu 2003) and of course there are many other unmeasured traits that might be important for adaptation and reproductive isolation between these species. Previous QTL mapping studies of

similar power between *M. guttatus* and *Mimulus nasutus* have found divergence in flower size, a trait related to mating system isolation, to be controlled by many QTL of small effect (Fishman et al., 2002, 2015; Hall et al,. 2006). Furthermore, because of limited recombination in early generation hybrids, QTL regions often span large sections of the genome which could collapse separate but physically proximate causal loci into a single locus. By analyzing genome-wide patterns of divergence between populations of *M. laciniatus* and *M. guttatus* we can either confirm the simplistic genetic architecture of divergence in this system or determine if there is a more complex mechanism at work.

While QTL mapping is a useful method for identifying chromosomal regions containing loci controlling for a quantitative trait (Mohan et al., 1997) they are generally too imprecise to identify individual genes (MacKay et al., 2009). To circumvent this issue, we employed a divergence mapping approach using the F_{st} statistic to create a more fine-scaled map of the genomic divergence occurring in this system. Through calculating the fixation index of the genes contributing to phenotypic traits known to cause ecological specialization and ultimately reproductive isolation, genomic regions of a few cM can be identified (Barrett & Hoekstra, 2011; Via et al., 2012).

Examining genome-wide divergence between species in naturally hybridizing populations is a good way to detect loci of small effect contributing to adaptation and reproductive isolation because of the many generations of recombination between species (Ortíz-Barrientos et al., 2002; Payseur & Rieseberg, 2016). We conducted a divergence-mapping approach using PoolSeq data to identify regions of genomic divergence and candidate loci contributing to trait differentiation and reproductive isolation between neighboring populations of *M. guttatus* and *M. lacinatus* (Hendricks et

al., 2016; Gould et al., 2017). Through this approach, we will be able to better answer the question: Is the genetic divergence in the face of gene flow between *M. guttatus* and *M. laciniatus* driven by simple genetic architecture characteristic of large-effect barrier loci acting through linkage groups to reduce the effect of gene flow or through a complex genetic architecture made up of many small-effect loci acting in a concerted fashion?

2 | Materials and Methods

2.1 Plant Material and Trait Measurement

The *M. guttatus* species complex is composed of morphologically diverse but mostly interfertile species, dispersed throughout western North America. Populations collected for divergence mapping were collected from a sympatric location near the town of Shaver Lake, located in Fresno County California (37.1064° N, 119.3194° W, 1,715 m). This site was selected because of the presence of sympatric *M. laciniatus* and *M. guttatus* populations occurring within a meter of each other (Ferris & Willis, 2018).). Seeds were collected from one hundred of each species for a total of two hundred maternal families (100 *M. guttatus* and 100 *M. laciniatus*). Phenotypic hybrids between *M. laciniatus* and *M. guttatus* have been previously described at the Shaver Lake population and preliminary genomic analyses confirm the presence of ongoing gene flow between the two species (K. Ferris personal observation; D. Tataru unpublished

data).Therefore, this population is a good location to study the genetic architecture of divergence with gene flow between these two *Mimulus* species.

Seeds were taken from *M. guttatus* and *M. laciniatus* individuals from the field site at Shaver Lake and transported back to Tulane University, LA. Seeds from 100 maternal families of each species were planted in flats of SunGro Metro Mix (SNGMM830) soil. Seeds were stratified at 4ºC for 10 days before being moved into growth chambers with 15 hour days and 21ºC growing conditions. Plots were thinned to one plant per pot once they germinated and replicates were moved into new pots. This gave us a total of 68 *M. laciniatus* and 75 *M. guttatus* seedlings for tissue collection. *M. guttatus* seeds were planted and grown out in a growth chamber from July 2020 to September 2020. *M. laciniatus* individuals were regrown from November 2021 to January 2022 after the loss of the initial *M. laciniatus* grow out due to a faulty DNA extraction.

Once individuals reached flowering, the most lobed leaf, which was typically the second true leaf, was collected and taped to sheets of printer paper for shape analysis using ImageJ. We used a convex hull measurement to quantify the degree of leaf lobing as described previously in Ferris et al (2015). To quantify lobing we measure the difference between the area of the second true leaf of each plant and the area of that same leaf's convex hull which is the shape produced by connecting the outermost points of each leaf. Then we divide the difference in areas by the area of the convex hull to produce a lobing ratio. To confirm species identity we also measured flowering time, the height of the plant at first flower (soil to plant meristem) and corolla size (width and length) with digital calipers. Phenotypic measurements are able to confirm species

identity as each trait measured differs between *M. laciniatus* and *M. guttatus*. Equal amounts of tissue were collected from each individual and preserved at -80 ºC. DNA extractions were performed on each population using a modified CTAB-chloroform protocol (Kelly & Willis, 1998).

2.2 Pool Selection and Sequencing

To examine genome-wide divergence between sympatric populations of *Mimulus* at Shaver Lake, we used a pool sequencing approach to measure allelic differentiation at the population level in a cost-effective manner (Gould et al., 2017). Pools were constructed on the basis of leaf lobing index, as this is a good predictor of species identify between *M. guttatus,* which has characteristically round leaves, and *M. laciniatus,* which has diverged from *M. guttatus* by developing highly lobed leaves. To confirm species identity of each plant we also examined other good species predictor traits : flowering time, flower size, and plant height, we used leaf shape in hopes of identifying a specific locus controlling for this adaptive trait.

Based on our phenotypic measurements we sorted the genotypes into two pools: *M. laciniatus* (*n* = 68) and *M. guttatus* (*n* = 75). Seven individuals were identified as *M. nasutus* and not included in either pool. . DNA concentration was determined for each sample using PicoGreen (Invitrogen) on FLX800 (BioTek) Microplate Reader. Pools were created by combining equal amounts of each individual's DNA via manual volumetric pooling. To ensure each sample had an equal concentration of DNA contributing to the uL being sent off for sequencing, the number of samples in a

population was divided by 1000 ng to obtain the volume (in uL) needed from each sample (1000 ng / 68 samples = 14.7 uL for the *M. laciniatus* pool and 1000 ng / 75 samples = 13.33 uL for the *M. guttatus* pool). These volumes were then divided by each individuals concentration (in ng) to obtain the volume required from each individual. The sum of each inidivuals contribution to the pool was divided by 1000 ng to obtain the total concentration of DNA in each pool (221.45 ng for the *M. laciniatus* pool and 142.17 ng for the *M. guttatus* pool). Pooled concentrations were confirmed and quantified using a fluorometric assay with a Qubit 2.0 Fluorometer (ThermoFisher Scientific, MA).

Samples of 50 uL of each pool were sent to the Duke University Sequencing Facility for library preparation and whole genome sequencing. Each pool was library prepared with DNA-seq KAPA Hyperprep. Full genome re-sequencing data for both pools was generated using Novaseq 6000 S-Prime 150bp PE reads on a single lane.

2.3 Short Read Mapping

Raw reads were checked for quality using FastQC version 0.11.7 and trimmed for quality and adaptor sequence using Trimmomatic version 0.39. To ensure that all adaptors had been removed, the trimmed reads were then rerun using FastQC before being aligned. The cleaned reads were then mapped to the well-annotated MgIM767 *M. guttatus* reference genome (Mguttatusvar_IM767_836_v1.0.fa) using the BWA 0.7.15 alignment tool (Li H., 2013).

The process involved using trimmed forward and reverse sequences as input. Subsequently, the resulting SAM file was converted to a BAM file using Samtools 1.5 (Danecek et al., 2021), followed by sorting and indexing. Then, the coverage statistics

were computed using NGSEP 4.3.2 (Tello et al., 2023), with the objective to identify the average coverage. Afterward, the proper pair-aligned sequences were retrieved with Samtools. The proper paired BAM file underwent indexing and computation of coverage statistics. Each alignment and subsequent processing step were independently performed for the sequencing data of both populations. Afterward, an additional step was executed to remove duplicated sequences from the proper pair BAM file of each population. To achieve this, the functionality MarkDuplicates of Picard (“picard toolkit”, 2019) was used and the indexing functionality of Samtools was used once again to index the BAM file without duplicate sequences.

2.4 | Variant Calling

To retrieve variants present in both populations, two methods were tested: (1) variant calling performed independently, accounting for ploidy corresponding to the number of sequenced individuals, and (2) variant calling performed simultaneously for both populations, accounting ploidy corresponding to the average sequencing depth (as recommended by the author). However, despite achieving alignment depths of 45x for the Lobed population and 62x for the Round population in proper pairs and considering the sequenced individuals numbered 68 for Lobed and 75 for Round populations, the anticipated alignment depth was expected to be at least double the number of sequenced individuals. Consequently, the results obtained from the initial method were discarded due to insufficient sequencing depth, which compromised the reliable retrieval of variants.

In the second method, MultisampleVariantsDetector function of NGSEP 4.3.2 was used. Proper pair alignments without duplications served as input files. A ploidy of 50 was specifically chosen to align with sequencing depth across both populations. Subsequently, the VCF file underwent filtration to retain SNPs, evaluating genotyping quality (GQ) and supporting read depth (DP). Optimal thresholds were determined at GQ >= 40 and DP >= 30. The resulting VCF file was annotated with VCFAnnotate of NGSEP 4.3.2, employing the guttatusvar_IM767_836_v1.1.gene.gff3 file. Afterward, the VCF file was filtered once more to isolate those SNPs present within intragenic regions.

2.5 | Diversity Statistics

The MultisampleVariantDetector in NGSEP provides the count of observed alleles in both populations, normalized by the designated ploidy for the execution (field ACN). This count served as the basis for calculating allelic frequencies and subsequently computing diversity statistics. For computing the F_{ST} value per SNP between both populations, a Jupyter Notebook under the name of “getFST.ipynb” was written. In this notebook, the metrics necessary for calculating F_{ST} were calculated and appended to the data frame.

The Jupyter Notebook named “getFSTpergene.ipynb” was developed to calculate the average F_{ST} per gene. The initial phase involves a cleaning process aimed at filtering out multiallelic SNPs and absent SNP data for one of the two populations. Toward the conclusion of the script, genes with an F_{ST} value of 1 are identified, identifying SNPs with fixed differences in allelic frequency between both populations.

An F_{ST} Manhattan plot was generated using Matplotlib. Within the Jupyter Notebook “plotFST.ipynb”, the function used for plotting is described. The plot illustrates the F_{ST} value per SNP across the genome. The Jupyter Notebook “getFisherExactTest.ipynb” was developed specifically to compute Fisher’s exact test. This process necessitated the use of the Stats library from Scipy.

The Jupyter Notebook “getG-test.ipynb” was created to compute the G-test between the SNPs of both populations, utilizing the Stats library from Scipy. Subsequently, for visualizing the distribution of p-values, violin plots and histograms were generated using the Matplotlib and Seaborn libraries.

2.6 | Identifying Candidate Genes

A significance threshold was calculated with the Bonferroni-corrected to 4.183e-07 alpha = 0.01. Then, we applied a Z-score transformation to observe what would be this corrected alpha in terms of F_{st} distribution giving a result of 0.6517. Peaks of significance were identified as any F_{st} value that exceeded the statistical significance threshold of 0.65. Peaks over this significance threshold were examined using Phytozome (https://phytozome-next.jgi.doe.gov/) and all annotated genes in the significant regions were recorded. We predicted the functional effects of these SNPs based on known annotation of genes, transcripts, and other genomic features as described in other plant literature.

3 | Results

3.1 | Sympatric populations are phenotypically distinct

The distribution of leaf shape between the two pools shows a distinct distribution. The distribution of the *M. laciniatus* pool is between 0.1 and 0.8 (**figure 1**) with a mean leaf lobing index of 0.383. The distribution of the *M. guttatus* pool is between 0.0 and 0.4 with a mean leaf lobing index of 0.113 (**figure 2**). The distribution phenotypic traits diverging in this system were plotted on a PCA plot to confirm that phenotype is a good predictor of genotype (**figure 3**). Specifically corolla length, corolla width, plant height, smallest anther, longest anther and stigma length were plotted for every individual contributing to the sequencing pool. Plants with similar phenotypes clustered together into two distinct areas of PC space. With only a few exceptions, plants clustered into separate *M. guttatus* and *M. laciniatus* groups (**figure 3**). A few individuals that had been labeled as *M. guttatus* genotypes did phenotypically cluster with the *M. laciniatus* individuals, however this could be due to phenotypic plasticity. All plants labeled as *M. laciniatus* clustered together in phenotypic PC space (**figure 3**).

3.2 | Genome Wide Patterns of Divergence

After sequencing each pool and aligning to the reference genome, the average alignment depth for the round pool was 62x as illustrated by **figure 4** while the average alignment depth for the lobed pool was 45x as illustrated by **figure 5**.

A total number of 1,262,754 SNPs were identified and plotted by fixation index to calculate the number of fixed differences between the two pools (**figure 6**). The genome wide average for Fst is 0.158. Levels of genetic differentiation between the two pools is highly variable across the genome, giving a heterogeneous pattern of divergence. This pattern is evident when examining the F_{st} peaks across the genome.

The average F_{st} values per chromosome can be found in **table 1.** Chromosome 6 had the highest chromosome average F_{ST} (0.172) which was much higher than the genome wide average (0.158). Chromosome 6 also contains two previously mapped pleiotropic QTLs for differences in flowering time and flower size between Shaver Lake *M. guttatus* and *M. laciniatus* (Ferris et al., 2017). However, chromosome 8 that contains the largest effect pleiotropic QTL for leaf shape, flowering time and flower size had only slightly elevated F_{ST} (0.16) compared to the genome wide average. More informative of course than the average F_{ST} per chromosome are the individual loci exhibiting elevated divergence of which there were 71 genes total.

3.3 | Candidate Gene Analysis

In order to identify genes that may be under divergent selection in this system, a significance threshold was calculated using the Bonferroni correlation and found to be 0.57. A total of 166,618 SNPs were found to be above this significance threshold and therefore considered candidates targets of divergent selection. Candidate genes were located by position in Phytozome and recorded. **Table 3** displays all genes that contain significant SNPs found on Chromosomes 2, 7, 8 and 12 with their functions as described by the literature.

Many of the genes found were notably related to the regulation of plant growth, developmental timing, and resistance to abiotic stressors such as cold, salt or metal contaminants. There were also many hits for genes related to transcription factors and other regulatory machinery implicated in the growth or timing of reproductive tissues or roots.

Of particular interest, the MYB domain on chromosome 7 interacts with four other genes that were significant for fixation index. MYB is a transcription factor that is known to regulate the developmental transition in the root endodermis in *Arabidopsis thaliana* (Liberman et al., 2015). This developmental transition involves the formation of a protective waxy barrier called the Casparian strip (Roppolo et al., 2014), which interestingly is one of the significantly differentiated genes found while examining F_{st} for chromosome 12. While root endodermis may not be a reproductively isolating trait, its differentiation could help *M. laciniatus* colonize environments with harsher soil conditions such as the granite outcrops of Shaver Lake.

The MYB gene not only controls for the Casparian strip but also has been shown to be important for fruit development in tomatoes and seedling development in *Arabidopsis* (Barg et al., 2005), anther and pistil development in petunia (Huang et al., 2015), root and leaf architecture including leaf lobing in tomato (Naz et al., 2013) and lateral organ primordia in maize through the repression of KNOX homeobox genes (Timmermans et al., 1999). This developmental differentiation through the repression of KNOX homeobox genes is of particular interest because Ferris et al. (2017) noted KNAT6, a KNOX gene, as a candidate gene in one of their major QTLs found; LG5.

MYB also interacts with histone deacetylase, which was found as a significantly differentiated gene as determined by fixation index on chromosome 8. This is also of interest not only because of histone deacetylase's vital role in leaf shape and leaf lobing, an important adaptive trait in this system, but because it does so through repression of KNOX genes (Luo et al., 2012). Additionally, its location on chromosome 8 positions it right in the middle of the major chromosome 8 QTL controlling for corolla length upper and lower as well as leaf shape as found by Ferris et al. 2018.

3.4 | Overlap between F_{ST} outliers and species specific trait QTL

To identify fixed differences that may occur within the known QTL regions, LOD peaks previously found by Ferris et al. (2018) were compared to the Fst values obtained through divergence mapping and are presented in **table 2**.

Finally the leaf shape QTL on chromosome 2 has three significant F_{st} values while the two leaf shape QTLs on chromosome 8 have three and one significant F_{st} values respectively. This information is presented in **figures 8a-e**.

4 | Discussion

In this study, we used PoolSeq data to calculate genome-wide trait divergence between populations of *M. guttatus* and *M. laciniatus* experiencing ongoing gene flow. By aligning the sequencing reads back to the *M. guttatus* genome and plotting the fixation index of each SNP, we were able to determine that the genetic architecture of trait divergence in this system is both complex and spread throughout the genome. This

result is consistent with the multilocus theory that predicts complex and dispersed trait divergence suggesting multiple underlying adaptations that are not in linkage disequilibrium with one another.

This result expands on previous findings that traits under divergent selection in this system have a genetically simple architecture (Ferris et al., 2017). This study had found that four to five major pleiotropic QTL of large-to-moderate effect on chromosomes 2, 5, 8, and 10 controlled five adaptive traits looked at were being controlled pleiotropically through these QTL (Ferris et al., 2017). Our divergence mapping study suggests that despite the genetic architecture of these particular traits being genetically simple, the genetic architecture of species divergence and local adaptation is more complex.

To compare the results of the QTL mapping study with this divergence mapping study, the chromosomes that explained a significant amount of the variance from the QTL mapping study were examined closely using our F_{st} data. F_{st} values were plotted per each chromosome; however, only chromosomes 2, 7, 8, and 12 were included (figures 4a-d) as they showed the largest peaks in divergence throughout the genome. This result is interesting as it is distinct from the previously identified QTL regions with the exception of chromosomes 2 and 8. It is therefore suggestive that chromosomes 2 and 8 are important regions of adaptation or reproductive isolation between populations of *M. guttatus* and *M. laciniatus* in the Shaver Lake region of California.

Despite identifying these regions of interest on chromosomes 2 and 8, it is evident from the Manhattan plot displaying F_{st} values for the entire genome that divergence is occurring not in large linkage groups as expected but in a complex

genetic architecture involving many loci dispersed throughout the entire genome. This complex genetic architecture of reproductive isolating barriers in populations experiencing gene flow has been shown multiple times before. For example both coloration and lip phenotype, important traits leading to assortative mating and reproductive isolation, have been shown to have a complex genetic architecture in Midas cichlid fishes suggesting polygenic selection in sympatric populations (Kuatt et al., 2020). Additionally, pollen sterility in *Helianthus* hybrid zones have been shown to be controlled by 16 loci (Rieseberg et al., 1999), and *Drosophila* studies has found polygenic control in both differences in courtship behavior and reproductive morphology (C.I Wu & Ting, 2004).

In this study we highlight the regions of the genome in the *M. guttatus* species complex that may play a role in species divergence. This work contributes to evolutionary theory as it adds to our understanding of the interplay between gene flow and divergence. Understanding the genetic basis of species divergence is vital for not only speciation researchers but for conservation efforts. If there are specific genomic regions associated with reduced gene flow, this information can be used to identify populations at risk and guide conservation strategies. Additionally, this work provides some insight into adaptation, as high F_{st} values may indicate regions of the genome associated with local adaptation or environmental pressures. This information is valuable for understanding how species evolve in response to their ecological contexts.

Future work should look further into the genes highlighted by this divergence mapping study to elucidate any potential role these genes could be playing in *M.*

guttatus or *M. laciniatus* adaptation to their prospective microhabitats in the face of ongoing gene flow.

5 | Conclusion

Two closely related interfertile monkeyflowers, *M. guttatus* and *M. laciniatus* occur within meters of each other and are known to hybridize, yet they maintain phenotypic and ecological divergence despite this ongoing gene flow. Previous studies have identified five major QTLs controlling for traits that assist with adaptation to the unique ecotypes in this environment. However, we conducted a F_{st} analysis and found the genetic architecture of divergence to be genetically complex. This homogenous pattern of genomics is in contrast to the expected simple architecture of genomic islands of divergence as suggested by the QTL analyses. In light of these results, we hypothesize that local adaptation between neighboring monkeyflowers is driven by divergent selection on many loci of small effect dispersed throughout the genome.

Works Cited

Barg, R., Sobolev, I., Eilon, T., Gur, A., Chmelnitsky, I., Shabtai, S., Grotewold, E., & Salts, Y. (2005). The tomato early fruit specific gene Lefsm1 defines a novel class of plant-specific SANT/MYB domain proteins. *Planta*, *221*(2), 197–211.

Barrett, R. D. H., & Hoekstra, H. E. (2011). Molecular spandrels: Tests of adaptation at the genetic level. *Nature Reviews Genetics*, *12*, 767–780.

Barton, N. H. (1983). Multilocus clines. *Evolution*, 37, 454–471

Barton, N., & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity*, *57*, 357–376.

Beavis, W. D. (1998). *QTL analyses: Power, precision, accuracy. Molecular dissection of complex traits*. CRC Press

Brandvain, Y., Kenney A., Flagel L., Coop G., and Sweigart A. (2014). Speciation and Introgression between Mimulus Nasutus and Mimulus Guttatus. *PLoS Genetics*.

Cheng, C., White, B. J., Kamdem, C., Mockaitis, K., Costantini, C., Hahn, M. W., & Besansky, N. J. (2012) Ecological genomics of Anopheles gambiae along a latitudinal cline: a population-resequencing approach. *Genetics*, *190*, 1417–1432.

Coyne, J., & Orr, H. (2004). *Speciation*. Sinauer Associates

Danecek, P., Bonfield, J., Liddle, J., Marshall, J., Ohan, V., Pollard, M., Whitwham, A., Keane T., McCarthy, S., Davies, R., Li, H. (2021) Twelve Years of SAMtools and BCFtools. *GigaScience* 10 (2).

Emelianov, I., Marec, F., & Mallet, J. (2004). Genomic evidence for divergence with gene flow in host races of the larch budmoth. *Proceedings. Biological Sciences / The Royal Society*, *271*(1534), 97–105.

Fabian D.K., Kapun M., Nolte V., Kolfer, R., Schmidt, P.S., Schlötterer, & Flatt, T. (2012). Genome-wide patterns of latitudinal differentiation among populations of Drosophila melanogaster from North America. *Molecular Ecology*, *21*, 4748– 4769.

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, *28*, 342-.

Feder J. L., Gejji, R., Yeaman, S, & Nosil, P. (2012) Establishment of new mutations under divergence and genome hitchhiking. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*, 461– .

Feder, J. L., Nosil, P., Wacholder, A. C., Egan, S. P., Berlocher, S. H., & Flaxman, S. M. (2014). Genome-wide congealing and rapid transitions across the speciation continuum during speciation with gene flow. *Journal of Heredity*, *105*(Suppl. 1), 810–820.

Fenster, C. B., & Ritland, K. (1994). Evidence for natural selection on mating system in Mimulus (Scrophulariaceae). *International Journal of Plant Sciences*, *155*(5), 588–596.

Ferris, K. G. (2016). A hot topic: The genetics of adaptation to geothermal vents in *Mimulus guttatus*. *Molecular Ecology*, *25*(22), 5605-5607.

Ferris, K. G., Barnett, L. L., Blackman, B. K., & Willis, J. H. (2017). The genetic architecture of local adaptation and reproductive isolation in sympatry within the mimulus guttatus species complex.” *Molecular Ecology*, *26*(1): 208–224.

Ferris, K. G., Rushton, T., Greenlee, A. B., Toll, K., Blackman, B. K., & Willis, J. H. (2015). Leaf shape evolution has a similar genetic architecture in three edaphic

specialists within the Mimulus guttatus species complex. *Annals of Botany*, *116* (2): 213–223.

Ferris, K. G., Sexton, J. P., & Willis, J. H. (2014). Speciation on a local geographic scale: The evolution of a rare rock outcrop specialist in Mimulus. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 369(1648). https://doi.org/10.1098/rstb.2014.0001*

Ferris, K. G., & Willis, J. H. (2018). Differential adaptation to a harsh granite outcrop habitat between sympatric Mimulus species. *Evolution; International Journal of Organic Evolution*, *72*(6), 1225–1241.

Fishman, L., Beardsley, P. M., Stathos, A., Williams, C. F., & Hill, J. P. (2015). The genetic architecture of traits associated with the evolution of self-pollination in Mimulus. *New Phytologist*, *205*, 907–917.

Fishman, L., Kelly, A, J., & Willis, J. H. (2002). Minor quantitative trait loci underlie floral traits associated with mating system divergence in Mimulus. *Evolution*, *56*, 2138–2155.

Flaxman, S. M., Wacholder, A. C., Feder, J. L. & Nosil, P. (2014). Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Molecular Ecology*, *23*, 4074–4088.

Flagel, L. E., Bansal, R., Kerstetter, R. A., Chen, M., Carroll, M., Flannagan, R., Clark, T., Goldman, B. S., Michel, A. P. (2014a). Western corn rootworm (*Diabrotica virgifera virgifera*) transcriptome assembly and genomic analysis of population structure. *BMC Genomics*, *15*, 195.

Flagel, L. E., Willis, J., & Vision, T. J. (2014b). The standing pool of genomic structural variation in a natural population of *Mimulus guttatus*. *Genome Biology and Evolution*, *1*, 53–64.

Friedman, J., & Willis, J. H. (2013). Major QTLs for critical photoperiod and vernalization underlie extensive variation in flowering in the *Mimulus guttatus* species complex. *New Phytologist*, *199*, 571–583.

Gavrilets, S. (2004). *Fitness landscapes and the origin of species*. Princeton University Press.

Gould, B. A., Chen, Y., & Lowry, D. B. (2017). Pooled ecotype sequencing reveals candidate genetic mechanisms for adaptive differentiation and reproductive isolation. *Molecular Ecology*, *26*(1), 163–177.

Hall, M. C., Basten, C.J., & Willis, J. H. (2006). Pleiotropic quantitative trait loci contribute to population divergence in traits associated with life-history variation in *Mimulus guttatus*. *Genetics*, *172*, 1829–1844.

Hendrick, M. F. (2016). *The genetics of extreme microgeographic adaptation: An integrated approach identifies a major gene underlying leaf trichome divergence in Yellowstone* Mimulus Guttatus. Blackwell Scientific Publications.

Huang, X., Yue, Y., Sun, J., Peng, H., Yang, Z., Bao, M., & Hu, H. (2015, February). Characterization of a fertility-related SANT/MYB gene (PhRL) from Petunia. *Scientia Horticulturae*, *183*, 152–159.

Kapun, M., Fabian, D. K., Goudet, J., & Flatt, T. (2016). Genomic evidence for adaptive inversion clines in *Drosophila melanogaster*. *Molecular Biology and Evolution*, *5*, 1317–1336.

Kautt, A. F., Kratochwil, C. F., Nater, A., Gonzalo Kliman, R. M., Andolfatto, P., Coyne, J. A., Depaulis, F., Kreitman, M., Berry, A. J., McCarter, J., Wakeley, J. & Hey, J. (2000). The population genetics of the origin and divergence of the Drosophila Simulans complex species. *Genetics*, *156*(4), 1913– .

Kautt, A. F., Kratochwil, C. F., Nater, A., Machado-Schiaffino, G., Olave, M., Henning, F., Torres-Dowdall, J., Härer, A., Husley, C. D., Franchini, P., Pippel, M., Myers, E. W., & Meyer, A. (2020). “Contrasting signatures of genomic divergence during sympatric speciation.” *Nature*, *588*(7836), 106–111.

Kelly A.J., Willis J.H., (1998) Polymorphic microsatellite loci in Mimulus guttatus and related species. *Molecular Ecology*, 7, 769–774.

Kliman, R. M., P. Andolfatto, J. A. Coyne, F. Depaulis, M. Kreitman, A. J. Berry, J. McCarter, J. Wakeley, and J. Hey. (2000). The Population Genetics of the Origin and Divergence of the Drosophila Simulans Complex Species. *Genetics* 156 (4): 1913–31.

Kofler, R., Betancourt, A. J., & Schlötterer, C. (2012). Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in Drosophila melanogaster. *PLoS Genetics*, *8*, e1002487.

Lenormand, T., & Otto, S. P. (2000). The evolution of recombination in a heterogeneous environment. *Genetics*, *156*, 423–438.

Levin, D. A. (2009). Flowering-time plasticity facilitates niche shifts in adjacent populations. *New Phytologist*, *183*, 661–666.

Li H., (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2 [q-bio.GN]

Liberman, L. M., Sparks E. E., Moreno-Risueno, M. A., Petricka, J. J., & Benfey, P. N. (2015). MYB36 regulates the transition from proliferation to differentiation in the

arabidopsis root. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(39): 12099–12104.

Lowry, D. B., Modliszewski, J. L., Wright, K. M., Wu, C. A., & Willis, J. H. (2008). The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B*, *363*, 3009–3021.

Lowry, D. B., & Willis, J. H. (2010). A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biology*, *8*, .

Luo, M., Yu, C.-W., Chen, F.-F., Zhao, L., Tian, G., Liu, X., Cui, Y., Yang, J.-Y., & Wu, K. (2012). Histone deacetylase HDA6 is functionally associated with AS1 in repression of KNOX genes in arabidopsis. *PLoS Genetics*, *8*(12), e1003114.

Lynch, M., Bost, D., & Wilson, S. (2014). Population-genetic inference from pooled-sequencing data. *Genome Biology and Evolution*, *6*, 1210–1218.

Machado, C. A., & Hey, J. (2003). The Causes of Phylogenetic Conflict in a Classic Drosophila Species Group. *Proceedings. Biological Sciences / The Royal Society*, *270*(1520), 1193–1202.

MacKay, T. F. C., Stone, E. A., Ayroles, J. F. (2009). The genetics of quantitative traits: Challenges and prospects. *Nature Reviews Genetics*, *10*, 565–577.

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

Magwene, P. M., Willis, J. H., & Kelly, J. K. (2011). The statistics of bulk segregant analysis using next generation sequencing. *PLoS Computational Biology*, *7*, e1002255.

Marques, D. A., Lucek, K., Meier, J. I., Mwaiko, S., Wagner, C. E., Excoffier, L., and Seehausen, O. (2016). Genomics of rapid incipient speciation in sympatric threespine stickleback.” *PLoS Genetics*, *12*(2), e1005887.

Maynard Smith, J. (1966). *Sympatric speciation*. *The American Naturalist*, *100*, 637–650.

Maynard Smith, J. (1977). Why the genome does not congeal. *Nature*, *268*, 693–696.

Modliszewski, J. L., & Willis, J. H. 2012. Allotetraploid Mimulus sookensis are highly interfertile despite independent origins. *Molecular Ecology*, *21(*21), 5280–5298.

Mohan, M., Nair, S., Bhagwat, A., Krishna, T. G., Yano, M., Bhatia, C. R., & Sasaki, T. (1997). Genome mapping, molecular markers and marker-assisted selection in crop plants. *Molecular Breeding*, *3*, 87–103.

Navarro, A., & Barton, N. H. (2003). Accumulating postzygotic isolation genes in parapatry: A new wwist on chromosomal speciation. *Evolution; International Journal of Organic Evolution*, *57*(3): 447–459.

Naz, A. A., Raman, S., Martinez, C. C., Sinha, N. R., Schmitz, G., & Theres, K. (2013). Trifoliate encodes an MYB transcription factor that modulates leaf and shoot architecture in tomato. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(6), 2401–2406.

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

Nosil, P., Feder, J. L., Flaxman, S. M., & Gompert, Z. (2017). Tipping points in the dynamics of speciation. *Nature, Ecology, and Evolution*, *1*, 0001.

Ortíz-Barrientos, D., Reiland, J., Hey, J. & Noor, M. A. F. 2002. Recombination and the divergence of hybridizing species. *Genetica*, *116*(2-3), 167–178.

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, *25*(11), 2337–2360.

“Picard Toolkit.” 2019. Broad Institute, GitHub Repository. https://broadinstitute.github.io/picard/; Broad Institute

Pinho, Catarina, & Hey, P. (2010). Divergence with gene flow: Models and data. *Annual Review of Ecology, Evolution, and Systematics*, *41*(1), 215–230.

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

Pylkov, K. V., Zhivotovsky, L. A., & Feldman, M. W. (1998). Migration versus mutation in the evolution of recombination under multilocus selection. *Genetical Research*, *71*, 247–256.

Raineri, E., Ferretti, L., Esteve-Codina, A., Nevado, B., Heath, S., & Perez-Enciso, M. (2012). SNP calling by sequencing pooled samples. *BMC Bioinformatics*, *13*, 239.

Riesch, R., Muschick, M., Lindtke, D., Villoutreix, R., Comeault, A. A., Farkas, T. E., Lucek, K., Hellen, E., Soria-Carrasco, V., Dennis, S. R., Safran, R. J., Sandoval, C. P., Feder, J., Gries, R., Crespi, B. J., Gries, G., Gompert, Z., & Nosil, P. (2017). Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology & Evolution*, *1*(4), 82.

Rieseberg, L. H., Archer, M. A., & Wayne, R. K. (1999, October). Transgressive segregation, adaptation, and speciation. *Heredity*, *83*(Pt 4), 363–372.

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

Roppolo, D., Boeckmann, B., Pfister, A., Boutet, E., Rubio, E. C., Dénervaud-Tendon, V., Vermeer, J. E. M., Gheyselinck, J., Xenarios, I., & Geldner, N. (2014). Functional and evolutionary analysis of the CASPARIAN STRIP MEMBRANE DOMAIN PROTEIN family." *Plant Physiology*, *165*(4), 1709–1722.

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Saetre, G.-P., Bank, C., Bränström, A, Brelsford, A. Blarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., . . . Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, *15*, 176–192.

Servedio, M. R., Van Doorn, G. S., Kopp, M., Frame, A. M., & Nosil, P. (2011). Magic traits in speciation: "Magic" but not rare? *Trends in Ecology and Evolution*, *26*, 389–397.

Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals-mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, *15*, 749–763.

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

Smadja, C. M., & Butlin, R. K. (2011). A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology*, *20*, 5123–5140.

Sobel, J. M., & Chen, G. F. (2014). Unification of methods for estimating the strength of reproductive isolation. *Evolution*, *68*, 1511–1522.

Stankowski, S., Sobel, J. M., & Streisfeld, M. A. (2015). The geography of divergence with gene flow facilitates multitrait adaptation and the evolution of pollinator isolation in Mimulus aurantiacus. *Evolution: International Journal of Organic Evolution*, *69*(12): 3054– .

Sweigart, A. L., & Willis, J. H. (2003). Patterns of nucleotide diversity in two species of Mimulus are affected by mating system and asymmetric introgression. *Evolution; International Journal of Organic Evolution*, *57*(11), 2490–2506.

Tataru, D., Wheeler, E. C., & Ferris, K. G. (2023). Spatially and temporally varying selection influence species boundaries in two sympatric Mimulus. *Proceedings. Biological Sciences / The Royal Society*, *290*(1992), 20222279.

Tello, D., Gonzalez-Garcia, L.N., Gomez, J., et al., (2023). NGSEP 4: Efficient and accurate identification of orthogroups and whole-genome alignment.*Molecular Ecology Resources* 23(3): 712-724.

Timmermans, M. C., Hudson, A., Becraft, P. W. & Nelson, T. (1999). ROUGH SHEATH2: A myb protein that represses Knox homeobox genes in maize lateral organ primordia. *Science*, *284*(5411): 151–153.

Turner, T. L., Bourne, E. C., Von Wettberg, E. J., Hu, T. T., & Nuzhdin, S. V. (2010). Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. *Nature Genetics*, *42*, 260–263.

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

Vickery, R. K., Crook, K. W., Lindsay, D. W., Mia, M. M., & Tai, W. (1968). Chromosome counts in section simiolus of the genus Mimulus (scrophulariaceae). Vii. New numbers for M. guttatus, M. Cupreus, and M. Tilingii. *Madroño*, *19*(6), 211–218.

Westram, A. M., Rafajlović, M., Chaube, P., Faria, R., Larsson, T., Panova, M., Ravinet, M. Blomberg, A., Mehlig, B., Johannesson, K., & Butlin, R. (2018). Clines on the seashore: The genomic architecture underlying rapid divergence in the face of gene flow. *Evolution Letters*, *2*(4): 297–309.

Wu, C.I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology* 14:851–65

Wu, C. I., & Ting, C. T. (2004). Genes and speciation. *Nature Reviews Genetics*, *5*, 114–122.

Wu, C. A., Lowry, D. B., Cooley, A. M., Wright, K. M., Lee, Y. W., & Willis, J. H. (2007). Mimulus is an emerging model system for the integration of ecological and genomic studies. *Heredity*, *100*, 220–230.

Wu, C.A., Lowry, D.B., Cooley A.M., Wright K.M., Lee Y.W., Willis J.H., (2008) Mimulus is an emerging model system for the integration of ecological and genomic studies. *Heredity*, 100, 220–230.

Xu, S. (2003). Theoretical basis of the Beavis effect. *Genetics*, *165*(4), 2259-2268.

Yeaman, S., & Otto, S. P. (2011). Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution; International Journal of Organic Evolution*, *65*(7), 2123–2129.

Yeaman, S., and Whitlock M., (2011). The genetic architechture of adaptation under migration-selection balance. *Evolution* 65 (7): 1897–1911.

Figures

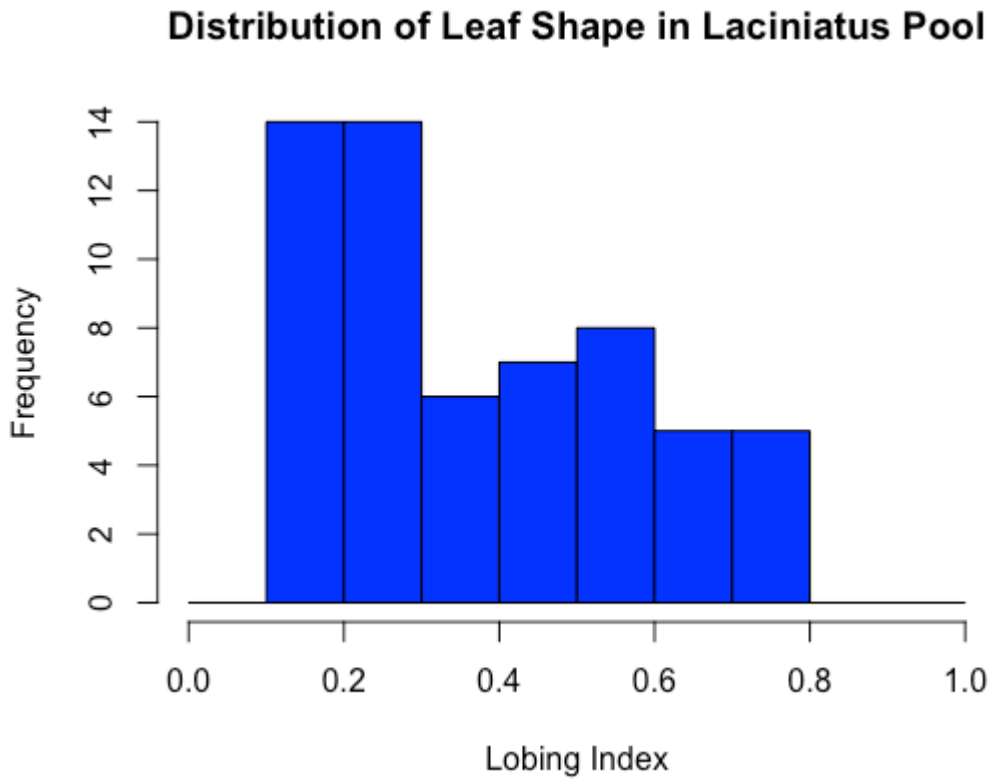


Figure 1: Distribution of leaf lobing index for the *M. laciniatus* pool

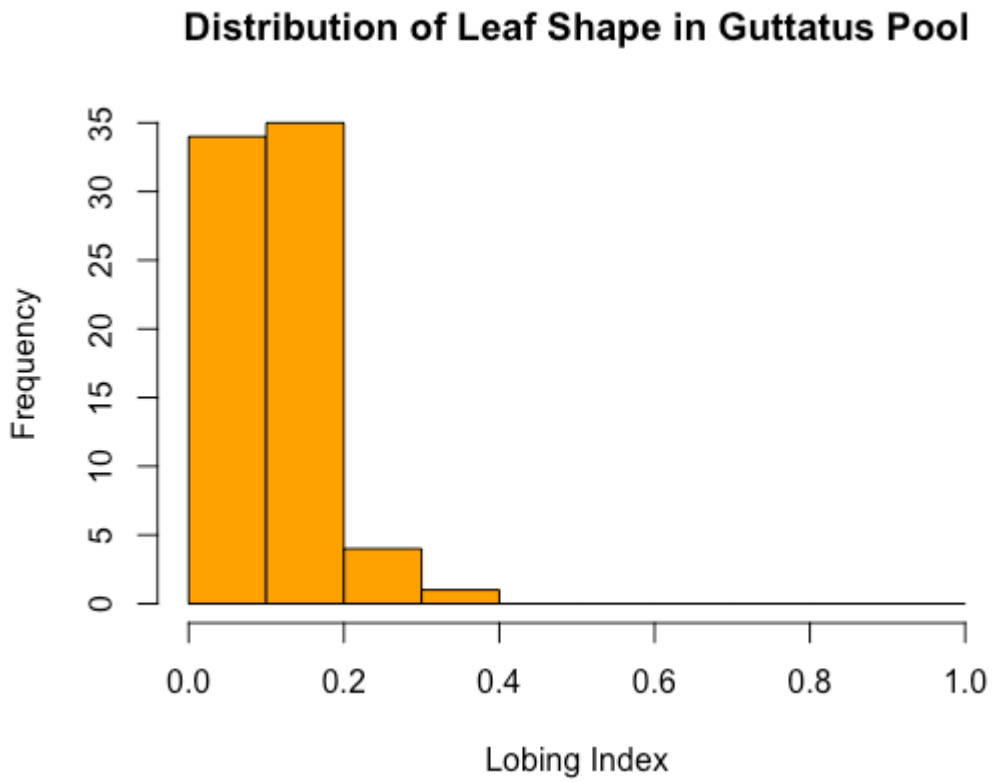


Figure 2: Distribution of leaf lobing index for the *M. guttatus* pool

Figure 3: PCA plot showing that the diverging phenotypes are a good predictor of genotype

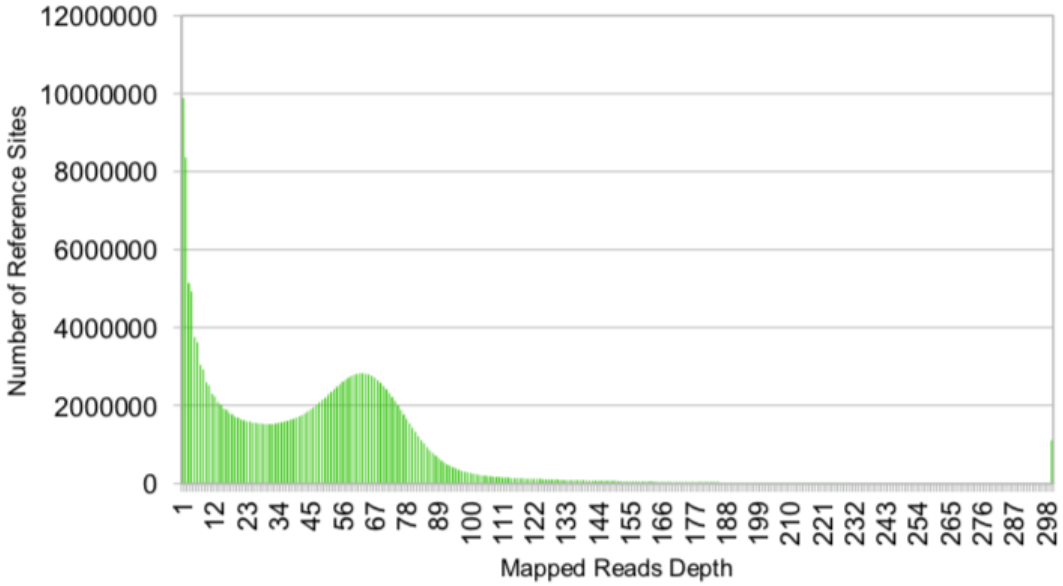


Figure 4: Average coverage for the Lobed pool

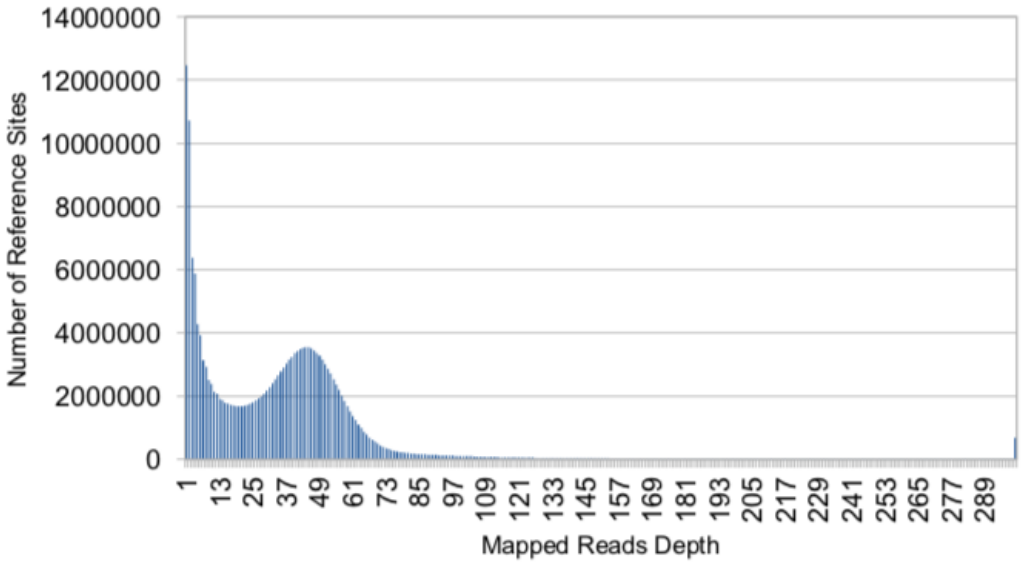


Figure 5: Average coverage of the Round pool

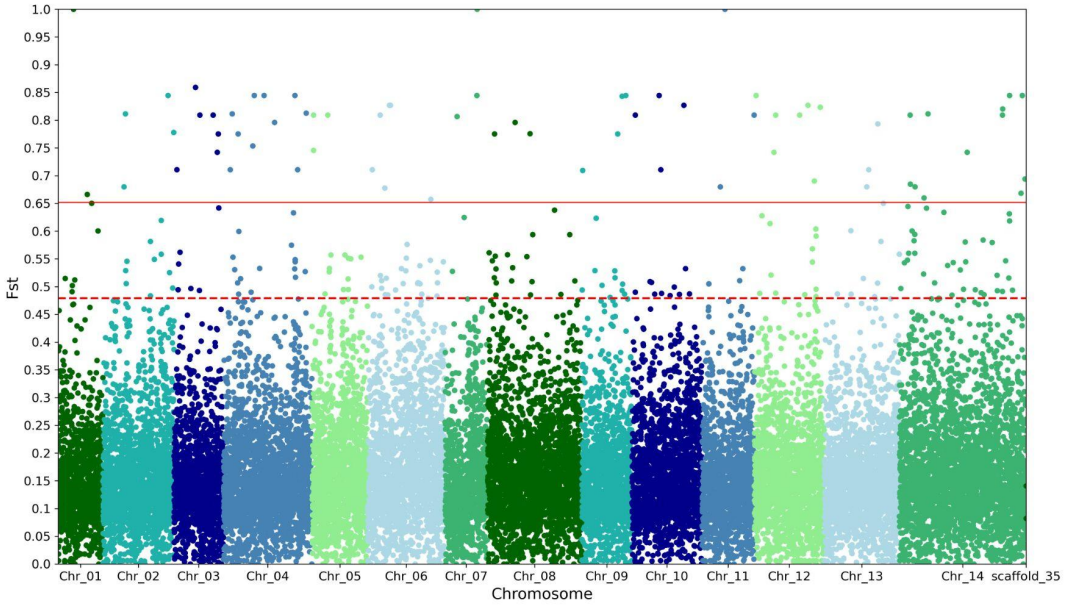


Figure 6: Manhattan plot showing the fixation index values from genome-wide divergence mapping. The X-axis displays the location of the SNPs by chromosome and the Y-axis is the fixation index value.

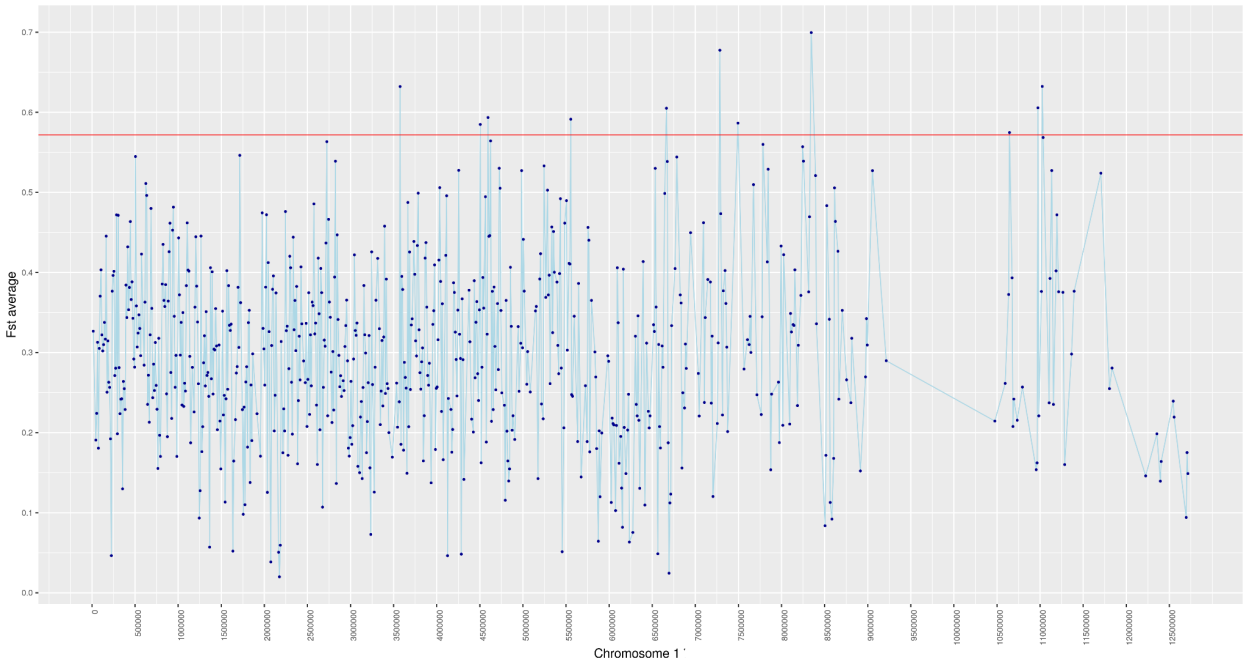


Figure 7a: F_{st} average per window for chromosome 1

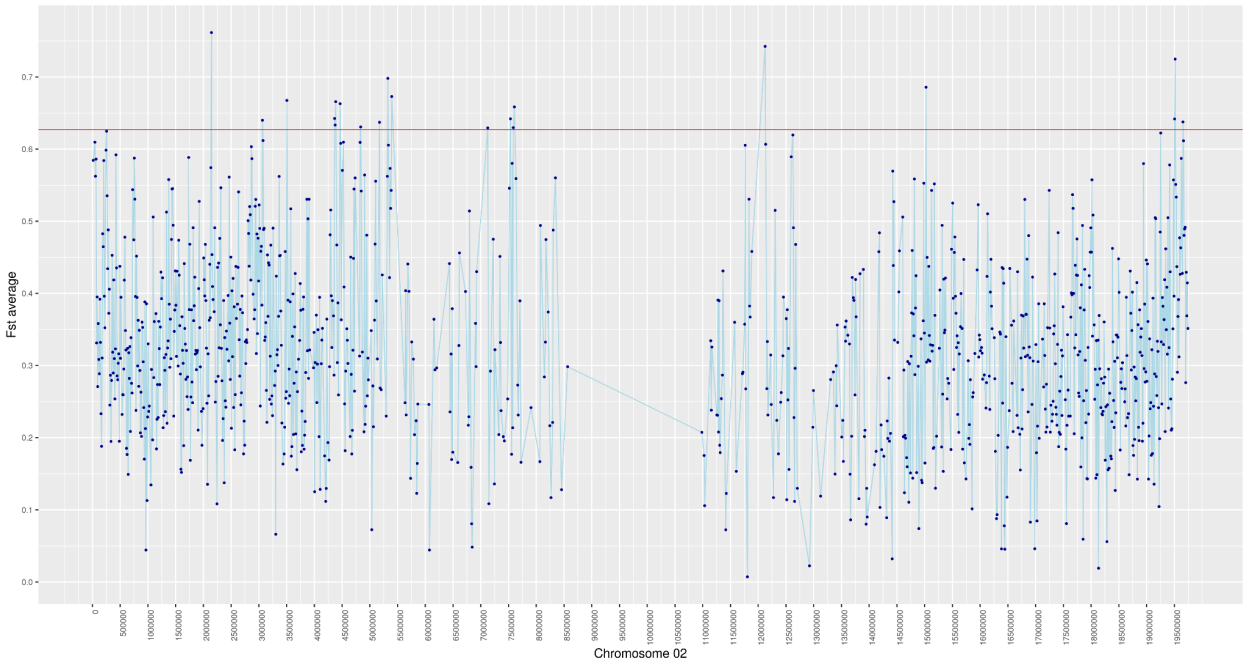


Figure 7b: F_{st} average per window for chromosome 2

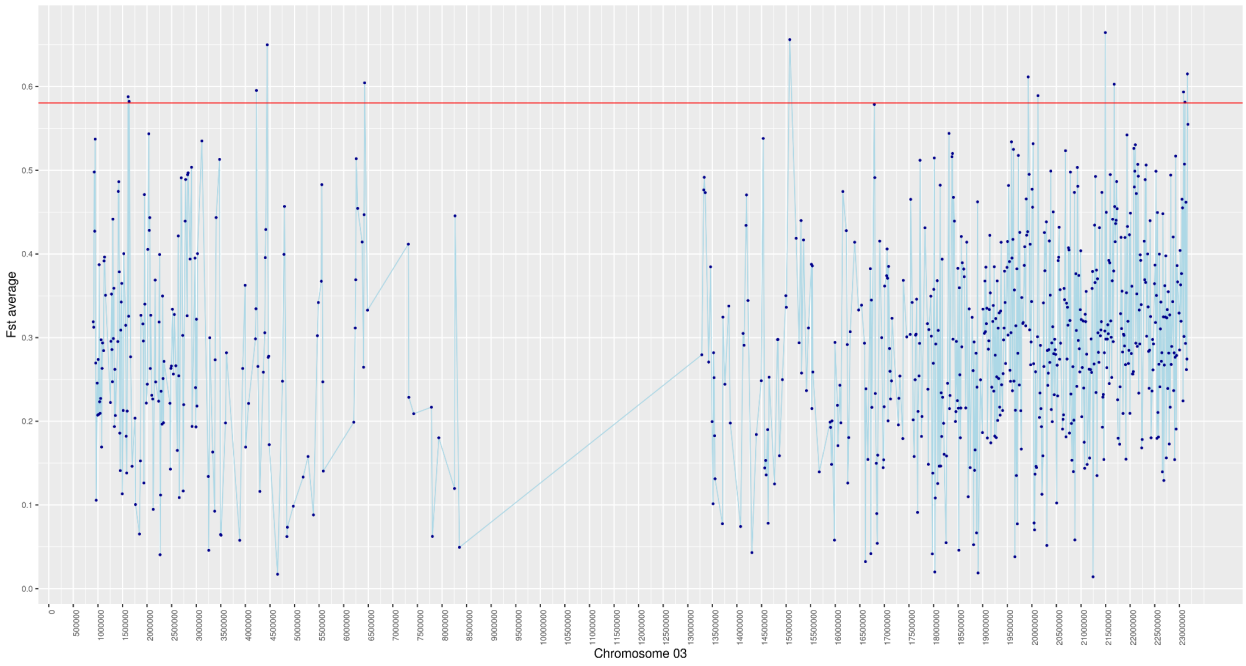


Figure 7c: F_{st} average per window for chromosome 3

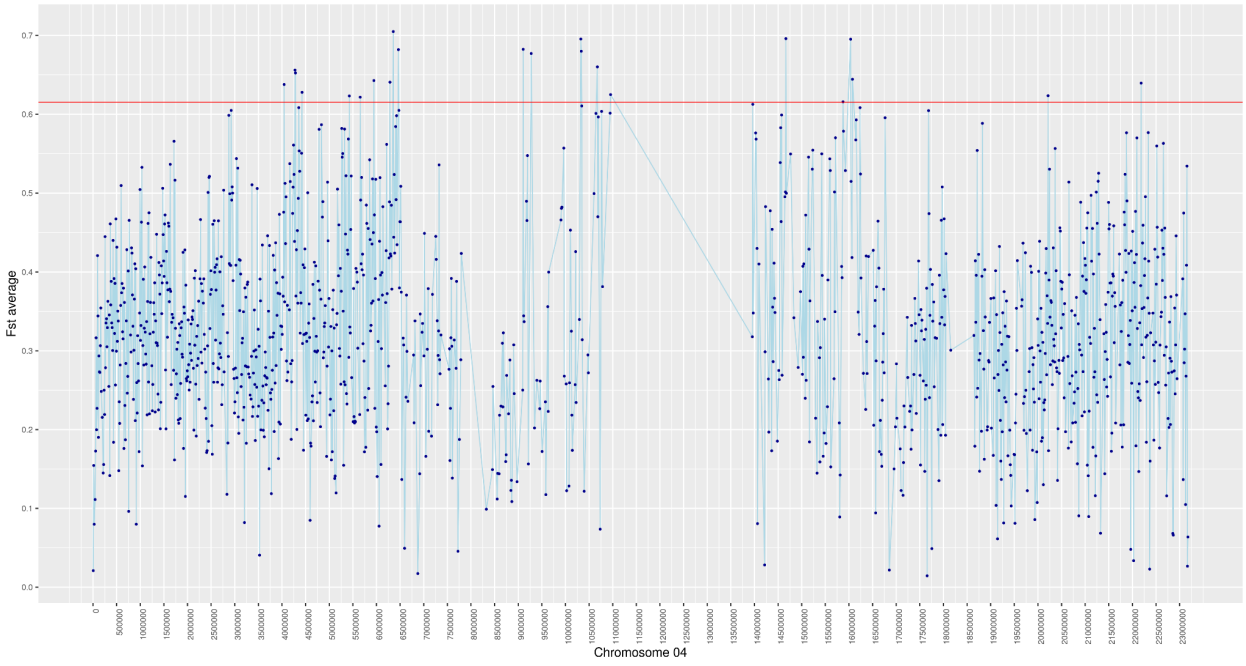


Figure 7d: F_{st} average per window for chromosome 4

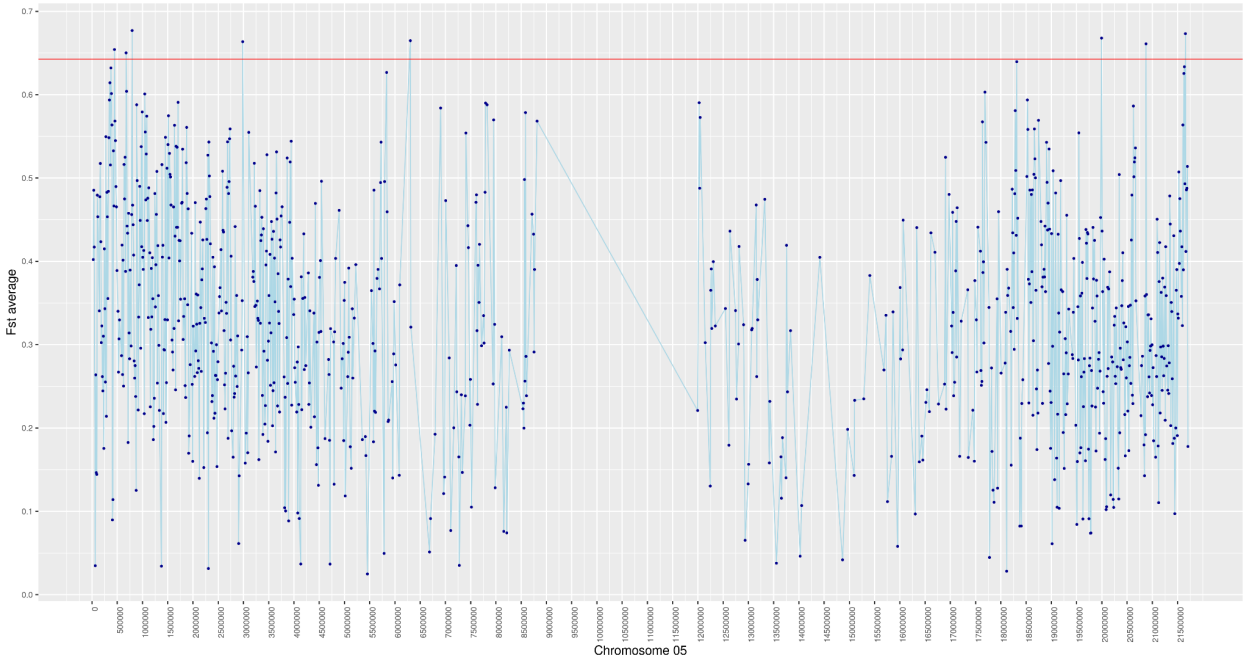


Figure 7e: F_{st} average per window for chromosome 5

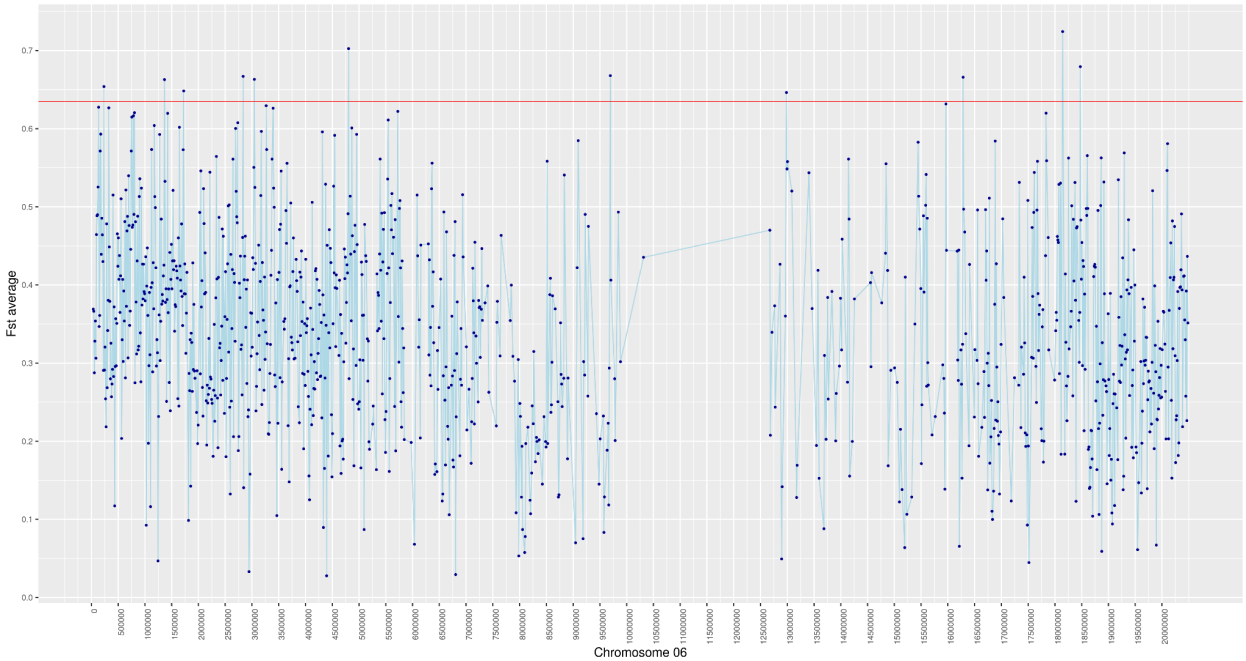


Figure 7f: F_{st} average per window for chromosome 6

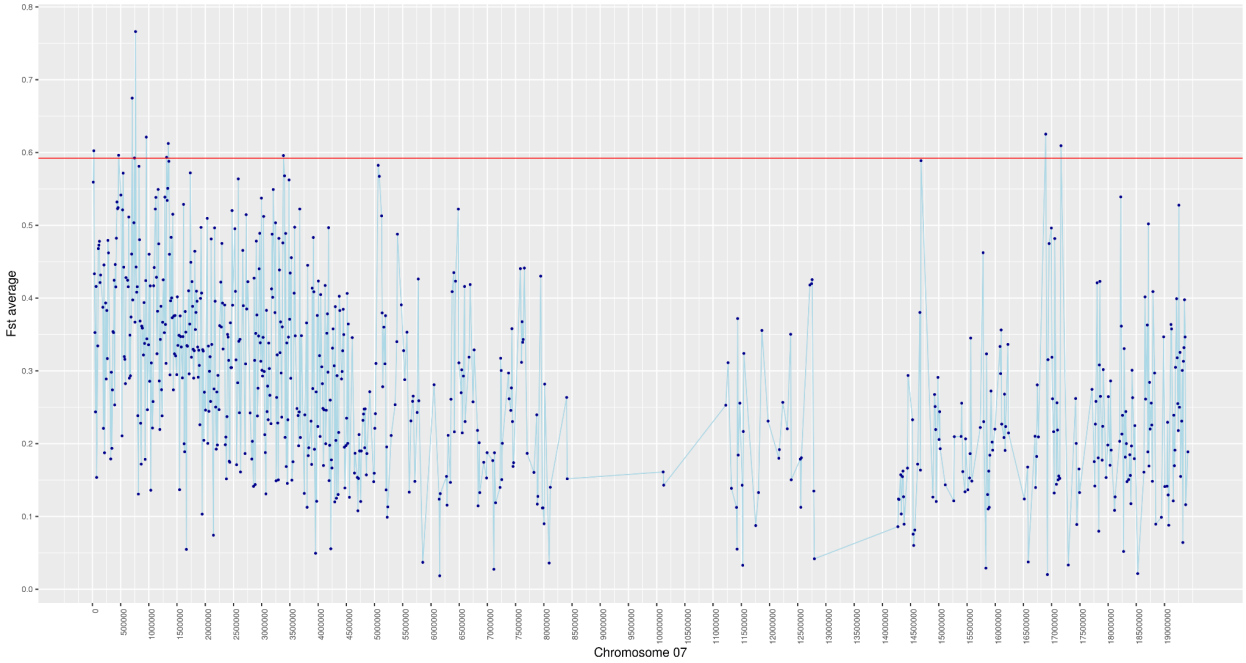


Figure 7g: F_{st} average per window for chromosome 7

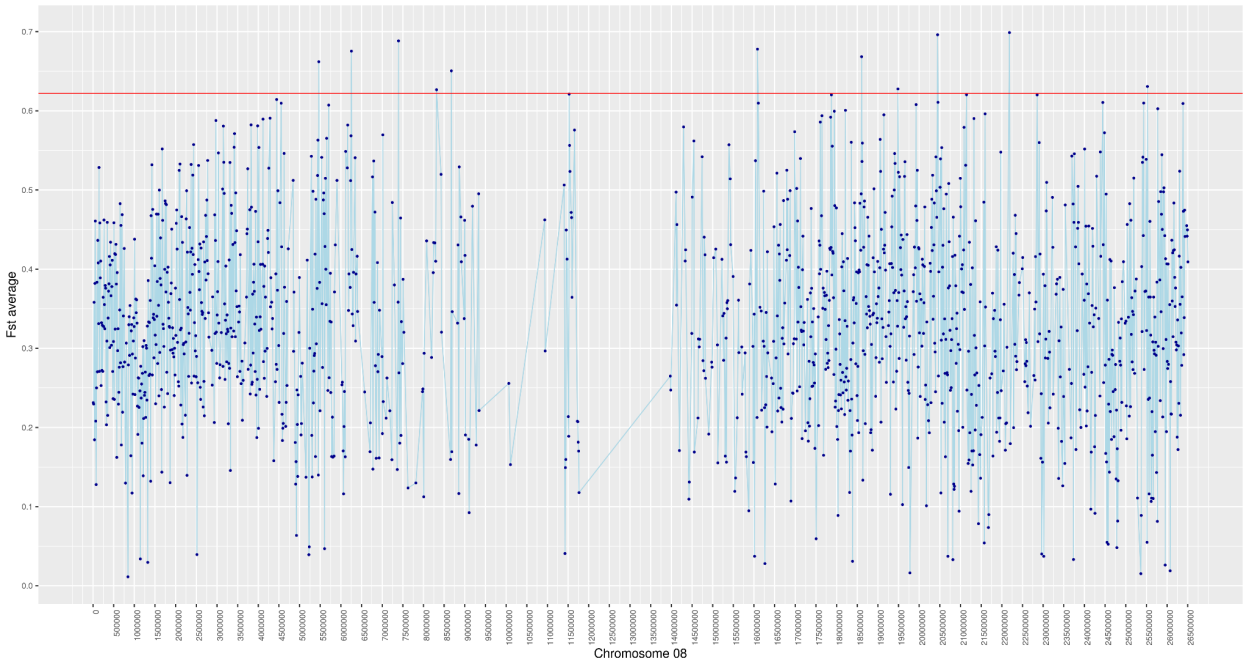


Figure 7h: F_{st} average per window for chromosome 8

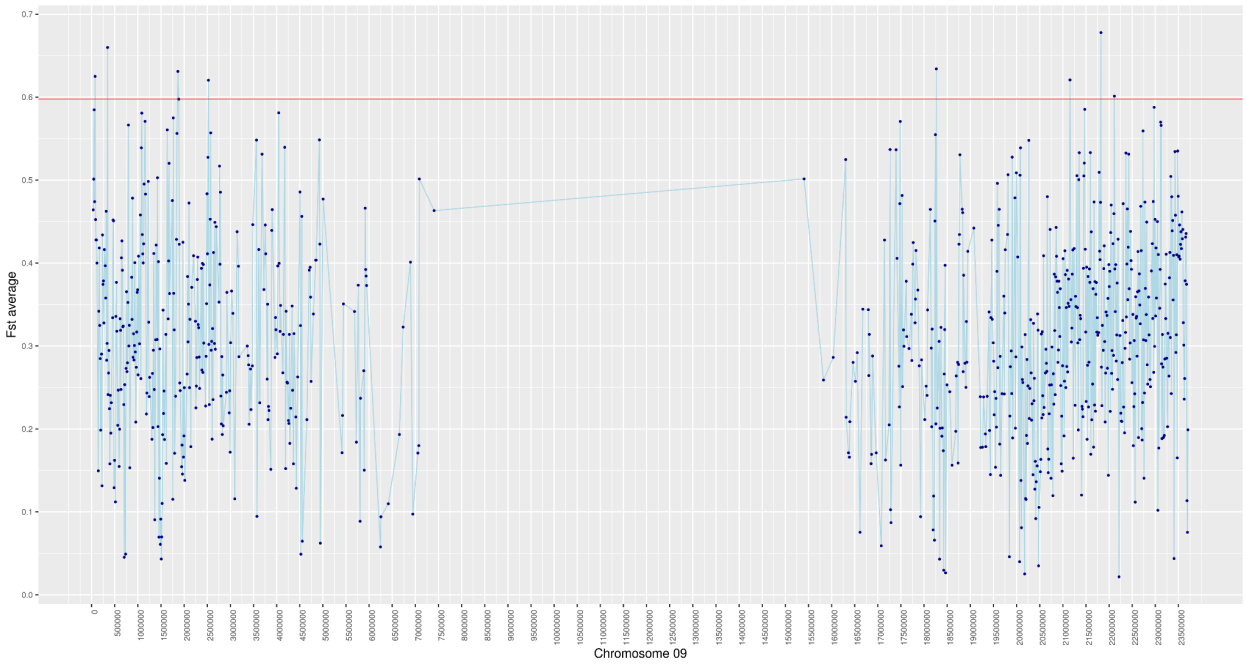


Figure 7i: F_{st} average per window for chromosome 9

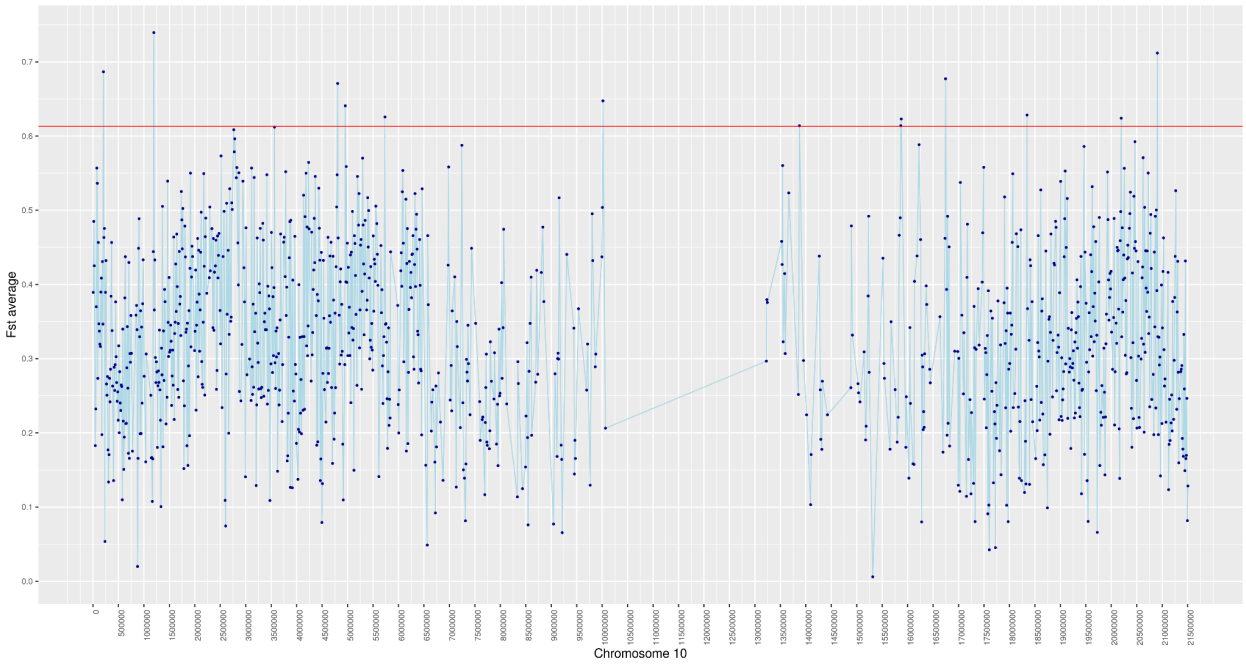


Figure 7j: F_{st} average per window for chromosome 10

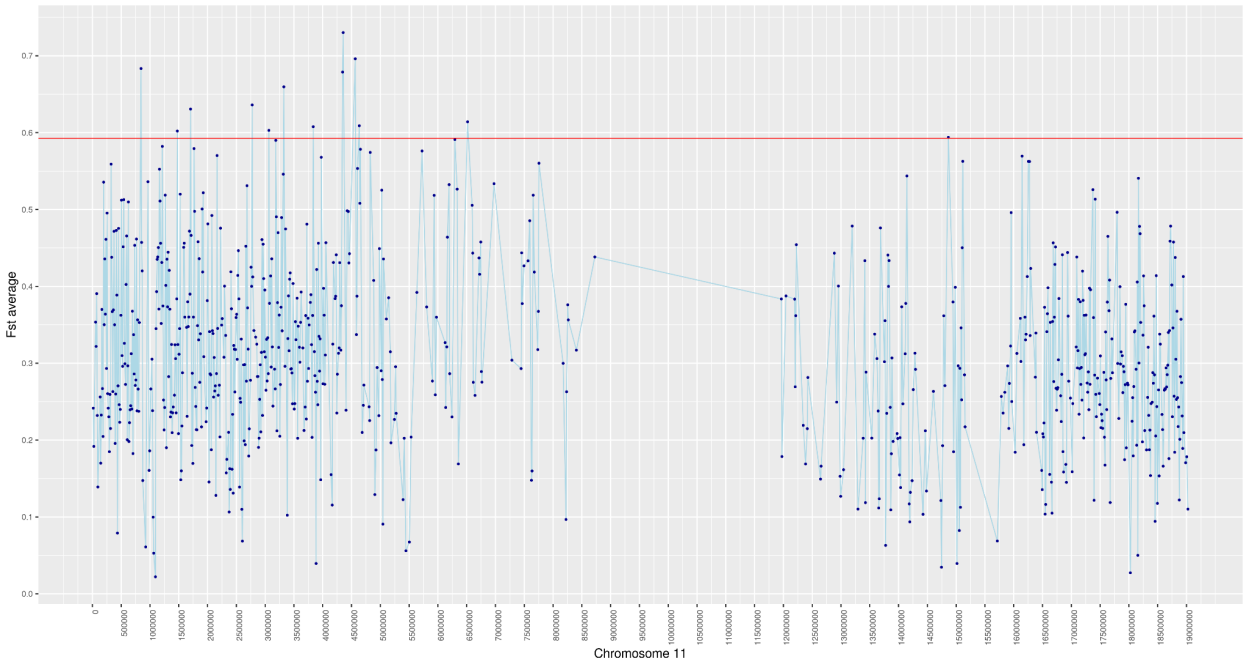


Figure 7k: F_{st} average per window for chromosome 11

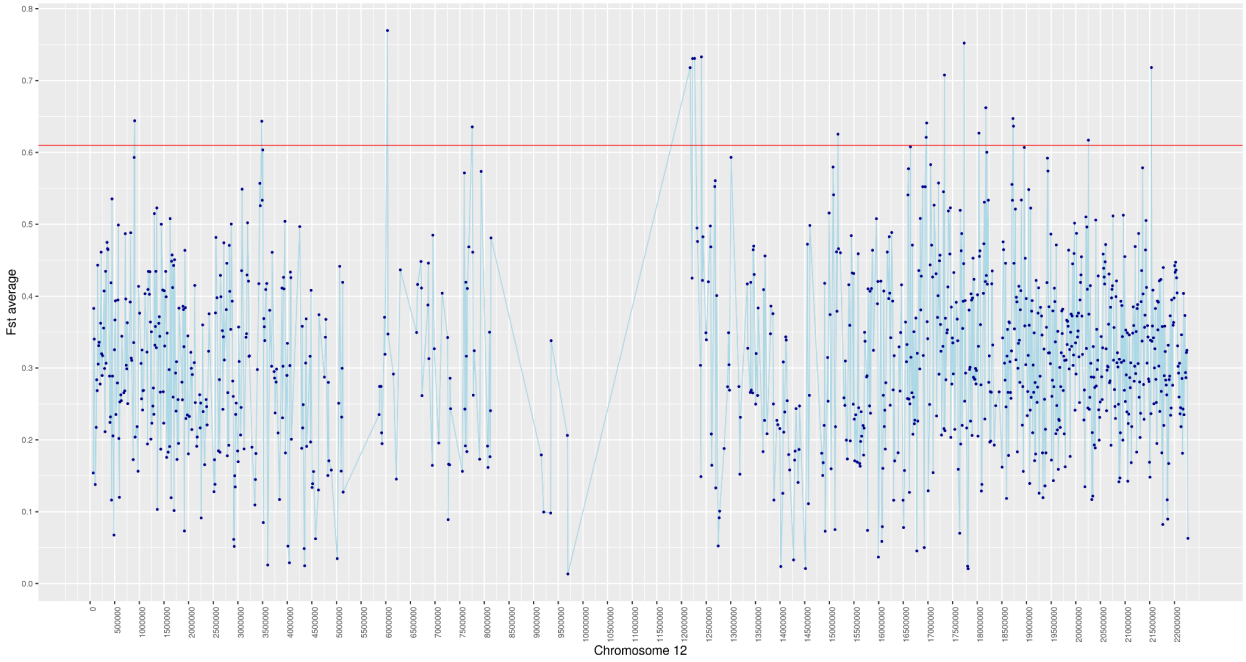


Figure 7l: F_{st} average per window for chromosome 12

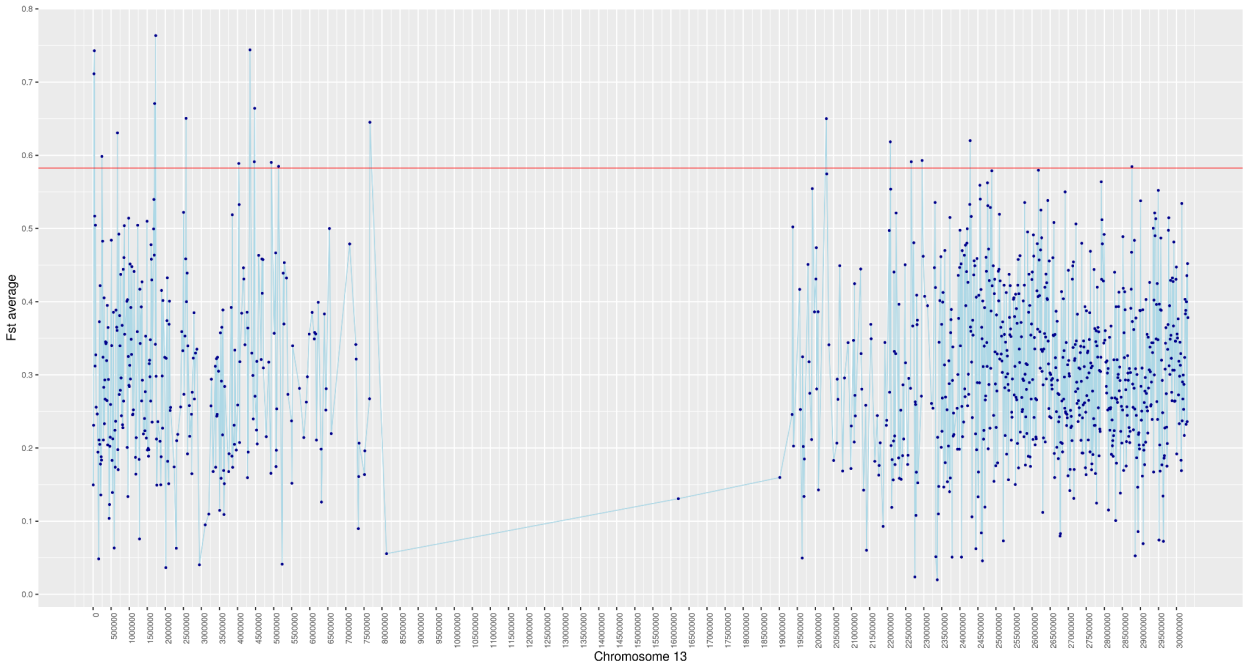


Figure 7m: F_{st} average per window for chromosome 13

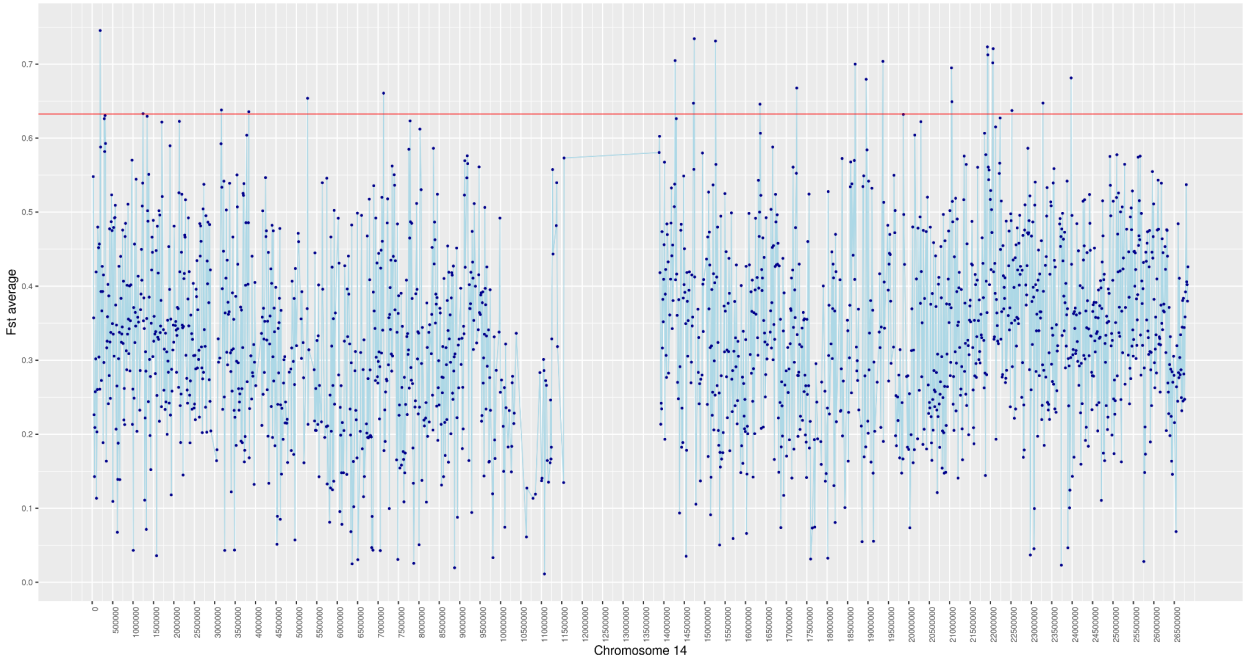


Figure 7n: F_{st} average per window for chromosome 14

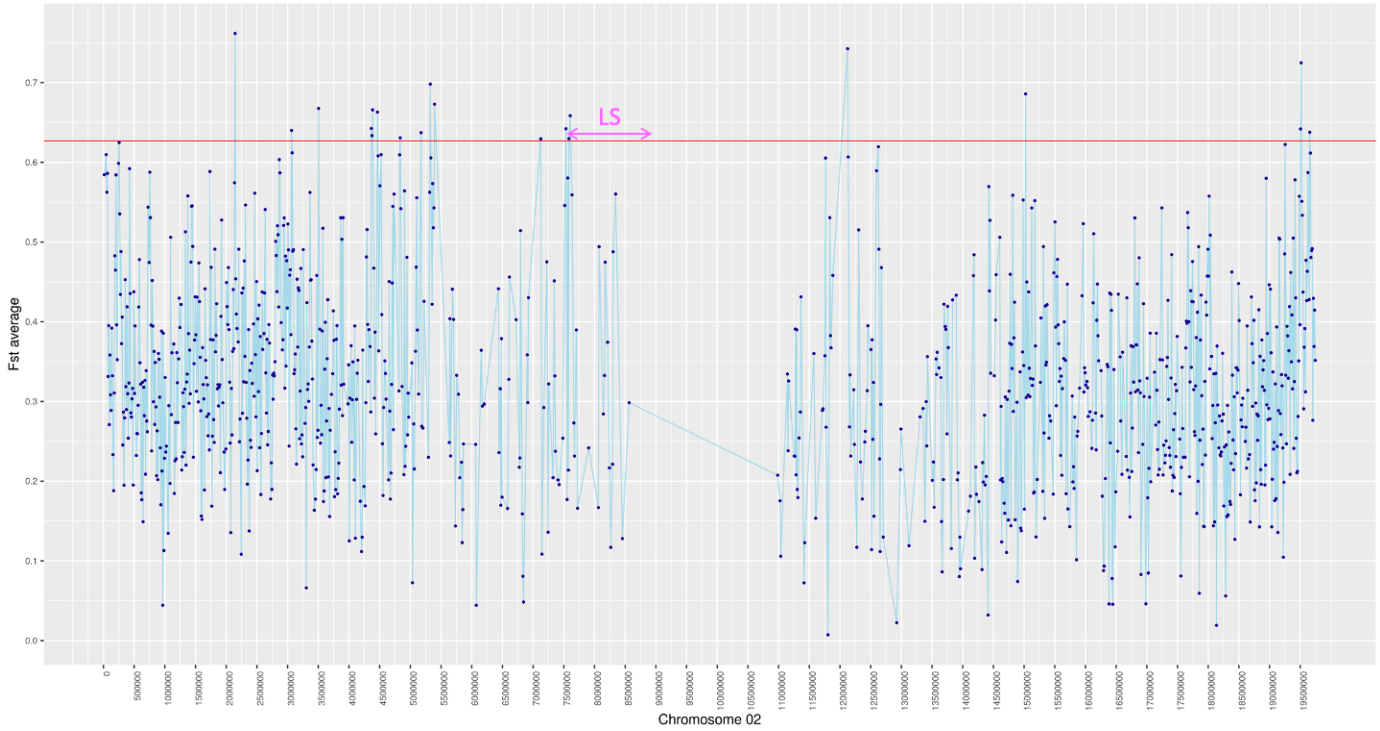


Figure 8a: F_{st} average per window for chromosome 2 labeled with QTL from Ferris et

al., 2018

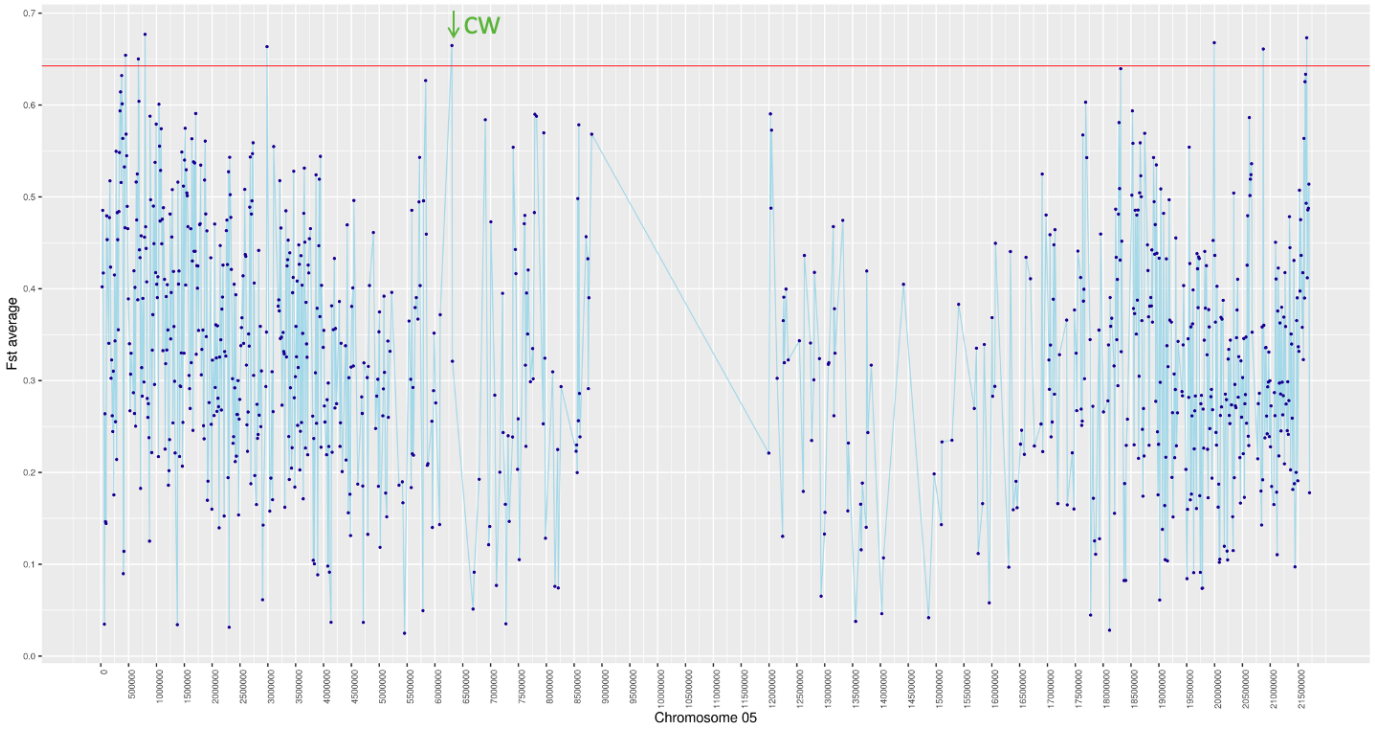


Figure 8b: F_{st} average per window for chromosome 5 labeled with QTL from Ferris et al., 2018

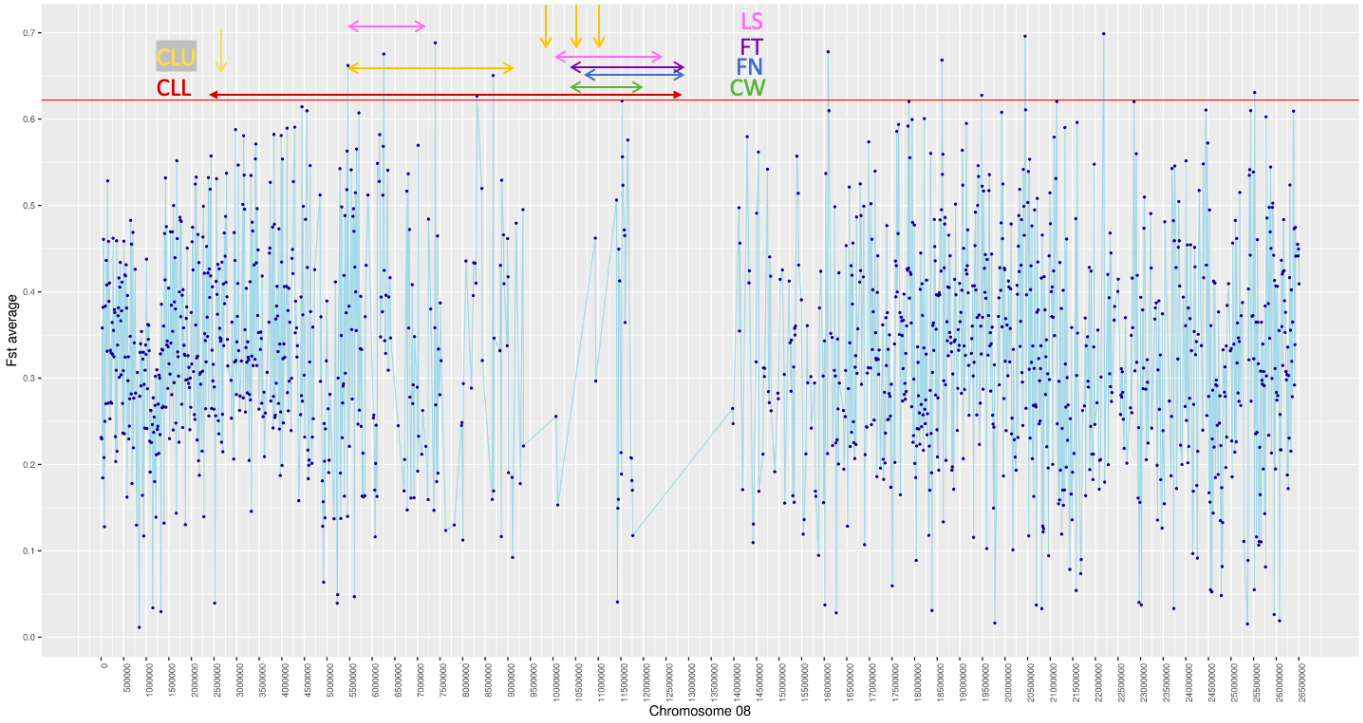


Figure 8c: F_{st} average per window for chromosome 8 labeled with QTL from Ferris et al., 2018

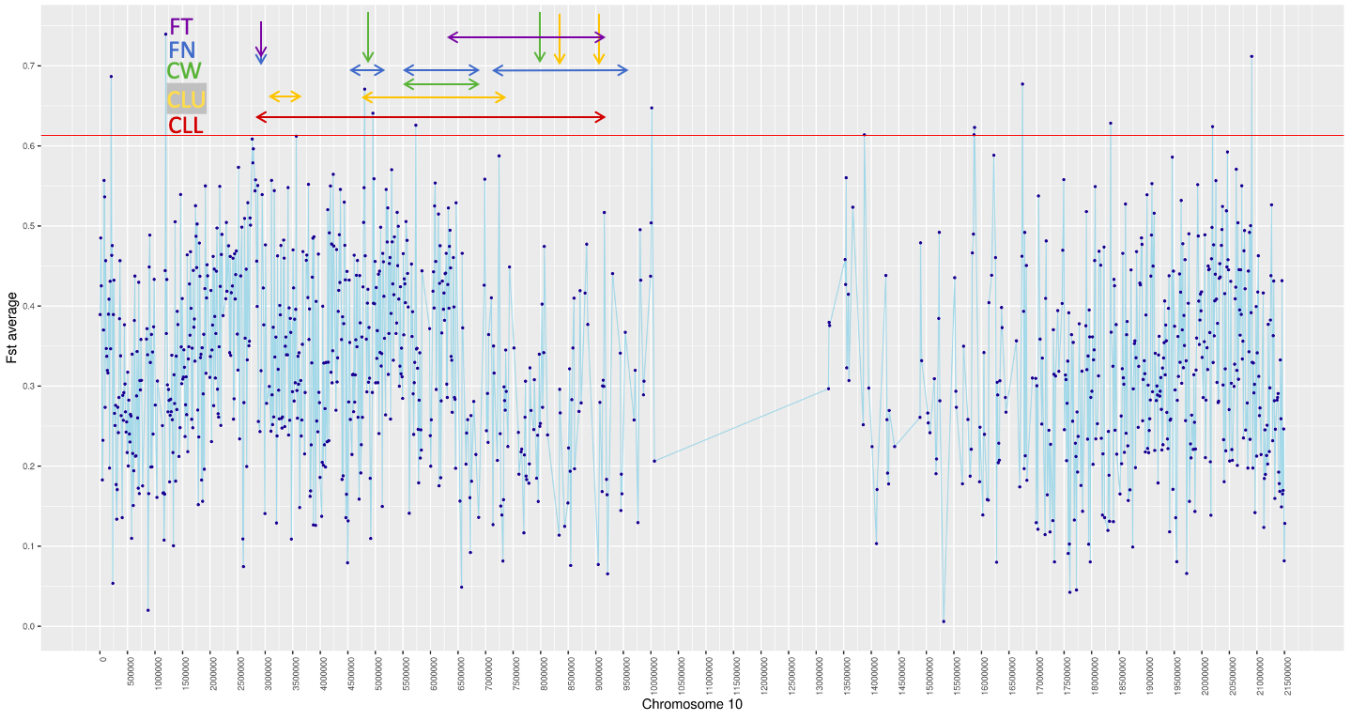


Figure 8d: F_{st} average per window for chromosome 10 labeled with QTL from Ferris et al., 2018

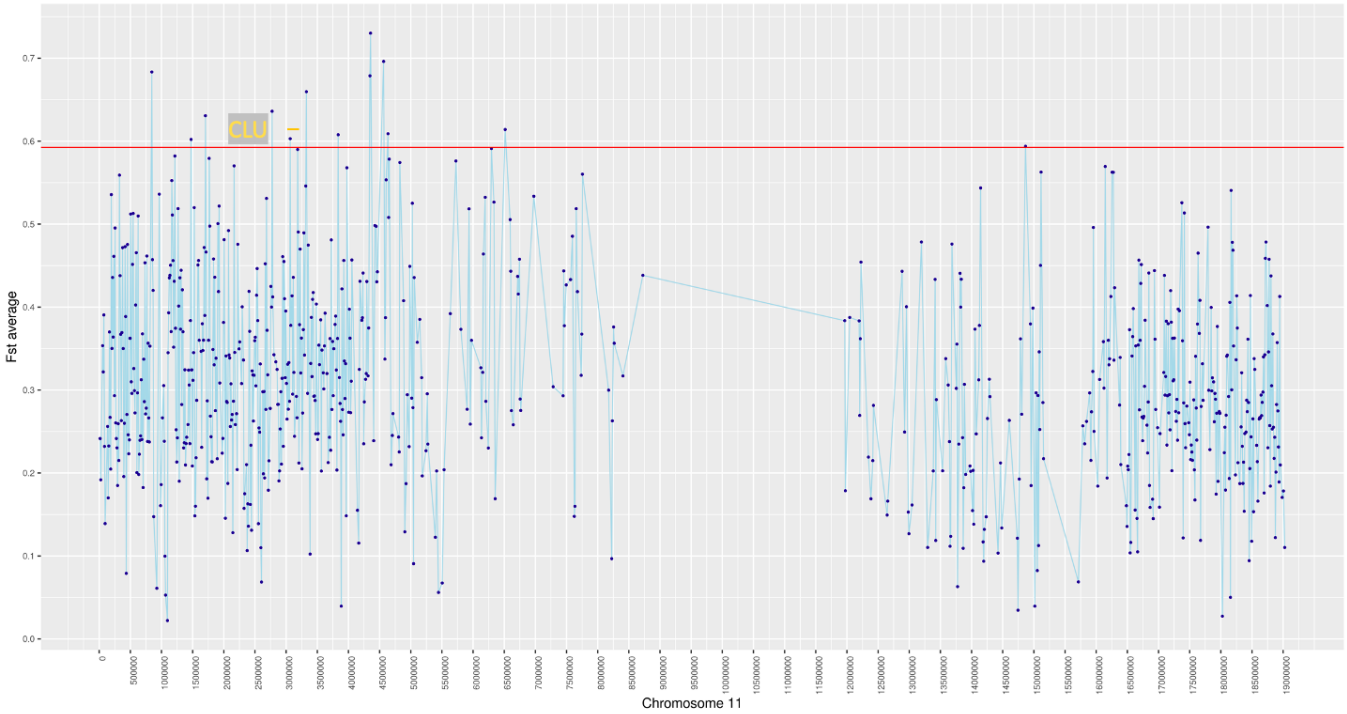


Figure 8e: F_{st} average per window for chromosome 11 labeled with QTL from Ferris et al., 2018

Tables

Table 1: Average F_{st} per chromosome

Chromosome number	Average F_{st}

1	0.1430
2	0.1642
3	0.1471
4	0.1583
5	0.1637
6	0.1719
7	0.1411
8	0.1610
9	0.1498
10	0.1639
11	0.1478
12	0.1557
13	0.1507
14	0.1684

Table 2: Significant Fst hits that overlap with QTLs identified in Ferris et al., 2018

	Chr 2	Chr 5	Chr 8	Chr 10	Chr 11
CLL	0	0	6	3	0
CLU	0	0	4	2	3
CW	0	1	1	2	0
FN	0	0	1	2 & 1	0
FT	0	0	1	1	0
LS	3	0	3 & 1	0	0

Table 3: Candidate Genes and Their Functions as Described by the Literature

Gene	Chromosome	Position	Main Function
Histidine decarboxylase / L-histidine carboxy-lyase	2	5220561..5223966 reverse	Amino acid metabolism during plant growth and development
Mads box protein	2	5227386..5228142 forward	Involved in reproductive tissues

Aspartyl proteases	2	5292941..5296791 reverse	Autophagy as fungal resistance in plants
Isopenicillin-N-epimerase	2	5318236..5320531 forward	Fungal resistance in plants
Glucan endo 1,3-beta-D-glucasidase	2	5339078..5341050 reverse	Plant defense by digesting wall components of the fungal pathogen
Serine/threonine protein kinase haspin	2	5355006..5360068 reverse	Cell-cycle cessation and differentiation of haploid germ cells
TBC1 domain, GTPase activating protein	2	5368086..5372412 reverse	Major control element in the coordination and definition of specific trafficking steps and intracellular compartments

Sulfate transporting ATPase // Xenobiotic transporting ATPase	2	5387479..5396027 forward	Increase tolerance to soil with metal contaminants
F-box associated domain	2	5399424..5400821 forward	Plant hormone response Required for flower formation in Arabidopsis
SANT/Myb domain // homeodomain-like	7	16705147..1671086 5 reverse	Related to fruit development in tomatoes and seedling development in arabidopsis Could be a fertility-related gene

Rab GTPase activator protein	7	16734940..16744287 reverse	Modulator for cell death progression during pathogen response and senescence process in plants Determinant of membrane identity and membrane targeting
Protein NLP8	7	16763113..16768118 forward	Master regulator of nitrate-inducible gene expression in higher plants
Seed storage 2S albumin super family protein related	7	16931305..16932203 reverse	Seed storage proteins by providing amino acids and other nutrients during

			germination and for seed defense
Protein TBC-3, isoform B	7	16935052..16939885 forward	Chitinase involved in cell-cell communication
Oxidoreductase, 20G-FE(II) // 2-oxoglutarate (20G) & Fe(II) dependent oxygenase	7	16952105..16955800 forward	Prevents the triggering of host defense when a Hs4F01 annexin-like effector is secreted into the cytosol
Anthocyanidin 3-O-glucoside 2”-o-xylosyltransferase	7	16989574..16991327 forward	Plays the major role in sinapoylation of anthocyanins
Protein kinase domain // EGF-like domain // EGF calcium binding domain	8	8002166..8007220 reverse	Involved in plant defense and heavy metal responses. Cell elongation and plant development

ATP-dependent RNA helicase DDX18/HAS1	8	8014877..8020137 reverse	Potentially plays a role in salt and cold stress in plants
WDSAM1 protein	8	8077771..8079514 reverse	Epigenetic regulation mechanisms involved in developmental transitions in plants, including seasonal changes in fruit trees

Histone deacetylase	8	8192795..8196913 forward	Gene silencing Required for callus formation from leaf explants in *Arabidopsis*

Protein IDA-like 3-related	8	8227190..8228189 forward	Stress response and floral abscission Suggested to be involved in regulation of *Arabidopsis* development
RAPSYN-related // Protein Tonsoku	8	8237476..8246108 reverse	Maintenance of meristem organization Heat shock memory in *Arabidopsis thaliana*
GCN5-related // Biogenesis of lysome related organelles complex 1 subunit 1	8	8277653..8279000 forward	Involved in the responses to heat stress, cold stress, and nutrient element deficiency

NAC domain containing protein 76-related	8	8313902..8316975 reverse	Implicated in important plant developmental processes like boundary cell formation in shoot apical mersitems secondary cell wall formation, and lateral root development
LAG1-DNA binding	12	9554129..9557087 forward	Involved in adipogenesis in mammals
Cucumisin	12	9580281..9582543 reverse	Plant serine protease
FHA domain // BRCA1 C terminus (BRCT) domain (BRCT)	12	9686307..9702650 forward	Regulate a variety of signaling pathways involved in plant growth and

			pathogen responses
E2F TRANSCRIPTION FACTOR-LIKE E2FE	12	9719187..9721492 forward	Genes associated with pollen fertility and seed formation
Casparian strip membrane protein	12	10301262..10302405 forward	Establish a plasma membrane and extracellular diffusion barrier in plants, and represent a novel way of epithelial barrier formation in eukaryotes
Phenylalanine T-RNA ligase	12	10801317..10802405 forward	Involved in the synthesis of phenylalanine T-RNA

Leucine-rich PPR motif-containing protein, mitochondrial (LRPPRC)	12	12172579..12177109 forward	Regulation of organelle genome expression
Sugar transport related protein	12	12197486..12199210 forward	Sugars serve not only as a metabolic energy source for sink tissues but also as signaling molecules, affecting gene expression through conserved signaling pathways to regulate plant growth and development