2019

Adaptation to a geothermal soil mosaic shapes genome-wide patterns of diversity and differentiation in Yellowstone monkeyflowers (M. guttatus)

Kory M. Kolis
University of Montana

ADAPTATION TO A GEOTHERMAL SOIL MOSAIC SHAPES GENOME-WIDE DIVERSITY AND DIFFERENTIATION IN YELLOWSTONE MONKEYFLOWERS (*MIMULUS GUTTATUS*)

By

KORY MICHAEL KOLIS

B.A. Gustavus Adolphus College, St. Peter MN, 2015

Thesis

presented in partial fulfillment of the requirements
for the degree of

Master of Science
in Organismal Biology, Ecology and Evolution

The University of Montana
Missoula, MT

May 2019

Approved by:

Scott Whittenburg,
Graduate School Dean

Lila Fishman, Chair
Division of Biological Sciences

Zachary Cheviron
Division of Biological Sciences

Scott Miller
Division of Biological Sciences

Kolis, Kory, M.S., Spring 2019 Organismal Biology, Ecology, and Evolution

Adaptation to a geothermal soil mosaic shapes genome-wide patterns of diversity and differentiation in Yellowstone monkeyflowers (*M. guttatus*)

Chairperson or Co-Chairperson: Lila Fishman

Abstract

Local adaptation across habitat mosaics can generate phenotypic divergence in the face of gene flow; however, adaptive divergence in reproductive traits may also create barriers to genetic exchange within and among distinct habitats. In plants, life-history, phenology, and mating system traits may lead to divergent selection over short (microgeographic) spatial scales. Changes to these traits are likely to directly affect patterns of gene flow and genomic diversity. In this study we combined field, common garden, and population genomic approaches to investigate phenotypic and genetic variation in *Mimulus guttatus* (yellow monkeyflowers) adapted to a complex geothermal soil mosaic in Yellowstone National Park (YNP). A previously-identified major locus underlying life-history divergence (*out6*) strongly sorts by habitat (thermal annual habitats [AH] vs. non-thermal perennial habitats [PH]) across YNP. Plants from AH and PH were also differentiated for self-pollination potential and flowering time traits in the common garden, consistent with adaptation to spring-flowering in thermal habitats. Genome-wide sequence data (ddRADSeq) reveals one highly differentiated (and ecologically extreme) AH population (AHQT), while the other AH and PH plants form four geographic populations. F'_{ST} between the geographic regions varied but remained relatively high (~0.10) across all comparisons. F'_{ST} between AH-AH sub-populations pairs were marginally more differentiated than PH-PH pairs (F'_{ST} = 0.13 vs. 0.10, respectively). Slightly elevated differentiation of thermal annual populations mirrors the isolation of AHQT from all

other populations and suggests that increased selfing and phenological assortative mating in thermal habits generate structure through reduced gene flow and increased drift. Individual inbreeding coefficients (F_{IS}) were positively associated with mean progeny stigma-anther distance and significantly elevated in thermal annual habitats. This is consistent with the inference that thermal habitats select for efficient self-pollination, with consequences for individual and population variation. Overall, multi-trait adaptation to geothermal soils occurs despite ongoing gene flow with nearby nonthermal populations, and parallel selective pressures in extreme thermal soils have reassembled similar adaptive phenotypes on distinct genetic backgrounds. To varying degrees, thermal annuals exhibit genomic signatures of elevated population differentiation that suggest they may be in the early stages of developing local reproductive isolation.

ACKNOWLEDGEMENTS

First, I would like to thank my adviser, Lila Fishman. Without Lila's guidance, advice, and dedication, I would not have been able to complete this thesis. I would also like to acknowledge my committee; Zac Cheviron, and Scott Miller for their advice, support, words of confidence, and interest in seeing me grow as a researcher.

I would like to thank all current and past Fishman lab members for their support, help, advice, and reassurance, particularly, Thom Nelson and Colette Berg. I truly could not have done it without you. Additional thanks to for the assistance in the field; Peter Breigenzer, Richard Hanes, Mariah McIntosh, and others. I could not have done this alone.

I would like to thank all of the Yellowstone National Park staff and affiliates that made the field component of my thesis possible, particularly Heidi Anderson and Annie Carlson for all the assistance getting in and around Yellowstone National Park.

Lastly, I would like to thank my family. There are no words to express the love I have for all of you. Mom, Dad, Stefan, Melissa, Ben, Teige, thank you.

TABLE OF CONTENTS

List of Tables

List of Figures

INTRODUCTION

Understanding when and how local adaptation influences genome-wide patterns of genetic variation and can shed light on longstanding evolutionary questions. Gene flow (migration of alleles in seeds or pollen) homogenizes neutral genetic variation genome-wide (Slatkin 1973; Lenormand 2002), and extrinsic barriers to gene flow (distance in continuous landscapes, gaps in suitable habit) allow populations to diverge neutrally by drift (Gavrilets 2003; Coyne & Orr 2004; Wright 1943). Across environmental gradients or mosaics, divergent selection further increases differentiation at loci contributing to local adaptation, even in the face of gene flow. This is particularly true if alternative alleles are beneficial in one environment, but costly in another (antagonistic pleiotropy), as opposing selection will then act to maintain the differences between the habitats (Hedrick 1986; Anderson et al. 2011). Thus, studies of adaptation in the face of gene flow often identify loci that are highly differentiated against a relatively homogenized genetic background (Hoekstra et al. 2004; Jones et al. 2018; Calsbeek & Smith 2003). However, adaptation to local environments can also influence neutral genomic differentiation if it alters reproductive traits that directly influence patterns of gene flow (Antonovics 1968; Wang & Bradburd 2014). At the extreme, this could to lead to sympatric speciation, but it is still not clear (empirically) whether and how local adaptation can lead to the evolution of reproductive isolation in the absence of geographic barriers to gene flow across habitat boundaries.

Plants often exhibit local adaptation to geologically-generated soil mosaics, where differences in water availability or mineral concentration may exert strong selection on diverse traits (Brady et al. 2005; Gardner & MacNair 2000). Thus, changes to soil

environments over short distances can generate strong divergent selective regimes well within the range of seed and pollen movement (Richardson et al. 2014). In such mosaics, we might expect differentiation of traits and the underlying loci, but little neutral genomic divergence. For example, *Arabidopsis thaliana* across elevational gradients are differentiated for loci related to water and temperature stress but have homogenized neutral backgrounds (Frachon et al. 2019). In contrast, divergent selection that causes shifts in mating system and flowering phenology in sweet grass (*Anthoxanthum odoratum)* on and off mine tailings, resulting in partial reproductive isolation (Antonovics 1968; Caisse & Antonovics 1987). Changes in reproductive traits has even been cited in maintaining strong enough barriers to gene flow to generate sympatric speciation (Savolainen et al. 2006; Babik et al. 2009). Local selection on plant reproductive traits offer a unique setting to understand how barriers to gene flow may or may not arise.

For plants, flowering phenology can directly affect patterns of gene flow, and is often locally adapted (Schemske & Bradshaw 1999; Goodwillie et al. 2006; Lowry & Willis 2010; Ågren et al. 2016). Correct timing of flowering is critical for reproductive success, as environments frequently have a narrow period of suitable conditions in which to flower. Across ranges, flowering cues are fine-tuned to local environmental signals, and often tend to vary predictively across latitudinal or elevational transects (Twyford & Friedman 2015; Koornneef et al. 1998). Additionally, moisture gradients can also influence flowering time, as plants are selected to reproduce before conditions for seed production are poor (perennials) or they die (annuals) (Sakaguchi et al. 2018; Wright et al. 2006). Changes to flowering phenology can generate assortative mating by time,

potentially generating intrinsic barriers to gene flow between populations in distinct but adjacent habitats (Hendry & Day 2005; Kenney & Sweigart 2016).

Similarly, a population's mating system (e.g. outcrossing or self-fertilizing [selfing]) often reflects local selective pressures (Goodwillie et al. 2006; Ågren et al. 2016). The transition from outcrossing to selfing is one of the most common evolutionary transitions in flowering plants (Barrett & Harder 1996; Barrett 2002; Charlesworth & Wright 2001). Selfing can arise through increased allele transmission (automatic selection advantage, Fisher 1941) overcoming the costs of inbreeding depression. However decades of empirical work suggest that selfing often evolves because it confers reproductive assurance (guaranteed seed set) in the absence of pollinators or (suitable) mates (Baker 1955; Cheptou 2011; Pannell et al. 2015). Selfing species are commonly found in ephemeral, harsh or marginal habitats (Munoz et al. 2016; Barrett 2002), reflecting both the reproductive assurance value of selfing for individuals and high colonization ability (Baker 1955; Cheptou 2011; Grossenbacher et al. 2015).

Evolutionary shifts toward increased self-fertilization reduce gene flow through pollen import and export, increase drift through reductions in effective population size, and also alter individual genetic diversity (Wright et al. 2008; Charlesworth & Wright 2001). Increases in selfing rates negatively impact nucleotide diversity and individual heterozygosity thereby increasing F_{is} (inbreeding coefficient) within populations (Glemin et al. 2006; Schoen & Brown 1991; Charlesworth & Willis 2009), as each selfing generation cuts heterozygosity by half. At the population-level, selfing species exhibit increased neutral differentiation between populations and lower overall genomic variation (Wright et al. 2008; Hamrick & Godt 1996; Wright et al. 2008; Sicard &

Lenhard 2011). Numerous empirical studies have demonstrated these population-level genetic consequences of plant mating system variation (Munoz et al. 2016; Kamran-Disfani & Agrawal 2014; Rausher 2013; Busch et al. 2011; Sicard & Lenhard 2011; Imai et al. 2016; Ness et al. 2010). For example, the selfer *Capsella rubella* exhibits elevated population differentiation (F_{ST}) and low nucleotide diversity compared to its outcrossing sister species, *C. grandiflora* (St Onge et al. 2011). However, work on the evolutionary origins and genetic consequences of selfing has primarily focused on established allopatric populations with contrasting mating systems where causes and effects of mating system adaptation are confounded by historical contingency and geographical isolation (Rausher 2013; Busch et al. 2011; Foxe et al. 2009; Lucek et al. 2019). Only in systems where both selfing and outcrossing occur can the genomic consequences of selfing fully be explored.

In this study, I investigated how divergent selection, neutral processes, and mating system influence phenotypic variation, population structure, and individual genomic variation in yellow monkeyflowers (*Mimulus guttatus*) adapted to a geothermal soil mosaic in Yellowstone National Park (YNP). *M. guttatus* is one of a few vascular plants that lives in extreme geothermal soils and has shifted its life-history to take advantage of the unique geothermal soil environment. Geothermal activity in YNP (springs, pools, dry vents) melts snow from adjacent soils throughout the winter, creating suitable *M. guttatus* habitat, but desiccate and heat the soils to > 45 °C in summer. *M. guttatus* requires saturated soils to persist, so populations in extreme thermal soils are obligately (and genetically; Lekberg et al. 2012; Hendrick et al. 2016; Nelson et al., in prep) annual and flower in the late spring or early summer before the soils dry out. However, the annual

thermal habitats are intermixed with cool rivers and bogs occupied by late-summer-flowering perennials (the presumed progenitors of thermal annuals). These extremes are connected (in many but not all cases) by intermediate habitats, such as perennially wet sites that receive warm water year-round or annual sites that only dry out in September. This complex habitat matrix generates a unique setting to examine life-history and mating system adaptation, population structure, and genomic variation on a microgeographic scale.

Previous work in this system has focused on three pairs of populations inhabiting extreme thermal annual habitats (AH) and nearby non-thermal perennial habitats (PH) (Lekberg et al. 2012; Hendrick et al. 2016). Under common garden conditions, AH plants (particularly from the extreme AHQT population) produce few to no rhizomes (the key trait that distinguishes annuals from perennials), are dwarf, flower under short days, and have reduced floral size, and stigma-anther distance (Lekberg et al. 2012). In contrast bog-dwelling PH plants produce abundant rhizomes, require long days to flower, and have relatively large corollas and stigma-anther distances. QTL mapping and F_{ST} outlier analysis using pooled whole genome sequencing of AH and PH pairs further identified a single genetic locus (*out6*; containing ~25 genes) that explains much of the life-history differentiation between the focal AH and PH populations (Nelson et al., in prep). The PoolSeq analyses also found variable levels of genome-wide differentiation among the three pairs. One population pair (AHQ) showed high differentiation (F_{ST} = 0.19) despite being less than 200m apart, whereas two others showed low differentiation (F_{ST} = 0.04 - 0.07) (Nelson et al., in prep). This suggests that adaptation to thermal habitats may

sometimes occur in the face of gene flow and sometimes contribute to barriers to gene flow.

Prior work on YNP *M. guttatus* thus provides a strong foundation to address questions about the causes and consequences of local adaptation in the absence of geographic isolation between habitats. Through characterizing phenotypic and genetic variation of *M. guttatus* across the complex habitat mosaic of the entire YNP system, we can examine how phenotypic differentiation arises across a habitat mosaic, and how such differentiation impacts population structure and individual genomic diversity. In this study, I examine 1) how *M. guttatus* traits vary (genetically) across habitats and life-histories, 2) how geography and habitat structure genomic variation, and 3) how changes in mating system impact individual-level diversity (F_{IS}).

MATERIALS & METHODS

Study system

The *Mimulus guttatus* species complex encompasses ecologically diverse but interfertile populations and named species (Brandvain et al. 2014). *M. guttatus* displays wide variation between (Wu et al. 2007; Kiang & Hamrick 1978; Beardsley et al. 2004) and within populations (Kooyers et al. 2019; Ferris et al. 2015). Additionally, the species shows repeated independent transitions in life-history (Lowry & Willis 2010; Twyford & Friedman 2015; Coughlan & Willis 2018), mating system (Fishman et al. 2002), and adaptation to harsh and ephemeral soils (Selby & Willis 2018; Ferris et al. 2015; Oneal et al. 2014). These diverse transitions and a short life-cycle makes *M. guttatus* a tractable study system and an ideal candidate to place evolutionary transitions in a comparative

framework. Additionally, a publicly available reference genome (www.Phytozome.jgi.doe.gov; Hellsten et al. 2013) facilitates genomic and genetic research. This system is both a well-characterized and diverse context to investigate how adaptation through changes in mating system, phenology, and geography impact gene flow patterns, and population structure.

M. guttatus occur throughout Yellowstone National Park, in a soil mosaic spanning cool perennial bogs and streamsides to harsh mineral crusts made hospitable to *Mimulus* only by snowmelt. Plants on the thermal soils can grow throughout the winter but must flower in early spring, before the soils desiccate and become unlivable. In contrast, the plants in the bogs and streams begin growing in late spring upon snowmelt and flower in late August, investing heavily in future reproduction through vegetative growth (stolons/rhizomes). *Mimulus guttatus,* both on and off the geothermal soils, are found in nearly all of the major geyser basins (exceptions: Mammoth Geyser Basin and Mud Volcano). The mosaics are often closely intermingled, with perennial and annual *M. guttatus* frequently within 100m of each other but can be highly discrete and distant (> 1 km). The majority of the thermal features in YNP occur throughout a relatively continuous matrix from the Upper Geyser Basin to the Lower Geyser Basin. However, several geyser basins, such as West Thumb, Shoshone, and Heart Lake in the southeastern region and Violet Creek in the northeastern region of the park, are more geographically isolated. Geographical structure (at a larger scale than the soil variation) allows us to examine the influence of habitat on phenotypic and genetic changes relatively independently.

Plant and environmental data collection

M. guttatus individuals were collected from 177 geo-referenced sites across Yellowstone National Park's geyser basins (n = 2-3 per location, N = 550; Fig.1) in the summer of 2017. Collection sites were chosen to evenly represent the two habitats and to sample individuals from distinct but geographically close annual and perennial habitats. However, due to the uniqueness of each geyser basin, sampling was not always even. The three geyser basins that make up the majority of the central populations are biased toward annual habitats, whereas the Northern region was biased toward perennial habitats, as the geothermal features in these areas are adjacent to cool rivers and springs. Wild plants were transplanted to the UM Greenhouse, DNA collected, and flowers hand-pollinated to generate inbred seeds. Three collection locations (AHQT, AHQN, RBC) were the focal populations in previous studies (Hendrick et al. 2016; Lekberg et al. 2011) and seed samples were collected prior to 2017. In August 2018, each collection location was revisited and categorized as an annual habitat (AH) or perennial habitat (PH) on the basis of soil moisture and the occurrence of live plants (annual habitats (AH) = dry, dead; perennial habitats (PH) = wet, live adults).

Common garden phenotyping

In 2018, phenotypic traits were measured in a common garden using two 1st generation inbred siblings from one wild-collected plant from each site (N = 177 sites, N = 363). Seeds were germinated in 24-well microplates on wet sand and stratified for two weeks at 4 ºC in the dark. The seeds were then transferred to 16-hour days to cue

germination. Multiple seedlings per parent were transplanted into two, two-inch pots in case of transplant related death and culled to one plant pot upon successful transplanting. On the day of their first flower (FF), we recorded days since transplant, height at FF, number of nodes to FF, corolla width, corolla length, and stigma-anther distance following the protocol from Fishman et. al 2002. Rhizome number and final height was recorded for all plants 85 days after transplantation.

Genotyping – PCR markers

DNA for the marker-based genotyping and RADSeq libraries were extracted using a standard CTAB-chloroform extraction protocol (Doyle 1987) modified for 96-well plates. Wild-collected (N = 345) and 1st generation inbred phenotyped plants (N = 363) were genotyped using the PCR-based markers mL1029, mF2118, and mF2137, which are diagnostic of *out6* T and N haplotypes identified in previous work. Markers were sized on an ABI 3130 automated sequencer with an in-lane size standard. The fragments were visualized with the inclusion of fluorescent labels by M13-tailing of the forward primer in the PCR amplification reaction. Genotype scores were called in GENEMAPPER software (Applied Biosystems, Waltham, MA).

Genotyping – High-throughput sequencing

We used a double-digest restriction-site associated DNA sequencing technique (ddRADSeq) to generate genome-wide sequence tags from wild-collected plants. The RADSeq libraries were generated with a modified BestRAD protocol (Ali et al. 2016). Restriction Enzymes (*PstI* and *BfaI)* were chosen after an *in silica* digestion indicated

that the combination produced highly genic and evenly dispersed tags. Individual DNAs were labeled with in-line barcoded oligos and pooled in sets of 48. Pools were barcoded using NEBNext indexing oligos with a degenerate barcode and libraries prepared using NEBNext Ultra II library preparation kits for Illumina (New England BioLabs, Ipswich, MA). The libraries were paired-end sequenced using two 1/2 lanes on a HiSeq4000.

Sequence data were demultiplexed using a custom Python script. Once demultiplexed, the adapters were trimmed, and low-quality read removed using trimmomatic (Bolger et al. 2014). Reads were then mapped to the *M. guttatus* V2 reference genome (https://phytozome.jgi.doe.gov) using BWA (Li & Durbin 2010). Lastly, reads were indexed with SAMtools (Li et al. 2009). Our sequencing and read processing filters produced 18x coverage (on average) across RAD loci for all individuals. Individuals with less than 5x coverage and several showing evidence of contamination (excess heterozygosity, known pipetting errors) were removed from the data set (N = 302). Heterozygosity can be underestimated in lower coverage sequence data, however, the coverage varied randomly across the plates, it should impact all comparisons equally. No qualitative differences were detected in analyses including only 20x coverage individuals compared to analyses including 5x individuals, so we opted to be more inclusive in our coverage restrictions.

SNPs were called using the Stacks pipeline "populations" (Catchen et al. 2013; Catchen et al. 2011). Across all analyses, the minimum frequency a locus must be present in a population to be included (0.8), and the minor allele frequency (0.1) were held constant (Supplemental Table 1 & 2). Lower minor allele frequency cut-offs did not change DAPC clustering results, though it did increase the total pools of SNPs.

The initial run included all individuals as a single population. Using the DAPC analyses (see below) to inform population clustering, "populations" was run with individuals divided into the 5 groups. To examine difference in genomic diversity and differentiation between habitat and *out6* genotype, individuals within each region were divided into either AH/PH or TT/NN, with AHQT being paired with geographically close PH/NN individuals. Heterozygotes were excluded from phenotypic contrasts due to low sample size. Lastly, to investigate whether habitat and out6 genotype association persisted at a local scale, individuals from the Upper Geyser Basin were, again, grouped into AH-PH, and TT-NN groups.

Phenotypic analyses

To test for an association between habitat type (AH/PH) and *out6* genotype (TT/Het/NN) in wild plants, I used a X^2 analysis implemented in JMP (SAS, Cary NC). I used ANOVAs to examine whether offspring phenotypes were associated with parental habitat and/or individual genotype (TT/NN; heterozygotes excluded due to small sample size). To account for the fact that I grew and phenotyped two offspring per parent, I included maternal identity as a random effect to control for the non-independence of siblings. Individuals from RBC (a subset of the Central regions), and a large majority of individuals from AHQT and AHQN (perennial sister population), were included in the RADSeq analyses (see below), but phenotyped in 2017 and thus excluded from the RADSeq-phenotype analyses, a small number of AHQT plants (3) were included in the common garden with the 1st generation inbred individuals and were included.

Population Genetic analyses

A discriminant analysis of principal components (DAPC) (Jombart et al. 2010) was used to identify and assess population structure in *M. guttatus* across YNP with the R program adegenet (Jombart 2008). DAPC uses genetic principal components and discriminant function analyses to cluster individuals into a number of discrete groups (clusters, K). Using the find.cluster script, I chose five clusters for two reasons. Five clusters was the lowest cluster number with a low BIC, and the BIC only marginally improved with increasing cluster number (Fig. 5B). The clustering was double-checked by running the DAPC script in adegenet (parameters: number of principal components retained = 20, number of discriminate axis retained = 4) to examine cluster overlap and to assess the assignment probabilities for each individual.

I performed principal components analyses used plink2.0 (Purcell et al. 2007). The first PC analyses contained all individuals 302 (Fig. 4). The first and second principal components were dominated by the separation of AHQ annual, thermal plants (AHQT) from all others. Additional PC analyses were performed on the populations with AHQT removed, and again on a small subset of geographically close individuals from the Upper Geyser Basin to further explore genetic differentiation (or lack thereof) on a more microscale region.

Once individuals were assigned to regional clusters (populations), genetic variation was assessed using the Stacks pipeline, with all statistics based only on variant sites. F'_{ST} (Meirmans 2006) was used as the differentiation statistic because F'_{ST} corrects for the maximum possible difference in allele frequencies based on the observed frequencies, and because of the correction, is well-suited for examining genetic drift,

migration, and population structure (Meirmans & Hedrick 2010). Additionally, uneven sample sizes of the regions and changes in heterozygosity are considered by correcting for total observed variation. While all differentiation statistics have their own caveats, F'_{ST} provides confidence among comparisons when individual-and population-level genomic diversity varies substantially. The inter-region F'_{ST} permutation analysis to test the differences between habitat pairs was performed using the R package "coin," using 10,000 permutations of the 24 among-region F'_{ST} statistics (Hothorn et al. 2008; Hothorn et al. 2006).

To explicitly examine local genetic structure, I built a partial mantel correlogram using a genetic dissimilarity matrix, a matrix of Euclidian distances between individuals, and an environment dissimilarity matrix of the 302 RADSeq individuals using the mpmcorrelogram (Matesanz et al 2011) and vegan package in R (Oksanen et al 2018). The analysis explores the autocorrelation between geographic and genetic distance which results from isolation by distance processes, as well as accounting for habitat differences (AH or PH). The habitat dissimilarity matrix was generated in vegan, with a matrix where AH-AH , AH-PH, and PH-PH pairs were coded as 0, 0.5, and 1 respectively.

To examine how genetic dissimilarity between individuals was impacted by habitat types, I ran separate least squares regression models using habitat, geography, and the interaction on distance classes of 0.5km for the first 2.5 km, and then every kilometer until the 5km, grouping all other distance classes (> 5 km) together (as they largely were not significant in the partial Mantel test).

Individual inbreeding coefficient (F_{IS}) within regions (as defined by the DAPC) was calculated in vcftools using the --het command (Danecek et al. 2011). F_{IS} values were

compared between habitats and *out6* genotypes using a t-test and an ANOVA (respectively) in JMP.

Nucleotide diversity (π) was calculated using the 17,715 loci containing SNPs that passed the filtering in the "populations" run of the five clusters. This restricted the analysis to loci that are well-represented across individuals and have high sequence coverage. From the 17,715 loci, pi was calculated per site for the different groupings (geography, habitat, *out6* genotype) using vcftools (Danecek et al. 2011).

To explore which measured traits contribute to individual-level genetic variation (F_{is} and homozygosity), I used forward and reverse model selection analysis, with the initial model including stigma-anther distance, *out6* genotype, habitat, days to first flower, and population, as well as all interactions. This model selection identified stigma-anther distance, habitat, and days to first flower as the model with the lowest Akaike information criterion (AIC) (Supplemental Table 3). These traits were then tested for independence, and no significant correlations among stigma-anther distance, habitat type, and days to first flower in this subset of individuals were identified (data not shown). A least-squares linear regression to model F_{is} was then performed including stigma-anther distance, habitat type, and days to first flower.

RESULTS

A major life-history locus (*out6*) is strongly associated with habitat and phenotype across YNP

Across the complex soil moisture and temperature matrix, habitat ephemerality (AH vs. PH) was strong predictor of *out6* genotype in wild-collected plants (Fig. 2B, $X^2_{(2, 320)} = 95.175$, $p < 0.0001$). Plants in seasonally dry habitats (AH) were > 90% TT (annual) genotypes and > 95% of NN (perennial) plants were in perennially wet sites (PH). As in previous work in three focal populations, *out6* genotype (TT vs. NN) is strongly associated with common garden rhizome production (Fig. 3A, $F_{(1,263)} = 35.46$ $p < 0.0001$). As expected, individuals homozygous for the annual *out6* haplotype (TT) produce far fewer rhizomes than the alternative homozygotes (6 vs 14), as do (unsurprisingly, given strong genotype-environment association) plants from AH habitats (Fig. 3A; $F_{(1, 308308)} = 74.279$, $p < 0.0001$).

Previous QTL mapping work, in the extreme AHQ population pair, indicated that *out6* strongly influences plant architecture and flowering time as well as life history (rhizome production). Consistent with prior findings, AH or TT plants from across YNP flower significantly closer to the ground than PH or NN plants (Fig. 3B, AH/PH, $F_{(1,322)} = 74.10$, $p < 0.0001$; TT/NN, $F_{(1,273)} = 19.16$, $p < 0.0001$). These initial height differences do not persist post-flowering in the greenhouse (AH/PH, $F_{(1,318)} = 0.9$, $p = 0.18$; TT/NN, $F_{(1,260)} = 0.5$, $p = 0.14$). However, Unlike in the AHQ F2 mapping population, wild-derived TT homozygotes flowered earlier. Under inductive (>16h days) greenhouse conditions, TT plants flowered, on average, four days earlier than NN plants (Fig. 3C, $F_{(1,274)} = 17.26$, $p < 0.0001$). Like rhizomes and height to first flower, this genotype-phenotype association parallel a significant association with parental habitat (Fig. 3C, $F_{(1,323)} = 7.35$, $p = 0.0075$), but this relationship is weaker than for the vegetative traits. AH plants flower on two days earlier, on average, than PH plants. This likely reflects high variance among AH

populations in intrinsic time to flower, and also a more complex genetic basis. Consistent with the *out6* genotypic effect in previous QTL mapping experiments, AH plants from AHQ were among the slowest to flower in this study as well. AH plants from several other locations were also slow to flower, but others were among the fastest plants, generating bimodality among AH flowering times (Supplemental Fig. 1). This suggests that the optimum time to flower may be highly variable across different AH habitats in the Park (e.g. plants using snowmelt on hot soil crust may be able to delay flowering relative to those growing in soils not hot enough to be snow-free all winter) and adaptively fine-tuned to local environments.

Floral traits associated with mating system (specifically stigma-anther distance) also showed habitat differences consistent with local adaptation. Stigma-anther distance was significantly lower in AH plants (Fig. 3D, AH/PH $F_{(1,316)} = 29.22$, $p < 0.001$), suggesting that low pollinator service in the spring-flowering AH habitats has selected for efficient self-pollination. This was paralleled by an association between *out6* genotype and stigma-anther distance (TT/NN $F_{(1,268)} = 15.56$, $p < 0.0001$). However, this relationship is likely due to linkage disequilibrium generated by the strong habitat-*out6* association, as *out6* was not associated with floral traits in prior QTL mapping studies. Plants from annual habitats have lower corolla width (Fig. 3E, AH/PH $F_{(1,317)} = 9.57$, $p < 0.002$) while TT and NN plants have similar sized corollas (Fig. 3E, TT/NN $F_{(1,269)} = 2.04$, $p = 0.15$). Lastly, corolla width and stigma-anther distance were uncorrelated ($p = 0.23$). The extreme annual population, AHQT, has both low stigma-anther distance and small corollas relative to nearby AHQN individuals (Lekberg et al. 2011), indicative of the transition to obligate self-pollination at this extremely early flowering site. Across the

Park, the (genetic) shift to lower stigma-anther separation in annual sites was recapitulated, but the shift in corolla width not. Like the variation among annual accessions in flowering time, this suggests that annual populations may vary in interactions with pollinators depending on genetically and environmentally determined flowering phenology.

Broad-scale patterns of genomic variation

Of 300 million paired-end RADSeq reads, 58 million (19%) were retained after alignment and filtering for read quality. In the filtered whole-YNP dataset (N = 302, mean coverage 18X, minimum coverage 5X), we scored 49,539 single-nucleotide polymorphism (SNPs) in 22,441 loci (mean length = 789bp ± 1bp). The SNPs covered all 14 scaffolds (chromosomes) on the *M. guttatus* V2 reference genome. The retained loci are highly genic, with 15,813 SNPs within annotated exons, and 23,228 SNPs within annotated genes. Only three RAD loci were in the <150kb *out6* region.

In an initial principal component (PC) analysis, plants from our focal (and ecologically extreme) AHQT population were unique and highly differentiated from all other individuals in YNP (Fig. 4), as they were in previous microsatellite and PoolSeq analyses (Lekberg et al. 2012; Hendrick et al. 2016). Analyses using the find.cluster module in adegenet indicated that clustering all YNP individuals into 5-10 populations (K = 5-10) minimizing the variance within clusters while maximizing separation among them (Fig. 5, inset). Visualization of the resulting discriminant analysis of principal components (DAPC; Fig. 5) shows that most conservative clustering (K = 5) captures the distinctness of AHQT (though it is closest to geographically adjacent clusters). At this

level, the other four clusters correspond to broad geographic regions (Central, Northern, Southern, and Southeastern (see Fig. 1), suggesting that (beyond AHQT) geography is the primary factor structuring *M. guttatus* genomic variation in YNP. DAPC plots using larger numbers of clusters split individuals further into local geographic regions, but never separate AH or PH groups within regions.

Patterns of nucleotide diversity (π) in and between the five DAPC-defined populations (Central, Northern, Southern, and Southeastern, plus AHQT) support these region assignments. As in previous studies, F'_{ST} between AHQT and each other region is elevated ($F'_{ST} = \sim 0.2$) and the geographic regions are moderately differentiated from each other ($F'_{ST} = \sim 0.1$) (Table 1). Genomic diversity in the four geographic regions are similar in the percent of variable sites (~0.22), homozygosity (~0.77), and π (~0.01) (Table 2). Conversely, AHQT has notably reduced genomic variation, with 20% fewer variant sites (25,792), ~10% increase in homozygosity (0.86 ± 0.0007), and ~45% decrease in π compared to the other four geographic regions (Table 2). This data, along with the DAPC and F'_{ST}, suggests that geography is broadly structuring population variation, but that AHQT (an ecologically extreme population), is experiencing large changes in genomic diversity.

Within-region genomic variation

With the exception of AHQT, geography (isolation by distance) appears to be the primary factor shaping neutral variation on a broad scale, with habitat-specific selection

structuring key phenotypes and *out6* in the face of ongoing gene flow. However, AH-PH divergence in morphological mating system (stigma-anther distance) and realized phenology (days to flower, but also other factors) could alter gene flow and genomic variation on a more local scale. A second PC analysis (with AHQT excluded) shows that AH and PH plants from Northern, Southern, and Southeastern regions show relatively little genomic differentiation (Fig. 6). Within all four geographic regions, AH and PH sub-populations had similar levels of polymorphism and uniformly low level of differentiation ($F'_{ST} = 0.03$; Table 3). However, AH and PH plants from the large and highly thermally active Central region strongly separate along PC1 (Fig. 6), suggesting that some substantial number of loci may sort along this axis. Within this region, we more closely examined a single geyser basin (Upper Geyser Basin- UGB) with many AH and PH sites in close proximity (Fig. 7). Even this site also shows no strong differentiation by habitat or *out6*-based life-history (Table 3), although geographically-dispersed AH individuals do cluster loosely on the first two PC axes.

The partial Mantel correlogram, which considers both spatial and habitat differences as potential causes for genetic dissimilarity, identified a strong pattern of localized spatial genetic structure. Within 5 km, spatial autocorrelation was positive and significant between individuals on both AH and PH habitats, and was strongest from 0km-0.5km ($r = 0.28$, $p < 0.001$, Fig. 8). The strong signature of spatial correlation drops considerably off after 4.5 km, and nears zero after 10km, where individuals are no more related to each other than expected by chance.

Further exploration of the influence of habitat on genetic dissimilarity between individuals showed that within the first 4km (0.5km clusters), habitat had a significant

effect on genetic dissimilarity ($p > 0.0005$). Individuals from annual habitats (AH-AH) were more dissimilar than AH-PH or PH-PH comparisons in the first kilometer, while PH-PH were more genetically similar than the AH-AH or AH-PH comparisons for the over larger distances (Fig. 9), suggesting that habitats and associated morphologies are impacting relatedness on a microgeographic scale.

Within-individual genetic variation

Although AH and PH plants within a region do not form distinct populations, habitat and individual mating system may differentially structure variation within and among individuals within regions (Fig. 10). Evolutionary transitions to obligate selfing are often associated with increased individual homozygosity and elevated differentiation (as at AHQT). Although AH and PH sub-populations (within regions) contain similar levels of diversity in aggregate (Table 3), inbreeding coefficient (F_{IS}) of AH plants is (on average) 27% higher than PH plants (0.32 vs. 0.25; N = 103 and 114, respectively (PH); $p < 0.007$) (Fig. 11). Similarly, TT individuals had significantly higher F_{IS} than both TN ($p < 0.001$) and NN individuals ($p < 0.03$), which were not significantly different from each other ($p = 0.32$). Fixation of region-specific variation within highly selfing lineages may further result in elevated differentiation among AH sites from different regions. To test for this possibility, I compared across-region AH-AH F'_{ST} values to AH-PH and PH-PH values and see slightly elevated differentiation among annual population pairs ($p < 0.03$) (Fig. 12). Again, this confirms that parallel shifts in morphology associated with AH habitats reflects ongoing local adaptation in the face of some gene flow with populations in alternative habitats, rather shared demographic history.

To further explore which traits influence individual inbreeding history, I used a model selection approach. Forward and reverse model selection identified parental habitat and progeny stigma-anther distance and days to flower as significant factors in predicting parental F_{IS} (AIC of -26.513 vs. the AIC of the all-inclusive model of -11.734). Annual habitat ($p < 0.008$), lower stigma-anther distance ($p < 0.004$), and later flowering ($p < 0.04$) were associated with increased F_{IS} ($r^2 = 0.23$; Fig. 13, Supplemental Table 3). Genotype at *out6* was not a significant predictor of inbreeding; this is consistent with *out6* not being causal locus (or QTL) for stigma-anther distance but merely associated with the genotype in wild individuals via strong parallel selection on both annuality and mating system in AH habitats.

DISCUSSION

This work extends the investigation of adaptation of *Mimulus guttatus* to the novel Yellowstone geothermal soil mosaic to genome-wide and landscape scales, demonstrating life-history and mating system divergence in the face of gene flow, as well as consequences of adaptation for genomic differentiation. Across the Park, genetic variation was primarily structured into geographical regions (rather than ecotypes), with the exception of the extreme thermal AHQT population, which is isolated from all others. Despite this evidence for ongoing gene flow, plants from ephemerally-wet thermal annual habitats (AH) and those in perennially wet habits (PH) were strongly (and adaptively) differentiated for rhizome number, height to first flower, and stigma-anther distance. The first two of these traits are under the strong genetic control of the *out6* locus (Nelson et al., in prep), which is highly sorted by habitat in this study. Local population structure is

dependent on geographic proximity, not habitat, with AH and PH sub-populations in the four geographical regions equally diverse and not differentiated. However, AH individuals had elevated F_{IS} (inbreeding coefficient) values consistent with an adaptive shift to increased autoagamous self-fertilization in the spring-flowering thermal annuals. Across all geographic regions, reduced stigma-anther distance, annual habitat, and slower flowering were significantly predictors of increased F_{IS}. These results provide a window into the early stages of the evolution of selfing, one of the most common transitions in flowering plants, as well as into the possibility (and impossibility) of speciation in the face of gene flow.

Across the vast majority of YNP's microgeographic soil mosaics, strong divergent selection on life history and mating system was not enough to generate genomic differentiation at the population level. Although individuals in thermal habitats are becoming reproductively isolated from each other and increasingly homozygous, these shifts are modest compared to the structure created by geographic distance. AHQT, on the other hand, is the exception that proves the rule. Although plants from this population shares a common thermal annual phenotype with other AH plants (few/no rhizomes, low height at first flower, low stigma-anther distance), they are highly differentiated genome-wide from all other populations ($F'_{ST} \sim 0.2$), and showed reduced π, heterozygosity, and number of variant sites. This, in part, reflects the extremeness of the site, which is at high elevation, highly snowmelt-dependent, and particularly early flowering (Lekberg et al. 2012). However, the most likely trigger for this differentiation is a 200-meter gap of no *M. guttatus* habitat between AHQT and the nearest other perennial population. This gap acts as a geographical barrier to gene flow, allowing both

neutral differentiation by and unimpeded specialization on thermal soils. AHQT illustrates that adaptation across a microgeographic habitat boundary can generate substantial reproductive isolation, but only if both extrinsic and intrinsic (phenology, mating system) barrier combine to effectively shut down gene flow.

Park-wide trait variation and covariation reveals adaptation to complex geothermal habitats

A major adaptive locus (*out6)* underlies the life-history transition from perenniality to annuality in YNP *M. guttatus* and shows signatures of a selective sweep consistent with antagonistic selection for alternative T and N haplotype in extreme habitats (Nelson et al. in prep). In this study, *out6* sorts by habitat with remarkable fidelity across the entire landscape (Fig. 2B). Furthermore, the effect of *out6* on rhizome production and growth form defined in a single cross were confirmed (in large part) across the entire phenotypically, environmentally and genetically complex habitat mosaic. Although there are a few individuals with mismatched phenotypes and genotypes, suggesting the possibility of additional loci affecting life history, this association is strong evidence that *out6* is a major developmental switch with strongly antagonistically pleiotropic effects.

Consistent with the evolution of selfing as a mechanism of reproductive assurance under conditions of mate/pollinator limitation (Dole 1992; Busch & Delph 2011; Lloyd 1992), stigma-anther distance was significantly lower in plants from annual habitats (Fig. 3D). Increased self-fertilization ability through decreasing distance between the anthers and stigma is a common morphological change, and has been experimentally shown to

rapidly evolve in pollen limited environments (Bodbyl Roels & Kelly 2011). Transitions from outcrossing to selfing are also frequently associated with movement onto harsh and ephemeral habitats, where reproductive assurance might be advantageous in the unpredictable environments (Barrett 2002; Dole 1992). In YNP, the thermal soils enforce a strong deadline of desiccation in the summer, but allow for flowering earlier in the season, with many plants flowering in early April-May, well before pollinators are present (personal observation), suggesting that pollen limitation early in the season selects for increased selfing efficiency. However, this might not be the case for all AH individuals, as a fair number of AH plants were collected flowering alongside perennial plants. In these cases, fast development and seed set may be beneficial, if the habitats become uninhabitable unpredictably (Barrett 2002; Munoz et al. 2016). Further research into how selfing-rates change across seasons and locations can offer insights into realized mating systems, and how selection on floral morphology may vary in time and space.

Although previous QTL mapping (Nelson et al, in prep) found the T *out6* alleles (from AHQT) were very strongly associated with late flowering, we found more complex variation in flowering phenology under common garden conditions. AH plants flowered significantly earlier (~3 days), on average, than PH plants but many AH accessions were slow to flower (Fig. 3C, Supplemental Fig. 1). Flowering phenology is a highly integrated trait combining daylength (critical photoperiod), temperature, and internal resources (Simpson & Dean 2002; Rubin et al. 2018; Friedman et al. 2014; Mitchell-Olds 2001). Because correct timing of flowering is critical to reproductive success, this trait is regularly identified as being locally adapted, with flowering cues frequently changing predictably over altitude and altitude (Ågren et al. 2016; Friedman et al. 2014; Hall &

Willis 2006). The geothermal soils vary in temperature and moisture patterns over meters, and we therefore expect the plants to both plastically and genetically vary to suit their local thermal regime.

The slow-flowering AH plants break a longstanding life-history correlation, where investment in rhizomes (perenniality) is associated with slow growth and late flowering while annuals grow and flower rapidly (Twyford & Friedman 2015; Munoz et al. 2016; Barrett & Schluter 2008; Dole 1992). Rather than living fast and dying young, some thermal annuals live slow and die young (but old for an annual); this suggests that the general life-history correlation is not an inherent trade-off, but the product of simultaneous selection for speed and annuality. While creating a hard deadline in the summer, geothermal soils can extend the beginning of the growing season by months, allowing for some annual plants to experience a much longer growing season (Dec-April, 5 months) then some of the less geothermally influenced annual habitats (May-June, 2 months). Finally, speed is only one component of flowering phenology, as flowering in the field also involves response of temperature and photoperiod cues. Previous work examining variation in critical photoperiod found that extreme AHQT plants had lost the perennial long-day requirement for flowering, allowing the plants to flower in early spring (Lekberg et al. 2012). To fully understand how changes in flowering phenology impacts patterns of gene flow, further work is needed in the field to examine realized flowering phenology (when and how it changes across habitats), and common gardens to identify the genetic basis and extent of critical photoperiod variation.

Previous work in AHQ and Rabbit Creek found that corolla width was significantly reduced in AH individuals (Lekberg et al. 2011). This is statistically recapitulated in our

park-wide comparisons of plants from different habitats (Fig. 3E), but there is wide overlap in flower size between annual and perennial plants (Fig. 3E). For example, the high elevation Southern region, where even AH plants can flower into July, had some of the biggest flowers in both annuals and perennials (data not shown). Reductions in flower size associated with the transition to selfing (e.g. "selfing syndrome") may arise through two non-exclusive mechanisms (Sicard & Lenhard 2011): adaptive changes in energy investment (Charlesworth & Charlesworth 1981), and relaxed selection and fixation of recessive "small flower" alleles by drift (Goodwillie et al. 2006; Fishman et al. 2002). Because exclusive selfers don't need to attract pollinators, reallocation of resources to primary reproductive traits (e.g. ovule number) may be favored instead. For the annual plants, populations might not be isolated long enough for drift to become apparent; selection may favor large flowers as to allow periodic outcrossing events to avoid inbreeding depression, and/or steady gene flow from perennials plants in late flowering sites may reintroduce alleles for large flowers back into some annual lineages.

Geography, rather than habitat, primarily structures neutral genetic variation

Changes in flowering phenology and mating system alter gene flow and can generate reproductive barriers and bias mating (Hamrick & Godt 1996; Hendry & Day 2005; St Onge et al. 2011; Andrew et al. 2012; Ostevik et al. 2016). The DAPC analysis (which maximizes between population variation, while minimizing within population variation) identified five population clusters, four of which were comprised of geographical close individuals (Northern, Central, Southern, Southeastern), while the fifth was AHQT, the extreme thermal annual population (Fig. 5). When examining just

the PC analyses, individuals remain grouped by region, but exhibit continuous variation between, suggesting gene flow along the edges (Fig. 6). Further investigation into within-region AH-PH (or TT-NN) comparisons of genetic variation and F'_{ST} found no genome-wide patterns of differentiation (Table 3). However, habitat is not without influence. Our F'_{ST} comparisons between regions' AH and PH plants suggest that while annual habitats are not differentiated from their perennial neighbors, the AH-AH show signs of increased differentiation then PH-PH (Fig. 12). While the difference is only marginally significant, it does suggest that there are differences in AH plants between regions, this pattern can be seen in the PC analyses where the annual, highly selfing individuals are more isolated than the perennial counterparts (Fig. 7, 9).

The partial Mantel correlogram further emphasizes that plants largely do not sort by habitat. Instead, related is strongly related to geographic distance (even among plants from different habitats) over short (<4 km) spatial scales (Fig. 8). This pattern is likely driven by limited dispersal of seeds and pollen (via pollinators) across the variable Yellowstone landscape. Additionally, the larger distances in the correlogram are likely underestimations of the true distances between individuals as distance was calculated using Euclidian methods and did not incorporate landscape resistance (such as dense forest or mountains) (Andrew et al. 2012) and how pollinators or seed dispersal events move across the topography. Contemporary work examining the role of bee pollinators have shown that pollen movement through single bees is often highly localized, not moving pollen more then 300m (Ilson et al. 2019) and is likely causing the genomic homogenization between the annual and perennial habitats. However, at microscale distances (<4km), pairs of individuals from annual habitats were dissimilar relative to

annual-perennial or perennial-perennial pairs, consistent with their elevated homozygosity due to selfing (Fig 9).

One contributor to low genome-wide differentiation in plant ecotypes, such as YNP *Mimulus,* could be plastic responses to inter-annual environmental variation that weakens barriers to gene flow. Flowering phenology is often plastic, and local temperatures and moisture can alter flowering time between individuals. Plants on and off serpentine soils mosaics, such as such in *Solidago virgaurea,* frequently display strong flowering asynchrony between ecotypes (Brady et al. 2005). However, despite the shifts in flowering time, no differentiation was observed in *Solidago* (Sakaguchi et al. 2018). Similarly, sand dune and non-sand dune ecotypes of *Helianthus petiolaris* use different pollinator communities and there is strong selection against immigrants and hybrids, but show only mild neutral genomic differentiation (Andrew et al. 2012; Ostevik et al. 2016). Thus, although mosaic environments such as serpentine soils are known for their diversity of endemic species, geography may be necessary to jumpstart speciation rather than maintain adaptive ecotypes with abundant gene flow (Safford et al. 2005; Brady et al. 2005).

The consequences of selfing: despite individual loss of heterozygosity, annual populations maintain genetic diversity

The transition to selfing has been shown to dramatically impact gene flow and genomic diversity (Charlesworth & Wright 2001; Charlesworth & Willis 2009; Charlesworth 2009). Unlike changes to mating time which partitions variation between

groups, obligately selfing plants, instead, make homozygous a subset of population variation. Each generation of selfing changes half of all heterozygous nucleotides to be homozygous, independently of all other selfing individuals. This, theoretically, allows for a large number of selfing plants to maintain intermediate diversity at the population level while having reduced heterozygosity at the individual level. As selfing evolves, however, species frequently experiences massive population bottlenecks and reduced effective populations size (St Onge et al. 2011; Laenen et al. 2018). One example is the selfer *Mimulus nasutus*, which exhibits low diversity that is a subset of the variation seen in its outcrossing progenitor *M. guttatus* (Brandvain et al. 2014). In AHQT, we see a similar pattern: the population has highly reduced π and total number of variant sites, as well as increased F'_{ST} with other populations (Table 1 & 2). This suggests that a strong bottleneck event in the founding of this population with infrequent gene flow events from the neighboring perennial populations. Additionally, the DAPC analysis groups AHQT with neighboring southern and northern regions, which suggests that AHQT has sampled a subset of the local geographic variation (Fig. 5).

In the other four regions, where there is no geographical isolation between AH and PH habitats, AH plants also show increased F_{IS}, however, this does not result in elevated differentiation of an aggregate AH population from nearby PH plants (F'_{ST} low within regions) (Fig. 11, Table 3). This suggests that annual habitats (and the morphological adaptation to them) increase selfing frequency across YNP. Taking a closer look at what factors are underling this relationship, our models show that reduced stigma-anther distance, annual habitats and slower flowering are significant factors in increases of inbreeding coefficients and homozygosity among individuals (Fig. 13). This

is consistent with the very early stages of the evolution of selfing, as inbreeding has direct consequences for F_{IS}. (Wright et al. 2008; Laenen et al. 2018; Charlesworth & Wright 2001; Lucek et al. 2019). Along with the lack of differentiation, this suggests that plants in highly thermal soils routinely self-pollinate and are individually adapted to selfing in their floral morphology and phenology, but that pollen exchange with nearby nonthermal populations periodically reintroduces neutral variation, such that no population divergence builds up. This is likely a product of the extremely microgeographic nature of the soil mosaic, which does not create any extrinsic barriers to mating. In addition, the variable soil moisture regime in most geyser basins may sometimes facilitate summer flowering by AH plants. In AH plants, our data suggests that strong habitat-specific selection, even accompanied by a transition to self-fertilization and shifts in phenology, is not enough to generate the evolution of reproductive isolation. Instead, we see adaptation with gene flow, except where micro-allopatry (as at AHQT) enforces isolation and allows consolidation of reproductive barriers by divergent selection as well as drift.

REFERENCES CITED

Ali, O. A., S. O'Rourke, S. J. Amish, M. H. Meek, G. Luikart, C. Jeffers, & M. R. Miller. 2016. RAD Capture (Rapture): Flexible and efficient sequence-based genotyping. *Genetics Society of America,* 202, pp.389–400

Anderson, J.T., Willis, J.H. & Mitchell-Olds, T. 2011. Evolutionary genetics of plant adaptation. *Trends in Genetics*, 27(7), pp.258–266.

Andrew, R.L., K. Ostevik, D.P. Ebert & L. Reiseberg. 2012. Adaptation with gene flow across the landscape in a dune sunflower. *Molecular Ecology*, 21(9), pp.2078–2091.

Antonovics, J. 1968. Evolution in closely adjacent plant populations V. Evolution of self-fertility. *Heredity*, pp.371–384.

Ågren, J., C.G. Oakley, S. Lundemo, & D.W. Schemske 2016. Adaptive divergence in flowering time among natural populations of Arabidopsis thaliana: Estimates of selection and QTL mapping. *Evolution*, 71(3), pp.550–564.

Babik, W., R.K. Roger, W. J. Baker, A.S.T. Papadopulos, M. Boulesteix, M. Ansett, C. Lexer, I. Hutton, & V. Savolainen. 2009. How sympatric is speciation in the Howea palms of Lord Howe Island? *Molecular Ecology*, 18(17), pp.3629–3638.

Baker, H.G. 1955. Self-compatibility and establishment after "long-distance" dispersal. *Evolution*, 9(3), pp.347–349.

Barrett, R., & Schluter C.D. 2008. Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, 23(1), pp.38–44.

Barrett, S.C.H. 2002. The evolution of plant sexual diversity. *Nature Reviews Genetics*, 3(4), pp.274–284.

Barrett, S.C.H. & Harder, L. 1996. The comparative biology of pollination and mating in flowering plants. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 351(1345), pp.1271–1280.

Beardsley, P.M., S.E. Schoenig, J.B. Whittall, & R.G. Olmstead et al., 2004. Patterns of evolution in western North American Mimulus (*Phrymaceae*). *American Journal of Botany*, 91(3), pp.474–489.

Bodbyl Roels, S.A. & Kelly, J.K., 2011. Rapid evolution caused by pollinator loss in *Mimulus guttatus*. *Evolution*, 65(9), pp.2541–2552.

Bolger, A.M., Lohse, M. & Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, pp.2114–2120.

Brady, K.U., A.R. Kruckeberg, & H.D., Jr. Bradshaw. 2005. Evolutionary Ecology of Plant Adaptation to Serpentine Soils. *Annual Review of Ecology, Evolution, and Systematics*, 36(1), pp.243–266.

Brandvain, Y., A.M. Kennedy, L. Flagel, G. Coop, & A.L. Sweigart. 2014. Speciation and Introgression between *Mimulus nasutus* and *Mimulus guttatus* C. D. *PLOS Genetics*, 10(6), pp.e1004410–15.

Breigenzer, P. 2018. Phenotypic and Genetic Analyses of Adaptation to Geothermal Soils in Yellow Monkeyflowers of Yellowstone National Park. Undergraduate Thesis. University of Montana. pp.1–17.

Busch, J.W. & L.F. Delph. 2011. The relative importance of reproductive assurance and automatic selection as hypotheses for the evolution of self-fertilization. *Annals of Botany*, 109(3), pp.553–562.

Busch, J.W., S. Joly, & D.J. Schoen. 2011. Demographic Signatures Accompanying the Evolution of Selfing in *Leavenworthia alabamica*. *Molecular Biology and Evolution*, 28(5), pp.1717–1729.

Caisse, M. & J. Antonovics, 1987. Evolution in closely adjacent plant populations. *Heredity*, 40, pp.1365–2540.

Calsbeek, R. & Smith, T.B. 2003. Ocean currents mediate evolution in island lizards. *Nature*, 426(6966), pp.552–555.

Catchen, J., P.H. Hohenlohe, S.B. Bassham, S. Amores, & W.A. Cresko. 2013. Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22(11), pp.3124–3140.

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

Charlesworth, B. 2009. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10(3), pp.195–205.

Charlesworth, D. & B. Charlesworth. 1981 Allocation of resources to male and female functions in hermaphrodites. *Biological Journal of the Linnean Society*, (15), pp.57–74.

Charlesworth, D. & Willis, J.H., 2009. Fundamental Concepts in Genetics: The genetics of inbreeding depression. *Nature Reviews Genetics*, 10(11), pp.783–796.

Charlesworth, D. & S.I. Wright. 2001. Breeding systems and genome evolution. *Current Opinion in Genetics & Development*, 11(6), pp.685–690.

Cheptou, P.O. 2011. Clarifying Baker's Law. *Annals of Botany*, 109(3), pp.633–641.

Coughlan, J.M. & J.H. Willis, 2018. Dissecting the role of a large chromosomal inversion in life history divergence throughout the *Mimulus guttatus* species complex. *Molecular Ecology*, 155(4), pp.419–15.

Coyne, J.A. & Orr, A.H. 2004. *Speciation*, Oxford University Press.

Danecek, P. A. Auton2, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, G. Lunter, G.T. Marth, S. T. Sherry, G. McVean, R.Durbin, & 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics*, 27(15), pp.2156–2158.

Dole, J.A., 1992. Reproductive assurance mechanisms in three taxa of the *Mimulus guttatus* complex (*Scrophulariaceae*). *American Journal of Botany*, 79(6), pp.650–659.

Doyle, J.J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, 19, pp.11–15.

Ferris, K.G. T. Rushton, A.B. Greenlee, K. Toll, B.K. Blackman, & J.H. Willis. 2015. Leaf shape evolution has a similar genetic architecture in three edaphic specialists within the *Mimulus guttatus* species complex. *Annals of Botany*, 116(2), pp.213–223.

Fisher, R.A. 1941. Average excess and average effect of a gene substitution. *Annals of Eugenics*, 11, pp.53–63.

Fishman, L. & J.H. Willis 2008. Pollen limitation and natural selection on floral characters in the yellow monkeyflower, *Mimulus guttatus*. *New Phytologist*, 177(3), pp.802–810.

Fishman, L., A.J. Kelly & J.H. Willis. 2002. Minor quantitative trait loci underlie floral traits associated with mating system divergence in *mimulus*. *Evolution*, 56(11), pp.2138–2155.

Foxe, J.P., T. Slotte, E.A. Stahl, B. Neuffer, H. Hurka, & S.I. Wright. 2009. Recent speciation associated with the evolution of selfing in *Capsella*. *Proceedings of the National Academy of Sciences*, 106(13), pp.5241–5245.

Frachon, L., C. Bartoli, S. Carrere, O. Bouchez, A. Chaubet, M. Gautier, D. Roby, and F. Roux. 2019. A Genomic map of climate adaptation in Arabidopsis thaliana at a micro-geographic scale. *Frontiers in Plant Sciences* 9, pp3-15

Friedman, J., A.D. Twyford, & J.H. Willis. 2014. The extent and genetic basis of phenotypic divergence in life history traits in Mimulus guttatus. *Molecular Ecology*, 24(1), pp.111–122.

Gardner, M. & M. MacNair. 2000. Factors affecting the co-existence of the serpentine endemic *Mimulus nudatus* Curran and its presumed progenitor, *Mimulus guttatus* Fischer ex DC. *Biological Journal of the Linnean Society*, 69(4), pp.443–459.

Gavrilets, S. 2003. Perspective: Models of speciation: What have we learned in 40 years? *Evolution*, 57(10), pp.2197–2215.

Glemin, S., E. Bazin, & D. Charlesworth. 2006. Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Proceedings of the Royal Society B: Biological Sciences*, 273(1604), pp.3011–3019.

Goodwillie, C., Ritland, C. & Ritland, K., 2006. The genetic basis of floral traits associated with mating system evolution in *Leptosiphon (Polemoniaceae*): an analysis of quantitative trait loci. *Evolution*, 60(3), p.491.

Grossenbacher, D., R.B. Runquist, E.E. Goldberg, & Y. Brandvain. 2015. Geographic range size is predicted by plant mating system. *Ecology Letters*, 18(7), pp.706–713.

Hall, M.C. & J.H. Willis. 2006. Divergent selection on flowering time contributes to local adaptation in mimulus guttatus populations. *Evolution*, 60(12), pp.2466–2477.

Hamrick, J.L. & M.J.W. Godt. 1996. Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 351(1345), pp.1291–1298.

Hedrick, P. 1986. Genetic polymorphism in heterogeneous environments: a decade later. *Annual Reviews of Ecological Systems*, pp.535–566.

Hellsten, U., K.M. Wright, & J. Jenkins. 2013. Fine-scale variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing. *Proceedings of the National Academy of Sciences*, 110(48), pp.19478–19482.

Hendrick, M.F., F.R. Finseth, M.E. Mathiasson, K.A. Palmer, E.M. Broder, P. Breigenzer, & L. Fishman. 2016. The genetics of extreme microgeographic adaptation: an integrated approach identifies a major gene underlying leaf trichome divergence in Yellowstone Mimulus guttatus. *Molecular Ecology*, 25(22), pp.5647–5662.

Hendry, A.P. & T. Day. 2005. Population structure attributable to reproductive time: isolation by time and adaptation by time. *Molecular Ecology*, 14(4), pp.901–916.

Hoekstra, H.E., J.G. Krenz, & N.W. Nachman. 2004. Local adaptation in the rock pocket mouse (*Chaetodipus intermedius*): natural selection and phylogenetic history of populations. *Heredity*, 94(2), pp.217–228.

Hothorn, T., K. Hornik, M.A. van de Wiel, & A. Zeileis. 2006. A Lego System for Conditional Inference. *The American Statistician*, 60(3), pp.257–263.

Hothorn, T., K. Hornik, M.A. van de Wiel, & A. Zeileis. 2008. Implementing a class of permutation pests: the coin package. *Journal of Statistical Software*, (28).

Imai, R., Y.Tsuda, S. Matsumoto, A Ebihara, & Y. Watano. 2016. The Relationship between Mating System and Genetic Diversity in Diploid Sexual Populations of *Cyrtomium falcatum* in Japan. *PLoS One*. 11(10) e0163183

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), pp.1403–1405.

Jombart, T., Devillard, S. & Balloux, F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, 11(1), p.94.

Jones, M.R. L.S. Mills, P.C. Alves, C.M. Callahan, J.M. Alves, D.J.R. Lafferty, F.M. Jiggins, J.D. Jensen, J. Melo-Ferreira, & J.M. Good. 2018. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science*, 360(6395), pp.1355–1358.

Kamran-Disfani, A. & A.F. Agrawal. 2014. Selfing, adaptation and background selection in finite populations. *Journal of Evolutionary Biology*, 27(7), pp.1360–1371.

Kenney, A.M. & A.L. Sweigart. 2016. Reproductive isolation and introgression between sympatric *Mimulus* species. *Molecular Ecology*, 25(11), pp.2499–2517.

Kiang, Y.T. & J.L. Hamrick. 1978. Reproductive Isolation in the *Mimulus guttatus M. nasutus* Complex. *American Midland Naturalist*, 100(2), p.269-276

Koornneef, M. 1998. Genetic control of flowering time in Arabidopsis. *Annual Reviews of Ecological Systems*, 49(1), pp.345–370.

Kooyers, N.J., J.M. Colicchio, A.B. Greenlee, E. Patterson, N.T. Handloser, & B.K. Blackman. 2019. Lagging Adaptation to Climate Supersedes Local Adaptation to Herbivory in an Annual Monkeyflower. *The American Naturalist*, 194(2), pp.000-000.

Laenen, B. et al. 2018. Demography and mating system shape the genome-wide impact of purifying selection in *Arabis alpina*. *Proceedings of the National Academy of Sciences of the United States of America*, 115(4), pp.816–821.

Lekberg, Y., B. Roskilly, M.F. Hendrick, C.A. Zabinski, C.M. Barr, & L. Fishman. 2012. Phenotypic and genetic differentiation among yellow monkeyflower populations from thermal and non-thermal soils in Yellowstone National Park. *Oecologia*, 170(1), pp.111–122.

Lenormand, T. 2002. Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, 17(4), pp.183–189.

Li, H. & R. Durbin. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, 26(5), pp.589–595.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, & 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), pp.2078–2079.

Lloyd, D.G. 1992. Self-and cross-fertilization in plants. II. The selection of self-fertilization. *The American Naturalist*, 153(3, Part 1), pp.370–380.

Lowry, D.B. & J.H. Willis 2010. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLOS Biology*, 8(9), pp.e1000500.

Lucek, K., N. Hohmann, & Y. Willi. 2019. Postglacial ecotype formation under outcrossing and self-fertilization in *Arabidopis lyrata*. *Molecular Ecology*, 28(5), pp.1043-1055.

Matesanz S., T.E. Gimeno, M. de la Cruz, A. Escudero and F. Valladares. 2011. Competition may explain the fine-scale spatial patterns and genetic structure of two co-occurring plant congeners. Journal of Ecology 99: 838-848.

Meirmans, P.G. 2006. Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution*, 60(11), pp.2399–2402.

Meirmans, P.G. & P.W. Hedrick. 2010. Assessing population structure: FST and related measures. *Molecular Ecology Resources*, 11(1), pp.5–18.

Mitchell-Olds, T. 2001. Arabidopsis thaliana and its wild relatives: a model system for ecology and evolution. *Trends in Ecology & Evolution*, 16(12), pp.693–700.

Munoz, F., C. Violle, & P.O Cheptou. 2016. CSR ecological strategies and plant mating systems: outcrossing increases with competitiveness but stress-tolerance is related to mixed mating. *Oikos*, 125(9), pp.1296–1303.

Ness, R.N., S.I. Wright & S.C.H. Barrett. 2010. Mating-System Variation, Demographic History and Patterns of Nucleotide Diversity in the Tristylous Plant *Eichhornia paniculata*. *Genetics,* 184, pp.381–392 I

Oksanen, J., F. Guillaume Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P.R. Minchin, R.B. O'Hara, G.L. Simpson, P. Solymos, M. Henry, H. Stevens, E. Szoecs and Helene Wagner (2018). vegan: Community Ecology Package. R package version 2.5-2. https://CRAN.R-project.org/package=vegan

Oneal, E., D. Lowry, K.M. Wright, Z. Zhu, & J.H. Willis. 2014. Divergent population structure and climate associations of a chromosomal inversion polymorphism across the *Mimulus guttat*us species complex. *Molecular Ecology*, 23(11), pp.2844–2860.

Ostevik, K.L., R.L. Andrew, S.P. Otto, & L. Rieseberg. 2016. Multiple reproductive barriers separate recently diverged sunflower ecotypes. *Evolution*, 70(10), pp.2322–2335.

Pannell, J.R., J.R. Auld, Y. Brandvain, M. Burd, J.W. Busch, P.O. Cheptou, J.K. Conner, E.E. Goldberg, A.G. Grant, D.L. Grossenbacher, S.M. Hovick, B. Igic, S. Kalisz, T.H. Petanidou, A.M. Randle, R. Rubio de Casas, A. Pauw, J.C. Vamosi & A.A. Winn 2015. The scope of Baker's law. *New Phytologist*, 208(3), pp.656–667.

Purcell, S. K., K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller, P. Sklar, P.I.W. de Bakker, M.J. Daly & P.C. Sham. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, 81(3), pp.559–575.

Rausher, M.D. 2013. Evolution of the selfing syndrome in Ipomoea. *Frontiers in Plant Science*, pp.1–8.

Richardson, J.L., M.C. Urban, D.I. Bolnick, & D.K. Skelly. 2014. Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology & Evolution*, 29(3), pp.165–176.

Rubin, M.J., K.M. Schmid, & J. Friedman. 2018. Assortative mating by flowering time and its effect on correlated traits in variable environments. *Ecology and Evolution*, 5(1), pp.1–11.

Safford, H.D., J.H. Viers, & S.P. Harrison. 2005. Serpentine endemism in the California flora: a database of serpentine affinity. *Madroño*, 52(4), pp.222–257.

Sakaguchi, S., K. Horie, N. Ishikawa, S. Nishio, J.P.P Worth, K. Fukushima, M. Yamasaki & M. Ito. 2018. Maintenance of soil ecotypes of *Solidago virgaurea* in close parapatry via divergent flowering time and selection against immigrants S. Bonser, ed. *Journal of Ecology*, 107(1), pp.418–435.

Savolainen, V., M. Anstett, C. Lexer, I. Hutton, J.J. Clarkson, M.V. Norup, M.P Powell, D. Springate, N. Salmin, & W.J. Baker. 2006. Sympatric speciation in palms on an oceanic island. *Nature*, 441, pp.210–213.

Schemske, D.W. & H.D. Bradshaw 1999. Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). *Proceedings of the National Academy of Sciences*, 96(21), pp.11910–11915.

Schoen, D.J. & A.H. Brown. 1991. Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proceedings of the National Academy of Sciences*, 88(10), pp.4494–4497.

Selby, J.P. & J.H. Willis. 2018. Major QTL controls adaptation to serpentine soils in Mimulus guttatus. *Molecular Ecology*, 27(24), pp.1–40.

Sicard, A. & M. Lenhard 2011. The selfing syndrome: a model for studying the genetic and evolutionary basis of morphological adaptation in plants. *Annals of Botany*, 107(9), pp.1433–1443.

Simpson, G.G. & C.D. Dean. 2002. Arabidopsis, the Rosetta stone of flowering time? *Science*, 292, pp.285–296.

Slatkin, M. 1973. Gene flow and selection in a cline. *Genetics*, 75(4), pp.733–756.

St Onge, K.R. T. Kallman, T. Slotte, M. Lascoux, & A.E. Palme. 2011. Contrasting demographic history and population structure in *Capsella rubella* and *Capsella grandiflora,* two closely related species with different mating systems. *Molecular Ecology*, 20(16), pp.3306–3320.

Twyford, A.D. & J. Friedman. 2015. Adaptive divergence in the monkey flower Mimulus guttatus is maintained by a chromosomal inversion. *Evolution*, 69(6), pp.1476–1486.

Wang, I.J. & G.S. Bradburd. 2014. Isolation by environment. *Molecular Ecology*, 23(23), pp.5649–5662.

Wright, J.W., M.L. Stanton & R. Scherson. 2006. Local adaptation to serpentine and non-serpentine soils in *Collinsia sparsiflora*. *Evolutionary Ecology Research*, 8(1), pp.1–21.

Wright, S. 1943. Isolation by Distance. *Genetics*, 28(2), pp.114–138.

Wright, S.I., R.W. Ness, J.P. Foxe, S.C.H Barrett. 2008. Genomic Consequences of Outcrossing and Selfing in Plants. *International Journal of Plant Sciences*, 169(1), pp.105–118.

Wu, C.A., D.B. Lowry, A.M. Cooley, K.M. Wright, Y.W. Lee, & J.H. Willis. 2007. Mimulus is an emerging model system for the integration of ecological and genomic studies. *Heredity*, 100(2), pp.220–230.

Yeaman, S. & M.C. Whitlock. 2011. The genetic architecture of adaptation under migration-selection balance. *Evolution*, 65(7), pp.1897–1911.

Table 1. Pair-wise F'_{ST} values among the five geographical regions defined using DAPC clustering.

Region (N individuals)	*Central (169)*	*Southeastern (29)*	*Southern (38)*	*AHQT (25)*
Northern (40)	0.12	0.10	0.11	0.24
AHQT (25)	0.16	0.23	0.24	
Southern (38)	0.10	0.10		
Southeastern (29)	0.09			

Table 2. Summary of the five regions total number of polymorphic sites, percent of polymorphic loci, observed and expected homozygosity (± SE), and π (± SD), which were calculated using variant sites.

Region (N individuals)	*Polymorphic Sites*	*Percent Polymorphic Sites*	*Observed Hom (± SE)*	*Expected Hom (± SE)*	*π (± SD)*
AHQT (25)	25,792	0.18	0.86 (0.0007)	0.82 (0.0009)	0.0056 (0.045)
Central (169)	32,713	0.23	0.78 (0.0005)	0.67 (0.0007)	0.0100 (0.060)
Southern (38)	31,815	0.23	0.73 (0.0008)	0.68 (0.0009)	0.0100 (0.062)
Northern (40)	32,093	0.23	0.76 (0.0007)	0.66 (0.0008)	0.0106 (0.063)
Southeastern (29)	29,936	0.21	0.80 (0.0008)	0.71 (0.0009)	0.0092 (0.056)

Table 3. Summary of the within population comparisons in sample size, F'st, percent polymorphic loci, and π (± SE) for both the annual/perennial and TT/NN comparisons (statistics were calculated using variant sites). AHQT was compared to a subset of the Central individuals to highlight the changes in genomic diversity. Plants from Upper Geyser Basin (UGB; a group within Central) were chosen to see if these patterns are similar on a microgeographic basis, with all plants being collected within 3 miles of each other.

Populations	*N A/P*	*F'st between habitats*	*Percent Polymorphic Sites A/P*	*π (± SD) A/P*	*N TT/NN*	*F'st Between genotypes*	*Percent Polymorphic Sites TT/NN*	*π (± SD) TT/NN*
Central	113/42	0.03	0.30 0.30	0.0093(0.06) 0.0102 (0.058)	75/13	0.05	0.17 0.17	0.0092 (0.06) 0.0100 (0.06)
Subset: UGB	31/21	0.04	0.18 0.19	0.0102 (0.06) 0.0089 (0.06) 0.0101 (0.066)	26/7	0.11	0.29 0.26	0.0091 (0.06) 0.0094 (0.06)
AHQT - AHQN	25/14	0.22	0.15 0.15	0.0056 (0.05) 0.0090 (0.06)	20/8	0.23	0.13 0.14	0.0056 (0.05) 0.0089 (0.06)
Southern	8/30	0.03	0.14 0.18	0.0081 (0.06) 0.0101 (0.06)	10/16	0.05	0.15 0.16	0.0084 (0.06) 0.0100 (0.06)
Northern	10/30	0.03	0.16 0.18	0.0097 (0.06) 0.0107 (0.06)	7/13	0.06	0.14 0.16	0.0093 (0.06) 0.0100 (0.06)
Southeastern	17/12	0.02	0.16 0.15	0.0087 (0.06) 0.0091 (0.06)	15/2	0.07	0.15 0.07	0.0090 (0.06) 00.063 (0.06)

Supplemental Table 1: Summary of the variable parameters as well as loci and SNP outputs from each unique "populations" run.

"Populations" run	*# of populations*	*# of polymorphic loci*	*# of variant sites*
Single population	1	22,441	49,539
5 regions	5	17,715	32,742
AH-PH	10	15,632	23,054
TT-NN	10	15,760	22,056
UGB AH-PH	2	24,666	57,529
UGB TT-NN	2	23,834	54,987

Supplemental Table 2: Summary of the number of polymorphic loci and variant sites by changing the minor allele frequency between "populations" runs.

"Populations" run	*Minor allele frequency*	*# of polymorphic loci*	*# of variant sites*
Single population	0.1	22,438	49,539
	0.075	22,441	56,975
	0.05	22,441	66,729
5 regions	0.1	17,715	32,742
	0.075	17,715	37,495
	0.05	21,065	60,649
AH-PH	0.1	15,632	23,054
	0.075	15,632	26,719
	0.05	15,632	31,480
TT-NN	0.1	15,760	22,056
	0.075	15,760	25,447
	0.05	15,760	30,055
UGB AH-PH	0.1	24,666	57,529
	0.075	24,666	65,963
	0.05	24,666	76,846
UGB TT-NN	0.1	23,834	54,987
	0.075	23,834	63,185
	0.05	23,834	68,254

Supplemental Table 3: Summary of the AIC and ΔAIC for the three different model for the three significant factors that predict individuals F_{IS}.

Model	*AIC*	*ΔAIC*	*K*	R^2
Habitat	23.46	0	2	0.114
Habitat + S-A Distance	27.69	4.23	3	0.188
Habitat + S-A Distance + Days to Flower	29.82	2.13	4	0.234

Figure 1. The collection locations of *M. guttatus* across YNP. The number of collection sites for a given geographic area is represented in the size of the pie chart, and colors represent the number of each habitat type (annual = red, perennial = blue) in that area. The regions assignments from the DAPC analysis (Fig. 5) is represented in the colored areas, with five distinct regions, a Northern (green), Central (blue), AHQT (maroon), Southern (light blue) and Southeastern (gold).

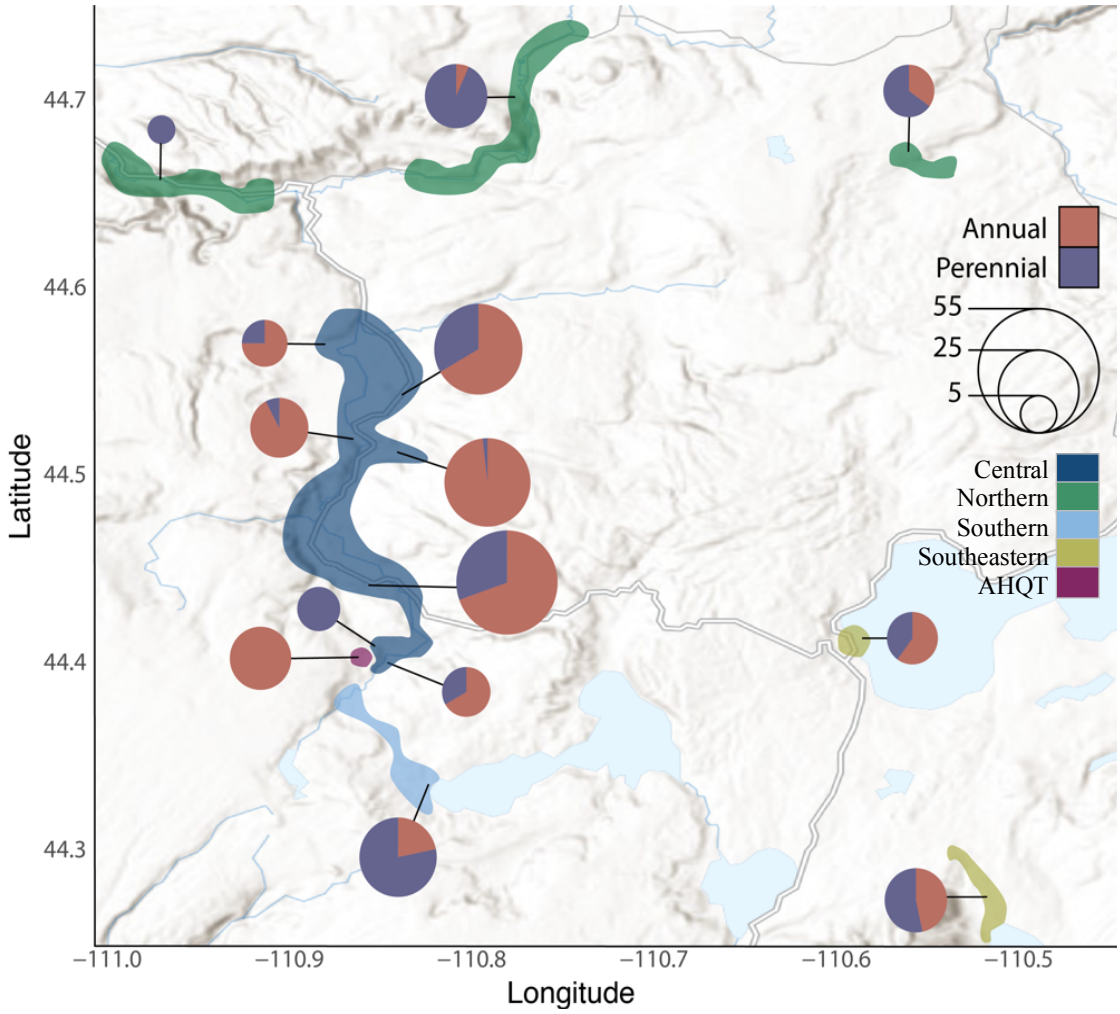
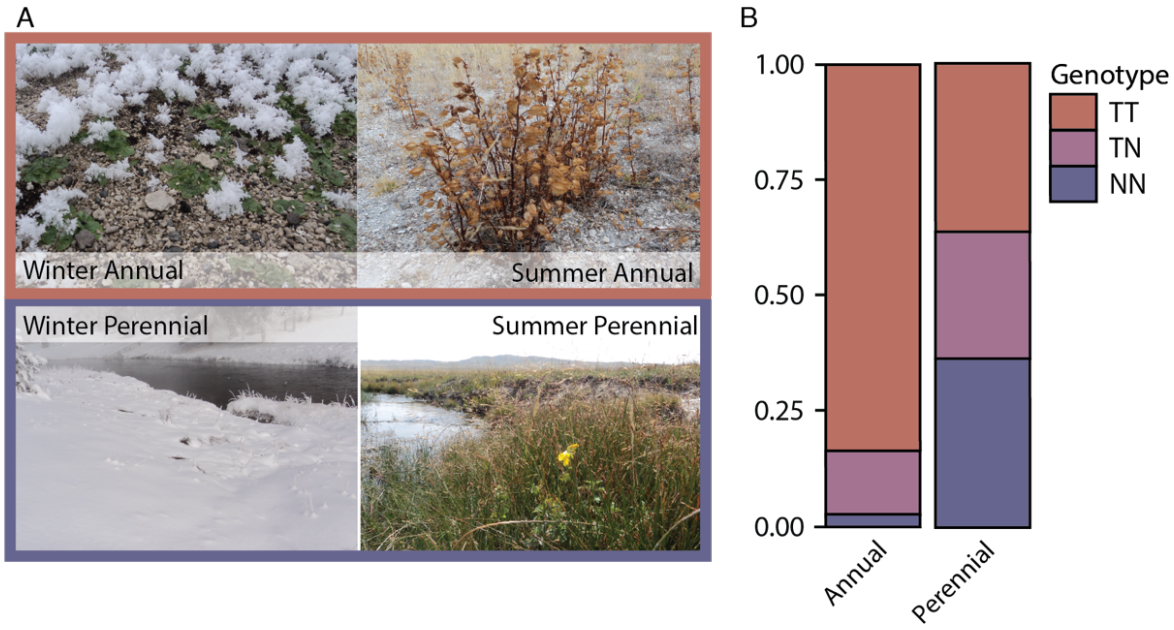


Figure 2. A) The differing ecology of annual (red) and perennial (blue) sites in the winter and summer. B) Genotypes at *out6* are non-randomly distributed across annual and perennial habitats ($X^2 = 95.175$, $N = 320$, $p < 0.0001$). Annual habitats contain over 85% TT genotypes, while NN genotypes are found almost exclusively in perennial habitats.



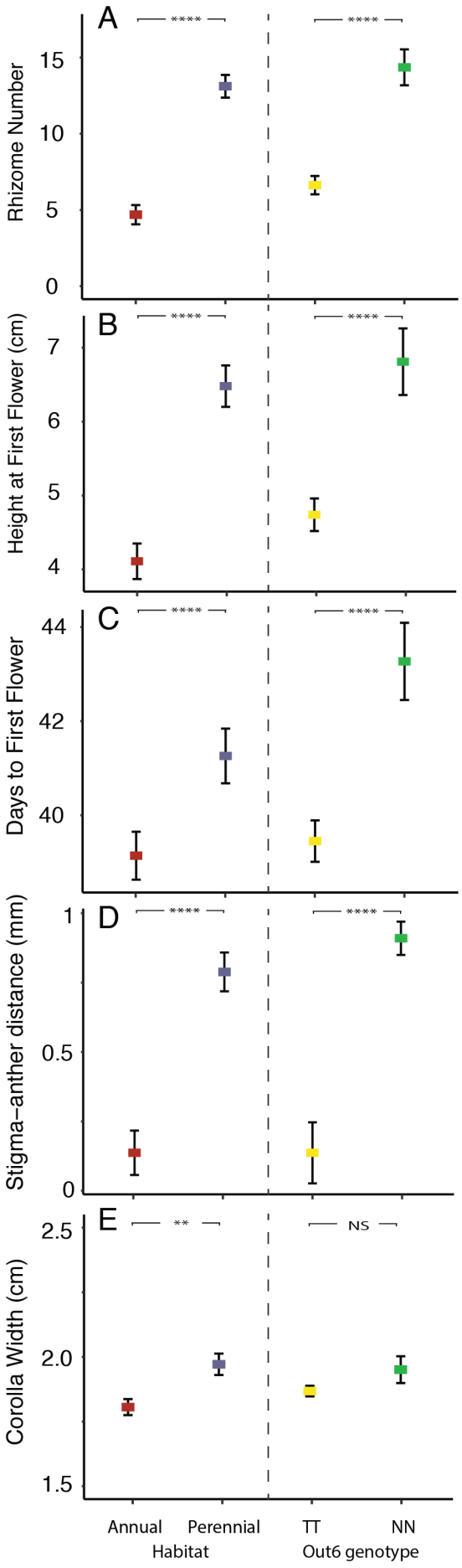


Figure 3. Mean (± SE) of phenotypic traits for the 1st generation inbred progeny from the wild collected accessions grown under greenhouse conditions. Each trait is analyzed in two different categories, annual/perennial habitats (N = 113/95; Red/Blue) (left) and TT/NN (N = 121/47; Yellow/Green) (right). *Out6* heterozygotes (NT) were excluded from the genotype analysis due to small sample size. Significance is denoted by asterisks, where **** is < 0.0001, *** is < 0.001, ** is < 0.01, * is < 0.05, NS is non-significant.

Figure 4. PC analysis of all wild-collected individuals (N = 302). Plants from the AHQT population (purple) are highly differentiated from the rest of the *M. guttatus* in Yellowstone National Park (grey).

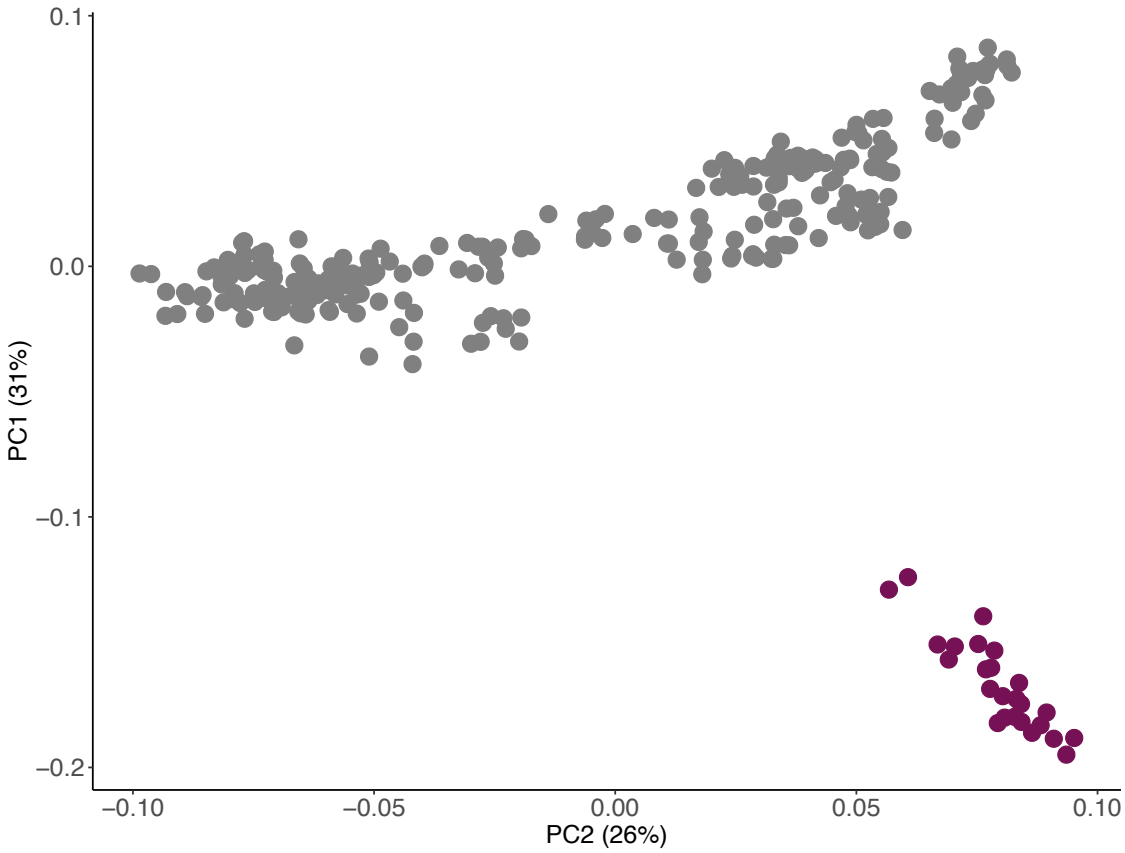


Figure 5. Discriminant analysis of principal components (DAPC) using 39% of the total variation. AHQT clusters with geographically close neighbors in the Central and Southern region (Fig. 1). Inset: The Bayesian Inference Criterion (BIC) across varying cluster numbers (K). K = 5 was chosen as the "true number of populations" as it is the lowest number of clusters defining a trough of BIC values from k = 5-11.

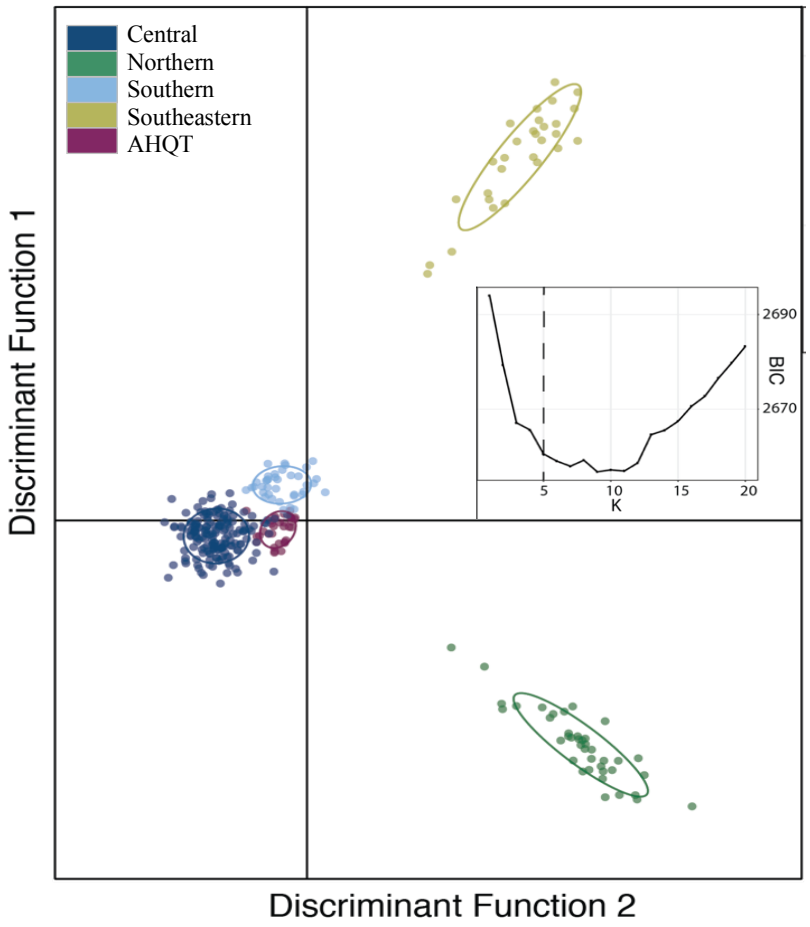


Figure 6. PC analysis excluding AHQT (N = 276) shows individuals cluster by region and geography. The first two PCs explain over 45% of the variation. Habitat types (perennial habitats (PH - triangles) and annual habitats (AH - circles) show little separation across geography, instead are distributed relatively evenly throughout.

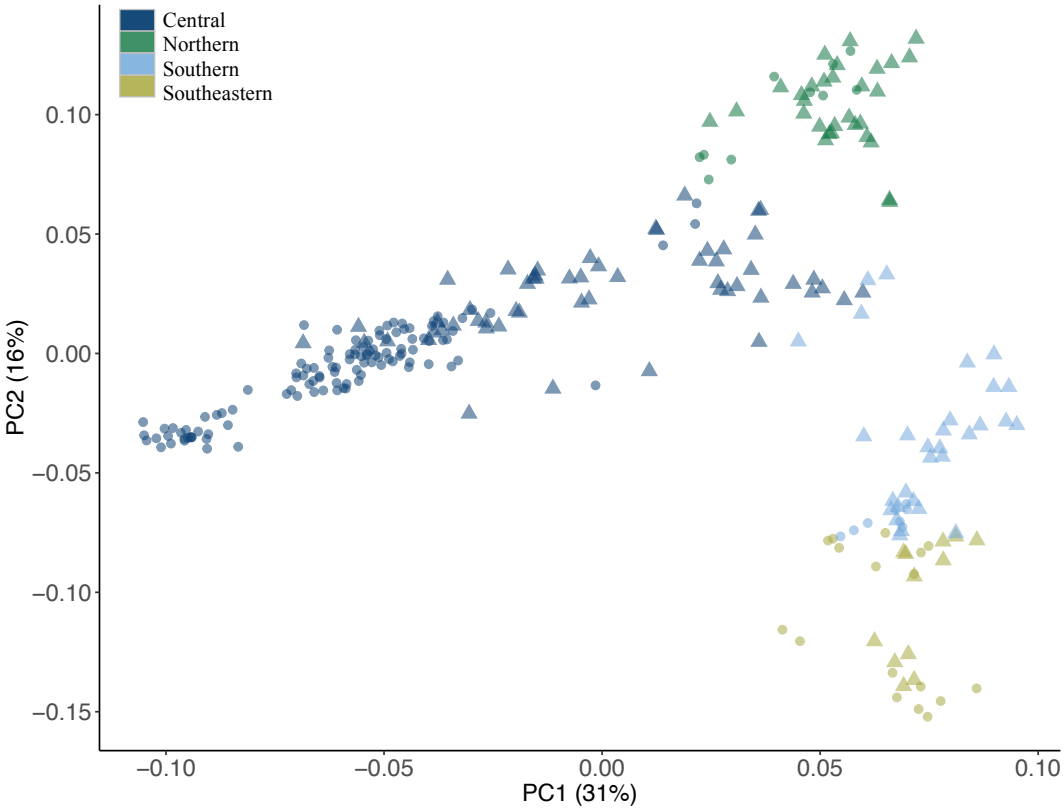


Figure 7. PC analysis of only individuals from Upper Geyser Basin (UGB, N = 52; Left) against geographic location (Right). Habitat (annual habitats = circle), perennial habitats (triangle)) and F_{IS} (low inbreeding = blue, high inbreeding = red) appear to structure a portion of variation.

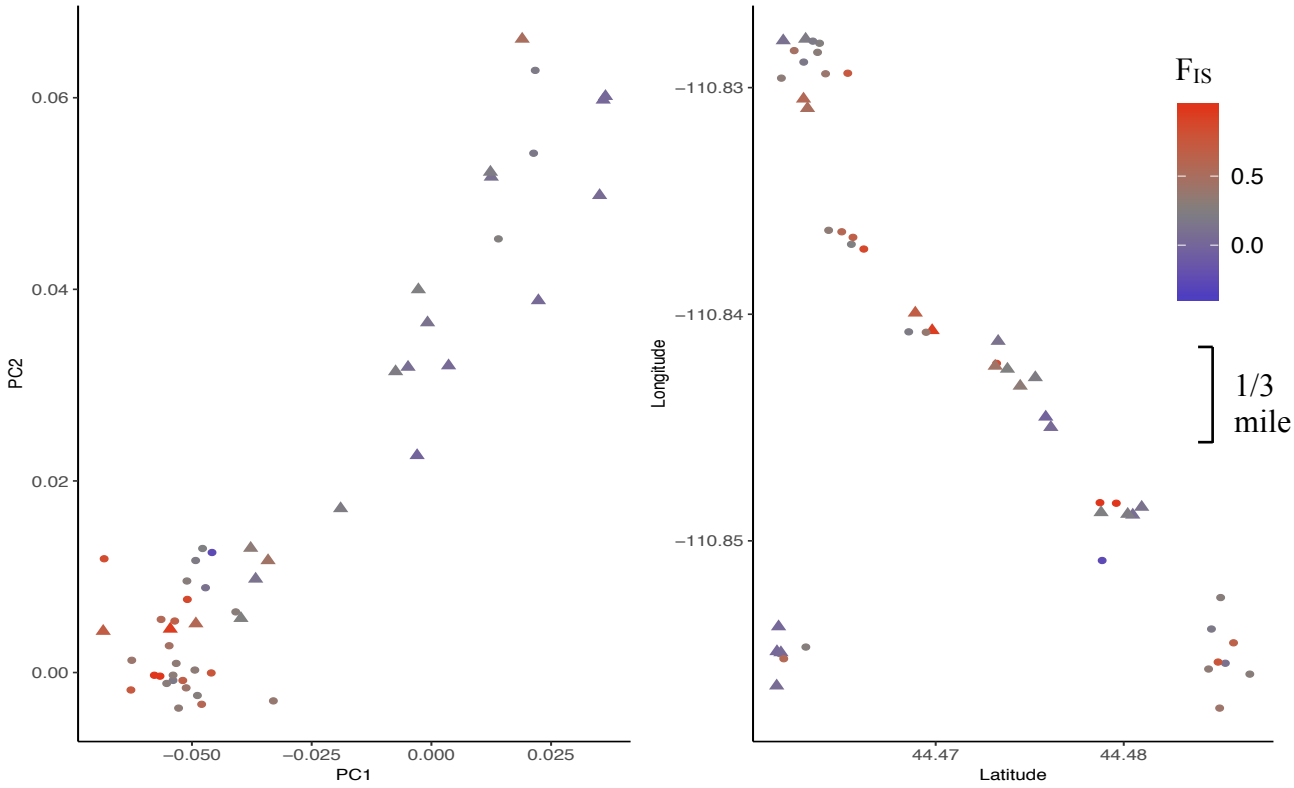


Figure 8. Partial Mantel correlogram depicting the association between genetic distance and geographic distance among pairs of individuals of each distance class (4km), while taking into account the habitat of the individuals (n = 302). Significant associations ($p < 0.05$) are represented by filled in squares. No values were significant between 30km and 60 km and are not shown.

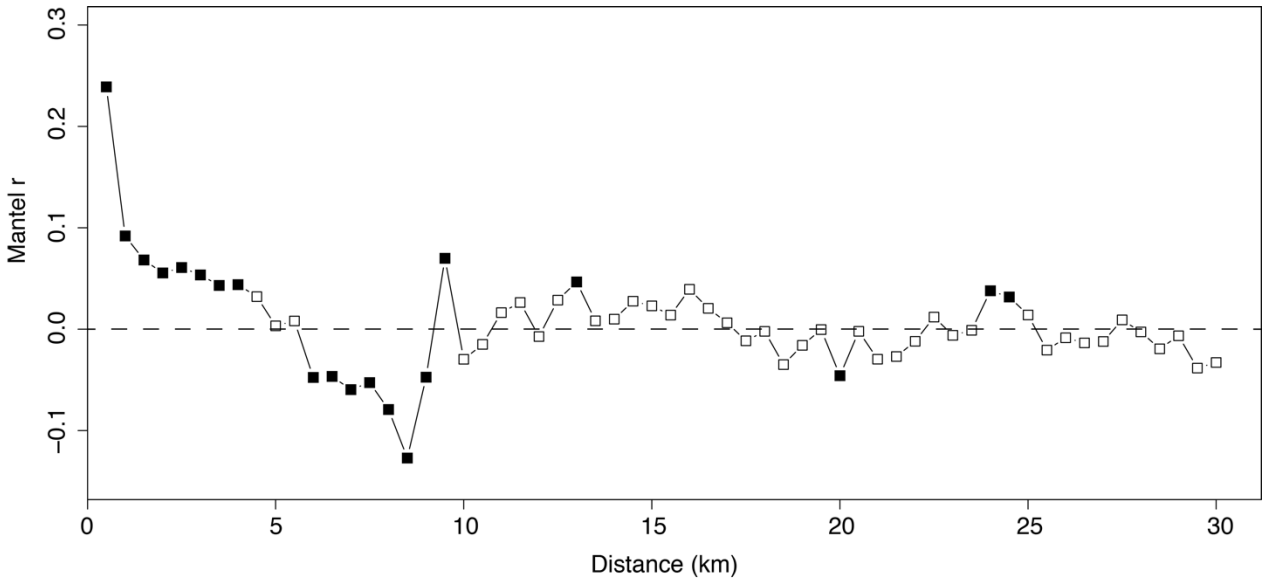


Figure 9. Least Squares Means (+/- SE) of genetic dissimilarity between induvial from AH-AH (red), AH-PH (purple) and PH-PH (blue) comparisons within 0.5 km distance classes of the first 5 km. PH-PH comparison in distance class 4 km-5 km was removed due to low sample size. Asterisks indicate significant differences in means for that distance class (* : $p > 0.0005$)

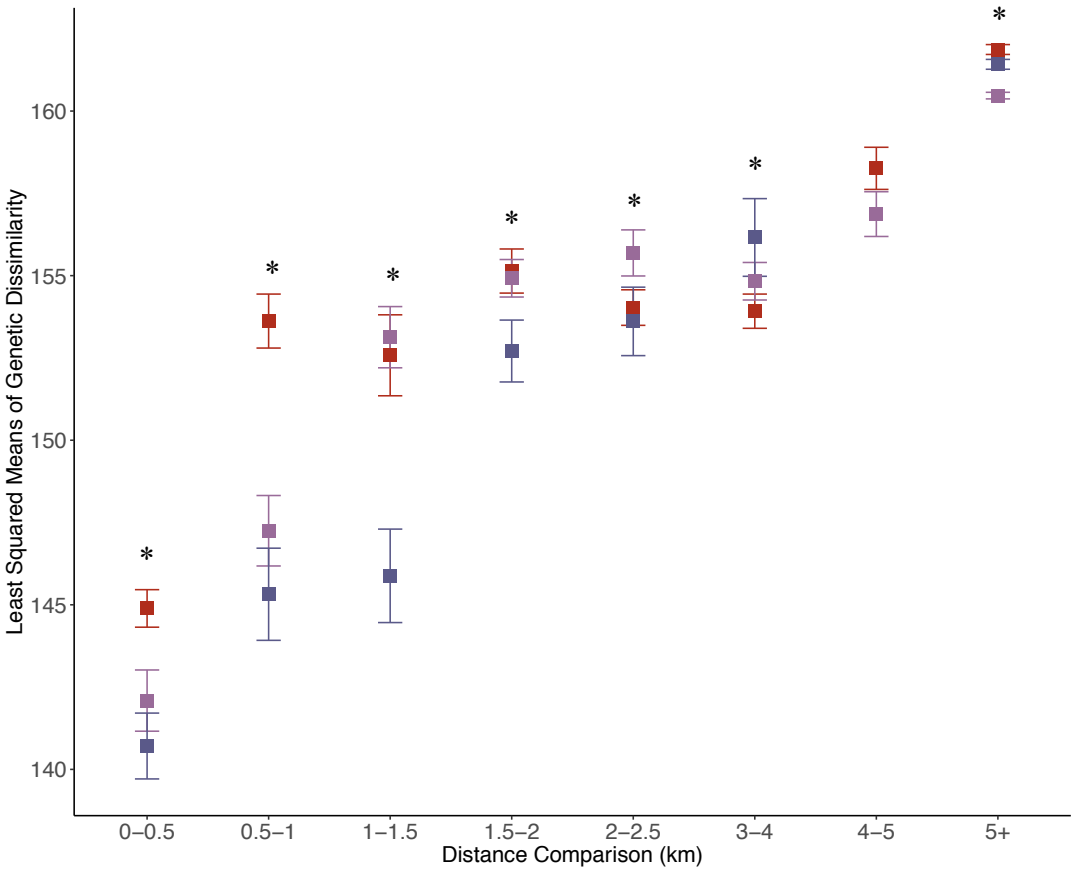


Figure 10. A PC analysis (excluding AHQT, N = 276) of the habitat (annual = circle, perennial = triangle) and F_{IS} (low inbreeding = blue, high inbreeding = red) of individual plants in YNP. PC 1 and 2 explain more than 45% of the variation. While geography was previously shown to structure genetic variation, individuals F_{IS} appears to structure some variation as well.

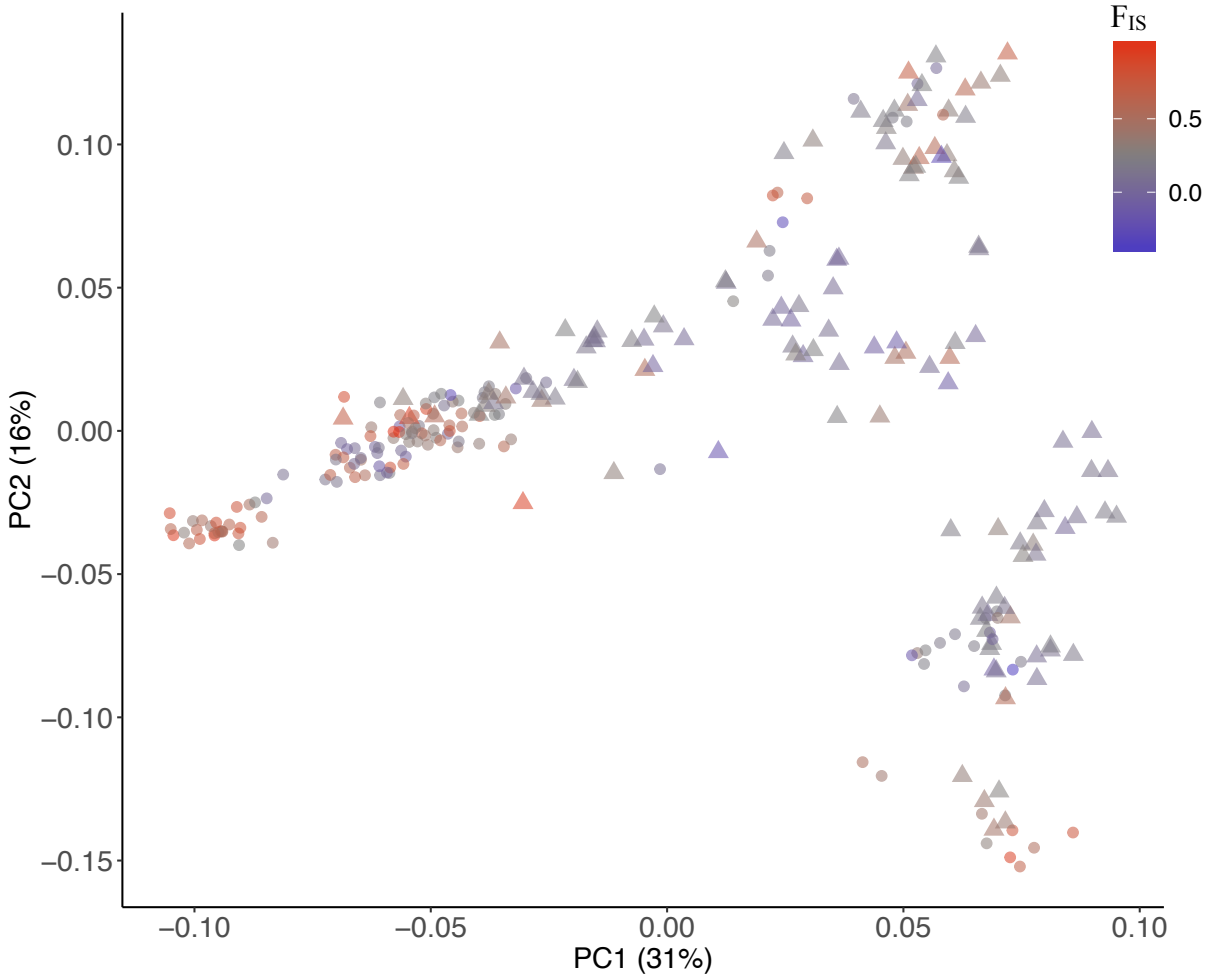


Figure 11. Mean F_{IS} (± SE) of wild collected individuals compared between habitat (AH-PH) and genotype (TT-TN-NN). Habitat was analyzed using a one-way t-test ($F(1, 217\ N) = 7.42$, $p < 0.007$). Genotype was analyzed using an ANOVA ($F(2, 200\ N) = 10.5$, $p < 0.001$. Post-hoc Tukey-Kramer HSD was run to test for directionality (TT-NN, $p < .03$; TT-TN, $p < 0.001$; TN-NN, $p < 0.3$). Habitat and genotype comparisons included the same individuals. Asterisk indicates significance, * = $p < 0.05$; ** = $p < 0.005$*.

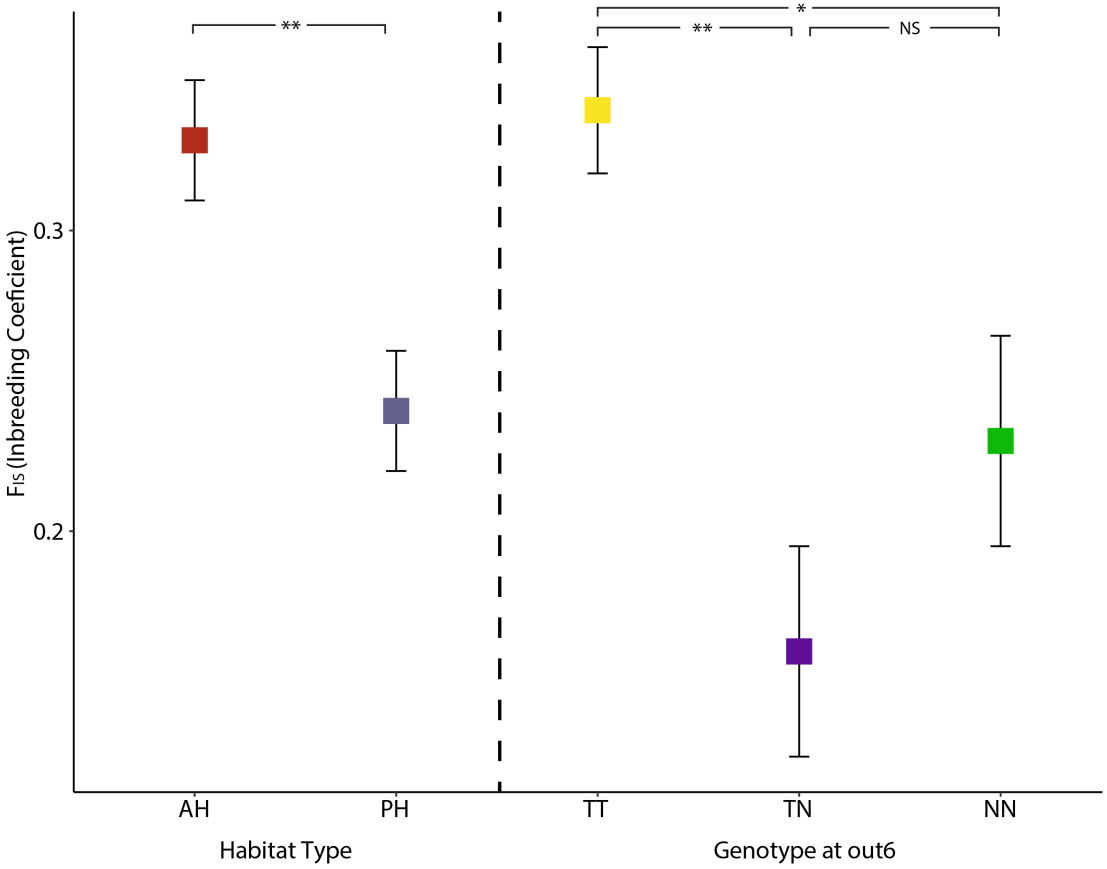


Figure 12. Mean among-region F'_{ST} for annual (A-A), perennial (P-P) or Annual-Perennial (A-P) pairs. A permutation analysis of these 24 samples (6 A-A; 12: A-P; 6 P-P) found marginally significant differences of F'_{ST} between A-A and P-P ($F(2, 24\ N) = 2.26$, $p < 0.03$). Asterias indicate significance * = $p < 0.05$; NS = $p > 0.05$.

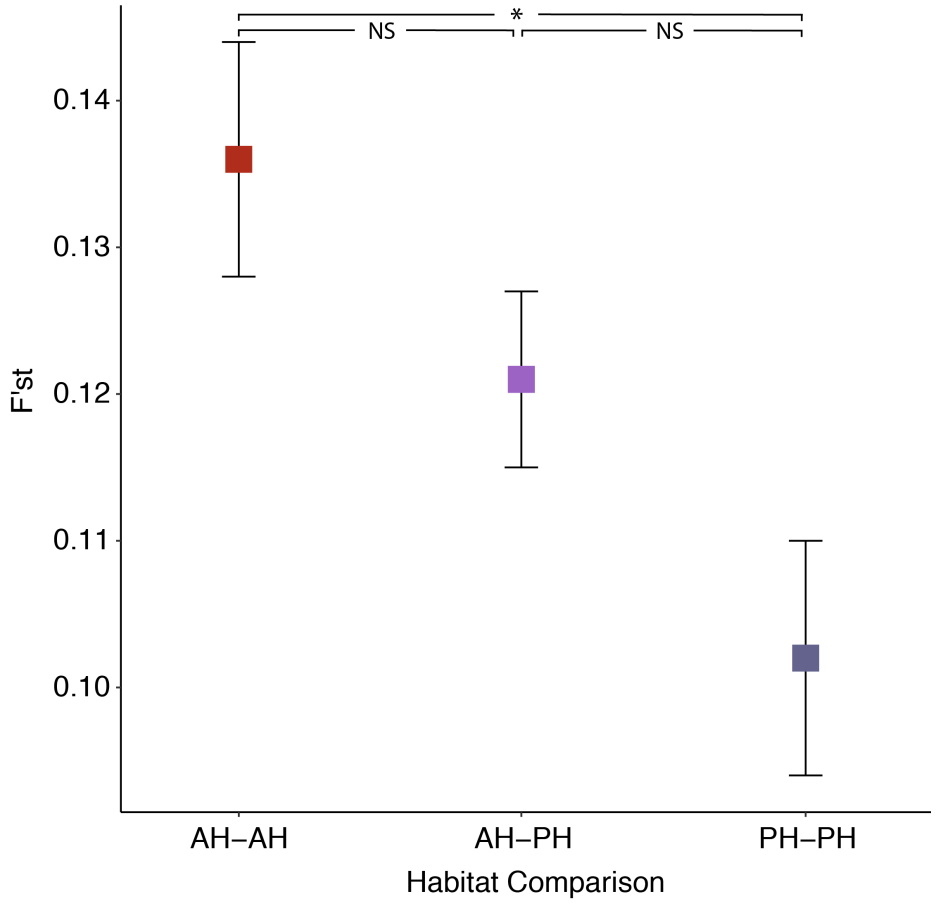
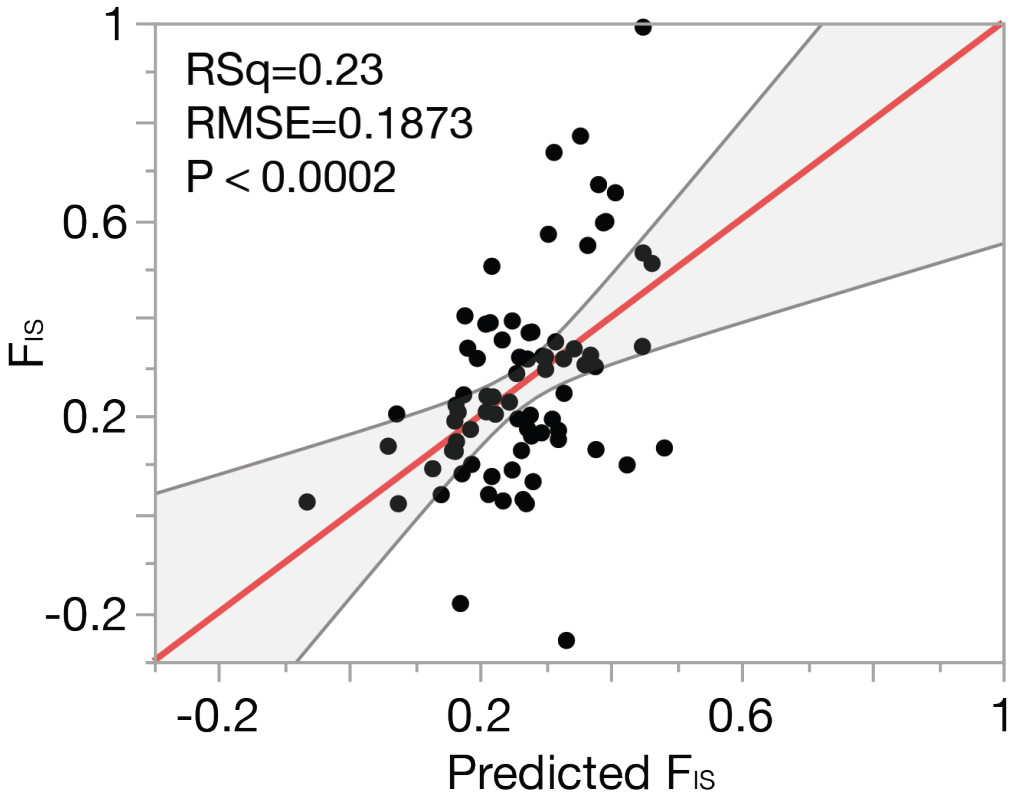


Figure 13. F_{IS} against model predicted F_{IS} with stigma-anther distance ($p < 0.004$), habitat ($p < 0.008$) and days to first flower ($p < 0.04$) as significant predictors. Grey lines indicate 95% confidence interval.



Supplemental Figure 1: Violin and box plot of number of days to flower for the 1st generation inbred progeny from the wild collected accession grown under greenhouse conditions. Days to flower is divided in two different categories, annual/perennial habitats (N = 60/52; Red/Blue) (left) and TT/NN (N = 80/22;) (right), with the two comparisons sample the same individuals. The heterozygotes (NT) were excluded from the genotype analysis due to small sample size. Significance is denoted by asterisks, where *** is < 0.001, and * is < 0.05.