One gene or many?
Different genetic mechanisms drive convergent evolution in monkey flowers

Matthew Farnitano

Under the supervision of Dr. John Willis

Undergraduate Thesis for Distinction in the Department of Biology
Duke University
Durham, North Carolina
May, 2016
# Table of Contents

Abstract .................................................................................................................................................. 2  
Introduction ........................................................................................................................................... 3  
  - The genetic basis of convergent evolution  
  - Approaches to the study of parallel evolution  
  - Serpentine soil as a model for parallel evolution  
  - Monkey flowers: an ideal model of serpentine adaptation  
  - Investigating parallel adaptation to serpentine soils in monkey flowers  
Methods .................................................................................................................................................. 9  
  - Populations used in this study  
  - Soils used in this study  
  - Reciprocal transplant experiment  
  - Hybrid formation  
  - Soil tolerance assays  
  - Genomic extraction  
  - Bulk sequencing  
  - Bulk segregant analysis and QTL mapping  
Results .................................................................................................................................................. 16  
  - Reciprocal transplant experiment  
  - *M. nudatus* X *M. guttatus* hybrid tolerance assay  
  - *M. glaucescens* X *M. guttatus* hybrid tolerance assay  
  - *M. pardalis* X *M. guttatus* hybrid tolerance assay  
  - QTL mapping in *M. nudatus* X *M. guttatus* hybrids  
  - QTL mapping in *M. pardalis* X *M. guttatus* hybrids  
Discussion .............................................................................................................................................. 28  
  - Serpentine soils are an equivalent, general selective environment for *Mimulus*  
  - No single large-effect gene is responsible for serpentine tolerance in *Mimulus nudatus*  
  - Germination and phenotype scoring procedures affect tolerance assays in *Mimulus pardalis*  
  - Hybrid incompatibility is a likely cause of a reverse QTL in *Mimulus pardalis*  
  - Robust dominance of serpentine tolerance characterizes *Mimulus glaucescens*  
  - Conclusions  
Acknowledgements ................................................................................................................................. 35  
Literature Cited ..................................................................................................................................... 36
Abstract

Evolutionary history is riddled with examples of convergent evolution, in which the same adaptation appears independently across multiple populations as a result of similar selective pressures. Convergent evolution can proceed by identical, similar, or unrelated genetic mechanisms. The relative frequencies of these scenarios and the evolutionary constraints that produce them are poorly understood, partly due to a lack of diverse comparative models. One such promising model is repeated adaptation to serpentine soil, a soil environment characterized by abundant heavy metals, low nutrient content, and poor water retention. Many species of *Mimulus* monkey flowers have populations that can tolerate these soils, but most *Mimulus* populations cannot. I compared the genetic signatures of serpentine tolerance across the genomes of four tolerant *Mimulus* populations, in order to determine whether these different species employed similar genetic mechanisms. Previous work has identified a single locus strongly connected with serpentine tolerance in *M. guttatus* tolerant populations. I found that *M. glaucescens* also likely has a single or small number of loci responsible for this adaptation. However, serpentine tolerance in *M. nudatus* appears to be controlled by many genes of smaller effect, rather than a single locus. This vastly different genetic mechanism is surprising given the close evolutionary and ecological relationships of these species. The results of this study show that, even under similar circumstances, evolution can employ very different genetic strategies.
Introduction

Convergent evolution, the independent acquisition of a particular trait by multiple lineages, is a central feature of natural selection. Repetition of the same adaptation is taken to be a sign of natural selection at work, since random drift is unlikely to produce such consistent results. In distantly related groups, evolutionary convergence demonstrates the power of natural selection to overcome limitations of heritage. Paradoxically, convergence in closely related groups can highlight the limitations of heritage on natural selection, by demonstrating that only one adaptive strategy is possible (Cooley and Willis, 2009). There are a number of factors that might limit the set of possible adaptations; these include the physical laws of nature, historical restrictions on development and structure, and demographic restrictions on available variation (Gould, 1989; Rosenblum et al., 2014). Fundamentally, all of these constraints are genetic: only variation that pre-exists in the genome, or that can be produced through mutation on the existing genetic background, can be selected for. The convergent evolution of closely related populations provides a system in which we can better understand the nature and influence of these genetic constraints. In addition, it provides insights into the age-old question of whether evolution would repeat itself if we were to “rewind the tape” of life. In order to delve into the genetic basis of convergence and constraint, I analyzed genetic parallelisms in the adaptation of monkey flower species (*Mimulus*: Phrymaceae) to harsh serpentine soils.

The genetic basis of convergent evolution

Convergence has been studied extensively at the phenotypic level, particularly through morphological and ecological comparisons (Thewissen et al., 2012). However, an understanding of evolutionary constraints requires knowledge of the molecular mechanisms driving
convergence at the genetic level. Similarity in the molecular mechanisms underlying convergence is termed parallel evolution, or genetic parallelism (Rosenblum et al., 2014). This similarity can happen at many different levels, from shared pathways to shared genes and even to shared mutations (Rosenblum et al., 2014). For example, multiple lineages of predatory snakes have acquired identical mutations in a sodium ion channel that confers resistance to tetrodotoxin, a toxin found in many amphibian prey species (Feldman et al., 2012). Meanwhile, different mutations in the same gene, the melanocortin receptor mc1r, have been implicated in coat color changes in white-phase black bears and in rock pocket mice (Nachman et al., 2003; Ritland et al., 2001). In contrast, another population of rock pocket mice has similar phenotypic changes in coat color caused by an entirely different gene, the Agouti signaling peptide (Linnen et al., 2013).

Approaches to the study of parallel evolution

A central question in studies of parallel evolution is the degree to which convergent phenotypes are caused by similar mutations, genes, or pathways. Experimental evolution is a promising tool to study these rates. In this approach, selection is reproduced in controlled laboratory conditions, allowing massive replication and accurate quantification at all stages (Achaz et al., 2014). In the longest-running experiment, twelve identical populations of E. coli evolved in the lab for 20,000 generations (Cooper et al., 2003). Analysis of gene expression patterns in two of these populations revealed that, of the 59 significant changes in gene expression, all 59 were changed in the same direction, demonstrating incredible parallelism at the scale of gene pathways. Similarly, an experiment by Wichman et al. (1999) cultured two populations of bacteriophage for 1000 generations under strong selective pressures. They found that about 50% of individual mutations occurred in both populations, showing a high degree of parallelism even at the mutation level. However, these same mutations evolved at different times.
and in different orders, demonstrating that parallel results do not necessarily represent identical histories (Wichman et al., 1999).

One limitation of using experimental evolution to infer frequencies of parallel evolution is that natural populations are subject to much more complex selective pressures, greater heterogeneity in the environment, and differing degrees of phylogenetic similarity prior to convergence. As a result, studies of natural populations in many different contexts are necessary for a complete understanding of parallel adaptation. A review of the literature by Conte et al. (2012) used 25 studies of natural populations to infer the probability of parallel adaptation at the gene level. They estimated that, on average, 32% of convergent phenotypes are the result of parallel changes in the same genes. However, this study could not eliminate publication bias as a factor, and the 25 studies they reviewed are unlikely to represent the full breadth of circumstances under which parallel adaptation can occur.

Situation-specific factors can have important effects on parallel evolution, as demonstrated in threespine stickleback fish. Sticklebacks are ancestrally a marine species, but a number of independent populations have become isolated in freshwater lakes (Schluter et al., 2004). Parallel adaptations have arisen in these freshwater populations, including a decrease in armor plating, pelvic reduction, and consistent changes in body shape. Each parallel trait, however, has a different genetic history. A single gene pathway underlies pelvic reduction across all populations, while many genes of small effect have resulted in body shape changes (Schluter et al., 2004; Shapiro et al., 2004). Armor plate reduction was caused by the parallel fixation of alleles already present at low levels in the ancestral marine populations (Colosimo et al., 2005). Stickleback research demonstrates how the number and effect size of related genes, as well as the presence of pre-existing variation, can alter the mechanisms of parallel evolution. More diverse
studies are needed to address the effects of these and other situational factors on evolutionary constraint and parallel evolution.

*Serpentine soil as a model for parallel evolution*

Another promising model for the study of parallel evolution in a natural context is serpentine soil. Serpentine is technically a mineral, but the term serpentine can refer simultaneously to the mineral itself, rocks composed primarily of this mineral, soils derived from these rocks, and habitats arising on these soils. Soils derived from these rocks are low-productivity, high-stress environments for most plant species. In addition, areas of serpentine exposure are characterized by high degrees of environmental heterogeneity and steep environmental gradients. As a result, serpentine localities form distinct habitats with high rates of endemics—species found only in a particular habitat or region. (Harrison and Rajakaruna, 2011).

Three primary characteristics of serpentine soils provide challenges to plant life. First, serpentine soils are high in magnesium, iron, and other heavy metals, at concentrations toxic to many plants (DeHart et al., 2014). Second, they are low in key inorganic nutrients, particularly calcium, potassium, and phosphorus (DeHart et al., 2014). Third, serpentine soils have low water retention rates, leading to drought stress (Hughes et al., 2001). In response to these threats, adaptations to serpentine soils have arisen in at least 105 unique families and 41 orders of plants, making serpentine habitat a powerful driver of convergent evolution (Harrison and Rajakaruna, 2011). In California, an estimated 246 plant species are endemic to serpentine habitats (Anacker et al., 2011). Adding to these numbers are bodenvag species, a term referring to species tolerant to serpentine soil but not found exclusively on serpentine sites (Hughes et al., 2001). A number of studies have specifically addressed the convergence of serpentine adaptation, demonstrating that serpentine tolerance has arisen independently in multiple populations of the same species.
(Berglund et al., 2004; Rajakaruna et al., 2003; Yokoo et al., 2009) as well as in closely related species (Cecchi et al., 2011; Ivalu Cacho et al., 2014). However, no study to date has addressed the genetic mechanisms underlying phenotypic convergence in any serpentine-tolerant group. Serpentine soil adaptation provides a unique opportunity to study parallel evolution at the genetic level, since multiple selective pressures—heavy metals, low nutrient content, and drought stress—are acting in concert in a highly repeatable fashion. Furthermore, the large number and diversity of endemic plant species means that parallel evolution can be compared at multiple hierarchical levels simultaneously.

**Monkey flowers: an ideal model of serpentine adaptation**

The genus *Mimulus* (Phrymaceae), a group of flowering plants known as monkey flowers, is a perfectly poised model system for this confluence of evolutionary questions. *Mimulus* has undergone a spectacular, very recent and ongoing adaptive radiation in the western United States, and as a result has been studied extensively by ecologists and evolutionary biologists who are interested in the speciation process (Twyford et al., 2015). The genus contains a number of serpentine endemics, as well as *bodenvag* species and species excluded from serpentine sites (Beardsley et al., 2004; Safford et al., 2005). Serpentine affinity appears to have arisen independently multiple times, including in closely related species with imperfect reproductive isolation (Nesom, 2014). Finally, a wealth of genetic and genomic tools have been developed in *Mimulus*, allowing the genetic mechanisms of adaptation to be understood (Twyford et al., 2015). Monkey flowers are uniquely situated to provide powerful insights into parallel adaptation in serpentine habitats.

In the widespread *bodenvag* species *Mimulus guttatus*, there has been some attempt to characterize the genetic basis of serpentine tolerance. Palm et al. (2012) demonstrated that
magnesium exclusion was not the operative mechanism, while Hughes et al. (2001) showed that
drought tolerance plays an important role. Recent work has identified a genomic signature, in the
form of a single major quantitative trait locus (QTL), associated with both serpentine soil
tolerance and tolerance to low Ca:Mg ratios, found in chromosome 13 of the *M. guttatus* genome
(Selby, 2014). This finding suggests that a single gene or gene cluster conveys serpentine
tolerance through some mechanism involving regulation of Ca:Mg ratios. In addition, Selby
(2014) found that this same QTL was present in two geographically distant populations of *M.
guttatus*. A number of species, some of which are adapted to serpentine soil, are thought to have
arisen through isolation from *M. guttatus*. It is unknown whether similar genomic signatures
exist in these species.

*Investigating parallel adaptation to serpentine soils in monkey flowers*

In order to investigate the molecular mechanisms underlying phenotypic convergence, I
compared adaptation to serpentine soil in *M. guttatus* with similar adaptations in three related
serpentine endemics: *M. nudatus, M. pardalis*, and *M. glaucescens*. First, I asked whether each
species was better adapted to its native serpentine soil relative to non-serpentine relatives. Next, I
investigated whether native serpentine soils associated with each species provided equivalent
selective environments. Third, I searched for genomic signatures associated with serpentine
tolerance in these species, and compared these genomic signatures across species to investigate
whether parallel mechanisms of evolution were at play. The purpose of this comparative study
was to shed light on the interaction of multiple selective pressures and multiple genetic
constraints on the degree of parallel evolution. Future studies will aim to extend these
comparisons across the broad range of *Mimulus* species and other serpentine-tolerant plant
groups, providing a broad view of the drivers of parallelism in evolution.
Methods

Populations used in this study

The *Mimulus guttatus* species complex is a group of closely related populations in the process of speciation, for which reproductive isolation and speciation definitions are in flux (Beardsley et al., 2004; Oneal et al., 2014). I used four species within this complex to compare genetic sources of serpentine soil tolerance: *M. guttatus*, *M. nudatus*, *M. pardalis*, and *M. glaucescens*.

Two of the species used, *Mimulus nudatus* and *Mimulus pardalis*, are serpentine soil endemics with highly restricted ranges (Gardner and Macnair, 2000; Oneal et al., 2014). A third species, *Mimulus glaucescens*, is also restricted in range but found on both serpentine and non-serpentine soils (Habecker, 2012). *Mimulus guttatus* is a widespread species found on both serpentine and non-serpentine soils (*a bodenvag* species), and has been hypothesized to be the progenitor of the other three species, though true phylogenetic relationships remain unclear (Nesom, 2014; Palm et al., 2012). Source populations for each species are described in Table 1, and their collection sites are shown in Figure 1. Three independent populations of *M. guttatus* were used. A serpentine-tolerant population from Lake County, California, labelled GUG, is found sympatrically with serpentine-tolerant species *M. nudatus*, labelled GUN (Gardner and Macnair, 2000). A non-serpentine population from McAfee Gulch, California, labelled MED, is found nearly sympatrically with serpentine-tolerant species *M. pardalis*, labelled MCG (Oneal et al., 2014); the *M. guttatus* population is excluded from serpentine soils at that site. IMO is a non-serpentine *M. guttatus* population from Iron Mountain, Oregon that is used as the reference population for genomic studies in *Mimulus* (Wu et al., 2008); here it is used in hybrid crosses.
and to facilitate genomic analysis. The final serpentine population, *M. glaucescens*, is found in Butte and Tehama Counties in California, and is labelled PMO (Habecker, 2012).

### Table 1. *Mimulus* populations used in this study.

<table>
<thead>
<tr>
<th>Population ID</th>
<th>Location</th>
<th>County, State</th>
<th>Species</th>
<th>Serpentine?</th>
</tr>
</thead>
<tbody>
<tr>
<td>GUG</td>
<td>Guenoc Winery</td>
<td>Napa, CA</td>
<td><em>M. guttatus</em></td>
<td>YES (bodenvag)</td>
</tr>
<tr>
<td>GUN</td>
<td>Guenoc Winery</td>
<td>Napa, CA</td>
<td><em>M. nudatus</em></td>
<td>YES (endemic)</td>
</tr>
<tr>
<td>IMO</td>
<td>Iron Mountain</td>
<td>Linn, OR</td>
<td><em>M. guttatus</em></td>
<td>NO</td>
</tr>
<tr>
<td>MCG</td>
<td>McAfee Gulch</td>
<td>Calaveras, CA</td>
<td><em>M.pardalis</em></td>
<td>YES (endemic)</td>
</tr>
<tr>
<td>MED</td>
<td>Tuolumne</td>
<td>Tuolumne, CA</td>
<td><em>M. guttatus</em></td>
<td>NO</td>
</tr>
<tr>
<td>PMO</td>
<td>Platt Mountain</td>
<td>Butte, CA</td>
<td><em>M. glaucescens</em></td>
<td>YES (bodenvag)</td>
</tr>
</tbody>
</table>

![Figure 1](image_url)  

**Figure 1. Locations of *Mimulus* populations used in this study.** Four serpentine populations and two non-serpentine populations were collected from California and Oregon. Two populations, GUG and GUN, were collected from the same serpentine site. Counties where populations were collected are shaded in gray.
Soils used in this study

Three serpentine soil types collected in California were used in this study. They were collected from all three locations where the chosen study species were found: Guenoc Winery, McAfee Gulch, and Tuolumne County. Each is labelled according to the serpentine population found at its location: GUN, MCG, and PMO, respectively. Soil samples were dried and autoclaved prior to use, to prevent the germination of foreign seeds during experiments.

Reciprocal transplant experiment

In order to determine whether the native serpentine soils of the four tolerant populations were equivalent sources of selective pressure, I used reciprocal transplants to measure the degree to which the populations were adapted to specific soils versus serpentine soil more generally. Seeds from the four serpentine Mimulus populations, GUG (M. guttatus), GUN (M. nudatus), MCG (M. pardalis), and PMO (M. glaucescens), were germinated in petri dishes filled with ultra-pure filtered water. The germination dishes were placed in a dark room at 4°C for one week to simulate overwintering conditions, then moved to the Duke University greenhouse. After two to five days in the greenhouse, germinants from the petri dishes were planted in soil collected from each of three serpentine localities (GUN, MCG, and PMO), as well as in potting soil to serve as a control. For each population, 48 germinants were planted in each soil type. Soils were watered with ultrapure filtered water every 1-3 days as needed to maintain uniformly moist soil.

After six weeks under greenhouse conditions, I recorded the number of surviving plants, and the number of plants that had produced at least one bud or flower. Chi-squared tests for independence were used to determine whether source population and soil type were independent in their effects on survival rates and on reproductive rates. Under the null hypothesis that
adaptation is generic to all serpentine soils, source population and soil type would have independent effects. If source population and soil type had dependent, or interacting, effects, this would suggest the alternative hypothesis that the populations are differentially adapted to their local serpentine soils.

_Hybrid formation_

Three serpentine-tolerant populations were hybridized with a closely related non-tolerant population to produce hybrids with segregating tolerance alleles. *M. nudatus* (GUN) was crossed with the *M. guttatus* reference population (IMO). *M. glaucescens* was also crossed with the *M. guttatus* reference population (IMO). *M. pardalis* (MCG) was crossed with its nearly sympatric *M. guttatus* population (MED). Crosses were performed by manually transferring pollen from one individual to the stamen of its mate, marking the target flower, and collecting the dried seed pod. First-generation hybrids were grown in the Duke University greenhouse and self-pollinated using the same manual pollination technique to produce second-generation (F2) hybrid seeds. Seeds were stored dry at room temperature until use.

_Soil tolerance assays_

Soil tolerance assays were conducted to identify hybrid individuals displaying tolerance to serpentine soil in *M. nudatus* X *M. guttatus*, *M. glaucescens* X *M. guttatus*, and *M. pardalis* X *M. guttatus* F2 populations. 258 F2 seeds from the GUN X IMO hybrid cross were planted on serpentine soil from the *M. nudatus* native range (GUN soil). 238 F2 seeds from the MCG X MED hybrid cross were planted on serpentine soil from the *M. pardalis* native range (MCG soil). 789 F2 seeds from the PMO X IMO hybrid cross were planted on serpentine soil from the *M. glaucescens* native range (PMO soil). The soils were kept in a dark room at 4°C for one week.
after planting to simulate overwintering conditions, then moved to the Duke University greenhouse. Soils were watered with ultrapure filtered water every 1-3 days as needed to maintain uniformly moist soil. After six weeks, I recorded the number of seeds that had not germinated, germinated but died, or germinated and survived.

In order to confirm whether the serpentine-tolerant parent populations were better adapted to local serpentine soil than non-tolerant parent populations, seeds from each parent of the cross were planted alongside the hybrid offspring in serpentine soil, interspersed randomly between the hybrid seeds and labelled. 15 seeds of each parent were used for the GUN X IMO and MCG X MED crosses; with the PMO X IMO cross, 131 PMO and 132 IMO seeds were used. In addition, F1 offspring were planted with the MCG X MED and the PMO X IMO crosses, in the same manner as the parents. 20 MCG X MED F1 seeds and 100 PMO X IMO F1 seeds were planted. After six weeks, I recorded the number of seeds that had not germinated, germinated but died, or germinated and survived.

A modified replicate of this serpentine tolerance assay was conducted for the MCG X MED cross, in which 396 F2 hybrids, 20 MCG parents, 19 MED parents, and 10 F1 hybrids were first germinated in water for 2-5 days and then planted in MCG serpentine soil as above. After six weeks, plants were scored as healthy, necrotic (showing signs of poor health), or dead.

*Genomic extraction*

I used a CTAB-based DNA extraction protocol to extract bulk DNA samples from serpentine-tolerant F2 hybrid populations. For each hybrid cross, leaf or bud tissue samples from 80-100 F2 individuals that were classified as serpentine-tolerant during soil tolerance assays were pooled into a single bulk tissue sample. For MCG X MED hybrids, only individuals
classified as healthy during the modified replicate experiment were used in bulk samples. This sample was frozen and ground up using liquid nitrogen and a mortar and pestle, then suspended in cetyltrimethylammonium bromide (CTAB) buffer to solubilize the cell wall and membranes, with beta-mercaptoethanol added as a reducing agent to remove phenolic compounds. Chloroform separation was used to fractionate cell contents, and RNase A was added to depolymerize RNA. Finally, isopropanol followed by ethanol were used to precipitate genomic DNA, which was isolated and re-suspended in deionized water. DNA content and purity was analyzed using QuBit and NanoDrop spectrophotometry.

Genomic DNA was extracted from pools of tissue from one to a few individuals of each parental strain used in the hybrid crosses. The same CTAB protocol was used for these extractions.

**Bulk sequencing**

Bulk DNA samples extracted from each F2 population and each parent were sequenced using high-throughput Illumina Hi-Seq sequencing-by-synthesis. Samples were prepared for sequencing using the Nextera DNA sample preparation workflow. Genomic samples were fragmented and tagged with unique adapter sequences, then purified and amplified by PCR, followed by a final purification step using AMPure XP magnetic beads. This process generates a library of indexed, paired-end DNA fragments of about 150 base pairs in length, prepared for Illumina sequencing. The libraries were sequenced using an Illumina Hi-Seq 2500 platform.

**Bulk segregant analysis and QTL mapping**

Sequencing reads were analyzed using a workflow devised by Corcoran and Prinz (2015) in order to align reads to the *M. guttatus* reference genome, identify single-nucleotide
polymorphisms, and map allele frequencies across the genome. Raw data was checked for read quality using the FastQC program (version 0.11.5, Babraham Informatics). The Trim Galore! Program (version 0.4.1, Babraham Informatics) was used to remove adapter sequences and low-quality bases. Reads were aligned to the *M. guttatus* IMO reference genome using the Burrows-Wheeler Aligner (Li and Durbin, 2009). The Samtools package (Li et al., 2009) was used to index and compress the alignments. Single-nucleotide polymorphisms (SNPs), places where some proportion of the bulk sample contained a different allele than the reference genome, were identified using GATK (McKenna et al., 2010).

Once SNPs were identified for each bulk sample, I used statistical methods described by Magwene et al. (2011) to map variation in allele frequencies across the genome, generating a quantitative trait loci (QTL) map. At each variable site, I calculated the proportion of called alleles that matched the tolerant parent genotype. A sliding window algorithm was used, in which allele frequencies were averaged across overlapping windows of 1 million base pairs, in order to reduce noise and produce a smooth distribution. For each site, a g-statistic was calculated as a measurement of statistical deviation from the typical distribution. A weighted average of g-statistics was then calculated for each window, which I refer to as the g-score. A threshold g-score for identifying significant sites was defined using the False Discovery Rate method described by Benjamini and Hochberg (1995), with FDR = 0.05; this method prevents aberrant estimations of error rates during multiple sampling. Quantitative trait loci (QTL) were identified as continuous strings of sliding windows with g-statistics above the threshold value. These QTL represent potential regions of association with the serpentine tolerant phenotype. I compared QTL identified in each serpentine population to determine whether the same or
different genomic regions were associated with serpentine tolerance across the four serpentine species.

Results

**Reciprocal transplant experiment**

During the reciprocal transplant experiment, four serpentine-tolerant populations (GUN, GUG, MCG, and PMO) were each tested for tolerance on three different serpentine soils and a potting soil control. Tolerance was measured in terms of survival rates (Figure 2) as well as presence of flower buds (Figure 3). Each population was not significantly more tolerant to its native soil type than to the non-native soils, since source population and soil type had independent effects on survival rate ($\chi^2=6.37, df=15, p=0.97$) as well as on the presence of flower buds ($\chi^2=5.09, df=15, p=0.99$). The PMO population had the highest survival rates overall, while GUN produced the most buds. The GUG population survived relatively poorly in PMO soil, but its bud rate was hardly affected. In all four populations, differences in tolerance were minimal across soil types, including between the serpentine and control soils.
Figure 2. Reciprocal transplant survival rates. Each graph indicates the proportions of surviving individuals after 6 weeks of growth in one of three serpentine soils or a potting soil control. For each population, 48 germinants were planted in each soil. Black bars indicate survival in the population’s native soil type, gray bars indicate a foreign serpentine soil, and white bars indicate the control (non-serpentine) soil.
Figure 3. Reciprocal transplant budding rates. Each graph indicates the proportions of germinants that produced at least one bud after 6 weeks in one of three serpentine soils or a potting soil control. For each population, 48 individuals were planted in each soil type. Black bars indicate budding rates in the population’s native soil type, gray bars indicate a foreign serpentine soil, and white bars indicate the control (non-serpentine) soil.

M. nudatus X M. guttatus hybrid tolerance assay

Both parental lines of a GUN X IMO hybrid cross were planted, along with second-generation (F2) hybrids, in native GUN serpentine soil, and survival and budding rates were observed after 6 weeks (Figure 4). The non-tolerant IMO parent population had no survivors, while a high proportion (80%) of the serpentine-tolerant GUN parent population survived. 60% of the GUN population began to bud during the experiment (three-fourths of the surviving
individuals). The survival rate of the F2 hybrids was about halfway between that of the two parents, at 39.9%. Only 8.1% of the F2 hybrids produced buds, representing less than one-fourth of surviving individuals.

Figure 4. Serpentine tolerance assay of parental lines and second-generation hybrid offspring for a *M. nudatus X M. guttatus* hybrid cross. 15 individuals each from the GUN population of serpentine-endemic *M. nudatus* and the IMO non-serpentine population of *M. guttatus* were planted alongside 258 individuals from the second generation (F2) of a hybrid cross between GUN and IMO parent populations. All 288 seeds were planted on serpentine soil from the GUN population collection site. After six weeks, the proportion of surviving plants (survival rate), and the proportion of plants with at least one bud or flower (budding rate) were recorded for each population. Sample sizes (n) are listed above each population.

*M. glaucescens X M. guttatus* hybrid tolerance assay

Both parental lines of the PMO X IMO hybrid cross were planted, along with first-generation (F1) and second-generation (F2) hybrids, in native PMO serpentine soil, and survival rates were recorded after 6 weeks (Figure 5). The serpentine PMO parent population had a
relatively high survival rate (79%), while none of the IMO parent population survived. The F1 population had a higher survival rate than either parent (93%), while the F2 population had a survival rate slightly lower than the serpentine parent (76%).

Figure 5. Serpentine tolerance assay of parental lines and hybrid offspring for a *M. glaucescens* X *M. guttatus* hybrid cross. Individuals from the PMO serpentine population of *M. glaucescens* and the IMO non-serpentine population of *M. guttatus* were planted alongside individuals from the first generation (F1) of a hybrid cross between PMO and IMO parent populations, and 789 individuals from the second generation (F2) of that cross. All seeds were planted on serpentine soil from the PMO population collection site. After six weeks, the proportion of surviving plants (survival rate) was recorded for each population. Sample sizes (n) are listed above each population.

**M. pardalis** X *M. guttatus* hybrid tolerance assay

Seeds from both parental lines of an MCG X MED hybrid cross were planted, along with first-generation (F1) and second-generation (F2) hybrid seeds, in native MCG serpentine soil. Overall survival rates (Figure 6A) and survival rates of successful germinants (Figure 6B) were
recorded after 6 weeks. The supposedly non-tolerant MED population had a higher overall survival rate, at 66.7%, than the supposedly tolerant MCG parent population, which had 40% overall survival. However, survival rates after germination were higher in the tolerant MCG parent, at 86%, compared to 77% tolerance after germination for the MED parent. This reversal is due to low germination success of the MCG parent. Survival rates of the F1 hybrids were much lower than either parent, at 15%, but this was entirely due to failed germination, since all three germinated F1 plants survived. The F2 hybrids had a survival rate roughly halfway between that of the two parents, at 52.1% survival; survival after germination was almost identical to the tolerant parent, at 86%, demonstrating a low germination rate for the F2 offspring as well.

Since parental tolerance was different from what was expected, a modified replicate assay was conducted using the same parental lines and hybrid offspring, but allowing germination in pure water before individuals were transplanted to MCG serpentine soil. In this assay, surviving individuals were scored as necrotic or healthy (Figure 6C). The MED non-tolerant parental population had an overall survival rate of 47%, but all of those individuals were classified as necrotic. The MCG tolerant parental population had a 90% survival rate, with only 5% necrosis. First-generation (F1) hybrids had complete survival, but high rates of necrosis (90%). Second-generation hybrids showed an 83% survival rate, with 53% of plants classified as necrotic.
Figure 6. Two variations on a serpentine tolerance assay, testing parental lines and hybrid offspring for a *M. pardalis* X *M. guttatus* hybrid cross. In both assays, individuals from the MCG population of serpentine-endemic *M. pardalis* and the MED non-serpentine population of *M. guttatus* were planted alongside first- and second-generation (F1 and F2) hybrids from a cross between these lines. In the first assay (Panels A-B), seeds were planted directly on serpentine soil from the MCG population collection site, and scored as not germinated, dead, or survived. In the second assay (Panel C), seeds were allowed to germinate in water before being transplanted to MCG serpentine soil; surviving plants were scored as necrotic or healthy. Sample sizes (n) are listed above each population.

**QTL mapping in *M. nudatus* X *M. guttatus* hybrids**

Illumina sequencing-by-synthesis reads were obtained from a bulk sample of 100 second-generation GUN X IMO hybrid individuals which had exhibited tolerance to GUN native serpentine soil during tolerance assays. Raw reads were aligned to the IMO reference genome, and 2.9 million polymorphic sites were identified. Of those sites, 2.1 million were chosen with sufficient read depth and heterozygosity. The average read depth for chosen sites was 10.23.

Data was analyzed for 14 chromosomes (scaffolds 1-14, Figures 7A-B) and 3 other contiguous regions from the reference genome that have not been mapped to a chromosome, but which are at least 1 million base pairs in length (scaffolds 15-17, Figures 7C-D). The proportion of MCG parental alleles in the F2 reads was averaged over windows of 1 million base pairs, using 1000 randomly selected, overlapping windows for each chromosome (Figures 7A and 7C). A g-statistic was calculated for each polymorphic site, and averaged across each window using a weighted averaging technique (Figures 7B and 7D). With the false discovery rate set at 0.05, and the parameters of the g-distribution approximated empirically, a g-score threshold of 6.95 was obtained.

G-scores for windows across the entire genome ranged from 0.377 to 8.05, with only a single window falling above the significance threshold (blue arrow, Figure 7A). That window
lies on the tail end of chromosome 1, and its significant g-score is at the peak of a sharp upward trend in g-scores (blue circle, Figure 7B). The proportions of tolerant-parent alleles at and around this window are below 50%, spiking slightly downward near the significant value. Other similar upward or downward spikes, which result in high but non-significant g-scores, occur at the tail end of certain other chromosome regions. No other continuous strings of high tolerant-parent allele frequencies are evident.
Figure 7. Frequency maps and g-score distributions for *M. nudatus* X *M. guttatus* tolerant hybrids. The proportions of reads identified with the GUN parental genotype, averaged over overlapping windows of 1 million base pairs, are mapped across the length of each chromosome or scaffold from the IMO reference genome (Panels A, C). The red horizontal line in Panels A and C indicates the expected null frequency of 0.5. The weighted-average g-scores, corresponding to the same overlapping windows, are mapped across the length of each chromosome or scaffold (Panels B, D). The red horizontal line in Panels B and D indicates the threshold g-score for significance, according to the False Discovery Rate of 0.05. A blue arrow highlights the single window, in chromosome 1, with significant deviation from the expected distribution (Panel A). The g-score marking this window as significant is highlighted with a blue circle in Panel B.

*QTL mapping in M. pardalis X M. guttatus hybrids*

Illumina sequencing-by-synthesis reads were obtained from a bulk sample of 100 second-generation MCG X MED hybrid individuals which had exhibited tolerance to MCG native serpentine soil during tolerance assays. Reads were also obtained for a bulk sample of MCG parental line individuals. Raw reads were aligned to the IMO reference genome, and 10.4 million sites were identified as polymorphic. Of those sites, 3.89 million were chosen which were homozygous in the MCG parent, and for which there was sufficient heterozygosity and sufficient read depth in the F2 population. The average read depth in the F2 population for chosen sites was 33.2 reads, while the average read depth for the parental population was 3.44 reads.
Data was analyzed for 14 chromosomes (scaffolds 1-14, Figure 8A-B) and 3 other contiguous regions from the reference genome that have not been mapped to a chromosome, but which are at least 1 million base pairs in length (scaffolds 15-17, Figure 8C-D). The proportion of MCG parental alleles in the F2 reads was averaged over windows of 1 million base pairs, using 1000 randomly selected, overlapping windows for each chromosome (Figures 8A and 8C). A g-statistic was calculated for each polymorphic site and averaged across each window using a weighted averaging technique (Figures 8B and 8D). With the false discovery rate set at 0.05, and the parameters of the g-distribution approximated empirically, a g-score threshold of 164.02 was obtained.

The g-scores for windows across the genome ranged from 24.34 to 224.24, with 369 windows (2.2%) falling above the significance threshold. These windows occurred in four stretches of continuous significance. The first consisted of just three windows at the tail end of chromosome 2, in an area where the proportion of tolerant-parent alleles remains below 50% (Figure 8A, left arrow, and Figure 8B, left circle). The second is a long continuous stretch on chromosome 13, corresponding with a long continuous trough where the proportions of tolerant-parent alleles are well below 50% and approach 20% (Figure 8A, right arrow, and Figure 8B, right circle). The third significant stretch is a region at the tail end of scaffold 16, which has an unknown chromosomal location (Figure 8C, left arrow, and Figure 8D, left circle). The proportions of tolerant-parent alleles decrease to dip below 50% in this third region. In the final significant region, very little variation can be seen in the proportions of tolerant-parent alleles, which hover near 40% (Figure 8C, right arrow, and Figure 8D, right circle). This region is at the tail end of scaffold 17, another alignment of unknown chromosomal location. There are no
continuous stretches with higher-than-average proportions of tolerant-parent alleles that approach significant g-scores.
Figure 8. Frequency maps and g-score distributions for *M. pardalis* X *M. guttatus* hybrids. The proportions of reads identified with the MCG parental genotype, averaged over overlapping windows of 1 million base pairs, are mapped across the length of each chromosome or scaffold from the IMO reference genome (Panels A, C). The red horizontal line in Panels A and C indicates the expected null frequency of 0.5. The weighted-average g-scores, corresponding to the same overlapping windows, are mapped across the length of each chromosome or scaffold (Panels B, D). The red horizontal line in Panels B and D indicates the threshold g-score for significance, according to the False Discovery Rate of 0.05. Blue arrows highlight the windows, in chromosomes 2 and 13 as well as scaffolds 16 and 17, with significant deviation from the expected distribution (Panels A, C). The g-scores marking these windows as significant are highlighted with blue circles in Panels B and D.

Discussion

My results suggest the presence of multiple genetic mechanisms driving convergent adaptation in closely related species of *Mimulus*. The strong signal indicating a single-locus driver of tolerance in *Mimulus guttatus* populations was noticeably absent in *Mimulus nudatus*, as was any other strong signal, indicating a combination of multiple small effects rather than a single gene. *Mimulus glaucescens* also had a unique pattern of phenotype segregation in hybrid crosses, indicating a different mechanism of serpentine tolerance. These differences show that convergent evolution can take multiple genetic paths, even in closely related and ecologically similar species.
Serpentine soils are an equivalent, general selective environment for Mimulus

The reciprocal transplant experiments, which found the effects of soil type and population to be independent, suggest that the adaptation to serpentine soils in each of the four serpentine populations is a general rather than a local phenomenon. The populations are adapted to serpentine soils as a whole rather than to location-specific characteristics of each soil. Therefore, serpentine soil as a general category can be seen as a uniform selective environment, acting equivalently on each population. This equivalency of selective pressures is necessary in order to consider the populations convergent (Brandon, 1990).

No single large-effect gene is responsible for serpentine tolerance in Mimulus nudatus

In order to determine whether the genetic mechanisms of this convergent adaptation are parallel or unique, I used hybrid crosses of presumably tolerant and non-tolerant populations to generate hybrid generations with segregating tolerant genotypes. In the M. nudatus X M. guttatus hybrid cross, the M. nudatus tolerant parent (GUN) had a high survival rate on serpentine soil, while the M. guttatus non-tolerant parent (IMO) could not survive on serpentine, as expected. The second-generation (F2) hybrids had a survival rate approximately halfway between the two parents. This suggests that multiple genes may be involved in tolerance, since a single dominant or recessive gene would be expressed 75% or 25% of individuals. Many genes of small effect, instead, would average out to about 50% expression overall, resulting in the observed tolerance rate (Mackay, 2001).

The budding rate of the M. nudatus X M. guttatus F2 population, in contrast, is much lower than the halfway point between the two parents. If reduction in budding was entirely due to reduction in survival, I would expect a proportional reduction in budding and a final budding
rate of about 30%. The observed budding rate, however, is less than 10%. This indicates that there are other factors beyond survival that affect budding rates. It could be that, due to elevated stress levels, the plants arrest the production of buds, or delay the timing of bud production. Similar effects on bud production and timing have been observed in plant species such as cotton and sunflowers in response to stress conditions (Hocking and Steer, 1989; Ungar et al., 1992). The specifics of serpentine soil chemistry, such as nutrient limitations, could also reduce bud formation (Kazakou et al., 2008). These mechanisms would then be candidates for adaptation in serpentine populations.

The QTL mapping of M. nudatus X M. guttatus hybrids supports the hypothesis that many genes of small effect, rather than a single or a few key genes, contribute to serpentine tolerance. Only a single significant window, out of 17000 tested across the genome, passed the threshold for significance, and that window corresponded to a reduction, rather than an increase, in the tolerant parent alleles. No single stretch of the genome even appeared to suggest selection for M. nudatus alleles. This provides a contrast from the previous results in M. guttatus serpentine populations, which identified a single QTL of strong effect in chromosome 13 (Selby, 2014). There is no hint of such a signal in the M. nudatus results.

Sequence coverage, and therefore statistical power, was substantially lower in this mapping experiment relative to the M. pardalis experiment, as demonstrated by the lower average g-scores and the correspondingly lower threshold (Mackay, 2001; Magwene et al., 2011). This presents the possibility that strong QTL do exist for serpentine tolerance, but this study was unable to detect them. A possible reason for the poor coverage is difficulty in aligning the M. nudatus genome to the reference M. guttatus genome. Improved alignment methods accounting for lower homology levels should be used to increase coverage, and statistical power,
in future studies. In addition, individual sequencing markers can be used to confirm the absence of QTL in chromosome 13 or other regions of interest.

*Germination and phenotype scoring procedures affect tolerance assays in Mimulus pardalis*

The original tolerance assay results for the *M. pardalis* X *M. guttatus* cross were unexpected. The *M. guttatus* parent population MED, which was assumed to be non-tolerant, had a higher survival rate in serpentine soil than the presumably tolerant MCG population of *M. pardalis*. This could indicate that both populations are actually tolerant to serpentine soil, and that previous documentation of MED as a non-tolerant population is simply incorrect. In this scenario, the hybrid cross and QTL analysis of these two populations would no longer be a useful test of the genetic basis of serpentine tolerance.

Germination rates can be affected by extraneous factors such as the length of time a seed has been dried and dormant. Therefore, a more accurate depiction of serpentine tolerance should exclude the effects of poor germination. When germination effects are excluded, survival rates were higher in the tolerant MCG parent than the non-tolerant MED parent, as expected. However, MED survival rates were still high, suggesting a degree of tolerance in both parents.

In response to these results, a second assay was conducted, using more detailed criteria for tolerance. This assay found that, while the MED parental line was frequently able to survive in serpentine soil, it appeared unhealthy and necrotic. This assay confirms that the MED population is not completely serpentine-tolerant, but instead is able to persist in a crippled state.

*Hybrid incompatibility is a likely cause of a reverse QTL in Mimulus pardalis*

QTL mapping results in the *M. pardalis* X *M. guttatus* hybrid cross found significant deviations only in regions with reduced frequencies of the putative tolerant parent. Three of
those four regions are small spikes at the very tail end of chromosomes or scaffolds. This is a pattern also seen in the *M. nudatus* QTL mapping, and is probably an artifact of the experiment. Reduced recombination rates at the tails, or reduced SNP densities in gene-rich regions leading to increased overlap of SNP windows, may magnify random variation in allele frequencies disproportionately (Magwene et al., 2011). One significant region, however, has a more robust divergence than the other chromosome tail regions. This region, on chromosome 13, is in the same location as the QTL identified previously in a *M. guttatus* locally adapted serpentine population (Selby, 2014).

This QTL co-localization, in the direction of the MED population, raises the possibility that the *M. guttatus* parental line used in this experiment harbors the same tolerant allele found in serpentine *M. guttatus* populations. My original MCG X MED tolerance assay supports this possibility, since the MED seeds appeared tolerant. The source of this tolerant allele might be standing variation for that tolerant allele in the *M. guttatus* overall population. Standing variation refers to the presence of a beneficial allele at low levels in an ancestral population, which can rise to fixation in multiple lineages when selection begins to act (Rosenblum et al., 2014).

Standing variation has been implicated as a potential driver of parallel evolution (Rosenblum et al., 2014). Another possibility is that the close proximity of the MED site to serpentine sites such as MCG could have driven selection for tolerance in the MED population, since some individuals may disperse to and from serpentine patches.

However, the more detailed results of the second MCG X MED tolerance assay appear to refute these scenarios, on the grounds that the MED parental line was not truly serpentine-tolerant, at least to the same degree as other locally adapted serpentine populations of *M. guttatus*. Instead, a gene or genes in the relevant region of chromosome 13 might be involved in
hybrid incompatibility. First-generation hybrids had a low germination rate, and seed inviability is a common result of hybrid incompatibility in *Mimulus* species as well as other plants (Martin and Willis, 2010). If an allele in chromosome 13 of *M. pardalis* was detrimental in the *M. guttatus* genomic background, selection to eliminate this allele would result in the observed QTL. A similar hybrid incompatibility locus acting between *Mimulus guttatus* and *Mimulus nasutus* has been identified in chromosome 13 (Sweigart and Flagel, 2015), increasing the likelihood of this explanation.

Robust dominance of serpentine tolerance characterizes *Mimulus glaucescens*

Both the first and second generation hybrid offspring from a cross of *M. glaucescens* with non-tolerant *M. guttatus* have persistent tolerance, at approximately the same levels as the *M. glaucescens* tolerant parent. Tolerance in the first generation shows that serpentine tolerance is a dominant trait in *M. glaucescens*, since each offspring should be heterozygous throughout the genome. If tolerance was controlled by a single dominant trait, then simple Mendelian genetics predicts a 75% survival in the second generation (Mackay, 2001). Since tolerance is substantially higher than this, there may be multiple redundant genes contributing to tolerance in *Mimulus glaucescens*. These results stand in contrast to the single-gene tolerance model of *M. guttatus*. They are also strikingly different from the results for segregation in *Mimulus nudatus*, which had tolerance halfway between the two parents. Therefore, my results indicate that a different combination of gene inputs is controlling serpentine tolerance in *M. glaucescens* relative to *M. guttatus* and to *M. nudatus*. In the future, QTL mapping of *M. glaucescens* X *M. guttatus* hybrids will test the validity of this hypothesis.
Conclusions

Between species of *Mimulus* monkey flowers, I have demonstrated that at least two, and likely three, very different mechanisms of serpentine adaptation are at work. One, previously identified in *Mimulus guttatus*, involves a single gene of large effect. The second, which I found in *Mimulus nudatus*, implicates many genes of smaller effect. The third potential mechanism, in *Mimulus glaucescens*, shows signs of redundancy in multiple genes. The fact that these very disparate mechanisms have arisen in response to the same selective pressures in very closely related species is striking. One possible reason for this difference is that serpentine soil is a complex set of multiple interacting pressures, including heavy metals and low nutrient concentrations. It may be that very few, or no, single genomic changes can convey tolerance to all of these selective pressures. For this reason, even the species that have a single QTL of large effect probably also have many genes of smaller effect contributing to refine their adaptation.

The presence of a single large-effect gene in some populations, but not others, indicates a genetic constraint, in that not every population was able to develop and utilize this beneficial variant. However, the evolution of tolerance despite lack of this strongly beneficial allele in some populations demonstrates the power of natural selection to overcome this constraint and utilize alternative genetic mechanisms.

In the future, the specific genes and mutations underlying the QTL on chromosome 13 should be identified, through fine-mapping techniques, haplotype analysis, and protein characterization. Then, sequencing can be used to confirm the presence or absence of these mutations in different populations of *Mimulus*, expanding even beyond the populations used here. With more data points, we can assess how prevalent this single locus is across species, and look for patterns of pre-adaptation or standing variation. Furthermore, knowing the mutations
allows us to assess the cellular mechanism by which this locus confers adaptation. Experimental manipulations of this locus can be used to produce serpentine adaptation in previously non-adapted populations, confirming its role in tolerance. This may have implications for bioremediation and our understanding of plant-soil interactions. Finally, my study can be seen as a preliminary launching point for a widespread and systematic effort to assess serpentine adaptation across *Mimulus* as well as other California flower species, producing a large dataset for answering questions about parallel and convergent evolution.

**Acknowledgements**

I am indebted to Annie Jeong for her kind mentoring and support, her willingness to drop everything to help me, and her enthusiasm for the science. In addition, Annie was invaluable with troubleshooting and brainstorming, and she contributed her own hand to some of the described experiments. I would like to thank Dr. Willis and the entire Willis Lab for theoretical, technical, and moral support, in addition to infrastructure and financial backing. In particular, Jenn Coughlan contributed *M. glaucescens* work and helped me throughout the process. Many thanks to the plethora of people who have given me advice and comments on the writing and revision of this thesis: my biology faculty reader, Dr. Emily Bernhardt; Dr. Alexander Motten and Dr. Elizabeth Pryer in the thesis writing class; and my thesis writing peers both within and outside of the Biology department. I also thank all the friends and family who have listened to me and encouraged me throughout this process. The Duke Undergraduate Research Support Office provided funding through a Dean’s Summer Research Fellowship, which contributed to my work.
Literature Cited


