UNIVERSITY OF CALIFORNIA

SANTA CRUZ

GENETIC STRUCTURE AND HYBRIDIZATION IN TWO RARE SERPENTINE *MONARDELLA* (LAMIACEAE)

A thesis submitted in partial satisfaction
of the requirements for the degree of

MASTER OF ARTS

in

ECOLOGY AND EVOLUTIONARY BIOLOGY

by

Brett Smith

December 2014

The thesis of Brett Smith
is approved:

_____
Assistant Professor Kathleen Kay, Chair

_____
Professor Giacomo Bernardi

_____
Professor Grant Pogson

_____
Tyrus Miller
Vice Provost and Dean of Graduate Studies

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES	*iv*
ABSTRACT	*v*
ACKNOWLEDGMENTS	*vii*
INTRODUCTION	*1*
MATERIALS AND METHODS	*12*
RESULTS	*22*
DISCUSSION	*41*
REFERENCES	*58*

LIST OF FIGURES AND TABLES

Figure 1: *Photos of* Monardella 9

Figure 2: *Map of Study Sites* 12

Figure 3: *Structure bar plot of* M. stebbinsii *and* M. follettii 26

Figure 4: *Mantel test for isolation by distance in* M. follettii 29

Figure 5: *Principal coordinates analysis of* M. stebbinsii *and* M. follettii 32

Figure 6: *Structure bar plot of all* Monardella *individuals* 34

Figure 7: *Maximum likelihood hybrid scores for putative hybrids* 37

Table 1: Monardella *conservation status* 7

Table 2: Monardella *collection locations* 14

Table 3: *Summary statistics for* M. follettii *and* M. stebbinsii 24

Table 4: *Analysis of molecular variance for* M. follettii 28

Table 5: *Analysis of molecular variance for* M. stebbinsii 30

Table 6: *Soil variable loadings on principal components* 39

Table 7: *SNPs significantly correlated with PC5* 41

ABSTRACT

Brett Smith

Genetic structure and hybridization in two rare serpentine *Monardella* (Lamiaceae)

Small populations and rare species offer unique opportunities to study fundamental evolutionary questions, but many rare species are threatened by disturbance and climate change. Molecular population genetics enable biologists to examine evolutionary processes while simultaneously assessing levels of genetic diversity, population structure, and gene flow that can help shape management plans for rare species. However, developing molecular markers can be an expensive, time-consuming process, especially if little is known about the genomes of the species of interest. A new technique, Genotyping by Sequencing (GBS), requires little to no prior information to develop SNP markers. Here I employ this technology to develop hundreds of markers in two rare, serpentine soil-endemic *Monardella* (Lamiaceae) plant species and a common congener. Using the SNP markers, I investigate population structure, genetic diversity, and hybridization between species. I also use a soil dataset to determine whether species and hybrid zones occur on divergent soils. I find low levels of genetic diversity and little population structure in the two rare species, as expected by rapid genetic drift in small populations. I find evidence of hybridization and introgression among species at sites where multiple species co-occur. Further, Bayesian assignment finds

mixed ancestry in one of the rare species. The soil data show that the soils inhabited by the two species are divergent, but not significantly different. Some hybrids occur on soils that seem to be intermediate between the two parental species, but others do not. I synthesize these data with ecological surveys to provide species management recommendations to the USDA Forest Service.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Kathleen Kay for her endless support, encouragement, and patience over the last three years. I am fortunate to have been her student in seminars, meetings, the lab, the field, and the classroom, where her endless knowledge and leadership through action never ceased to amaze. I would also like to thank my thesis committee members, Dr. Giacomo Bernardi and Dr. Grant Pogson, for their support and invaluable contributions to this thesis. I have a much more complete comprehension of my results after my conversations with them. I am thankful for the guidance of Yann Surget-Groba, whose work in the Kay Lab helped pave the way for my own. I would like to thank Suzie Woolhouse for allowing me to use her soil data, and for showing me the ropes in Plumas. I am thankful for my labmates, Jenn Yost, Megan Peterson, Tim Miller, Shelley Sianta, and Julie Herman, who helped me over the years. They made the good times better, and made sure the bad times were brief. I am grateful for Norah Saarman, who is my Python buddy, and who was always generous with her time and mental energy in solving my problems. I would not have been able to begin collecting plants without the support of Jim Belsher-Howe, and Michelle Coppoletta of Plumas National Forest, who were great advocates for this work. I would also like to thank my girlfriend Mary Zúñiga for her truly endless supply of love and support. Likewise, I am grateful for my

family and friends who ensured that I remained healthy and mentally strong through this project.

I was financially supported through many organizations during my time at UC Santa Cruz. The majority of the work was supported through a joint agreement with Plumas National Forest – Mt. Hough Ranger district, and I am very grateful for their support. The Ecology and Evolutionary Biology department at UCSC gave me four quarters of TA funding and three summers of funding, for which I am very thankful. Additional support was provided by Northern California Botanists and the Santa Clara Valley chapter of the California Native Plant Society, and I am thankful they believed in the value and importance of this work.

INTRODUCTION

Variation within and among populations affects demographic processes and governs how populations respond to environmental changes through time. Modern biologists are equipped with tools to measure variation in the physiology, morphology, behavior, and genetics of populations and how this variation might allow organisms to exploit new resources, remain resilient to changing environments and new diseases, and compete with invaders. Population genetic markers can be used to measure neutral, functional, and potentially adaptive variation at the population level.

A population's persistence through time is determined by its ability to adapt and compete in a changing competitive and abiotic environment. Genetic diversity in populations allows for resilience against disease and parasites, physiological stress, and environmental change (Ellstrand and Elam 1993), and provides the raw material on which selection can act. Populations with low genetic diversity may become disadvantaged, because they lack the variation to survive a selective event. In our current era of rapid environmental change and increased disturbance, populations lacking genetic diversity could quickly become extinct through strong, fluctuating selection.

Small and patchily distributed populations are especially susceptible to a number of genetic diversity-lowering phenomena including drift, inbreeding, and low levels of gene flow. In small populations, drift acts rapidly to randomly

fix alleles due to chance. Importantly, drift can cause the stochastic fixation of deleterious alleles in a small population. Further, small populations of sexually reproducing individuals may be subject to inbreeding, the increase of homozygosity through nonrandom mating. Drift and inbreeding may be alleviated through gene flow, but patchily distributed populations may experience insufficient levels of migration to combat low genetic diversity. Each of these phenomena may lower a species' genetic diversity, and can subsequently erode a species' future potential for adaptation, a necessity for the preservation of a species (Soulé 1980, Moritz 2002).

Generally, small populations are expected to have low genetic diversity, which can directly lower fitness. Often the result of rapid reductions in population size (i.e. bottlenecks) or colonization events, small populations are susceptible to rapid genetic drift. In plants with self-incompatible mating systems (SI), low diversity of SI alleles can result in lower availability of compatible mates and hasten extinction in small populations. Much empirical work on small populations has supported this theory (e.g. Byers and Meagher 1991, Young and Pickup 2010), but some invasive species are successful despite reduced levels of genetic diversity (Tsutsui et al 2000, Amsellem et al 2000).

Deviations from predicted levels of genetic diversity in small populations can occur due to high levels of gene flow from other populations.

Recently-bottlenecked populations may also exhibit higher levels of genetic diversity than expected from their size alone, as alleles have not yet randomly been fixed under genetic drift. For example, Luan and colleagues found high genetic diversity in small ($n < 50$) populations of the threatened *Nouelia insignis*, but these populations were only recently fragmented, perhaps not providing enough time for genetic drift to significantly influence genetic diversity (2006).

In small populations with limited mate availability, inbreeding is inevitable and can result increased homozygosity through nonrandom mating. Inbreeding has long been shown to decrease fitness and increase the chance of extinction in experimental (Frankham 1995), natural (Saccheri et al 1998, Crnokrak and Roff 1999) and simulated populations (O'Grady et al 2006). However purging, the process by which deleterious alleles are quickly eliminated by natural selection in populations undergoing inbreeding, may confer some benefits to inbred populations. Evidence of purging is inconsistent across studies, but some have demonstrated increases in mean population fitness through purging in fruit flies (Frankham et al 2001) and others (reviewed Crnokrak and Barrett 2002). Eventually, in small populations without gene flow, any gains in fitness from purging will be diminished by the loss of genetic diversity through genetic drift and repeated inbreeding.

Gene flow in small populations can augment genetic diversity and ameliorate inbreeding depression, but can erode local adaptation and homogenize divergent populations. Nearby populations of a species may exchange genetic material at some rate dependent upon dispersal distances, proximity in space, environmental conditions, and interspecific interactions such as pollinator and seed disperser availability in plants. Genetic rescue, the corresponding increase in fitness after an introduction of genetic diversity through gene flow, can enable lasting persistence in populations suffering from inbreeding depression and low genetic diversity. Genetic rescue has been seen in experimental populations and used as an effective management tool for rare species (Sexton et al 2011, Willi et al 2007). Species management through artificial gene flow should be carefully planned and thoroughly informed by data, as gene flow may also result in outbreeding depression. Outbreeding depression is the lowering of fitness correlated with the loss of local adaptation in divergent populations, and it can precipitate the extinction of a species (Greig 1979, Ellstrand and Elam 1993).

Interspecific hybridization and subsequent introgression can rapidly influence small populations and rare species. Introgression has been found to be beneficial and maintain genetic diversity for disease resistance and stress tolerance in otherwise perpetually inbreeding species (Ingvarsson and Whitlock 2000, Ebert et al 2002). In contrast, introgression has also been

shown to create less fit offspring in locally adapted populations (Keller et al 2000). In the face of hybridization and introgression, small populations and rare species may face competition from hybrids and their common progenitors, diminished reproductive barriers among populations, and complete assimilation of a rare species into a more common congener through gene swamping (Levin et al 1996, Rieseberg and Swensen 1996, Kleindorfer et al 2014). However, even when hybridization is successful and the individual is better adapted than its sympatric progenitor, a lack of niche space and genetic barriers to backcrossing may prevent introgression and hybrid establishment in a population (Grant 1981). Instead, introgression is more likely where niche space is available, such as areas of anthropogenic disturbance (Grant 1981, Lamont et al 2003). Hybridization and introgression can be especially problematic in legally protected species, where taxonomic uncertainty can severely complicate management strategies (Rieseberg 1991, Allendorf et al 2001). The legal protection of hybrids in the United States has been fiercely debated since the passage of the Endangered Species Act in 1973, and no official policy is codified to this day (Ellstrand et al 2010).

Each population of each species may be under a number of unseen selective forces and demographic events that can influence genetic diversity, inbreeding, and gene flow. Genetic diversity, inbreeding, and gene flow are important to understanding population divergence, speciation, and evolution,

but quantification of past and present levels can prove difficult. Though quantification of adaptive variation might be of most interest to a biologist, most molecular markers only provide data from neutral genetic variation in the target individuals. Neutral genetic variation accumulates through time due to neutral mutation throughout the genome, and by definition, it is not under selection. Because of their slow buildup through time, these neutral changes can be especially useful in investigating genetic structure and gene flow among populations and species (Setoguchi et al 2010). Further, neutral genetic variation can reflect genetic diversity and inbreeding within a population (Freeland et al 2010). Molecular population genetic analyses can approximate values of genetic diversity, inbreeding, gene flow, and hybridization using neutral markers from the genome of the populations and species of interest. Such tools allow researchers to illuminate the evolutionary histories and population dynamics of species for which these data are not already available and may be otherwise unattainable.

One such group of species, the plant genus *Monardella* (Lamiaceae), has a poorly understood evolutionary history and has been largely ignored by scientists who might have performed more classical analyses (e.g. morphometric systematics or crossing studies) of the genus. A Web of Science search on September 10, 2014 for “Monardella” yielded only 9 results. *Monardella* has over 30 described species, all in western North

America. In Plumas County, California, there are two serpentine soil endemics that are restricted to limited habitat within Plumas National Forest, *M. stebbinsii* and *M. follettii*.

As serpentine soil-endemics, *M. stebbinsii* and *M. follettii* are restricted to the patchily-distributed, limited habitat entirely within the bounds of Plumas and Lassen National Forests. This equates to about 15 populations and <1,500 individuals of *M. stebbinsii* and 25 populations and 5,000 – 10,000 individuals of *M. follettii* (CNPS 2014). The Forest Service and conservation organizations list both species on their conservation lists (Table 1), and both species face disturbance from logging and erosion. Though both species inhabit serpentine soil, *M. stebbinsii* tends to occur on steep scree slopes with open canopies and very thin, serpentinite-derived soils, whereas *M. follettii* typically occurs on less extreme slopes of peridotite-derived soil (Coppoletta and Woolhouse 2010). In general serpentine soils are shallow, retain little water, and are characterized by low Ca:Mg ratios, high levels of toxic metals, and low concentrations of essential plant nutrients (Brady et al 2005). Despite the harsh conditions, serpentine soil adaptation is widespread in plants, and can contribute to speciation (Baldwin 2005, reviewed in Kay et al 2011).

Table 1. *Conservation Status*

	Forest Service	NatureServe	CNPS
M. follettii	Critically imperiled	G1	1B.2
M. stebbinsii	Critically imperiled	G2	1B.2

Beyond their basic ecology and life history, not much is known about the rare *Monardella* or the genus at large, but some have grouped species into alliances based on morphology and geographic distribution. Elvin and Sanders (2009) place *M. follettii* in the Odoratissimae alliance, in which species share glabrous (smooth) leaves and suffrutescent habit (erect stems woody near the base and herbaceous at the top). Elvin and Sanders (2009) fit *M. stebbinsii* in the Australae alliance, a group of taxa that they posit to be relictual mountaintop dwellers that share a unique morphology despite their allopatric distribution. Others have argued that *M. stebbinsii* is not closely related to any other member of the genus (Hardham and Bartel 1990). There is also one widespread congener, *M. sheltonii*, that occurs prolifically throughout Plumas and Lassen National Forests, occasionally sympatrically on serpentine soils with the rare species. Elvin and Sanders (2009) assign *M. sheltonii* to the Villosae alliance on the basis of its wide distribution in Western North America.

Mechanisms of reproductive isolation in *Monardella* are unclear, and morphologically intermediate individuals among taxa appear to be common. In a description of the natural history of *M. follettii* and *M. stebbinsii*, researchers hypothesized that the only reproductive isolation between species results from spatial isolation and temporal differences in phenology (Coppoletta and Woolhouse 2010). However, these conclusions result from

limited pollinator observations and are untested in controlled experiments. At sympatric sites the suite of pollinators visiting the flowers likely carry pollen between heterospecific individuals. Furthermore, flower morphology across species is nearly identical (Figure 1), suggesting little divergence in pollinator specialization. Indeed, for as long as these species have been described, it has been suggested that hybridization and introgression in this area is common (reviewed in Hardham and Bartel 1990).

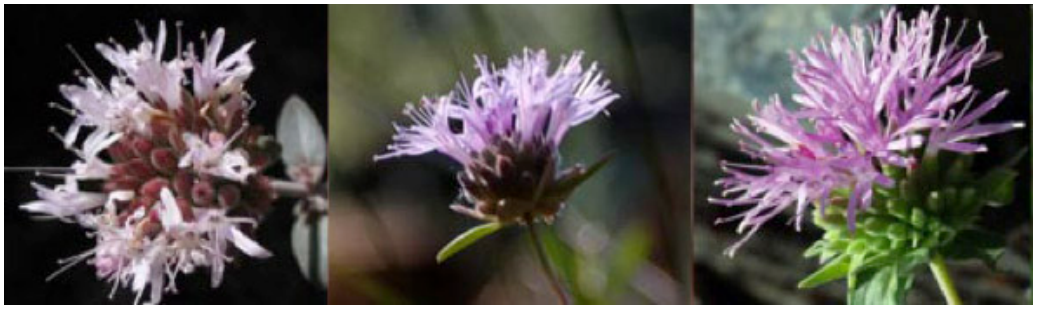

Figure 1. *From left to right, the* Monardella *flowers of 3 species in Plumas National Forest:* M. stebbinsii, M. follettii, *and* M. sheltonii (photo from Barry Breckling).

In the summer of 2000 much of the serpentine habitat in Plumas National Forest burned in the Storrie Fire, a blaze started by a railroad worker for Union Pacific. After a settlement with Union Pacific, the Forest Service authorized an assessment of the rarest plants in the areas affected by the fire. Although only some populations burned, the rare *Monardella* in Plumas National Forest are of concern to managers due to their small numbers and patchily distributed population. Small population sizes, patchy distribution, a

lack of knowledge of the genus, and potential hybridization in the rare species make them prime candidates for a conservation genetics analysis. The genetic diversity, genetic structure, and hybridization data derived from such an analysis would equip land managers to make informed conservation decisions.

The general goals of a conservation genetic analysis are to quantify levels of genetic diversity, resolve genetic structure among populations, and to detect inbred populations. In some cases, a researcher might want to determine the extent of hybridization and introgression in populations of rare species. With these data, managers can supplement genetically depauperate or inbred populations, preserve natural genetic structure among divergent healthy populations, and prevent introgression or assimilation in rare species.

In this framework, a number of predictions and hypothetical management responses can be drawn for different demographic and genetic scenarios in the rare *Monardella* of Plumas National Forest. In the simplest scenario, all species are reproductively isolated, population census size is directly related to genetic diversity but inversely related to the extent of inbreeding of a population, and populations are isolated relative to distance among them. In this scenario, managers would use nearby genetically diverse populations to augment conspecific genetically depauperate populations. This scenario could be complicated by a recent bottleneck or source-sink gene

flow in which small populations have a large amount of genetic diversity despite their small census size. In this case, managers would closely monitor the populations, and potentially reexamine genetic diversity in the future. Hybridization may occur in sympatric populations, and introgression into the rare species may occur in subsequent generations. If introgression occurs, managers might remove the *M. sheltonii* from sympatric sites to prevent assimilation. However if only F1 hybrids are present, a management response might not be necessary. Finally, gene flow among populations and species might correlate with the soil attributes in that only pollen from individuals on a similar soil type would produce viable seeds. In this case, managers could remove *M. sheltonii* from the area of sympatry, or could simply classify other populations as higher conservation priorities.

Using a Genotyping by Sequencing approach, I completed a conservation genetics analysis of *Monardella* in Plumas National Forest to test these predictions. I use these data to inform land managers of the genetic structure of the populations and species in the area, extent of inbreeding and levels of genetic diversity in each sampled population, and estimates of intra- and interspecific gene flow within and between populations. These data can be synthesized with population demographics into well-informed management plans for the focal species.

Methods

Sampling

I extensively sampled populations throughout the range of *M. follettii* and *M. stebbinsii* including numerous putative hybrid sites at which more than one species of *Monardella* was present (Table 2, Figure 2). I chose sites to match a previous ecological and demographic assessment (Coppoletta and Woolhouse 2010). At each of six sites for *M. folletii* and four sites of *M. stebbinsii*, I sampled 20-30 individuals (or a smaller number that represents every individual in the population). Any putative hybrids were labeled as such. Additionally, I sampled individuals from two sites of putatively pure *M. sheltonii*, based on geographic distance from other documented populations and whether individuals exhibited morphology representative of the published species description. To characterize the soils at the conservation sites, I used a soil dataset from Woolhouse (2012) in which she sampled from 5 locations near *Monardella* plants within each site (Table 2). At hybrid sites, I took soil samples from the rhizosphere of several individuals within a site, taking care to sample near individuals of parental species as well as putative hybrids. In the lab I dried the soil and removed large rocks from the samples before sending them for analysis.

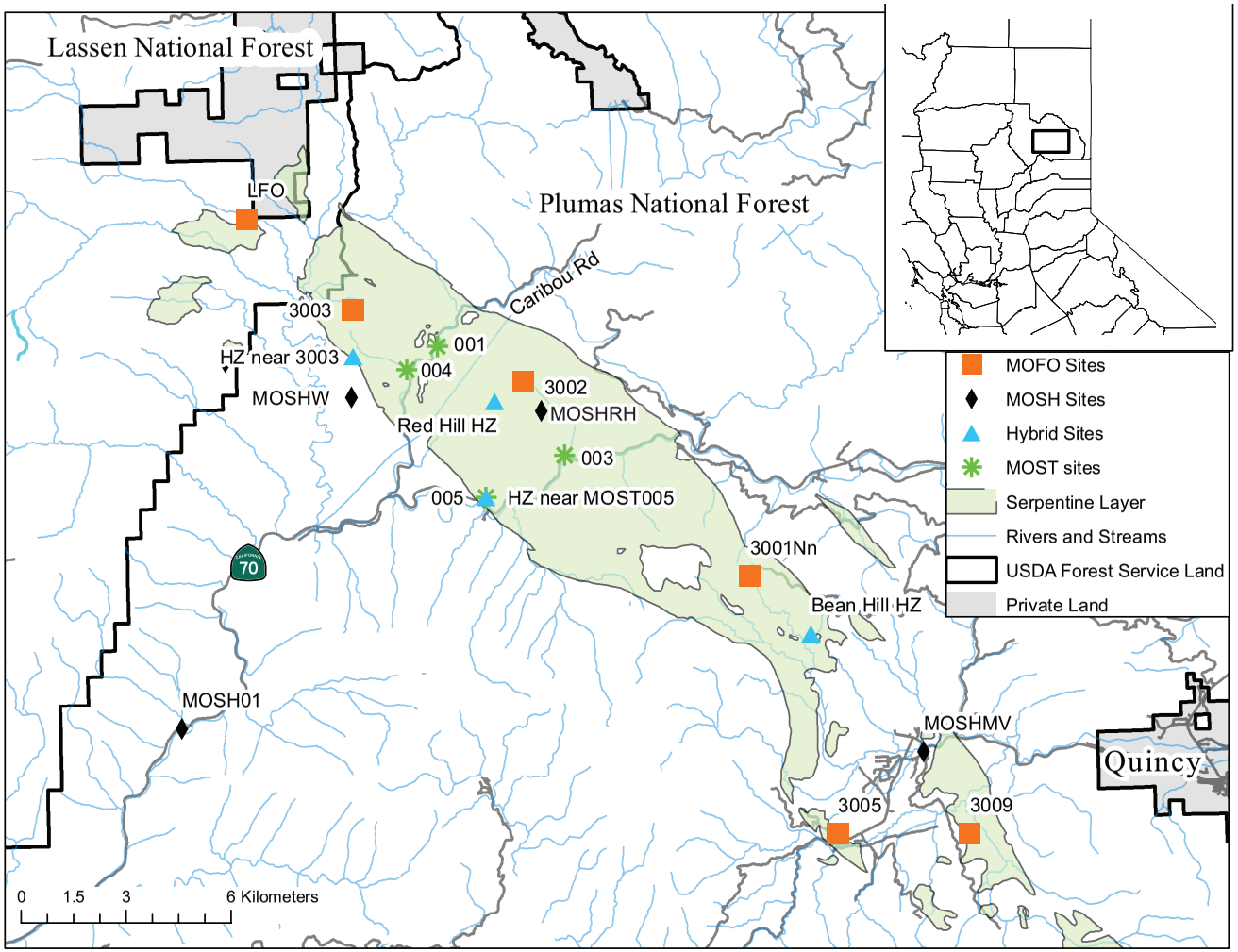

Figure 2. *Locations of sampled* Monardella *occurrences in Plumas*

Table 2. Monardella *collection locations*

Occurrence Name[a]	Species	Easting[b]	Northing[b]	Altitude (m)	Soil source[d]	UCSC Herbarium accession #[e]
LFO	*M. follettii*	646928	4438715	1347	N/A	8318
MOFO 3009	*M. follettii*	668044	4421220	1219	W	8316
MOFO 3005	*M. follettii*	664281	4421210	1158	W	JEPS63417[r]
MOFO 3003	*M. follettii*	650425	4434853	1463	W	8310
MOFO 3002	*M. follettii*	655298	4434138	1767	W	CAS886559[g]
MOFO 3001 Nn	*M. follettii*	661756	4428572	1584	W	8312
Bean Hill hybrid	*M. follettii & M. sheltonii*	663512	4426929	1432	S	8313
Hybrid zone near MOFO 3003	*M. follettii & M. sheltonii*	650418	4436166	1472	S	8317
MOSH W	*M. sheltonii*	650384	4433674	1340	N/A	[e]
MOSH 01	*M. sheltonii*	645534	4424207	623	S	CHSC52052[h]
MOSH MV	*M. sheltonii*	666730	4423554	1103	S	8320
MOSH RH[c]	*M. sheltonii*	657923	4434171	1308	N/A	[e]
MOST 005	*M. stebbinsii*	654226	4430827	792	W	8311
MOST 004	*M. stebbinsii*	651964	4434468	822	W	CHSC34000[h]
MOST 003	*M. stebbinsii*	656475	4432021	853	W	8314
MOST 001	*M. stebbinsii*	652839	4435138	762	W	[e]
Red Hill hybrid zone	*M. stebbinsii & M. follettii*	654460	4433568	1828	S & W	N/A
Hybrid Zone near MOST 005	*M. stebbinsii & M. sheltonii*	654226	4430827	792	S	N/A

[a] *Occurrence names correspond to Coppoleta and Woolhouse (2010).*
[b] *NAD 1983 UTM 10S*
[c] *Approximate location*
[d] *Soil data from Woolhouse 2012 (W) and Smith 2014 (S)*
[e] *Because of very small population sizes, herbarium specimens were not taken from some populations. If possible, existing accessions are listed for these populations; see below.*
[f] *Jepson Herbarium, University of California, Berkeley. Berkeley, California.*
[g] *California Academy of Sciences. San Francisco, California.*
[h] *California State University, Chico. Chico, California*

Genetic Marker Discovery

I extracted genomic DNA from the leaves or flower buds of each individual using a modified CTAB protocol (Doyle and Doyle 1987). I tested genomic DNA purity on a NanoDrop (Thermo Fisher Scientific, Wilmington, Delaware), ensured a lack of degradation using agarose gel electrophoresis, and quantified DNA concentrations using a Qubit (Invitrogen, Carlsbad, California). I experienced considerable difficulty extracting DNA from *Monardella* tissue, likely caused by the high concentration of terpenoids and other secondary chemicals in the tissue. I suggest future researchers avoid using kits (e.g. Qiagen) or other faster methods of DNA extraction, as these were all unsuccessful even with extensive troubleshooting. Even with my refined protocol, some extractions still produced sub-par DNA, and therefore I genotyped 20 individuals per occurrence. There seemed to be no pattern to the individuals that produced subpar DNA extractions, and the 20 individuals chosen for analysis were selected for their high quality DNA. I ran simulations in SPOTG, a conservation genetics planning tool (Hoban et al 2013, Excoffier and Lischer 2010, Laval and Excoffier 2004), using the actual marker numbers, individual counts, and population numbers recovered as described in the results. The simulations suggest I had sufficient sampling for my analyses.

I sent samples to the Institute for Genomic Diversity (IGD) at Cornell University (Ithaca, New York) for library construction using a genotyping by sequencing (GBS) protocol (Elshire et al 2011) and sequencing using an Illumina HiSeq platform (San Diego, California). The GBS method can discover thousands of genome-wide loci in non-model organisms, and is the simplest of the reduced representation library methods developed thus far (Davey et al 2011). GBS has been quickly adopted, and has shown power to resolve phylogenies and examine genetic diversity in non-model organisms (Lu et al 2013a, White et al 2013). Compared to microsatellites, GBS does not require costly development and testing, and it provides hundreds to thousands of genome-wide markers. Microsatellites are also typically not transferable across species, and the putative *Monardella* hybrid zones required homologous markers shared across these species. AFLPs, although common markers in conservation genetics, are anonymous, dominant presence-absence markers that are less informative than the nucleotide sequence data provided by GBS.

At the IGD, genomic DNA samples were digested with *Pst*I, barcoded adapters were ligated to each sample, and samples were pooled and cleaned up in a size-exclusion column before amplifying the library via PCR and sequencing the library (see Elshire et al 2011 for a full protocol). Ninety-five samples plus one negative control were multiplexed in each sequencing lane,

for a total of 285 samples. These samples comprised 6 populations of 20 individuals of *M. folletii*, 4 populations of 20 individuals of *M. stebbinsii*, 22 individuals of *M. sheltonii* across populations, and 63 putative hybrids.

I used the TASSEL/UNEAK bioinformatics pipeline to generate bi-allelic SNP calls from the raw sequence data (Lu et al 2013a). In brief, the Universal Network Enabled Analysis Kit (UNEAK) pipeline sorts raw data into files for each individual in the library, trims the sequences to 64 bp, compiles exactly matching reads as tags, pairwise aligns sequences to find tags differing by only 1 bp, creates networks of these nearly matching tags, and filters networks that are too complex (Lu et al 2013a). Tags that pass through the pipeline are output in HapMap files for further analysis. I employed strict filtering parameters on sequence quality and a low but acceptable minimum coverage threshold of 3 to call a SNP (Lu et al 2013b). I ran the analysis once for the entire dataset with all individuals, once using only *M. stebbinsii* individuals, and once using only *M. follettii* individuals. Because of variation in the distribution of GBS target sequences, only a subset of loci are shared among species. Thus I recovered larger datasets of markers for my single species analyses and a smaller shared dataset for the three species combined. For the all-species dataset, I removed any locus that was not sequenced in at least 90% of individuals and any individual missing more than 20% of the data. For the individual-species dataset, I removed any locus that

was not sequenced in at least 80% of individuals and any individuals missing more than 20% of the data. I tried several alternative values for these filtering cutoffs. However, my results were not qualitatively sensitive to these changes, therefore I only report results for these conservative filtering levels.

Genetic analysis

There are several commonly used methods for detecting genetic structure among populations, and each can detect different patterns. First I used a Bayesian assignment analysis, implemented in the software STRUCTURE (Pritchard et al 2000), to identify genetic clusters within my dataset and to infer the genomic composition of each individual in terms of genetic clusters. I ran STRUCTURE three times, once for the overall dataset including all individuals, and once each for the *M. follettii* and *M. stebbinsii* datasets. I ran the simulations using the admixture model with 50,000 burn-in steps followed by 100,000 steps. I estimated the hyperparameter λ for each dataset before running the simulations, and subsequently fixed it at the estimated value, as suggested for SNP datasets by Pritchard et al (2000). For the overall dataset, I identified the most likely number of clusters (*K*) for a prior defined range of 1-14 to cover the total number of populations sampled. For the *M. follettii* and *M. stebbinsii* datasets, I used prior defined ranges of 1-6 and 1-4, respectively. I determined the most likely number of genetic clusters by looking at the rate of change in the probability of successive numbers of

clusters (Evanno et al 2005) as implemented in STRUCTURE HARVESTER (Earl and Von Holdt 2012).

To further examine genetic structure within the species I analyzed the individual species datasets using other common metrics implemented in the software GenAlEx 6.5 (Peakall and Smouse 2012). I ran locus-by-locus Analysis of Molecular Variance (AMOVA) (Excoffier et al 1992) using the codominant allelic input with 9999 permutations, and calculated pairwise F-statistics with 999 permutations for significance testing (Wright 1969). I corrected p-values for F-statistics using a Holm-Bonferroni adjustment (Holm 1979). Missing data were interpolated to avoid biased sources of variation, and an AMOVA run in ARLEQUIN showed the same partitioning of genetic variance. I calculated genetic distances (Peakall et al 1995), which are not directly reported but were used for a Principal Coordinates Analysis (PCoA). A PCoA examines a dissimilarity matrix to find the major axes of variation in a dataset (Orlóci 1978). Using the F-statistics derived from the AMOVA, I tested for Isolation by Distance using a paired Mantel Test with 9999 permutations as implemented in GenAlEx.

In order to understand genetic diversity within species, I also used GenAlEx to calculate summary statistics for the two rare species datasets. The statistics include the number of private alleles, private allele frequency, expected heterozygosity, and observed heterozygosity. These summary

statistics are derived per individual and averaged across all loci. Private alleles are alleles that are only present in one occurrence or group, and inform my understanding of how unique and diverse each occurrence is. Of most interest here are the values of expected heterozygosity (H_E), which is a measure of the genetic variation of an occurrence.

To test the extent of hybridization and introgression in the putative hybrid zones, I used the software HINDEX (Buerkle 2005) as implemented in the R (R Core Team 2014) package INTROGRESS (Gompert and Buerkle 2010). I used subsets of the all-species SNP dataset. Using the STRUCTURE results as a guide, I chose putatively pure sites for each species to input as the parental individuals for each analysis. After examination of the STRUCTURE results, only three of the four putative hybrid zones (Bean Hill, Red Hill, and the HZ near MOFO 3003) appeared to contain hybrids. For the analysis of the Red Hill zone, I used all putatively "pure" *M. stebbinsii* and *M. follettii* individuals as parental populations. For the Bean Hill and HZ near 3003, I combined all *M. sheltonii* individuals into one parental population due its limited representation in the dataset. In the Bean Hill and HZ near 3003 analyses, I only used individuals from FO3001Nn as the parental population for *M. follettii* to match the lower sample size in *M. sheltonii*.

Soil Analysis

I sent soil samples to A & L Western Laboratories (Modesto, California) for mineral and ion analysis. I quantified organic matter, estimated nitrogen

release, phosphorus (weak bray and sodium bicarbonate-P), extractable cations (potassium, magnesium, calcium, sodium), hydrogen, sulfate-s, pH, cation exchange capacity, percent cation saturation (computed), soluble salts and excess lime, nitrate-nitrogen, zinc, manganese, iron, copper, boron, and nickel.

To determine whether the species and hybrid zones occur on divergent soils, I ran a Principal Components Analysis (PCA) of the soil variables in R (R Core Team 2014). I used all soil variables except for Ca and Mg, which were combined in a Ca:Mg ratio, a common ratio of interest in studies of serpentine-endemic plants (Brady et al 2005). All variables were standardized before analysis. Confidence ellipsoids for were drawn at 95%.

Serpentine tolerance is an adaptation to an extreme, toxic environment, and researchers have long sought to understand the genetic basis of the trait (Brady et al 2005). To determine whether there were any correlations between SNPs in our genetic dataset and the variables in my soil dataset, I used Latent Factor Mixed Models as implemented in the software LFMM (Frichot et al 2013) Using latent factors, in this case *K*, to control for genetic structure, LFMM uses a Bayesian model to estimate correlations between a matrix of allele frequencies and a matrix of environmental variables (Frichot et al 2013). These models are especially useful in my de novo analysis because they can estimate correlations directly from the observed

genotypes, whereas other recent methods (e.g. BAYENV Coop et al 2010) require a set of neutral control loci. I averaged soil PC scores for each population for the first five principal components as determined by PCA. Then I created a matrix of wherein every individual from a population shared the population mean score for the first five principal components as determined by PCA. Some populations (FO3001Nn, HZ near FO3003, and LFO) did not have soil samples, and were not included in the analysis. I used the same genetic data as input in the all-species STRUCTURE analysis detailed above. In all I tested 174 individuals at 158 loci. For each of K = 2-6 latent factors I ran five replicates of the model for 50,000 burn-in sweeps and 100,000 iterations. To score associations with the environmental variable, the program outputs z-scores and p-values for each SNP tested for every replicate. I averaged z-scores (z) for multiple runs, and adjusted P-values using R scripts included with LFMM. The user must choose the best models from the range of tested K using the deviance information criterion (DIC) and genomic inflation factor (λ). In line with guidelines set forth in Frichot et al (2013) I chose models with the lowest (DIC) and genomic inflation factor (λ) ≈ 1, (where λ = median(z^2)/0.456).

RESULTS

SNP calling

From nearly 900 million sequencing reads, I identified 158 loci with SNPs in 215 individuals, including 72 *M. stebbinsii*, 93 *M. follettii*, 12 *M.*

sheltonii, and 38 putative hybrids. For the *M. stebbinsii* and *M. follettii* dataset, I identified 675 loci with SNPs in 78 individuals and 365 loci with SNPs for 100 individuals, respectively. I identified different numbers of loci in the different datasets because I ran the analysis three times, once for each group of individuals.

Genetic Diversity

Genetic diversity is low in both of the rare species. I find low heterozygosities and few private alleles in *M. follettii* occurrences. (Table 3a). In *M. stebbinsii*, heterozygosities are also low in all occurrences (Table 3b). Furthermore, H_O is lower than H_E in all occurrences, suggesting that inbreeding is lowering genetic diversity in *M. stebbinsii*. The private allele counts and frequencies show the genetic differentiation that has occurred in the different populations. Private alleles are found at much higher numbers and frequency in *M. stebbinsii* occurrences compared to *M. follettii* occurrences.

Table 3a. *Summary statistics*[a] *for* Monardella follettii

Occurrence	MOFO 3001Nn	MOFO 3002	MOFO 3003	MOFO 3005	MOFO 3009	LFO	Mean
H_O[b]	0.124 (±0.009)	0.157 (±0.01)	0.181 (±0.011)	0.157 (±0.01)	0.135 (±0.009)	0.153 (±0.010)	0.151 (±0.004)
H_E[c]	0.136 (±0.008)	0.149 (±0.008)	0.160 (±0.008)	0.149 (±0.008)	0.135 (±0.008)	0.149 (±0.008)	0.146 (±0.003)
Private Allele Frequency	0.005 (±0.004)	0.008 (±0.005)	0.008 (±0.005)	0.011 (±0.005)	0.005 (±0.004)	0.008 (±0.005)	
Total Number of Private Alleles	2	3	3	4	2	3	

Table 3b. *Summary statistics*[a] *for* Monardella stebbinsii

Occurrence	MOST001	MOST003	MOST004	MOST005	Mean
H_O[b]	0.17 (±0.007)	0.159 (±0.006)	0.15 (±0.006)	0.179 (±0.007)	0.165 (±0.003)
H_E[c]	0.208 (±0.007)	0.218 (±0.007)	0.202 (±0.007)	0.208 (±0.007)	0.209 (±0.003)
Private Allele Frequency	0.053 (±0.009)	0.03 (±0.007)	0.037 (±0.007)	0.012 (±0.004)	
Total Number of Private Alleles	36	20	25	8	

[a] *Values are means with standard errors in parentheses.*
[b] *Observed heterozygosity.*
[c] *Expected heterozygosity.*

Genetic Structure

I found three genetic clusters when analyzing the *M. follettii* dataset (Figure 3a). All individuals share a majority of their genetic makeup from a single cluster (red, bottom cluster in Figure3a) with two other clusters shared disproportionately across the six populations. Interestingly, the sites (3003, 3002) near *M. stebbinsii* occurrences tend to have a greater proportion of the green, topmost genetic cluster. If there were strong genetic structure among occurrences, in this analysis I would expect to see a number of clusters equal

to the number of occurrences (6 in this case), with each occurrence belonging nearly 100% to only one cluster. Instead, three clusters are shared almost equally among occurrences. This initial result suggests that there has been little differentiation between the different occurrences sampled for this investigation.

For the *M. stebbinsii* dataset, there are two genetic clusters (Figure 3b). First, three occurrences (MOST004, MOST001, MOST005) appear to entirely derive their ancestry from only one of the two genetic clusters, whereas MOST003 appears be of split ancestry. Second, genetic structuring seems to occur with some spatial correlation as MOST001 and MOST004 occur along Caribou Rd, but MOST003 and MOST005 occur along Highway 70. These results suggest that some genetic differentiation has occurred among these occurrences.

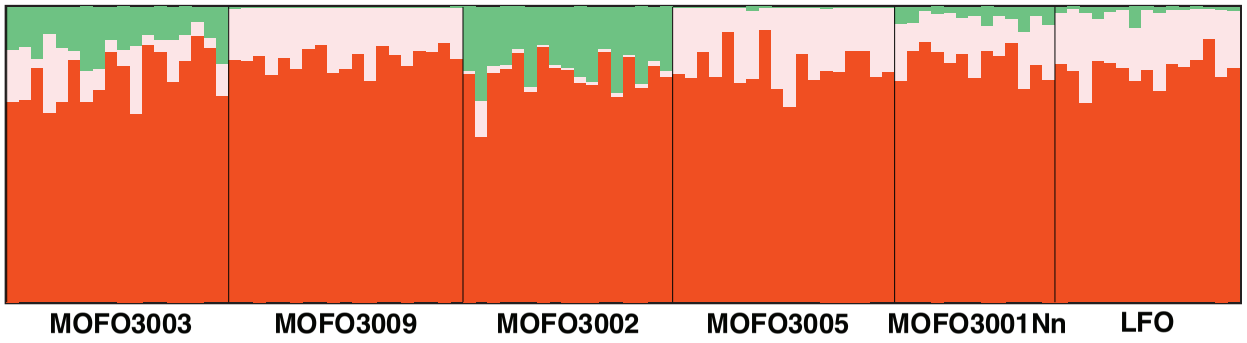

Figure 3a

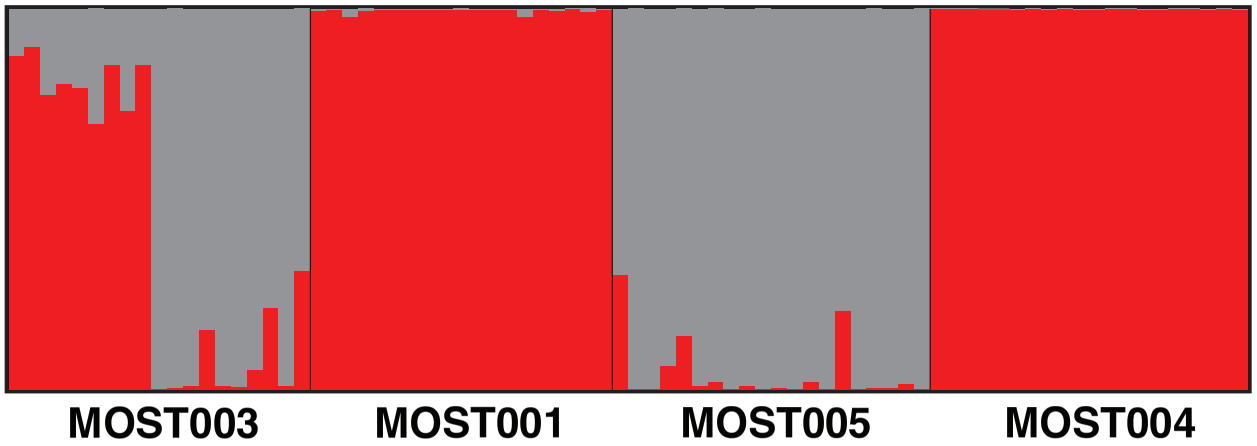

Figure 3b

Figure 3.*Genetic structure of* M. follettii *individuals (a) and* M. stebbinsii *(b) as estimated by Bayesian assignment. Each vertical bar represents one individual, and the proportion of each color within each bar corresponds to the proportion of the individual's genome assigned to each of the two genetic clusters. Occurrences are labeled below the bars and (in 3a) the species present at each site are indicated above the bars.*

In *M. follettii*, the AMOVA results (Table 4a) and associated overall and pairwise F_{ST} values (Table 4b and 4c, respectively) closely mirror the STRUCTURE result. The AMOVA shows that of all the genetic variation in the dataset, most (98%) of this variation occurs within individuals, with almost none attributed to variation among occurrences (2%). In my analysis, I see very low values, suggesting little structure between populations. F_{IT} and F_{IS} are not significantly different from zero in the *M. follettii* analysis. This suggests the plants are outbreeding, and that they are in Hardy-Weinberg equilibrium. Using the pairwise F_{ST} values from above with a matrix of pairwise geographic distances, a Mantel test shows a weak pattern of isolation by distance in *M. follettii* (Figure 4). This positive relationship between genetic and geographic distance is significant at $P = 0.051$ with an r^2=0.23. However, all F_{ST} values are very low and most are not significantly different than zero, such that even geographically distant occurrences are only slightly differentiated.

Table 4a. *Analysis of Molecular Variance (AMOVA) for* Monardella follettii

Source of Genetic Variation	Df[a]	SS[b]	MS[c]	Est. Var.[d]	%
Among Pops	5	235.928	47.186	0.597	2%
Among Indiv	94	2571.395	27.355	0.000	0%
Within Indiv	100	2800.157	28.002	28.002	98%
Total	199	5607.480		28.598	100%

[a] *Degrees of freedom,* [b] *Sum of squares,* [c] *Mean squares,* [d] *Estimated Variance*

Table 4b. *AMOVA F-statistics*

F-Statistics	Value	P-value[a]
F_{ST}	0.021	0.000
F_{IS}	-0.012	0.745
F_{IT}	0.010	0.292

[a] *P-values determined by randomization*

Table 4c*. Pairwise* F_{ST} *values for* Monardella follettii *occurrences*

	MOFO 3001Nn	MOFO 3002	MOFO 3003	MOFO 3005	MOFO 3009	LFO
MOFO3001Nn						
MOFO3002	0.023					
MOFO3003	0.016	0.005				
MOFO3005	0.019	0.030	0.021			
MOFO3009	0.021	0.038*	0.022	0.012		
LFO	0.016	0.029	0.015	0.023	0.025	

**Significant at p < 0.05. P-values were based on 9999 simulations and corrected for multiple tests using a Holm-Bonferroni adjustment (Holm 1979).*

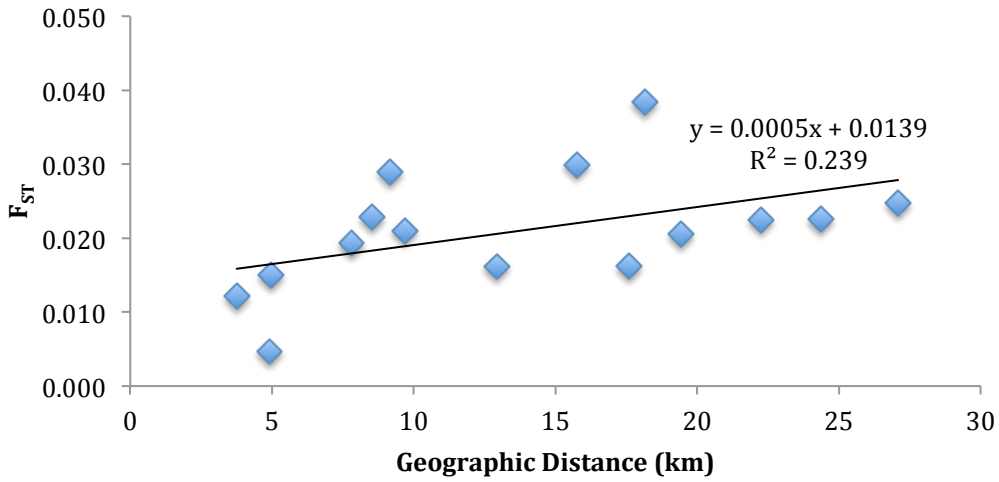

Figure 4. *Mantel Test for Isolation by Distance in* M. follettii*. Points represent pairwise occurrences as their geographic distances and F_{ST} values. The best fit line and its equation are shown.*

The AMOVA for the *M. stebbinsii* dataset reveals similar patterns to the STRUCTURE results for the species (Table 5a). Most (73%) of the variance occurs within individuals, but a much higher proportion occurs among individuals (19%) and among occurrences (8%) than in *M. follettii*. The overall *M. stebbinsii* F-statistics (Table 5b) show a small but significant amount of genetic structure among occurrences (F_{ST} = 0.082) and a significant, fairly large inbreeding coefficient (F_{IS} = 0.210). Pairwise F_{ST} values are small but significant between all pairs of occurrences (Table 5c). Reflecting the STRUCTURE results, occurrences paired between Highway 70 and Caribou Rd have higher F_{ST} values than occurrences paired within Caribou Rd or Highway 70, again suggesting some geographic differentiation. However, I do

not find any significant isolation by distance for *M. stebbinsii* occurrences with a Mantel Test on pairwise F_{ST} values and geographic distances.

Table 5a. *Analysis of Molecular Variance (AMOVA) for* M. stebbinsii

Source	df[a]	SS[b]	MS[c]	Est. Var.[d]	%
Among Pops	3	1024.598	341.533	6.495	8%
Among Indiv	74	6532.429	88.276	15.315	19%
Within Indiv	78	4496.366	57.646	57.646	73%
Total	155	12053.393		79.456	100%

[a] Degrees of freedom, [b] Sum of squares, [c] Mean squares, [d] Estimated Variance

Table 5b. *F-Statistics for* M. stebbinsii

F-Statistics	Value	P-value[a]
F_{ST}	0.082	0.000
F_{IS}	0.210	0.000
F_{IT}	0.274	0.000

[a] Determined by randomization

Table 5c. *Pairwise Occurrence F_{ST} Values for* M. stebbinsii

	MOST001	MOST003	MOST004	MOST005
MOST001				
MOST003	0.073*			
MOST004	0.069*	0.071*		
MOST005	0.118*	0.026*	0.125*	

**Significant values ($p < 0.01$). P-values are based on 9999 simulations, and were corrected for multiple tests using a Holm-Bonferroni correction (Holm 1979).*

In both species I find that principal coordinates analyses based on genetic distances reveal patterns similar to my other analyses. In *M. follettii*, the PCoA shows little differentiation among occurrences and the first three axes explain only 10% of the genetic variation (Figure 5a). However in *M. stebbinsii*, MOST004 and MOST001 (the Caribou Rd occurrences) cluster tightly together and are separated by Coordinate 1 from MOST005, whereas some individuals from MOST003 cluster near the Caribou Rd occurrences

and others are grouped with MOST005 (Figure 5b). The first three axes in this analysis only explain 17.27% of the variation, further supporting that most of the variation in *M. stebbinsii* is within occurrences and individuals.

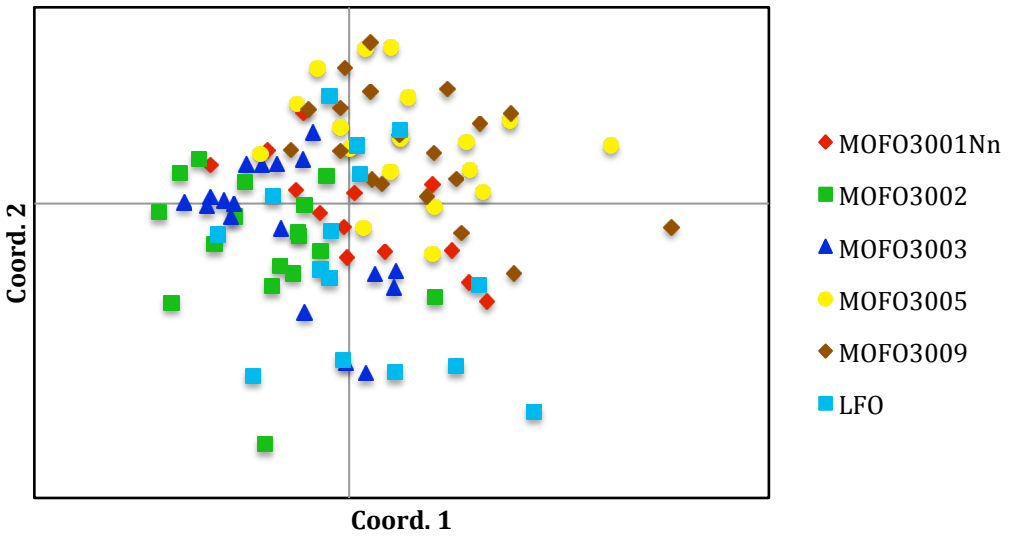

Figure 5a

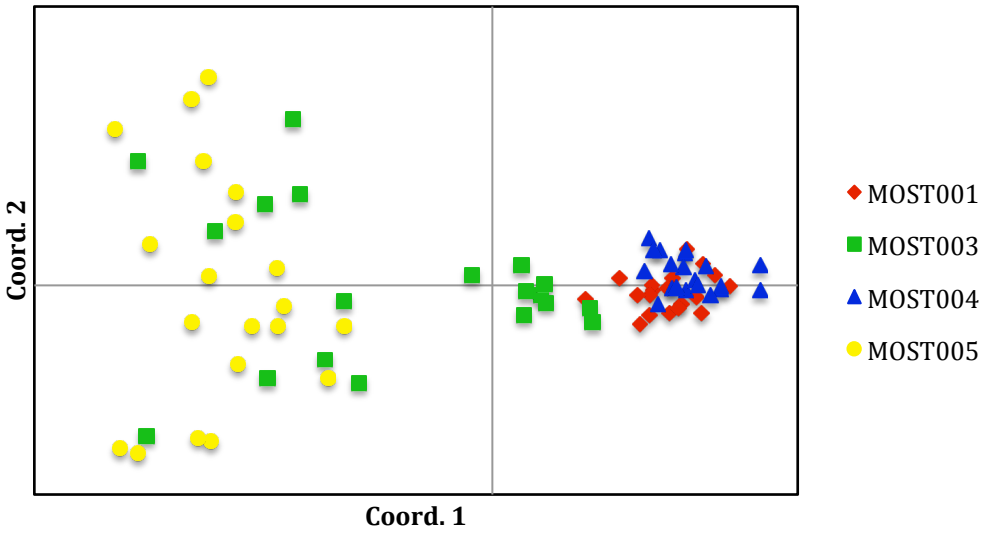

Figure 5b

Figure 5. *Principal Coordinates Analysis of* M. follettii *(a) and* M. stebbinsii *(b) genetic distances Individuals are positioned in space according to their genetic distance from other individuals. In* M. follettii *coordinates 1 and 2 explain 3.8% and 3.1% of the genetic variation in the data set, respectively. Coordinate 3 (not shown) explained an additional 2.9%. In* M. stebbinisii *the first and second coordinates explain 8.67% and 4.48% of the variation, respectively. The third coordinate (not shown) explains an additional 4.12% of the variation. Individuals from different occurrences are shown with different symbols and colors.*

Among Species Structure and Hybridization

The STRUCTURE data show variation in hybridization among the different mixed occurrences (figure 6a). There are two genetic clusters identified in the Bayesian assignment analysis across all three species (Figure 6a). Though the Evanno et al (2005) method showed only one peak at K=2, I also present the K=3 model (Figure 6b) because Pritchard et al (2000) suggest casually interpreting the posterior probabilities output by STRUCTURE while considering biological reality. In the K=2 model, individuals of *M. sheltonii* comprise one of the two clusters, and individuals of *M. stebbinsii* comprise the other. *Monardella follettii* individuals are a mix of the two genetic clusters, varying from 8-63% of the *M. stebbinsii*-like cluster where *M. follettii* occurs alone, and from 1-57% and 53-97% where it occurs with *M. sheltonii* and *M. stebbinsii*, respectively, in putative hybrid zones (see Bean Hill, Red Hill, and HZ near FO3003 hybrid zones, Figure 2). I do not find the same pattern when *M. stebbinsii* and *M. sheltonii* occur together (HZ near MOST005, Figure 2). Instead, I find that individuals in this hybrid zone show full ancestry from the *M. stebbinsii*-like cluster. Though STRUCTURE found only two clusters in the data, it is clear that each of the three species is unique in its genetic identity, i.e. species are either entirely composed of one unique cluster (*M. sheltonii* and *M. stebbinsii*) or a consistent mix of the two clusters (*M. follettii*).

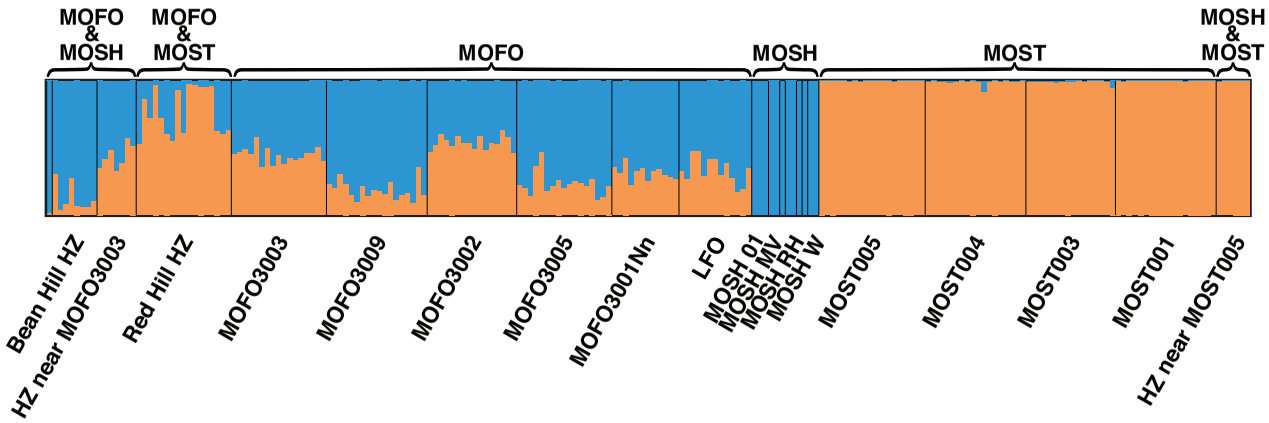

Figure 6a

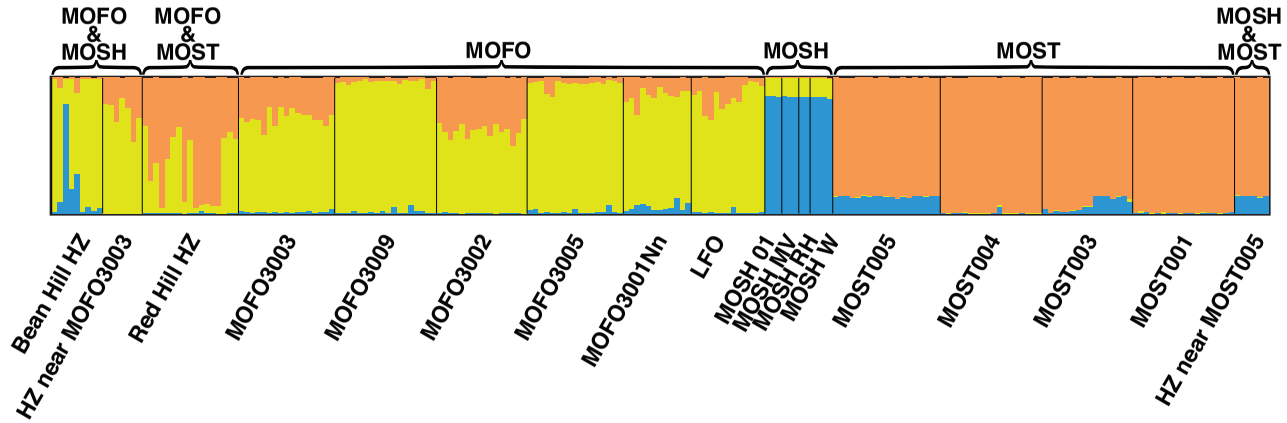

Figure 6b

Figure 6. *Genetic structure of all individuals in the K=2 model (a) and all individuals in the K=3 model (b) as estimated by Bayesian assignment. Each vertical bar represents one individual, and the proportion of each color within each bar corresponds to the proportion of the individual's genome assigned to each of the two genetic clusters. Occurrences are labeled below the bars and the species present at each site are indicated above the bars.*

The K=3 model shows the similar patterns of assignment for *M. stebbinsii* and *M. sheltonii*, but *M. follettii* individuals are assigned mostly to a third cluster. *Monardella sheltonii* shows a consistent 13% assignment to the *M. follettii*-like cluster. All individuals in MOST005 and nearly half of individuals in MOST003 show 13% assignment to the *M. sheltonii*-like cluster. Individuals from MOFO3005 and MOFO3009 show nearly 100% assignment to the *M. follettii*-like cluster, but the *M. follettii* populations show 3-40% assignment to the *M. stebbinsii*-like clusters. In hybrid zones in which *M. follettii* is present, I find that individuals are assigned to the *M. follettii*-like cluster in varying amounts from 18-98% where it occurs with *M. sheltonii*, and 3-68% when it occurs with *M. stebbinsii*. Again, the individuals in the putative hybrid zone near MOST005 follow the same assignment patterns as *M. stebbinsii* in the nearby MOST005 population.

Monardella follettii and *M. sheltonii* individuals co-occur near Bean Hill. Here, *M. follettii* individuals display very *M. sheltonii*-like morphology; they have inflorescences with more flowers, are paler green in color, and are generally larger than other *M. follettii* individuals. These observations are reflected at the genetic level in their majority assignment to the *M. sheltonii*-like cluster. The Bean Hill individuals clearly show some assignment to the *M. stebbinsii*-like cluster, but less than individuals in other spatially isolated *M. follettii* occurrences. Along FS Road 26N26 south of MOFO3003, *M. follettii*

individuals occur adjacent to some *M. sheltonii* individuals. All individuals here have the *M. follettii* gestalt, and they show similar assignment patterns to the individuals in the nearby MOFO3003 occurrence. This suggests these individuals may not be hybrids. Finally, there are some *M. sheltonii* individuals at the MOST005 occurrence. At this site, *M. stebbinsii* individuals are not morphologically intermediate, but they do occur adjacent to *M. sheltonii* individuals. The evidence from STRUCTURE suggests that no hybridization occurs at this site, with genotyped individuals entirely assigned to the *M. stebbinsii*-like genetic cluster.

The results from INTROGRESS show similar patterns to the Bayesian assignment estimates from STRUCTURE. Most individuals appear to be at some stage of introgression between *M. follettii* and *M. stebbinsii* in the Red Hill hybrid zone, where all but two individual show mixed ancestry (figure 7a). Further these results suggest the individuals seem to generally show more ancestry from *M. stebbinsii* than *M. follettii*, a result seen morphologically and in the STRUCTURE results. In the Bean Hill hybrid zone, most individuals show near entire ancestry from *M. follettii* (figure 7b), though some individuals appear to derive significant proportions (0.92, 0.23, 0.28) of their ancestry from *M. sheltonii*. All individuals show nearly complete ancestry from *M. follettii* at the putative hybrid zone near the MOFO 3003 occurrence (figure 7c).

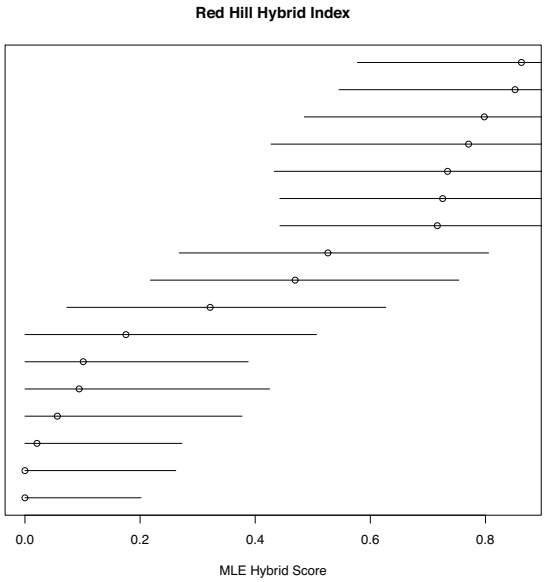

Figure 7a

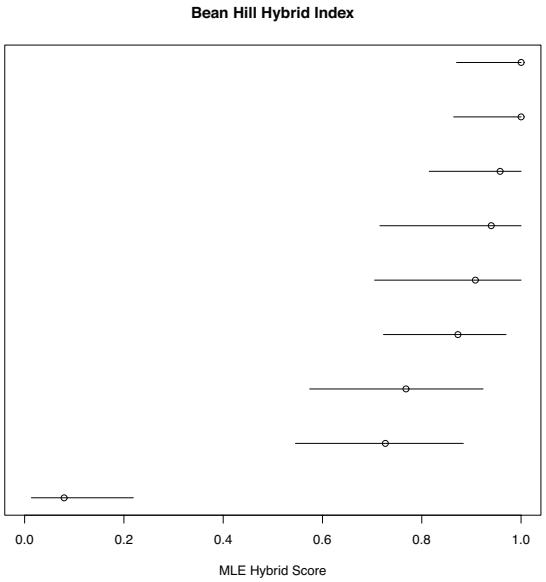

Figure 7b

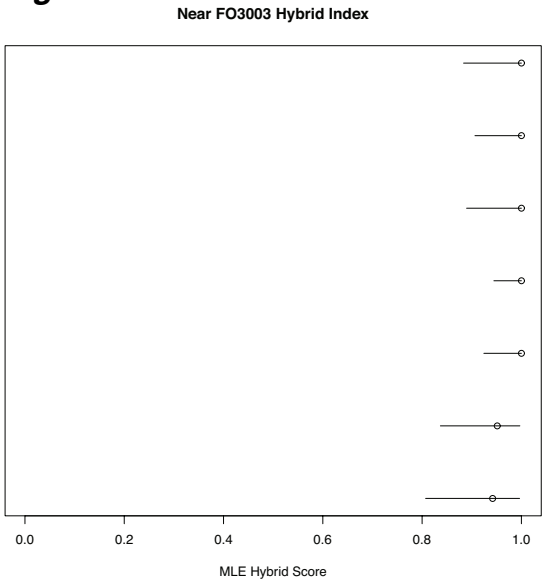

Figure 7c

Figure 7. *Maximum likelihood estimates with confidence intervals at 95% of ancestry in putative hybrid zones at Red Hill (a) Bean Hill (b) and near FO3003 (c). Hybrid index refers to the maximum likelihood estimate of proportion ancestry from* M. follettii *in putative hybrid zones. In figure 7a, 0 values show complete ancestry from* M. stebbinsii*, whereas in figures 7b and 7c, 0 values show complete ancestry from* M. sheltonii*. Note the different scale in figure 7a.*

Soil

The first 5 principal components (PC) explain 73% of the variance in the soil dataset (Table 6), with PC1 and PC2 explaining 26.0% and 15.9% of the variance, respectively. Organic matter, estimated nitrogen release, nickel, and potassium show high negative loadings on PC1, whereas pH showed the only high positive loading on first axis. Nitrate and pH show high positive loading on PC2, and sodium and phosphorus show large negative loadings on the second axis. The two rare species show some overlap when plotted on the first two axes (Figure 8). The Red Hill hybrid zone clusters completely within the *M. stebbinsii* cluster, showing little divergence from at least one of the parental species. The Bean Hill hybrid zone shows no overlap with the *M. follettii* ellipse, but exhibits less differentiation from the limited *M. sheltonii* samples.

Table 6. *Loadings of soil variables on the first five principal components.*

	PC1	PC2	PC3	PC4	PC5
Organic Matter	-0.4567	-0.0431	0.0926	-0.1762	-0.0059
Estimated Nitrogen Release	-0.4565	-0.0432	0.0932	-0.1761	-0.0085
Ni	-0.3601	0.2555	-0.1072	0.0735	-0.0178
K	-0.3567	0.0017	-0.2292	0.2074	0.1880
Cation Exchange Capcity	-0.3149	0.1608	-0.0034	0.3394	-0.2442
pH	0.3082	0.4236	0.0371	0.0293	0.1748
Mn	-0.2611	-0.1738	0.1506	-0.3259	-0.1988
S	0.1402	-0.2724	0.1778	-0.4776	0.0586
Fe	-0.1318	0.1910	0.5204	-0.1342	0.0777
Ca:Mg	-0.0873	-0.3120	-0.0412	0.1060	0.6573
Na	0.0837	-0.4221	-0.1001	0.0412	-0.2578
HCO3-P	-0.0617	-0.3616	0.2021	0.4072	0.3308
Cu	0.0584	-0.0933	0.4142	0.4889	-0.3535
Zn	0.0569	-0.0060	0.5913	0.0372	0.0946
NO3-N	-0.0416	0.4121	0.1359	-0.0459	0.2884
Proportion of Variance	0.2601	0.1589	0.1285	0.0998	0.0870
Cumulative Proportion	0.2601	0.4190	0.5475	0.6473	0.7343

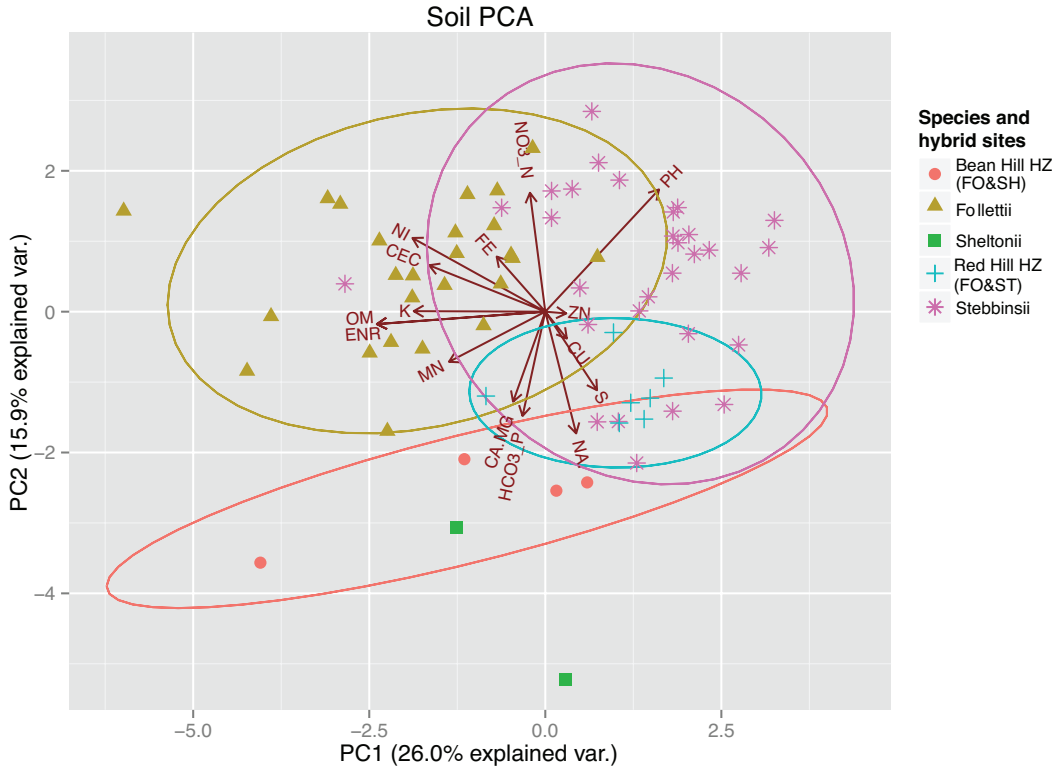

Figure 8. *Principal components analysis of soil variables from* Monardella *occurrences. Colors and shapes represent different species or co-occurring species pairs. Arrows show loadings of soil variables on the first two principal components. Ellipses show 95% confidence intervals for each set of points.* Monardella sheltonii *had too few samples to draw an ellipse.*

My analysis with latent factor mixed models shows that some SNPs are correlated with soil PC5. All DIC values were nearly identical, only ranging 27500-27503 across all *K* values. Of the models tested in LFMM, *K*=2 shows the lowest DIC across all PCs. However the software authors caution against choosing *K* on DIC alone, and suggest using STRUCTURE as a guide to examine other *K*. Therefore I also examine the results from *K*=3. For all models tested, only PC 5 in *K*=3 shows $\lambda \approx 1$, the genomic inflation factor at which the false discovery rate is sufficiently low. The *K*=3 model for PC 5

reveals 3 significant SNPs after p-value adjustment (table 7). For SNPs S1_44039024 and S1_38589050, BLAST queries in the angiosperm database using the blastn algorithm return only matches to unannotated complete genome sequences of tomato or *Arabidopsis*. However the SNP S1_48221057 sequence matches many species with over 90% identity to a putative gene in the NPH3 family, a group of photoreceptor genes (Motchoulski and Liscum 1999).

***Table 7**. SNPs significantly correlated with PC5.*

SNP	Adjusted P (*K*=3)	Sequence
S1_44039024	1.60×10^{-6}	TACGACGCGGGAT**[G/A]**CATGAGATCCT-TGAATTAAATTTATAAATGAATGTTCCTG
S1_48221057	3.64×10^{-6}	TGGACTGCCAGAAGCTCTCGCTCGAAGC-TTGCACACACGCGGCCCA**[G/A]**AACGAGA
S1_38589050	0.0007	AAAATCATGAATTTTAGAAGAACAATTGA-GAAAAAATAAC**[T/C]**GACTTCAAACTACA

DISCUSSION

Discussion

GBS and SNP generation

I chose GBS for its low cost and generation of many markers among multiple, related species. The GBS library method promises tens- or hundreds of thousands of tags (Elshire et al 2011), therefore my yield of SNPs from all three datasets was comparatively low. I am unsure what caused such a modest tag recovery with the GBS library protocol, but I experienced similarly low tag counts in previous RAD library preparations in which my samples were only 40-plexed. Due to the extensive problems I had isolating clean

DNA, I surmise that the DNA samples may have contained too many contaminants that inhibit restriction enzyme activity. Though verified as high quality by Nanodrop, Qubit, and agarose gel, the samples may not have been thoroughly digested by restriction enzymes.

I also encountered some missing data in our SNP calls in addition to the problems with DNA quality. Depending on the analysis, these data were either ignored or interpolated from other individuals in the population. A more standard practice with NGS data is to impute missing genotypes from assembled consensus genome sequences (Howie et al 2009). In this case, it was not feasible to sequence multiple genomes from the three species of interest. However even with the low tag yields and filtered data, simulations on SPOTG suggest that our sampling and data recovery were sufficient for our purposes (Hoban et al 2013, Excoffier and Lischer 2010, Laval and Excoffier 2004). Researchers have embraced reduced representation libraries as documented in the literature, where the three-year-old GBS method and six-year-old restriction site associated DNA (RAD) tag method have been cited 300 and 800 times, respectively, according to the BIOSIS citation index. For me, the GBS method worked well enough to provide within-species guidance for conservation. The method also provided enough SNPs to capture hybridization and introgression among species, which might have been missed using microsatellites or other markers.

Genetic diversity

The rare *Monardella* exhibit small population sizes and patchy distributions, attributes which are typically indicative of low levels of genetic diversity. Typical levels of heterozygosity vary widely based on marker type (Ryynänen et al 2007). For the types of markers we used, large populations of angiosperms exhibit heterozygosities ranging from 0.18-0.30 (Vandepitte et al 2012a, Vandepitte et al 2012b). Compared to these levels, all populations of *M. stebbinsii* and *M. follettii* have low genetic diversity. For rare, threatened species these results are concerning, but are not unexpected given the small population sizes found in nature (Paschke et al 2002, Ellstrand and Elam 1993). If the populations have been small for several generations, which we believe they have, genetic drift should account for some of the reduced heterozygosity. *Monardella stebbinsii* exhibits a significant inbreeding coefficient and a patchy distribution with significant structuring among occurrences. Inbreeding and low levels of gene flow are known to lower genetic diversity in plants (Honnay and Jacquemyn 2007), and are likely to account for some of the decreased genetic diversity in *M. stebbinsii*. Conversely *M. follettii* occurrences do not show signals of these phenomena, and generally have larger census population sizes. Though population sizes are larger than those found in *M. stebbinsii*, they are still small at 30 to ~300 individuals, likely explaining the low genetic diversity found in all occurrences.

Some individuals were filtered out of the dataset in the TASSEL/UNEAK pipeline, which could result in an artificially low calculation of genetic diversity. However, the inclusion of these individuals in SNP calling could have greatly reduced the quality of the overall SNP dataset.

Genetic structure

Monardella follettii and *M. stebbinsii* exhibit differing patterns of genetic structure. In *M. follettii*, we only find one significant pairwise F_{ST} value, whereas *M. stebbinsii* shows significant pairwise F_{ST} in all comparisons. Further, in the STRUCTURE analysis *M. stebbinsii* shows assignment to two genetic clusters, which correlate with geographic locations along Highway 70 and Caribou road. I also observe *M. stebbinsii* genetic clustering along geographic lines of in the PCoA, where the Highway 70 and Caribou Rd populations cluster across Coordinate 1 in the PCoA. While *M. follettii* is assigned to three genetic clusters, all individuals in all populations show majority assignment to only one cluster, suggesting little differentiation among populations. Given the much smaller distribution of *M. stebbinsii*, one might expect less genetic structure between populations. However *M. stebbinsii* populations are smaller than *M. follettii* populations, and likely undergo more rapid genetic drift. Further, a significant F_{IS} in *M. stebbinsii* suggests extensive inbreeding in the species, which can also lead to genetic structure (Barrett and Kohn 1991). Significant structure in *M. stebbinsii* suggests little gene flow

among populations, which could lead to adaptation to the local microhabitat, but I did not test this explicitly.

Many studies of rare, fragmented plant populations show patterns of structuring similar to those seen in *M. stebbinsii* (reviewed in Aguilar et al 2008), but notable exceptions occur in which geographically separated populations exhibit little genetic structure as in *M. follettii* (Mandel et al 2013). The range of *M. stebbinsii* is entirely surrounded by the much larger range of *M. follettii*, but only *M. stebbinsii* shows significant structuring among populations. The habitat of *M. stebbinsii* is very rare, only existing in ravines immediately surrounding Red Hill. Conversely, the habitat of *M. follettii* is relatively common and less extreme, giving rise to many stepping stone populations that likely allow for gene flow among *M. follettii* populations. The divergent patterns of genetic structure could be the result of different pollinators visiting the species, if the pollinators behave differently in the distance traveled between plants. The only pollination data available for the two species were not collected concurrently, and therefore any observed differences are inconclusive (Woolhouse 2012). Even if the exact mechanism for the differences in genetic structure cannot be disentangled with our analyses, these differences are still helpful for conservation planning in order to potentially preserve local adaptation in the species (McKay et al 2005).

Taxa in the California serpentine flora have long been proposed to have arisen through either speciation on newly available serpentine habitat, neoendemism, or restriction of ancient taxa to marginal habitat, paloendemism (Stebbins 1942, Raven and Axelrod 1978, Kruckeberg 1986, Kay et al 2011). A neoendemic species, which has recently and insularly speciated, should exist in only a limited range, likely within only one mostly homogenous serpentine outcrop. Further, a recently diverged taxon might exhibit low genetic structure among populations, low genetic diversity, incomplete barriers to reproduction. Conversely, a paleoendemic should either appear on multiple, distant serpentine outcrops, or have sister taxa on said outcrops. Likewise, due to the age of the lineage, a paleoendemic might show significant genetic structure, moderate to high genetic diversity, and complete barriers to reproduction, assuming these have not been eroded through inbreeding or very small population sizes. On a phylogeny, neoendemics should appear on a short branch sister to nonserpentine taxa, whereas paleoendemics should appear on a longer branch sister to serpentine or nonserpentine taxa. In the cases of *M. follettii* and *M. stebbinsii*, my genetic data do not indicate a paleoendemic origin. Rather the genetic data, the propensity to hybridize, and the limited geographic distributions suggest that the plants have always been rare, and have not been recently restricted in range. However, we cannot confidently conclude any origin of the

species without a robust phylogeny, a possible next step in understanding the speciation process in *Monardella.*

Among Species Structure, Hybridization, and Introgression

The among-species STRUCTURE analysis finds only two genetic clusters, corresponding to *M. stebbinsii* and *M. sheltonii,* and assigns *M. follettii* and hybrid individuals to a mix of these two clusters. The Delta-K method (Evanno et al 2005) for choosing the best model for our STRUCTURE analysis found only one peak at K=2, strongly suggesting it is the best model. However, Pritchard et al (2000) suggest using a casual approach to interpreting posterior probabilities output by STRUCTURE, and to consider biological reality when choosing a model. Thus, the K=3 model might make the most sense. In the K=3 model, individuals from each species are assigned to a genetic cluster that mostly corresponds to its own species, but there is notable mixing of clusters in most individuals. In *M. sheltonii*, which shows a small mixed assignment to the *M. follettii*-like genetic cluster. There are also notable mixed assignments in the K=3 model in occurrences where more than one species are present or nearby.

Both models have their own challenges, and choosing one over the other remains difficult. Accepting the K=2 model could suggest that *M. follettii* derives its entire genome from a combination of *M. sheltonii* and *M. stebbinsii*, and that the proportions of this combination vary across individuals and

occurrences. Examples of hybrid speciation exist (e.g. Aïnouche et al 2004), but are rare and require either allopolyploidy or overcoming a host of problems including genetic incompatibilities, the rapid establishment of reproductive barriers, and others (Otto and Whitton 2000, Abbott et al 2013). In this case, these assumptions are difficult to accept because, if *M. follettii* were polyploidy or sufficiently reproductively isolated from the other species, the backcrossing patterns with *M. sheltonii* and *M. stebbinsii* we see (discussed below) are unlikely to occur due to complications with meiosis in triploids. However adaptive introgression, in which portions of a genome are assimilated into a hybrid lineage allowing individuals to exploit new resources and diverge over time, has been considered as a possible path to speciation (Martin et al 2006). In the case of *M. follettii,* serpentine soil tolerance could give a *M. sheltonii* x *M. stebbinsii* hybrid the necessary adaptation to persist and spread across the serpentine outcrops of Plumas National Forest, but our results do not test this hypothesis. In contrast, the K=3 model might require a different set of explanations. The K=3 model might suggest *M. follettii* evolved independently from the other two species, but all species lack sufficient reproductive barriers to prevent hybridization and introgression. The K=3 model also allows for sustained introgression, varying in degree by population (e.g. consistent ~35% assignment to the *M. stebbinsii*-like cluster in MOFO3002, but not in MOFO3005). On the other hand, the K=3 model may

not be very different from the K=2 model, as there are no *M. follettii* individuals with a pure genome. In both models, evidence of hybridization and introgression in some putative hybrid zones is readily apparent.

Hybridization in *Monardella* is very common across the genus, which taxonomists have previously noted in genus descriptions (Elvin and Sanders 2009, Sanders et al 2013), and our tests show widespread hybridization in the *Monardella* of Plumas National Forest. Careful observation in the field reveals evidence of morphologically intermediate individuals in some populations where species of *Monardella* co-occur. In the Red Hill and Bean Hill Hybrid Zones, the hybrid indices show a range of H-values suggesting advanced generation hybrids and introgression (Anderson 1936, Buerkle 2005). We do not find evidence of hybridization using hybrid indices in the other putative hybrid sites, instead all individuals appear to show full ancestry from *M. follettii* or *M. stebbinsii*. At the putative hybrid site near MOFO3003, hybrids probably do not occur frequently because only a very small area of contact exists between the two species. The final site near MOST005 is the only site where *M. sheltonii* and *M. stebbinsii* co-occur, and the individuals tested show little evidence of hybridization. These two species may be incapable of producing viable hybrids due to a number of pre- and post-zygotic isolating mechanisms, but determining the extent and mechanisms of isolation requires further experimentation.

Hybridization often occurs in area of anthropogenic disturbance, mixed habitats, or marginal populations where postzygotic barriers (e.g. a lack of niche space due to competition) would otherwise prevent the establishment of hybrids (Grant 1981, Rieseberg 1997, Wang et al 1997). The *Monardella* hybrid zones in Plumas National Forest seem to follow this pattern. The principal components analysis of soil found that the Bean Hill hybrid zone exhibits mixed soil somewhere between *M. follettii* and *M. sheltonii* on the first two PC axes. *Monardella follettii* and *M. stebbinsii* occur on divergent, but not significantly different soils. Accordingly, I would not expect a mixed soil between the two species to be significantly different from either parental species, and the soil at Red Hill hybrid zone is very similar to the *M. stebbinsii* soil found at other sites.

The hybrid populations also occur in areas of high disturbance. The hybrid zone at Red Hill sits immediately below several radio and cell phone towers that make up the communications center for Plumas National Forest. The towers are frequently visited by repairmen and Forest Service workers, and are fiercely defended during forest fires. The characteristic red soil at the population has been stained a particularly deep crimson from repeated coats of fire retardant dropped from fire fighting helicopters. Meanwhile, the plants at Bean Hill sit adjacent to an area subjected to hydraulic mining during the California Gold Rush, and a heavily-used Forest Service road bisects the

population. At these sites, disturbances could have opened niche space that allows hybrids to form and establish. However, some of the remaining populations examined for hybridization, as well as other populations with multiple species present, are equally disturbed, but show no evidence of hybridization. Further, hybridization can occur in undisturbed areas (Rieseberg et al 2003, Arnold et al 2012). Therefore other factors, such as shared pollinators, may be more important than disturbance and soil characteristics in mediating hybridization in these populations.

Soil-SNP correlations

Serpentine soil is toxic for most plants, and the genetic basis for adaptation to the harsh environment is of keen interest to ecologists (Brady et al 2005). With data from across the genomes of three species living on divergent soils, we expect that some SNPs might be directly or indirectly (through linkage) correlated with principal components that describe the soil. Our analysis in LFMM finds three SNPs that are correlated with soil PC axes. One of these is part of a gene known as root phototropism protein 3 (Motchoulski and Liscum 1999), and its significance in the context of serpentine soil remains to be understood. We strongly caution any interpretation of these results without further experimentation and verification, because genome-wide association methods are prone to false positives (Frichot et al 2013).

Conservation planning

Monardella stebbinsii and *M. follettii* are two of the rarest plants in Plumas National Forest, and the genetic parameters derived from these analyses can help shape species management policy. We offer our recommendations under the assumptions that genetic diversity is essential for the long-term evolutionary potential of the populations (Honnay and Jacquemyn 2007), inbreeding can increase extinction risk (O'Grady et al 2005), and maximizing genetic diversity through transplantation should not compromise local adaptation (McKay et al 2005). We also assume that unchecked hybridization can result in the assimilation of a rare species into a more common congener (Rieseberg and Gerber 1995, Ellstrand et al 2010), therefore populations with putative hybrids should not be utilized as source stock for assisted gene flow projects. For the sake of clarity, these recommendations are offered separately for the two rare species.

Management recommendations for M. follettii

Low genetic distances and equally low levels of genetic diversity among all occurrences suggest that no existing occurrence would likely experience a genetic benefit or genetic cost from supplementation. The parity in genetic diversity and low genetic differentiation among occurrences suggest managers of *M. follettii* do not need to be especially selective with occurrence seed sourcing and transplantation after catastrophic disturbance or simply for demographic supplementation. A weak, but marginally

significant, pattern of isolation by distance suggests that managers could use nearby occurrences as seed and plant sources when possible, but using more distant seed sources is unlikely to be problematic. However, we caution that because of the propensity of *Monardella* to hybridize, managers should avoid seed and plant sourcing from occurrences with more than one species present. Finally, although we did not find significant inbreeding in the adults, we did not examine the genotypes of seeds, and plants may be producing less viable, inbred seed. Managers should take care to closely monitor any seed-based restoration.

Monardella follettii occurs in a wildfire-prone environment, and occurrences of the species burn in the occasional blaze. Occurrences that have recently experience fire and salvage logging (MOFO3003, Lassen) do not appear to have lower or higher genetic diversity than other occurrences in the *M. follettii* range. This suggests that these occurrences have neither detectably benefitted nor suffered at the genetic level due to fire. However, we cannot conclude based on genetic data alone that other ecological or demographic benefits or damages could occur after fire. For fire management issues, we suggest botanists and other decision makers examine pre- and post-fire demographic data for *M. follettii* and *M. stebbinsii* and their important, mutualistic pollinators in response to the recent Storrie, King, and Chips fires (see Woolhouse 2012 for pollinator identities).

Management recommendations for M. stebbinsii

Overall patterns in *M. stebbinsii* are very similar to those seen in *M. follettii*, but higher levels of genetic structure and significant levels of inbreeding motivate a different set of management recommendations. First, the occurrences might benefit from pollen supplementation among occurrences because the data show that the plants are inbred. Because the habitat is very fragile, if this is attempted, we suggest placing flowering inflorescences in vases as close as possible to the target occurrence without disturbing the soils. If such measures are undertaken, we suggest sourcing from nearby populations. Further, unique alleles in all occurrences and genetic structure among occurrences suggest all occurrences should be closely monitored and preserved. Due to the difficulty in monitoring the occurrences and the fragile nature of the habitat, managers should take care to protect the delicate *M. stebbinsii* occurrences by following the recommendations given in Coppoletta and Woolhouse (2010). These recommendations include minimizing ground disturbance, evaluating any Forest Service activities on a site-by-site basis, and using spatial models to protect suitable *M. stebbinsii* habitat.

In the case of restoration after disturbance at any occurrence, our genetic structure results suggest that seeds should be sourced from nearby occurrences, i.e. restoration at a Caribou Rd. site should use seeds from another Caribou Rd. occurrence. If managers believe removing flowering

inflorescences from a nearby occurrence to be largely damaging to the source occurrence, low genetic distances among all occurrences suggest that managers could source pollen from a larger, more distant occurrence. We caution against using seeds from any site with a co-occurring species of *Monardella* as hybridization between congenerics appears to occur.

Based on anecdotes related to us over the course of this investigation, we believe most *M. stebbinsii* occurrences did not burn in the recent King, Chips, and Storrie wildfires. Accordingly, we do not make any conclusions relating our results to fire. Instead, we echo Coppoletta and Woolhouse's recommendations to try to avoid disturbing soils while fighting fires. Erosion may be the biggest threat facing *M. stebbinsii*, so any fire-related management of *M. stebbinsii* should focus on preserving soil stability during and after wildfires.

Summary and Conclusions

Small populations of rare species in disturbance prone environments are susceptible to extinction, and are of considerable concern to conservation biologists. The low genetic variation and high levels of inbreeding expected in small populations can diminish a species evolutionary potential. Molecular population genetics offers tools to estimate genetic diversity, inbreeding, and other phenomena, and can answer fundamental questions on the evolutionary consequences of small populations. When combined with ecological and

demographic data, managers can use population genetics parameters to create effective conservation strategies.

Our study of *Monardella* is one of the first uses of GBS in conservation genetics, and the method proves efficient for the generation of informative, genome-wide SNPs from multiple species with no known genomic information. In the current era of increased disturbance, many threatened species might benefit from examination using conservation genetics, and given the right conditions, GBS offers a rapid way to generate informative data. In the rare *Monardella*, the data yield may be lower than expected, but the GBS method still discovers a sufficient number of SNPs for analyses of genetic diversity, structure, and hybridization. For the same budget and timeline, we may not have been able undertake the hybrid analyses in these species using microsatellites, because developing microsatellites that amplify in multiple species can be difficult and expensive (Allendorf et al 2010). Instead, our use of GBS enables us to derive valuable information about hybridization and introgression, which in turn shapes management plans for the Forest Service.

We find that both rare species have a troubling paucity of genetic diversity, but

patterns of genetic structure differ between the two species. In *M. follettii*, there is very little genetic structure among occurrences. An analysis of genetic

diversity shows low levels in all *M. follettii* occurrences, but no evidence of inbreeding. In contrast, we find significant genetic structure in *M. stebbinsii*, and find species-wide inbreeding. Highway 70 and Caribou Rd occurrences are genetically differentiated at slight, but statistically significant, levels. Genetic diversity is low in all *M. stebbinsii* occurrences, and there is significant inbreeding across occurrences. The species appear to hybridize and backcross readily, which confirms earlier field observations and taxonomic descriptions of the genus at large. Each species inhabits divergent soil, and hybrids at Bean Hill inhabit soils that are divergent from either parental species.

In the broader scope of small population evolutionary biology, our work uses new techniques to examine unstudied species in the context of many theoretically and empirically established biological concepts, and we find patterns of low genetic diversity, variable genetic structure, and hybridization in marginal habitats that are generally consistent with those in the literature (Honnay and Jacquemyn 2007, Ellstrand and Elam 1993, Rieseberg 1997). The work described here introduces *Monardella* as a compelling genus in which to study the implications of hybridization and introgression in the speciation process, a fundamental area of research in evolutionary biology (e.g., Darwin 1859, Abbott et al 2013). Full elucidation of the species organization within the genus requires a full phylogenetic analysis and

experimentation designed to disentangle species relationships. However, the genus is challenging as a study system due to the difficulty of extracting usable DNA as well as variable germination rates in greenhouse and field experiments (Woolhouse 2012). Finally, this work combined with ecological data and field experiments (Coppoletta and Woolhouse 2010, Woolhouse 2012) offers a case study in the utility of a multi-pronged approach to conservation planning. Incorporating these data ensures small populations remain resilient against the many challenges and negative feedbacks in the extinction vortex (Gilpin and Soulé 1986).

REFERENCES

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, Butlin RK, Dieckmann U, Eroukhmanoff F, Grill A, Cahan SH, Hermansen JS, Hewitt G, Hudson AG, Jiggins C, Jones J, Keller B, Marczewski T, Mallet J, Martinez-Rodriguez P, Möst M, Mullen S, Nichols R, Nolte AW, Parisod C, Pfennig K, Rice AM, Ritchie MG, Seifert B, Smadja CM, Stelkens R, Szymura JM, Väinölä R, Wolf JBW, Zinner D. **2013**. Hybridization and speciation. *Journal of Evolutionary Biology* **26**: 229–46.

Aguilar R, Quesada M, Ashworth L, Herrerias-Diego Y, Lobo J. **2008**. Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. *Molecular Ecology* **17**: 5177–88.

Aïnouche M, Baumel A, Salmon A. **2004**. *Spartina anglica* CE Hubbard: A natural model system for analyzing early evolution changes that affect allopolyploid genomes. *Biological Journal of the Linnaean Society* **82**: 475–484.

Allendorf FW, Hohenlohe PA, Luikart G. **2010**. Genomics and the future of conservation genetics. *Nature Reviews: Genetics* **11**: 697–709.

Allendorf FW, Leary RF, Spruell P, Wenburg JK. **2001**. The problems with hybrids: setting conservation guidelines. *Trends in Ecology & Evolution* **16**: 613–622.

Amsellem L, Noyer JL, LE Bourgeois T, Hossaert-McKey M. **2000**. Comparison of genetic diversity of the invasive weed *Rubus alceifolius* Poir. (Rosaceae) in its native range and in areas of introduction, using amplified fragment length polymorphism (AFLP) markers. *Molecular Ecology* **9**: 443–55.

Anderson E. **1936**. Hybridization in the American Tradescantias. *Annals of the Missouri Botanical Garden* **23**: 511–525.

Arnold M, Ballerini E, Brothers A. **2012**. Hybrid fitness, adaptation and evolutionary diversification: lessons learned from Louisiana Irises. *Heredity* **108**: 159–66.

Baldwin BG. **2005**. Origin of the serpentine-endemic herb *Layia discoidea* from the widespread *L. Glandulosa* (Compositae). *Evolution* **59**: 2473.

Barrett SCH, Kohn JR. **1991**. Genetic and evolutionary consequences of small population size in plants: implications for conservation. In: Falk DA, Holsinger KE, eds. *Genetics and Conservation of Rare Plants*. Oxford: Oxford University Press, 3–30.

Brady KU, Kruckeberg AR, Bradshaw Jr. HD. **2005**. Evolutionary ecology of plant adaptation to serpentine soils. *Annual Review of Ecology, Evolution, and Systematics* **36**: 243–266.

Buerkle CA. **2005**. Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes* **5**: 684–687.

Byers D, Meagher T. **1992**. Mate availability in small populations of plant species with homomorphic sporophytic self-incompatibility. *Heredity* **68**: 353–359.

CNPS Rare Plant Program. **2014**. Inventory of Rare and Endangered Plants (online edition, v8-02). *California Native Plant Society, Sacramento, CA*.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK. **2010**. Using environmental correlations to identify loci underlying local adaptation. *Genetics* **185**: 1411–23.

Coppoletta M, Woolhouse S. **2010**. *Conservation Assessment for* Monardella stebbinsii. Quincy, CA.

Crnokrak P, Barrett SCH. **2002**. Perspective: purging the genetic load: a review of the experimental evidence. *Evolution* **56**: 2347–58.

Crnokrak P, Roff DA. **1999**. Inbreeding depression in the wild. *Heredity* **83**: 260–70.

Darwin C. **1859**. *On the Origin of Species by Means of Natural Selection, Or, The Preservation of Favoured Races in the Struggle for Life*. New York City: Signet Classic.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. **2011**. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews: Genetics* **12**: 499–510.

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

Earl DA, VonHoldt BM. **2012**. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**: 359–361.

Ebert D, Haag C, Kirkpatrick M, Riek M, Hottinger JW, Pajunen VI. **2002**. A selective advantage to immigrant genes in a *Daphnia* metapopulation. *Science* **295**: 485–8.

Ellstrand NC, Biggs D, Kaus A, Lubinsky P, McDade LA., Preston K, Prince LM, Regan HM, Roriver V, Ryde OA., Schierenbeck KA 2010. Got Hybridization? A multidisciplinary approach for informing science policy. *BioScience* **60**: 384–388.

Ellstrand N, Elam D. **1993**. Population genetic consequences of small population size: implications for plant conservation. *Annual Review of Ecology and Systematics* **24**: 217–242.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. **2011**. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One* **6**: e19379.

Elvin M a., Sanders AC. **2009**. Nomenclatural changes for *Monardella* (Lamiaceae) in California. *Novon: A Journal for Botanical Nomenclature* **19**: 315–343.

Evanno G, Regnaut S, Goudet J. **2005**. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611–20.

Excoffier L, Lischer HEL. **2010**. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**: 564–7.

Excoffier L, Smouse PE, Quattro JM. **1992**. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479–491.

Frankham R. **1995**. Inbreeding and extinction: a threshold effect. *Conservation Biology* **9**: 792–799.

Frankham R, Gilligan D, Morris D, Briscoe D. **2001**. Inbreeding and extinction: effects of purging. *Conservation Genetics* **2**: 279–285.

Freeland JR, Gillespie J, Ciotir C, Dorken ME. **2010**. Conservation genetics of Hill's thistle (*Cirsium hillii*). *Botany* **88**: 1073–1080.

Frichot E, Schoville SD, Bouchard G, François O. **2013**. Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution* **30**: 1687–99.

Gilpin ME, Soulé ME. **1986**. Minimum viable populations: processes of species extinction. In: Soulé ME, ed. *Conservation biology: the science of scarcity and diversity*. Sunderland, Massachusetts, USA: Sinauer, 19–34.

Gompert Z, Buerkle CA. **2010**. Introgress: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources* **10**: 378–84.

Grant V. **1981**. Natural Hybridization and its Products. *Plant Speciation*. New York: Columbia University Press, 193–282.

Greig JC. **1979**. Principles of genetic conservation in relation to wildlife management in southern-Africa. *South African Journal of Wildlife Research* **9**: 57–78.

Hardham CB, Bartel JA. **1990**. *Monardella stebbinsii* (LAMIACEAE), a new serpentine endemic species from the Northern Sierra Nevada, Plumas County, California. *Aliso* **12**: 693–699.

Hoban S, Gaggiotti O, Bertorelle G. **2013**. Sample Planning Optimization Tool for conservation and population Genetics (SPOTG): a software for choosing the appropriate number of markers and samples (RB O'Hara, Ed.). *Methods in Ecology and Evolution* **4**: 299–303.

Holm S. **1979**. A simple sequentially rejective multiple test procedure. *Scandinavian journal of Statistics* **6**: 65–70.

Honnay O, Jacquemyn H. **2007**. Susceptibility of common and rare plant species to the genetic consequences of habitat fragmentation. *Conservation Biology* **21**: 823–31.

Howie BN, Donnelly P, Marchini J. **2009**. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genetics* **5**: e1000529.

Ingvarsson PK, Whitlock MC. **2000**. Heterosis increases the effective migration rate. *Proceedings of the Royal Society B: Biological Sciences* **267**: 1321–6.

Kay KM, Ward KL, Watt LR, Schemske DW. **2011**. Plant Speciation. In: Harrison S, Rajakaruna N, eds. *Serpentine: The Evolution and Ecology of a Model System*. University of California Press, 71–97.

Keller M, Kollmann J, Edwards P. **2000**. Genetic introgression from distant provenances reduces fitness in local weed populations. *Journal of Applied Ecology* **37**: 647–659.

Kleindorfer S, O'Connor JA, Dudaniec RY, Myers SA, Robertson J, Sulloway FJ. **2014**. Species collapse via hybridization in Darwin's Tree Finches. *The American Naturalist* **183**: 325–341.

Kruckeberg AR. **1986**. An Essay: The stimulus of unusual geologies for plant speciation. *Systematic Botany* **11**: 455–463.

Lamont BB, He T, Enright NJ, Krauss SL, Miller BP. **2003**. Anthropogenic disturbance promotes hybridization between Banksia species by altering their biology. *Journal of Evolutionary Biology* **16**: 551–7.

Laval G, Excoffier L. **2004**. SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics* **20**: 2485–7.

Levin DA, Francisco-Ortega J, Jansen RK. **1996**. Hybridization and the extinction of rare plant species. *Conservation Biology* **10**: 10–16.

Lu F, Glaubitz J, Harriman J, Casstevens T, Elshire R. **2013b**. TASSEL 3.0 Universal Network Enabled Analysis Kit (UNEAK) pipeline documentation. : 1–12.

Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, Buckler ES, Costich DE. **2013a**. Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genetics* **9**: e1003215.

Luan S, Chiang T-Y, Gong X. **2006**. High genetic diversity vs. low genetic differentiation in *Nouelia insignis* (Asteraceae), a narrowly distributed and endemic species in China, revealed by ISSR fingerprinting. *Annals of Botany* **98**: 583–9.

Mandel JR, Milton EF, Donovan LA, Knapp SJ, Burke JM. **2012**. Genetic diversity and population structure in the rare Algodones sunflower (*Helianthus niveus ssp. tephrodes*). *Conservation Genetics* **14**: 31–40.

Martin NH, Bouck AC, Arnold ML. **2006**. Detecting adaptive trait introgression between Iris fulva and I. brevicaulis in highly selective field conditions. *Genetics* **172**: 2481–9.

McKay JK, Christian CE, Harrison S, Rice KJ. **2005**. "How Local Is Local?"— A review of practical and conceptual issues in the genetics of restoration. *Restoration Ecology* **13**: 432–440.

Moritz C. **2002**. Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology* **51**: 238–254.

Motchoulski A, Liscum E. **1999**. *Arabidopsis* NPH3: A NPH1 photoreceptor-interacting protein essential for phototropism. *Science* **286**: 961–964.

O'Grady JJ, Brook BW, Reed DH, Ballou JD, Tonkyn DW, Frankham R. **2006**. Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. *Biological Conservation* **133**: 42–51.

Orlóci L. **1978**. *Multivariate Analysis in Vegetation Research*. The Hague: Dr. W. Junk B. V.

Otto S, Whitton J. **2000**. Polyploid incidence and evolution. *Annual Review of Genetics* **34**: 401–37.

Paschke M, Abs C, Schmid B. **2002**. Relationship between population size, allozyme variation, and plant performance in the narrow endemic *Cochlearia bavarica*. *Conservation Genetics* **3**: 131–144.

Peakall R, Smouse PE. **2012**. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics* **28**: 2537–9.

Peakall R, Smouse P, Huff D. **1995**. Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss *Buchloe dactyloides*. *Molecular Ecology* **4**: 135–147.

Pritchard JK, Stephens M, Donnelly P. **2000**. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–59.

R Core Team. **2014**. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

Raven PH, Axelrod DI. **1978**. *Origin and Relationships of the California Flora*. Berkeley: University of California Press.

Rieseberg LH. **1991**. Hybridization in rare plants: insights from case studies in *Cerocarpus* and *Helianthus*. In: Falk D, Holsinger K, eds. *Genetics and Conservation of Rare Plants*. New York: Oxford University Press, 171–181.

Rieseberg LH. **1997**. Hybrid origins of plant species. *Annual Review of Ecology, Evolution, and Systematics* **28**: 359–89.

Rieseberg LH, Gerber D. **1995**. Hybridization in the Catalina Island mountain mahogany (*Cercocarpus traskiae*): RAPD evidence. *Conservation Biology* **9**: 199–203.

Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, Durphy JL, Schwarzbach AE, Donovan LA, Lexer C. **2003**. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**: 1211–6.

Rieseberg LH, Swensen SM. **1996**. Conservation genetics of endangered island plants. In: Avise JC, Hamrick J, eds. *Conservation Genetics: Case Histories from Nature*. New York: Chapman and Hall, 305–334.

Ryynänen HJ, Tonteri A, Vasemägi A, Primmer CR. **2007**. A comparison of biallelic markers and microsatellites for the estimation of population and conservation genetic parameters in Atlantic salmon (*Salmo salar*). *The Journal of Heredity* **98**: 692–704.

Saccheri I, Kuussaari M, Kankare M. **1998**. Inbreeding and extinction in a butterfly metapopulation. *Nature* **392**: 491–494.

Sanders AC, Elvin MA, Brunell MS. **2013**. *Monardella*. *Jepson eFlora*: Monardella.

Setoguchi H, Mitsui Y, Ikeda H, Nomura N, Tamura A. **2010**. Genetic structure of the critically endangered plant *Tricyrtis ishiiana* (Convallariaceae) in relict populations of Japan. *Conservation Genetics* **12**: 491–501.

Sexton JP, Strauss SY, Rice KJ. **2011**. Gene flow increases fitness at the warm edge of a species' range. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 11704–9.

Soulé ME. **1980**. Thresholds for survival: maintaining fitness and evolutionary potential. In: Soulé ME, Wilcox BA, eds. *Conservation Biology: An Evolutionary-Ecological Perspective*. Sunderland: Sinauer Associates, Inc, 151–170.

Stebbins Jr. GL. **1942**. The genetic approach to problems of rare and endemic species. *Madroño* **6**: 241–258 CR – Copyright © 1942 California Bota.

Tsutsui ND, Suarez AV, Holway DA, Case TJ. **2000**. Reduced genetic variation and the success of an invasive species. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 5948–53.

Vandepitte K, Gristina AS, De Hert K, Meekers T, Roldán-Ruiz I, Honnay O. **2012a**. Recolonization after habitat restoration leads to decreased genetic variation in populations of a terrestrial orchid. *Molecular Ecology* **21**: 4206–15.

Vandepitte K, Honnay O, Mergeay J, Breyne P, Roldán-Ruiz I, De Meyer T. **2012b**. SNP discovery using Paired-End RAD-tag sequencing on pooled genomic DNA of *Sisymbrium austriacum* (Brassicaceae). *Molecular Ecology Resources* **13**: 269–75.

Wang H, McArthur ED, Sanderson SC, Graham JH, Freeman DC. **1997**. Narrow hybrid zone between two subspecies of Big Sagebrush (*Artemisia tridentata*: Asteraceae). IV. reciprocal transplant experiments. *Evolution* **51**: 95–102.

White TA, Perkins SE, Heckel G, Searle JB. **2013**. Adaptive evolution during an ongoing range expansion: the invasive bank vole (Myodes glareolus) in Ireland. *Molecular Ecology* **22**: 2971–85.

Willi Y, van Kleunen M, Dietrich S, Fischer M. **2007**. Genetic rescue persists beyond first-generation outbreeding in small populations of a rare plant. *Molecular Ecology* **274**: 2357–64.

Woolhouse S. **2012**. The Biology and Ecology of Six Rare Plants From Plumas National Forest, Northern California, United States.

Wright S. **1969**. *Evolution of the Genetics of Populations. Vol. 2, The Theory of Gene Frequencies*. Chicago: University of Chicago Press.

Young AG, Pickup M. **2010**. Low S-allele numbers limit mate availability, reduce seed set and skew fitness in small populations of a self-incompatible plant. *Journal of Applied Ecology* **47**: 541–548.