# POPULATION GENETICS AND GENETIC STRUCTURE IN SAN JOAQUIN WOOLY THREADS (*Monolopia congdonii* (A. Gray) B.G. Baldwin

**Bureau of Land Management – UC Berkeley**
**Grant/Cooperative Agreement Number: L12AC20073**

**Susan Bainbridge**
**Jepson Herbarium**
**1001 VLSB #2465**
**University of California**
**Berkeley, CA 94720**

**Ryan O'Dell**
**Bureau of Land Management**
**California**
**&**
**Bruce Baldwin**
**UC Berkeley**

## INTRODUCTION

Conservation of genetic variability is one of the most important and concrete measures ecologists and land managers can implement to maintain viability for specific organisms vulnerable to climate change and to increase success of restoration. Genetic diversity via adaptation and gene flow can help organisms increase chances of surviving environmental changes (Anderson et al. 2012, Reusch et al. 2005, Jump et al. 2008, Doi et al. 2010). Genetic diversity is also threatened by climate change by erosion due to range shifts and/or reductions (Aguilar et al. 2008, Alsos et al. 2012). Genetic diversity is also an important criterion for restoration and conservation planning, and prioritizing protection of populations.

For plant taxa with a declining or limited geographical range or declining populations, perspective on regional and local genetic structure and history is important for developing effective management and restoration planning. Genetic diversity and the processes that shape or maintain that diversity are considered integral for population viability. Inherently rare taxa or those that have had a reduction in population size or an increase in fragmentation may have low levels of genetic variability, in addition to other factors. This increases their vulnerability to inbreeding depression and changes in environmental conditions. At the same time, these taxa are also more vulnerable to unintended consequences of restoration or attempts at enhancement, such as insufficient genetic sampling; swamping of rare genotypes or alleles; or hampered local adaptation by introduction of non-local material.

Understanding landscape patterns of genetic diversity is useful but understanding the mechanism is important too for making conservation and restoration decisions. More than often, the assumption is made that adjacent populations are not isolated or divergent, and geographic distance is the most important factor in structuring diversity. Differentiation between populations can be due to geographic distance and dispersal limitation, but also by local adaptation and by factors related to colonization history such as founder and priority effects (Orsini et al. 2013) or the function of two or more causes of isolation. Local adaptation maybe among the most important consideration for making restoration decisions, but it is difficult detect or predict without strong or apparent ecological differences in habitat.

Self-fertilization is a common mating system in plants and is known to reduce genetic diversity within populations and often overall (Charlesworth and Pannell 2001, Hamrick and Godt 1996, Nybom 2004, Glemin et al. 2006) and potentially put populations at greater risk of extinction. Selfing also changes the partitioning of genetic variation among populations thereby increasing genetic differences between populations and population structure (Charlesworth and Pannell 2001, Hamrick and Godt 1996, Nybom, 2004). Taxa with a self-fertilizing mating system also generally have much greater homozygosity than outcrossing populations and reduced effective population sizes (Pollak 1987, Schoen and Brown 1991).

Despite these differences, potential for local adaptation is not significantly different for selfing, annual plants compared to outcrossing perennials (Leimu and Fischer 2008). Short-lived

and self-compatible species tend to be more strongly differentiated at a smaller scale than long-lived and outcrossing species and therefore the former are expected to show stronger adaptation to local conditions (Linhart and Grant 1996).

Mechanisms other than what is expected from reduced effective population size or what would be expected from inbreeding, contribute to population structure in selfing plant taxa. Low pollen migration rates in selfing taxa relative to outbreeding taxa (e.g., reduced flower size and pollinator attraction, lower investment in pollen) reduces gene flow between populations. This reduces probabilities of establishment of new genes into a population, and thus also limits disruption of adaptive gene complexes in the local environment or outcrossing depression. In addition, reduced recombination in selfing species increases the homogenizing effects of background selection and selective sweeps (Charlesworth et al. 1993, Smith and Haigh 1974). Furthermore, selfing species can establish new populations or subpopulations from a single seed (Schoen and Brown 1991) and it is very likely to be homozygous, and thus populations can establish from a one set of alleles. Populations with reduced genetic diversity due to bottlenecks caused by demographic fluctuations may also occur but can be difficult to detect due to the inherent lack of heterozygosity in selfing populations.

Selfing populations may be more prone to metapopulation dynamics or more frequent colonization and extinction (Ingvarsson 2002). Although, variation between taxa, such as differences in local extinction rates or recolonization patterns must be considered and the relative importance of these in determining genetic diversity considered on a case-by-case basis (Parnell and Chatsworth 1999). In addition, taxa with regional genetic structure are more prone to loss of genetic diversity due to range shifts (Alsos et al. 2012). Thus, landscape scale conservation and conservation of processes that maintain metapopulations and local genetic structure and regional conservation are important for maintaining populations and genetic diversity in selfing organisms.

## STUDY ORGANISM

*Monolopia congdonii* is a federally-listed endangered taxon. It is a therophyte with a short life cycle. It germinates in the winter and flowers in March and April. It occurs in one of the most xeric regions of cismontane California, the South Inner Coast Ranges. Population sizes, size of individual plants, and reproductive output maybe highly dependent on annual climate. It is self-compatible, and unlike other *Monolopia*, ligules on ray flowers are highly inconspicuous. Reproduction is largely through self-fertilization (Mazer and Hendrickson 1993). Population sizes range from <10 individuals to tens of thousands of individuals. Distribution in some populations can be dense and essentially continuous, and distribution at other populations is clustered. Distribution pattern may be due to a variety of factors including disturbance, distribution of appropriate soil conditions, or dispersal and seed limitation. Populations may have decline due to habitat loss due largely to agricultural conversion, vegetation changes, such as invasion of

habitat by non-native plants, decline kangaroo rat populations that help maintain open habitat conditions, and other human facilitated changes. However, small population size may due in part to natural habitat conditions or annual climate conditions.

Habitat is open alkali saltbush scrub and characterized by *Atriplex polycarpa* and non-native annual grassland. Based on museum records (1881 to present, Consortium of California Herbaria 2016), the historic range includes the South Inner Coast Range, from Fresno to Kern County, and adjacent western edge of San Joaquin Valley floor in Fresno County and the valley floor in Kern County; plains on the eastern side of the San Joaquin Valley in Kern and Tulare Counties; the Carrizo, Elkhorn Plains, and the Cuyama Valley. Habitat loss and degradation due to agriculture and urbanization has been extensive on the valley floors. *Monolopia congdonii* has not been observed in most of its historic range on the valley floor, or east side of the San Joaquin Valley, with the exception of two sites about 7 miles southwest of Bakersfield.

Thus, although it may have not been historically rare, range has been greatly reduced, it has specific habitat requirements that limit its distribution, and remaining populations are vulnerable to further habitat loss and habitat degradation. In addition, it is vulnerable to loss of suitable habitat by climate change. Protection of additional populations, habitat and ecological processes, in addition to restoration and management of protected populations are critical to maintaining viability of this taxon.

We development microsatellite markers and used chloroplast markers on samples throughout the range of *Monolopia congdonii*. Results from both genomes were used to assess levels of diversity and patterns of spatial genetic structure. Our results provide insight into the spatial partitioning of genetic variation and will assist in conservation and management of this species

## GENERAL QUESTIONS

1) What are the patterns of genetic diversity in populations and across the range of *Monolopia congdonii*? Are there significant between-year differences in genetic diversity as would be predicted by a therophyte? Are populations of *M. congdonii* genetically impoverished as it might be suggested by its life history traits (self-fertilization, annual)? Or is there evidence of other factors confounding this such as population bottlenecks due to demographic history or founder effects?

2) Is there genetic structure in *M. congdonii* and is structure based on populations, regions, geography, or other factors such as ecology? Is there evidence of cryptic speciation? Are there discrete genetically based Conservation Units and/or Management Units that can help guide conservation efforts?

3) On a local scale, how is genetic diversity structured and on what scale does it vary? Does pattern of genetic diversity show evidence of gene flow between areas or suggest that its distribution was historically more continuous? Does population structure on a local scale follow

the predictions for a highly self-fertilizing taxon? What is the evidence and nature of gene flow between populations? Is genetic divergence of populations predicable based on distance?

## METHODS

**Sampling.** We collected and used in this analysis a total of 881 tissue samples across the range of known extant *Monolopia congdonii* populations (Table 1, Figure 1). The majority of the tissue samples used for this project were collected in 2013. Collection sites as listed in Table 1 and used for summary statistics were all greater than 0.25 miles apart and follow the definition of an element occurrence by the California Natural Diversity Database. They do not necessarily represent biological populations. In 2015, we revisited a subset of sites to collect samples for assessing between-year variation, and in 2016 additional tissue sampling occurred in newly discovered or rediscovered sites. This sampling of new sites included many new sites from the Panoche and Silver Creek watersheds that were needed to inform introduction efforts, in addition to three new sites in the Cuyama Valley. Collections from the Panoche and Silver Creek watersheds were lumped into collection sites (as listed in Table 1) for analysis to ensure sample sizes were large enough for analysis.

The extent of individuals at each collection site was roughly delineated in the field and plants were collected along transects that sampled the greatest extent of individuals within and the edges of distribution, and with the goal of collecting a minimum of 30 individuals across a given site. Because size and density of collection sites were highly variable, distance between individuals collected at a site was variable. We recorded UTM coordinates with a GPS device so that distance between samples could be estimated. We purposefully sampled very few individuals in close proximity to each other; however, some sites were of such limited extent that some sampled individuals were <0.5 meters apart. We also collected samples from isolated outlying patches.

In most cases, we aimed to conduct microsatellite analyses on 20 to 30 individuals per site, but in some cases the number of individuals was too small, and resources were not sufficient to analyze every collection site or 30 samples from each site. If subsampling of our field collections was necessary, we plotted the UTM coordinates and randomly subsampled using a grid. Fewer samples per site (9 to 34) were analyzed in the 2016 collections from Panoche and Silver Creeks, in order to maximize distribution of samples analyzed in that region.

In addition, twenty-seven herbarium accessions (Table 2) were used to sample cpDNA variation from portions of the range of *M. congdonii* that were not field sampled for microsatellites and are possibly extirpated: Mendota Plain, east of Coalinga, Middlewater Plain, and portions of the valley floor and eastern San Joaquin Valley that are west and east of Bakersfield, CA (Eastern Kern Region). Sites in Middlewater Plain have since been

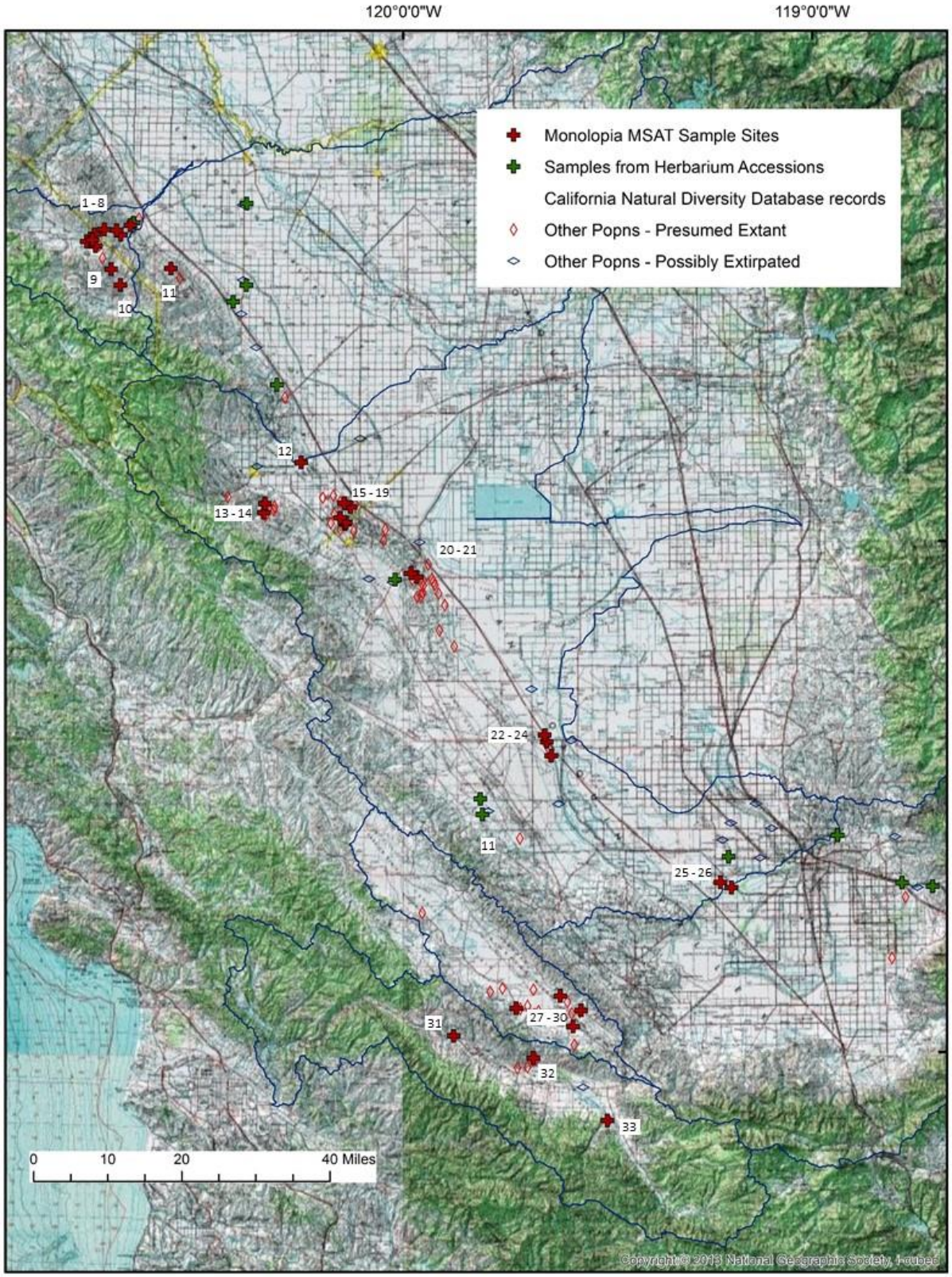
**TABLE 1.** ***Monolopia congdonii*** **sites sampled in 2013-2016 and used in this analysis by major watershed. Sample sizes (N) are number of individuals used for microsatellite by year (), and sample sizes for cpDNA analysis [].**

| POPULATION NUMBER | SITE | LATITUDE | LONGITUDE | N |
|---|---|---|---|---|
| **PANOCHE - SAN LUIS RESERVOIR** | | | | |
| 1 | Panoche | 36.603612 | -120.686982 | 33(2013)/28 (2015) [7] |
| 2 | Lower Panoche Creek | 36.622879 | -120.663444 | 11 (2016) [3] |
| 3 | Panoche T | 36.614375 | -120.727173 | 8 (2016) [2] |
| 4 | Middle Panoche | 36.60472578 | -120.7507908 | 34 (2016) [5] |
| 5 | Silver Creek - Pipeline | 36.58149687 | -120.7482801 | 10 (2016) |
| 6 | Panoche Water Crossing | 36.59401169 | -120.7556958 | 9 (2016) [2] |
| 7 | Silver Creek Ranch | 36.59057494 | -120.7694646 | 32 (2016) [2] |
| 8 | Panoche-Right Angle Cyn | 36.61282049 | -120.6966122 | 10 (2016) [2] |
| 9 | Silver Creek 1080 | 36.53645304 | -120.7092005 | 9 (2016) [2] |
| 10 | Silver Creek 1320 | 36.50596914 | -120.6866987 | 9 (2016) |
| **UPPER DRY** | | | | |
| 11 | Monocline | 36.53876 | -120.564203 | 32 (2016) [5] |
| **TULARE LAKE BED** | | | | |
| 12 | Pleasant Valley ER | 36.160662 | -120.246438 | 39 (2013) [6] |
| 13 | Jacalitos 1 | 36.079893 | -120.333886 | 18 (2013) [3] |
| 14 | Jacalitos 2 | 36.06212186 | -120.3362698 | 29 (2013) [5] |
| 15 | Kettleman BLM 9100 | 36.08090687 | -120.1420705 | 23 (2013) [3] |
| 16 | Kettleman BLM 6800 | 36.074007 | -120.126789 | 27 (2013) [8] |
| 17 | Kettleman BLM 7300 | 36.052649 | -120.152116 | 29 (2013)/10 (2015) [4] |
| 18 | Kettleman BLM 8300 | 36.04311568 | -120.1378228 | 25 (2013) [3] |
| 19 | Kettleman BLM 8500 | 36.04000351 | -120.1407208 | 24 (2013)/10(2015) [3] |
| 20 | Arroyo Conejo | 35.9439680 | -119.9806930 | 19 (2013) [2] |
| 21 | Arroyo Conejo Crossing | 35.9366520 | -119.9665210 | 28 (2013) [4] |
| 22 | Lost Hills North | 35.6269900 | -119.6595590 | 40 (2013) [3] |
| 23 | Lost Hills Arco | 35.6140950 | -119.6547500 | 8 (2103)/17(2015) [3] |
| 24 | Lost Hills South | 35.5867090 | -119.6439670 | 40 (2103) [6] |
| 25 | West Kern Water Bank | 35.3369740 | -119.2401430 | 35 (2013) [5] |
| 26 | East Kern Water Bank | 35.3265150 | -119.2143300 | 32 (2013) [4] |
| **CARRIZO and ELKHORN PLAINS** | | | | |
| 27 | Traver Ranch | 35.0562370 | -119.5947450 | 32(2013)/18(2015) |
| 28 | Elkhorn Road | 35.1161640 | -119.6247830 | 14(2013) [5] |
| 29 | KCL Campground | 35.0925300 | -119.7300880 | 36 (2013) [6] |
| 30 | Cochora Ranch | 35.0877450 | -119.5754890 | 34 (2013) [4] |
| **CUYAMA VALLEY** | | | | |
| 31 | Lower Cuyama | 35.0385338 | -119.8792054 | 23 (2016) [6] |
| 32 | Horse Canyon | 34.9952032 | -119.6892399 | 23 (2016) [3] |
| 33 | Santa Barbara Canyon | 34.8720358 | -119.5140441 | 23 (2016) [4] |

**Table 2. *Monolopia congdonii* herbarium accessions sampled for additional geographic representation in cpDNA haplotype analysis.**

| MENDOTA PLAIN | ACCESSION |
|---|---|
| | JEPS5254, 6 mi s of Mendota; Jepson #16,987 |
| | UC762544; 17 mi s of Mendota; Hoover #4240; 1940 |
| | JEPS5254, Jepson 16987, Fresno County: 6 mi s of Mendota; Mendota Plain |
| | GH427296; Ferris 10333; 26 mi s of Mendota |
| | RSA27117; Ferris 10333; 26 mi s of Mendota |
| | JEPS5254; Jepson 16987; Fresno Co., 6 mi s of Mendota |
| | JEPS5254; Jepson 16987; Fresno Co., 6 mi s of Mendota |
| | JEPS5252; Jepson 16979; Fresno Co., Hayes Station |
| | UC762503; Hoover 3291; Fresno Co., btwn Arroyo Hondo and Cantua Ck |
| | JEPS5259; Hoover 4240; Fresno Co., 17 mi s of Mendota |
| | UC762544; Hoover 4240; Fresno Co., 17 mi s of Mendota |
| EASTERN VALLEY | ACCESSION |
| | UC1229833; n of Rosedale; Bacigalupi #4368 |
| | UC1229833; n of Rosedale; Bacigalupi #4368 |
| | UC569076; Munz #13,657; Caliente Ck, 15 mi east of Bakersfield |
| | JEPS5260, Krames s.n., Kern County: S.P.R.R. right-of-way between Edison and Bena |
| | JEPS5260, Krames s.n., Kern County: S.P.R.R. right-of-way between Edison and Bena, 4/24/35 |
| | CAS606375, Bacigalupi 4368, Kern County: 1/4 mile N of Rosedale Hwy (Bakersfield to Buttonwillow and McKittrick), 10.1 miles W of Kern River Bridge at western outskirts of Bakersfield, 3/10/54 |
| | UC762503, Hoover 3291, Fresno County: between Arroyo Hondo and Cantua Creek |
| COALINGA | ACCESSION |
| | UC614634, Constance 2098, Fresno County: 12 mi ne of Coalinga Coalinga-Fresno road; San Joaquin Valley |
| | GH427292; Constance 2098; 12 mi ne Coalinga |
| | UC614634; Constance 2098; 12 mi ne Coalinga |
| | UC614634; Constance 2098; 12 mi ne Coalinga |
| | GH427292; Constance 2098; 12 mi ne Coalinga |
| | RSA176565; Raven 16978; Kings Co., 3mi ne Reef City; n=11 voucher |
| MIDDLEWATER PLAIN | ACCESSION |
| | CAS606466, Twisselmann 582, Kern County: Middlewater Plain, near the Middlewater Pump Station. (Sand dunes 1/4 mi NW of the station on the Williams Lease road.), 2/25/1952 |
| | CAS390498, Twisselmann 972, Kern County: Middlewater Plain: 2 mi S of Middlewater Pump Station, 4/2/1954 |
| | CAS606466, Twisselmann 582, Kern County: Middlewater Plain, near the Middlewater Pump Station. (Sand dunes 1/4 mi NW of the station on the Williams Lease road.), 2/25/1952 |

**FIGURE 1. Location of *Monolopia congdonii* sample sites, major watershed boundaries (USDA 2013), and California Natural Diversity Database occurrences records (CNDDB 2016).**



rediscovered. Information from these areas was needed to improve understanding of historical gene flow and phylogeography.

**DNA Extraction:** Leaf samples were cleaned using a small brush and stored at 4 degrees C until used for DNA extraction. Genomic DNA was extracted from leaf tissue using the DNeasy Mini plant kit (Qiagen, Valencia, CA). Samples were ground dry using glass beads in a Mini-Bead-Beater-16 (BioSpec Products, Inc., Bartlesville, Oklahoma) or ground directly in the API buffer.

**Microsatellite Development.** Genomic DNA from two individuals of *M. congdonii,* one from the vicinity of Lost Hills (Kern County) and one from Panoche (Fresno County), were sent to the Savannah River Ecological Laboratories (SREL) at the University of Georgia for microsatellite marker development. Samples were prepared for paired-end shotgun sequencing on an Illumina platform. Microsatellite repeats were identified from resulting sequences and used for further consideration if found in both pairs of sequences. Potential primers were selected with the following initial criteria: adequate flanking regions; occurred no more than 3 times in the sequences; tri- or tetranucleotide repeats; and length in the range of 300 to 600 bp.

Twenty-four of these primers that were polymorphic when tested on twenty-two individuals of *M. congdonii* from throughout the extant range of the species and two individuals of its sister species, *M*. *major*, were tested. AutoDimer (Vallone and Butler 2004) and Multiplex Manager (Holleley and Geerts 2009) were used to determine potential interactions between primers and complimentary sets of primers that could be amplified in single reactions and reliably sized. Ultimately after testing various combinations and PCR conditions, the twenty-four primers could be amplified and sized using three multiplex reactions. Primers were also used in singleplex reactions for four samples to validate results of the multiplex reactions. Designing and testing the multiplex reactions required some up-front investment allowed for more samples and markers to be included in the study (Hayden et al. 2008, Culley et al. 2013).

**Microsatellite Analysis.** Forward primers were tagged at the 5' with one of four tags developed by Blacket et al. (2013): Tag A (GCCTCCCTCGCGCCA), Tag B (GCCTTGCCAGCCCGC), Tag C (CAGGACCAGGCTACCGTG), and Tag D (CTAGTTATTGCTCAGCGGT CGGAGAGCCGAGAGGTG). PCR amplifications were performed in a total reaction volume of 8-10 µL containing 5 µL Qiagen Multiplex master mix, 0.2 µM untagged R primer, 0.05 µM tag-modified F primer, 0.2 µM fluorescently labeled tag with either 6-FAM, PET, NED or VIC, 1 µL of undiluted DNA template, and DNase free water. Amplifications for two sets of loci were conducted using a touchdown PCR protocol beginning with an initial denaturation step of 15 minutes at 95° C; followed by 20 cycles at 95° C for 30 s, annealing at a temperature of 65° C for 60 s (decreased by 0.5° C per cycle), 72° C for 1 min; and 20 cycles at 95° C for 30 s, 55° C for 30 s, and 72° C for 60 s. The final extension was 72 ° C for 10 min. The third set of loci were similarly amplified but with 18 cycles of

touchdown using a start annealing temperature of 66° C for 60 s and the 72° C extension for 45 s 18 cycles, followed by 22 cycles of annealing at 56° C.

Amplification products were confirmed on 1.7% gels and were loaded on an ABI 3730 DNA analyzer (Applied Biosystems, Foster City, California, USA) with 0.15 µL GS-500 LIZ size standard (Applied Biosystems) to allow allele length sizing. Electrophoretic results were initially scored using Genemapper 5.0 (Applied Biosystems) followed by visual confirmation. Loci that did not amplify for a given sample in the multiplex reaction were amplified in singleplex reactions. In order to check for errors and spurious results, approximately 15% of the multiplex reactions were duplicated. In addition, alleles rare to the samples (< 10 individuals) were verified in singleplex reactions. The presence of null alleles was tested using Micro-Checker version 2.2.3 (van Oosterhout et al., 2004).

**Chloroplast DNA.** Four chloroplast DNA (cpDNA) regions were screened for polymorphism using eight *M*. *congdonii* samples. Primer pairs used for amplifying and sequencing the cpDNA regions were as follows: 1587MADIA and 607R (Panero and Crozier 2003) for the 3'end of *ndh*F and the *ndh*F-*ycf*1 intergenic spacer (hereafter, *ndh*F), F71 (Jordan et al. 1996) and R1516 (Kelchner and Clark 1997) for the *rpl*16 intron, *psb*AF and *trn*HR (McGlaughlin and Friar 2011) for the *psb*A--*trn*H intergenic spacer, and trnK-3914F (Johnson and Soltis 1994) and 884R (Panero and Crozier 2003) for the 5' *trn*K intron and 5' portion of *mat*K. PCR amplification of the four cpDNA genes involved an initial denaturation at 95˚C for 5 min followed by 35 cycles of denaturation at 95˚C for 1 min, primer annealing at 48˚C for 45 sec (+2 sec/cycle), and primer extension at 72˚C for 1 min (+9 min in final cycle), and a final extension at 72˚C for 10 min. Both DNA strands were sequenced for each sample. Exonuclease I and shrimp alkaline phosphatase were used to remove excess nucleotides from PCR products using the PCR Product Pre-Sequencing Kit (70995, United States Biochemical Corp., Cleveland, Ohio, USA). Sanger sequencing of PCR products was conducted at the UC Berkeley DNA Sequencing Facility (Barker Hall) or the Evolutionary Genetics Lab (Department of Integrative Biology) with the same primers used for PCR. Sequences were aligned using Geneious 6.1 (http://www.geneious.com, Kearse et al. 2012). Haplotype relationships were resolved by parsimony using PAUP* 4.0 (Swofford 2002), as implemented in Geneious, using sequences of other species of *Monolopia* as the outgroup to discern nucleotide substitutions or length mutations from ancestral states within *M. congdonii*.

One hundred fifty-one samples were sequenced using just the two cpDNA regions that showed variation within *M*. *congdonii* (*ndh*F and *rpl*16 intron). These samples included material from throughout the range of *M. congdonii*; namely, from the six a priori sample regions where field collections were made for microsatellite analysis, augmented by twenty-seven additional samples from herbarium specimen DNA extractions representing other occurrences, including putatively extirpated populations (Table 1). The samples from herbarium specimens were from: Mendota Plain, east of Coalinga, Middlewater Plain, portions of the valley floor in Kern County and eastern San Joaquin Valley that are west and east of Bakersfield. Populations in

Middlewater Plan have since been rediscovered. Information from these areas was needed to improve understanding of historical population structure.

**Analysis of Microsatellite Markers.** GenAlex version 6.5 (Peakall and Smouse 2006) was used to calculate sample sizes (N), number of alleles (Na), number of effective alleles (Ne), Shannon's information index (I), expected, observed, or unbiased expected heterozygosity (Ho, He, and uHe), fixation index (F), percentage of polymorphic loci (P), the inbreeding coefficient within individuals relative to the total (Fit) across all sampled individuals. In addition, deviations from Hardy–Weinberg equilibrium (HWE) were estimated using GenAlEx version 6.5 and P values for tests of deviation from HWE were adjusted using a sequential Bonferroni correction (Rice 1989).

**Population Structure with Microsatellite Markers.** The number of genetic clusters within the dataset was estimated and compared using Bayesian methods utilized in INSTRUCT (Gao et al. 2007) and STRUCTURE 2.3.4 (Pritchard et al. 2000). The former does not assume Hardy-Weinberg Equilibrium and it is a more appropriate method for this taxon. Two to thirty-three ancestral populations (K) were tested with 5 simulations run for each. Burn-in was set at $10^5$ and the number of iterations was 2 x $10^5$ under the admixture model with independent allele frequencies. Results were analyzed for optimal population size using the delta K method (Evanno et al. 2000) using Adegenet (Jombart 2008). In addition, the percent of individuals with significant admixture was compared for each method. Identified clusters were investigated by a hierarchical AMOVA in GenAlEx (Peakall and Smouse 2006).

Discriminant Analysis of Principal Components (DAPC) was used to assess genetic structure using the pre-defined sample sites (Jombart et al. 2008, 2010). The method applies PCA to the data before discriminant analysis in order to uncorrelated variables. It has the advantages of 1) independence from any assumptions about the population model, and 2) displaying relationships among clusters and hierarchy. DAPC was performed in the Adegenet package for R (function dapc) using the predefined sample sites 1-33. In all analyses, 50 principal components of PCA were retained in the data. Nei's genetic distance was used to make a dendrogram with bootstrap support. The R package POPPR and function *aboot* (Kamvar et al. 2017) was used to randomly sample loci with replacement to provide support for the clades. This procedure avoids any 'hyperdiverse' loci from overestimating diversity.

Spatial genetic analyses were performed using GenAlEx (Peakall and Smouse 2006). Isolation by distance (IBD) was tested for the entire dataset by sample and by collection site. The Mantel test was used on a matrix of Fst/(1-Fst) versus geographic distance (log), with 999 random permutations in as recommended by Rousset (1997).

Current migration rates among collection sites was estimated in BayesAss (version 3.0; Wilson and Rannala 2003). BayesAss 3.0 is a Bayesian method that uses a Markov chain Monte Carlo (MCMC) algorithm and multilocus genotypes to estimate the proportion of migrants per generation for each sampling site (Wilson and Rannala 2003). Bayesass does not require data sets to conform to Hardy-Weinberg Equilibrium. Five independent replicate runs of $10^7$ iterations with $3\times10^6$ iterations discarded as burn-in. The program Migrate 3.2.1 (Beerli and Felsenstein 2001) is more sensitive to historical migration than BayesAss. The Maximum Likelihood approach in Migrate was also used on a subset of populations (discussed below).

BOTTLENECK 1.2.02 (Piry et al. 1999) was used to test for evidence of recent bottlenecks by assessing allele frequencies relative to the distribution expected under mutation-drift equilibrium, and by tests for excess heterozygosity relative to the number of alleles in the population. With mutation-drift equilibrium, rare alleles (frequency <0.1%) are numerous due to genetic drift. A recent bottleneck would greatly reduce the number of rare alleles (Cristescu et al. 2010). A bottleneck causes a loss of alleles at a faster rate than a loss of gene diversity also known as ''heterozygosity excess'' (Nei 1987). All three models were tested: infinite alleles (IAM); stepwise mutation (SMM); and the two-phase model (TPM), which is a combination of the first two and 50% of each was used.

## RESULTS

**Loci Used in Analysis.** A total of 24 microsatellite markers were scored in the 881 individuals analyzed from samples made in 2013, and 2015-2016. All loci were segregating across the collection sites, but only 19 were retained for population analysis due to spurious results with the other 5 markers. A total of 134 alleles were found in the nineteen loci and 881 samples. All but one collection site had at least one monomorphic loci, but all loci were polymorphic across the range of the species. The number of alleles for a given loci across all samples was as low as 2 (Mc7) and a maximum of 12 (Mc29) but the average was 7.05 ($\pm$ 0.86) alleles (Table 3). See section below for discussion of heterozygosity and Fixation Index for the loci.

**Genetic Diversity, Richness and Private Alleles.** Statistics for genetic diversity for the thirty-three collection sites are reported in Table 4 and Figure 2. The nineteen loci were polymorphic in one or more sites. Across all samples, the mean number of alleles ($N_a$) per loci was 7.05 ($\pm$ 0.743), and the mean of the effective number of alleles ($N_e$) was 3.22 ($\pm$ 0.399). The overall expected heterozygosity (He) is 0.577 ($\pm$ 0.058); the observed heterozygosity is 0.1 ($\pm$ 0.0); and the overall Fixation index (F) is nearly 1 (0.986 $\pm$ 0.005). The Shannon Information Index (I) is 1.21 ($\pm$ 0.138); and the between collection site genetic differentiation (Fst) was 0.573 (0.22).

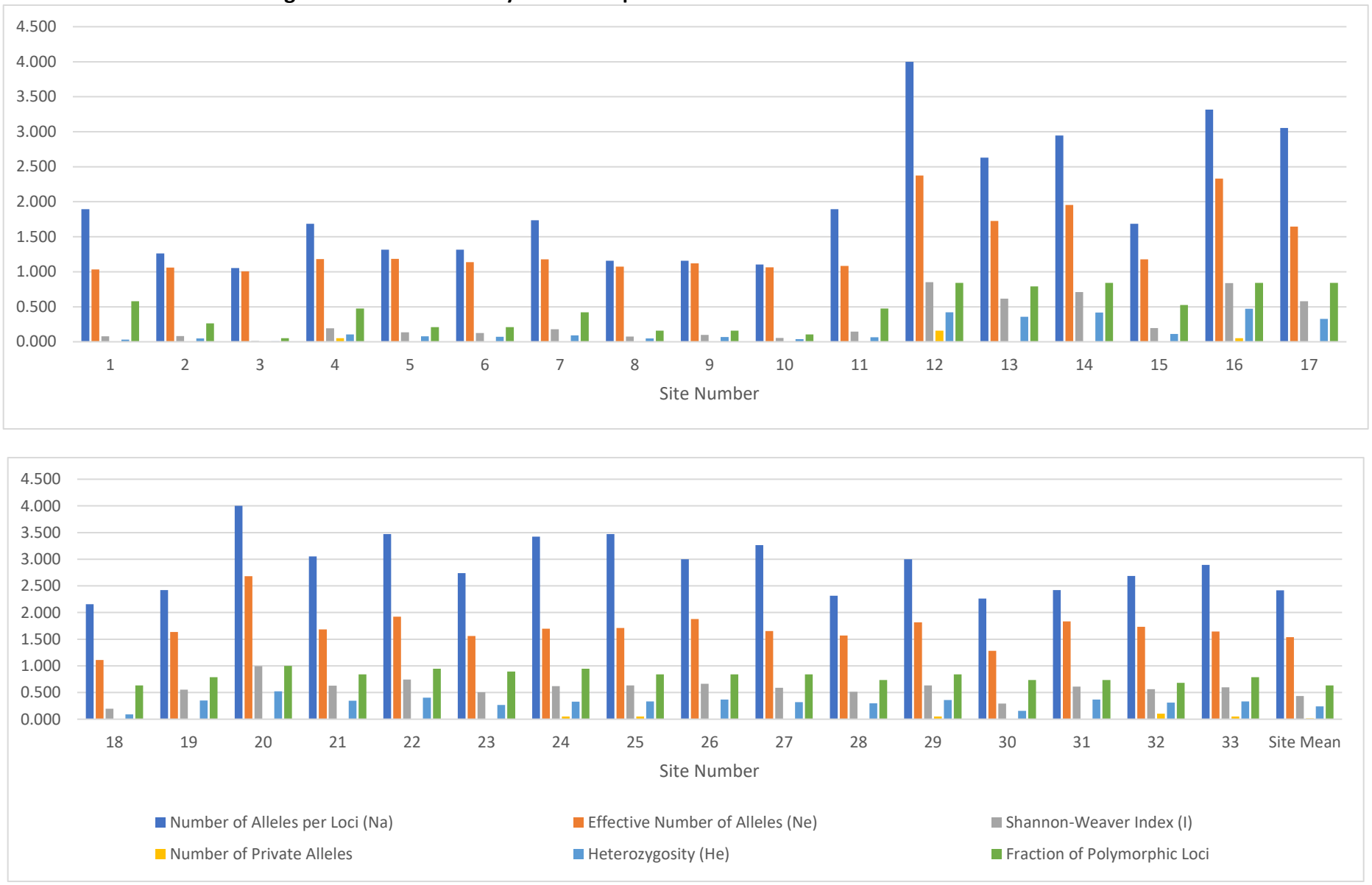
**Table 3. Sample sizes (N), number of alleles (Na), number of effective alleles (Ne), Shannon Information Index (I), observed heterozygosity (Ho), expected heterozygosity (He), unbiased expected heterozygosity ($uH_e$) and Fixation Index (F) for the loci used in analysis.**

| Locus | N | Na | Ne | I | Ho | He | uHe | F |
|---|---|---|---|---|---|---|---|---|
| **Mc3** | 877 | 9 | 5.98 | 1.97 | 0.02 | 0.83 | 0.83 | 0.98 |
| **Mc4** | 880 | 8 | 4.87 | 1.77 | 0.01 | 0.79 | 0.80 | 0.99 |
| **Mc6** | 877 | 11 | 7.30 | 2.15 | 0.00 | 0.86 | 0.86 | 0.99 |
| **Mc7** | 880 | 2 | 1.06 | 0.13 | 0.00 | 0.06 | 0.06 | 0.98 |
| **Mc10** | 878 | 3 | 1.56 | 0.60 | 0.00 | 0.36 | 0.36 | 1.00 |
| **Mc11** | 881 | 7 | 2.87 | 1.33 | 0.02 | 0.65 | 0.65 | 0.97 |
| **Mc14** | 878 | 10 | 5.48 | 1.90 | 0.01 | 0.82 | 0.82 | 0.99 |
| **Mc15** | 881 | 5 | 1.27 | 0.45 | 0.00 | 0.21 | 0.21 | 0.99 |
| **Mc16** | 867 | 11 | 3.80 | 1.63 | 0.01 | 0.74 | 0.74 | 0.99 |
| **Mc18** | 877 | 10 | 3.58 | 1.58 | 0.01 | 0.72 | 0.72 | 0.99 |
| **Mc19** | 876 | 3 | 2.02 | 0.73 | 0.00 | 0.51 | 0.51 | 1.00 |
| **Mc21** | 879 | 9 | 2.89 | 1.31 | 0.06 | 0.65 | 0.65 | 0.91 |
| **Mc22** | 881 | 10 | 2.53 | 1.26 | 0.00 | 0.61 | 0.61 | 1.00 |
| **Mc29** | 875 | 12 | 2.86 | 1.32 | 0.00 | 0.65 | 0.65 | 1.00 |
| **Mc34** | 877 | 4 | 1.24 | 0.41 | 0.00 | 0.19 | 0.19 | 1.00 |
| **Mc35** | 870 | 4 | 3.61 | 1.33 | 0.00 | 0.72 | 0.72 | 1.00 |
| **Mc39** | 879 | 3 | 1.18 | 0.30 | 0.00 | 0.16 | 0.16 | 0.98 |
| **Mc43** | 880 | 6 | 3.47 | 1.34 | 0.01 | 0.71 | 0.71 | 0.99 |
| **Mc45** | 873 | 7 | 3.53 | 1.47 | 0.00 | 0.72 | 0.72 | 1.00 |
| **Mean** | 877 | 7.05 | 3.22 | 1.21 | 0.01 | 0.58 | 0.58 | 0.99 |
| **SE** | 0.86 | 0.74 | 0.40 | 0.14 | 0.00 | 0.06 | 0.06 | 0.00 |

**Table 4. MICROSATELLITE sample sizes (N), number of alleles per locus (Na), number of effective alleles (Ne), Shannon's information index (I), expected, observed, or unbiased expected heterozygosity (Ho, He, and uHe), fixation index (F), and percent of polymorphic loci (%P) by sample site and average.**

| Site | N | Na | Ne | I | Ho | He | uHe | F | P% |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 61 | 1.89 (0.24) | 1.03 (0.01) | 0.08 (0.02) | 0 | 0.03 (0.01) | 0.03 (0.01) | 0.99 (0.01) | 57.89% |
| **2** | 11 | 1.26 (0.1) | 1.06 (0.03) | 0.08 (0.03) | 0 | 0.05 (0.02) | 0.05 (0.02) | 0.79(0.11) | 26.32% |
| **3** | 8 | 1.05 (0.05) | 1.01 (0.01) | 0.01 (0.01) | 0.01 (0.01) | 0.01 (0.01) | 0.01 (0.01) | -0.07 (0.01) | 5.26% |
| **4** | 34 | 1.68 (0.2) | 1.18 (0.08) | 0.19 (0.07) | 0 | 0.1 (0.04) | 0.11 (0.04) | 1 (0) | 47.37% |
| **5** | 10 | 1.32 (0.17) | 1.18 (0.13) | 0.13 (0.07) | 0 | 0.08 (0.04) | 0.08 (0.04) | 1 (0) | 21.05% |
| **6** | 9 | 1.32 (0.17) | 1.14 (0.09) | 0.13 (0.07) | 0 | 0.07 (0.04) | 0.08 (0.04) | 1 (0) | 21.05% |
| **7** | 32 | 1.74 (0.26) | 1.18 (0.1) | 0.18 (0.08) | 0 | 0.09 (0.04) | 0.09 (0.04) | 1 (0) | 42.11% |
| **8** | 10 | 1.16 (0.09) | 1.07 (0.04) | 0.08 (0.04) | 0 | 0.05 (0.03) | 0.05 (0.03) | 1 (0) | 15.79% |
| **9** | 9 | 1.16 (0.09) | 1.12 (0.07) | 0.1 (0.05) | 0 | 0.07 (0.04) | 0.07 (0.04) | 1 (0) | 15.79% |
| **10** | 9 | 1.11 (0.07) | 1.06 (0.05) | 0.05 (0.04) | 0 | 0.04 (0.03) | 0.04 (0.03) | 1 (0) | 10.53% |
| **11** | 32 | 1.89 (0.3) | 1.08 (0.03) | 0.14 (0.05) | 0 | 0.07 (0.02) | 0.07 (0.02) | 0.89 (0.08) | 47.37% |
| **12** | 39 | 4 (0.51) | 2.37 (0.34) | 0.85 (0.14) | 0 | 0.42 (0.07) | 0.43 (0.07) | 0.99 (0) | 84.21% |
| **13** | 16 | 2.63 (0.27) | 1.72 (0.12) | 0.62 (0.09) | 0.03 (0.01) | 0.36 (0.05) | 0.37 (0.05) | 0.93 (0.03) | 78.95% |
| **14** | 29 | 2.95 (0.35) | 1.96 (0.15) | 0.71 (0.1) | 0 | 0.42 (0.05) | 0.42 (0.06) | 1 (0) | 84.21% |
| **15** | 23 | 1.68 (0.17) | 1.18 (0.07) | 0.2 (0.06) | 0 | 0.11 (0.04) | 0.11 (0.04) | 0.96 (0.03) | 52.63% |
| **16** | 27 | 3.32 (0.38) | 2.33 (0.24) | 0.84 (0.12) | 0 | 0.47 (0.06) | 0.48 (0.06) | 1 (0) | 84.21% |
| **17** | 39 | 3.05 (0.35) | 1.65 (0.12) | 0.58 (0.09) | 0 | 0.33 (0.05) | 0.33 (0.05) | 0.99 (0.01) | 84.21% |
| **18** | 25 | 2.16 (0.24) | 1.11 (0.03) | 0.2 (0.04) | 0.01 (0.01) | 0.09 (0.02) | 0.09 (0.02) | 0.88 (0.04) | 63.16% |
| **19** | 34 | 2.42 (0.22) | 1.63 (0.08) | 0.55 (0.07) | 0.01 (0.01) | 0.35 (0.04) | 0.35 (0.04) | 0.97 (0.01) | 78.95% |
| **20** | 19 | 4 (0.4) | 2.68 (0.3) | 1 (0.12) | 0.03 (0.01) | 0.52 (0.06) | 0.54 (0.06) | 0.91 (0.05) | 100.00% |
| **21** | 28 | 3.05 (0.34) | 1.68 (0.12) | 0.63 (0.09) | 0.02 (0.01) | 0.34 (0.05) | 0.35 (0.05) | 0.96 (0.02) | 84.21% |
| **22** | 40 | 3.47 (0.35) | 1.92 (0.17) | 0.74 (0.09) | 0.01 (0) | 0.41 (0.05) | 0.41 (0.05) | 0.98 (0.01) | 94.74% |
| **23** | 25 | 2.74 (0.28) | 1.56 (0.18) | 0.5 (0.09) | 0.01 (0) | 0.27 (0.05) | 0.27 (0.05) | 0.98 (0.01) | 89.47% |
| **24** | 40 | 3.42 (0.36) | 1.7 (0.15) | 0.62 (0.1) | 0.02 (0.01) | 0.33 (0.05) | 0.33 (0.05) | 0.96 (0.02) | 94.74% |
| **25** | 35 | 3.47 (0.36) | 1.71 (0.14) | 0.63 (0.11) | 0.01 (0.01) | 0.33 (0.06) | 0.34 (0.06) | 0.98 (0.01) | 84.21% |
| **26** | 32 | 3 (0.31) | 1.88 (0.19) | 0.66 (0.1) | 0 | 0.37 (0.06) | 0.38 (0.06) | 0.99 (0.01) | 84.21% |
| **27** | 53 | 3.26 (0.42) | 1.65 (0.13) | 0.59 (0.1) | 0.01 (0) | 0.32 (0.05) | 0.32 (0.05) | 0.96 (0.02) | 84.21% |
| **28** | 13 | 2.32 (0.24) | 1.57 (0.11) | 0.51 (0.09) | 0 | 0.3 (0.05) | 0.31 (0.05) | 1 (0) | 73.68% |
| **29** | 36 | 3 (0.41) | 1.82 (0.18) | 0.64 (0.1) | 0 | 0.36 (0.05) | 0.37 (0.06) | 1 (0) | 84.21% |
| **30** | 34 | 2.26 (0.26) | 1.28 (0.1) | 0.3 (0.08) | 0 | 0.16 (0.04) | 0.16 (0.05) | 0.97 (0.02) | 73.68% |
| **31** | 23 | 2.42 (0.28) | 1.83 (0.16) | 0.61 (0.1) | 0 | 0.37 (0.06) | 0.38 (0.06) | 0.99 (0.01) | 73.68% |
| **32** | 23 | 2.68 (0.4) | 1.73 (0.19) | 0.56 (0.12) | 0.03 (0.03) | 0.31 (0.06) | 0.32 (0.06) | 0.9 (0.06) | 68.42% |
| **33** | 23 | 2.89 (0.3) | 1.64 (0.11) | 0.6 (0.09) | 0.06 (0.03) | 0.33 (0.05) | 0.34 (0.05) | 0.87 (0.05) | 78.95% |
| **Mean for Sites** | 26.58 (0.51) | 2.42 (0.06) | 1.54 (0.03) | 0.43 (0.02) | 0.01 (0) | 0.24 (0.01) | 0.24 (0.01) | 0.96 (0.01) | 70.76% |

**Figure 2. Genetic Diversity Across Sample Across Collection Sites from Microsatellite Data.**



The number of polymorphic loci at a site was as low as one (Site 3, Lowest Panoche) and as many as nineteen (Site 20, Arroyo Conejo) with a mean of thirteen to fourteen per site. The lowest genetic diversity based on mean number of alleles, number of polymorphic loci, expected heterozygosity, Shannon Information Index, and Fixation Index was for the Panoche T site (Collection Site 3, Na=1.05, Ne =1.01, PPL=0.05%, He=0.01, I=0.01, F=0.05). The highest mean number of alleles (Na=4.00, Ne=2.68) was at Arroyo Conejo and Pleasant Valley Ecological Reserve (Collection sites 20 and 12), although the highest percent of polymorphic loci was for Arroyo Conejo, Lost Hills North and Lost Hills South (collection sites 20, 22, 24, PPL=100%, 94.74% and 94.74%). The lowest Fixation Indices and lowest levels of homozygosity were found in the Santa Barbara Canyon, Horse Canyon, Arroyo Conejo and Pleasant Valley Ecological Reserves (F= 0.87 to 0.93, Ho= 0.03 to 0.06). Allelic richness (Na) and collection site divergence (Fst) are inversely related (Figure 3). Inbreeding coefficients (Fis) range from 0.88 to 1.0.

A total of eleven private alleles were found among the 33 collection sites. Middle Panoche, Kettleman BLM 6800, Lost Hills South, West Kern Water Bank, KCL Campground and Santa Barbara Canyon (Sites 4, 16, 24, 25, 29 and 33) each have a single private allele. There were two private alleles at Horse Canyon (Site 32). Finally, there were three private alleles at the Pleasant Valley Ecological Reserve (Site 12). These private alleles were rare in each of these populations (frequency = 0.025 to 0.111).

Allele frequencies at six of the collection sites suggest there may have been recent population bottlenecks, but heterozygote excess was not detected at any of the sites. Allele frequencies at the Panoche Creek and Silver Creek Collection sites 5-6, 8, and 9-10 do not fit the normal L-shaped distribution and indicated very rare alleles (frequency <0.1) are absent. All other collection sites in those watersheds have an L-shaped distribution, although they also have a very high number of common alleles (Figure 4). The Cuyama Valley collection site 31 also had a shifted pattern of alleles suggesting a recent bottleneck.

**Genotype Diversity.** Statistics for genotype diversity are reported in Table 5 and Figure 5. The total number of genotypes across the sample range is 193 with a diversity index of 0.94, but the effective number of genotypes is 17 (Geff= 17.10, G/N = 0.22). Evenness across the species was 0.088. Half the individuals share only five genotypes and 181 genotypes were found in ten or lower individuals (Figure 6). The Shannon-Weiner diversity index for genotypes is 1.63. The number of genotypes per site averages 7.5 ($\pm$ 5.7). It ranges from a low of 1 genotype (several sites in the Panoche/Silver Creek watershed) to 23 genotypes (Pleasant Valley Ecological Reserve). Isolation by distance (IBD) was significant (Mantel, $R^2$=0.52, P=0.001).

Nineteen other genotypes were restricted to a single population, but none of the populations is represented only by unique or rare genotypes. The highest number of genotypes was found in the Pleasant Valley Ecological Reserve ($G_{num}$=23) and Arroyo Conejo 1 ($G_{num}$=17) populations. Arroyo Conejo 1 is exceptionally diverse as almost every individual represented a unique genotype ($G_{num}$ =17, Geff 14.44, G/N=0.894, Gd=0.982). Pleasant Valley Ecological

**Figure 3. Relationship between allelic richness and genetic distinctiveness for twelve sites sampled for microsatellite data. The relationship is identical for richness versus mean pairwise Fst and He.**

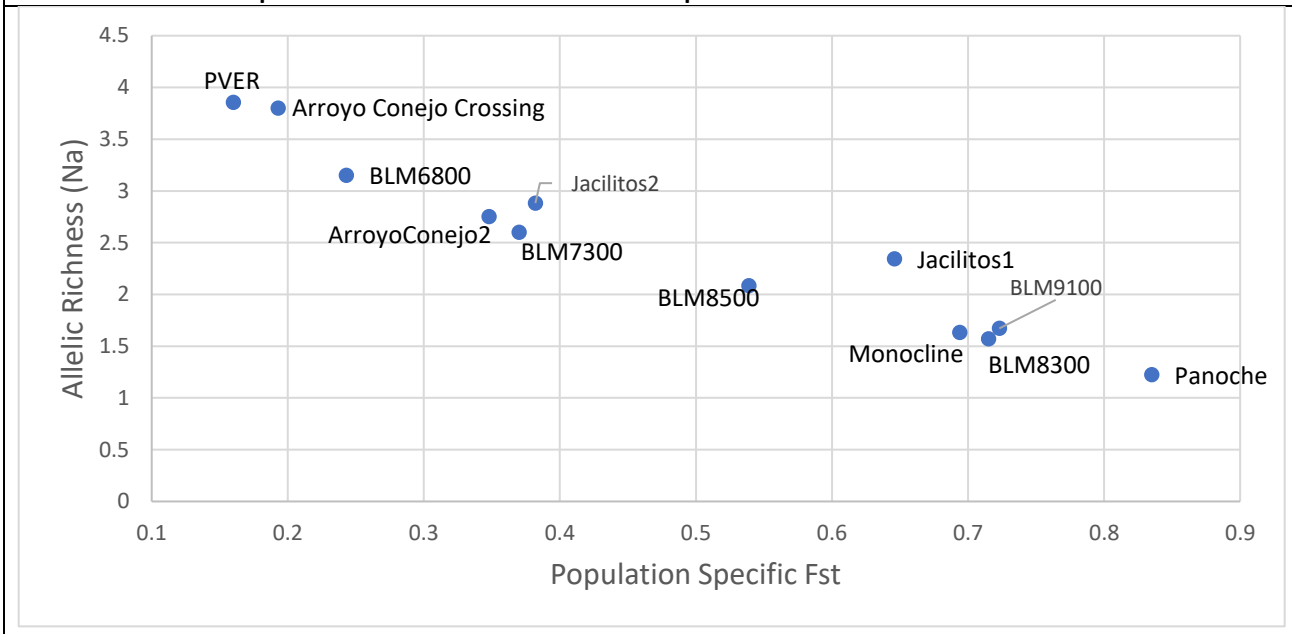



**Figure 4. Allelic pattern in the Monocline Ridge, Panoche and Silver Creek watersheds.**

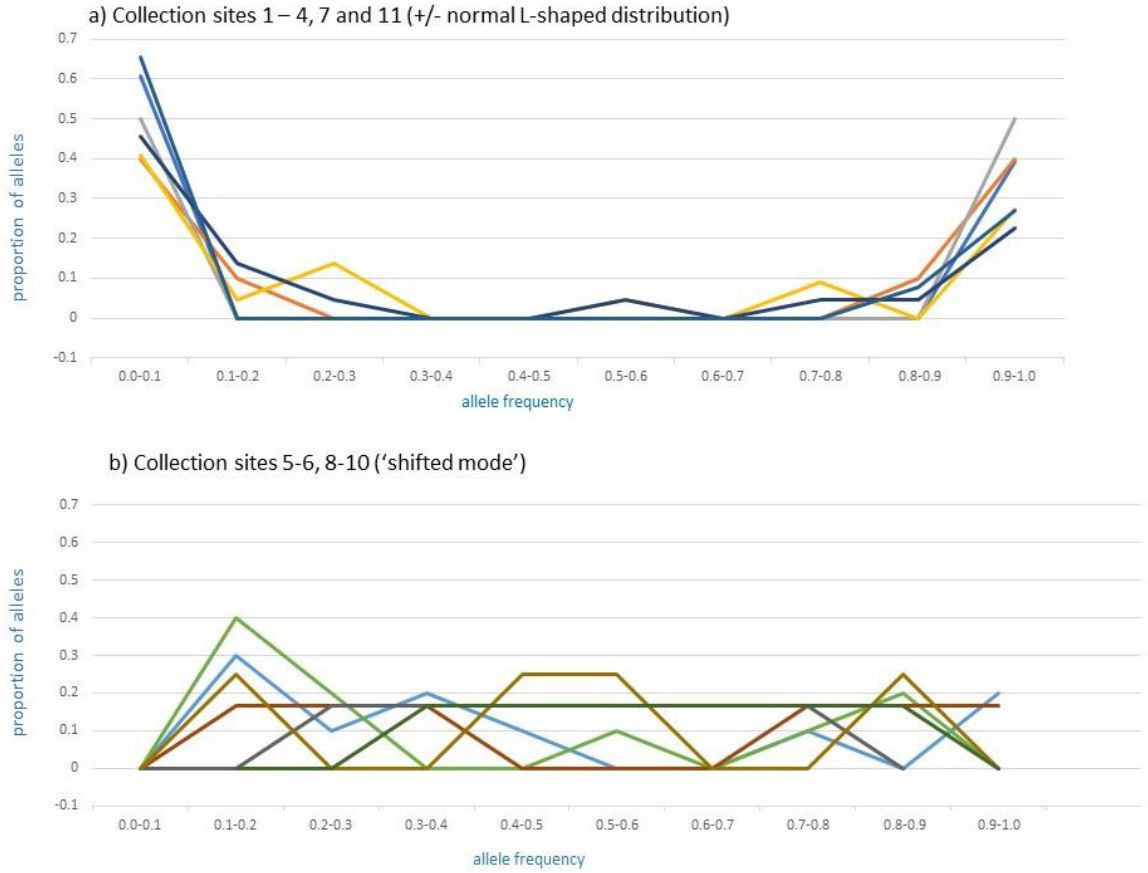


**Table 5. Microsatellite genotype samples sizes (N), number of genotypes, number of effective genotypes, evenness, and Shannon-Wiener diversity by site.**

| Site | N | Number of Genotypes | Number of Effective Genotypes | Genotypic Diversity | Evenness | Shannon-Wiener | Corrected Shannon-Wiener | Uncorrected Genotypic Diversity |
|---|---|---|---|---|---|---|---|---|
| **1** | 61 | 4 | 1.11 | 0.1 | 0.28 | 0.11 | 0.18 | 0.1 |
| **2** | 11 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| **3** | 8 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| **4** | 34 | 3 | 1.59 | 0.38 | 0.53 | 0.28 | 0.31 | 0.37 |
| **5** | 10 | 2 | 1.22 | 0.2 | 0.61 | 0.14 | 0.23 | 0.18 |
| **6** | 9 | 2 | 1.25 | 0.22 | 0.62 | 0.15 | 0.24 | 0.2 |
| **7** | 32 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| **8** | 10 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| **9** | 9 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| **10** | 9 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| **11** | 32 | 4 | 1.39 | 0.29 | 0.35 | 0.26 | 0.32 | 0.28 |
| **12** | 39 | 23 | 18.33 | 0.97 | 0.76 | 1.32 | 1.57 | 0.95 |
| **13** | 18 | 11 | 4.41 | 0.83 | 0.44 | 0.83 | 1.22 | 0.77 |
| **14** | 29 | 13 | 8.49 | 0.91 | 0.65 | 1.01 | 1.16 | 0.88 |
| **15** | 23 | 3 | 1.19 | 0.17 | 0.4 | 0.15 | 0.25 | 0.16 |
| **16** | 27 | 14 | 7.52 | 0.9 | 0.58 | 0.99 | 1.17 | 0.87 |
| **17** | 39 | 11 | 5.02 | 0.82 | 0.46 | 0.82 | 0.94 | 0.8 |
| **18** | 25 | 4 | 1.28 | 0.23 | 0.32 | 0.22 | 0.35 | 0.22 |
| **19** | 34 | 7 | 2.36 | 0.59 | 0.34 | 0.54 | 0.64 | 0.58 |
| **20** | 19 | 17 | 14.44 | 0.98 | 0.85 | 1.2 | 1.99 | 0.93 |
| **21** | 28 | 12 | 5.76 | 0.86 | 0.48 | 0.92 | 1.09 | 0.83 |
| **22** | 40 | 12 | 3 | 0.68 | 0.25 | 0.72 | 0.9 | 0.67 |
| **23** | 25 | 7 | 1.7 | 0.43 | 0.24 | 0.43 | 0.65 | 0.41 |
| **24** | 40 | 13 | 2.91 | 0.67 | 0.18 | 0.79 | 1.08 | 0.66 |
| **25** | 35 | 11 | 5.44 | 0.84 | 0.49 | 0.84 | 0.97 | 0.82 |
| **26** | 32 | 13 | 6.48 | 0.87 | 0.5 | 0.95 | 1.11 | 0.85 |
| **27** | 53 | 11 | 3.32 | 0.71 | 0.3 | 0.72 | 0.82 | 0.7 |
| **28** | 11 | 4 | 2.86 | 0.71 | 0.57 | 0.57 | 0.7 | 0.65 |
| **29** | 36 | 10 | 3.86 | 0.76 | 0.39 | 0.74 | 0.87 | 0.74 |
| **30** | 34 | 5 | 1.36 | 0.27 | 0.27 | 0.27 | 0.38 | 0.27 |
| **31** | 23 | 7 | 3.65 | 0.76 | 0.52 | 0.67 | 0.78 | 0.73 |
| **32** | 23 | 13 | 8.67 | 0.92 | 0.67 | 1.03 | 1.25 | 0.88 |
| **33** | 23 | 4 | 1.44 | 0.32 | 0.36 | 0.28 | 0.38 | 0.31 |
| **Mean** | 7.55 | 3.82 | 0.5 | 0.56 | 0.51 | 0.65 | 0.48 | 7.55 |

**Figure 5. Comparison of Genotypic Diversity Across Collection Sites from Microsatellite Data.**

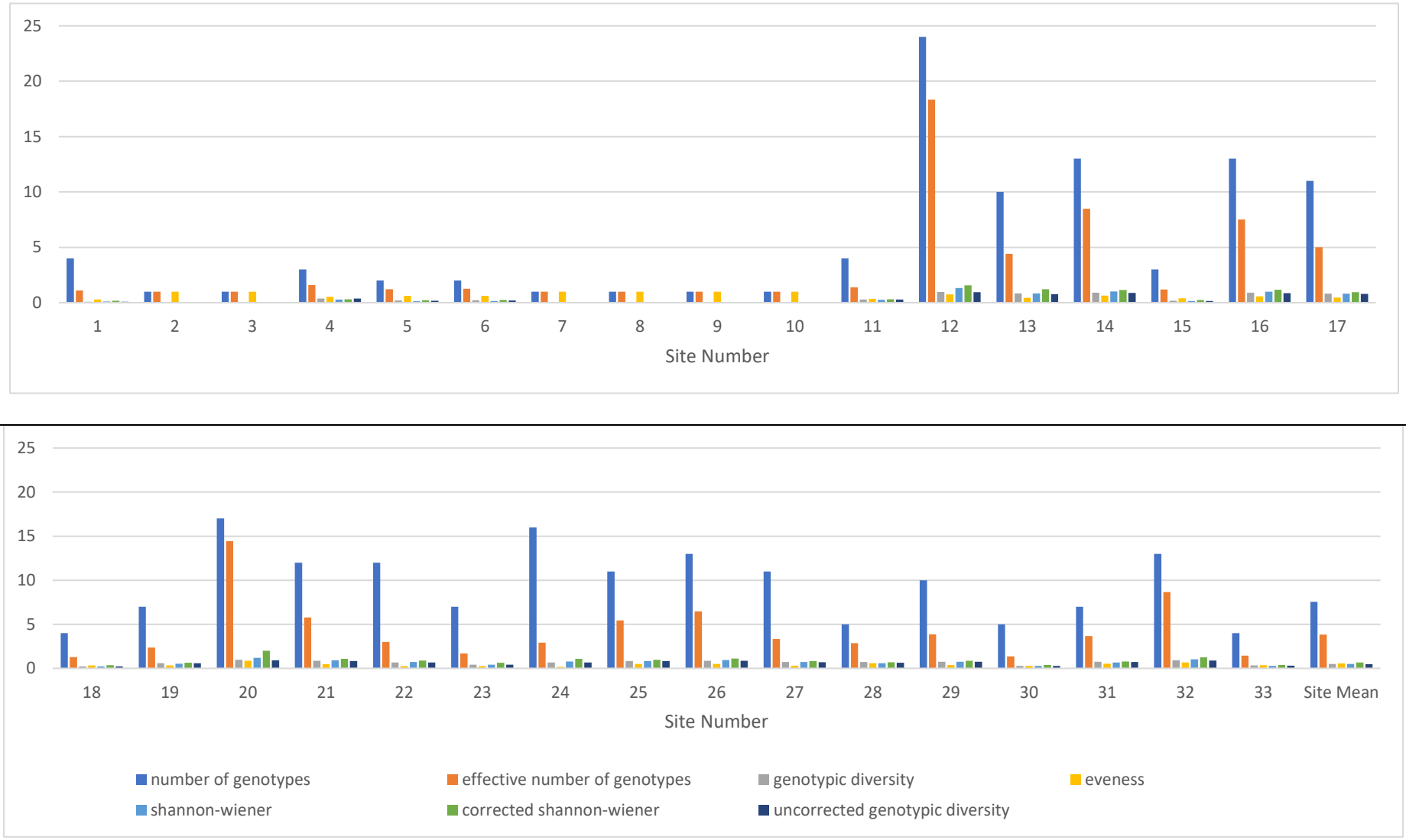
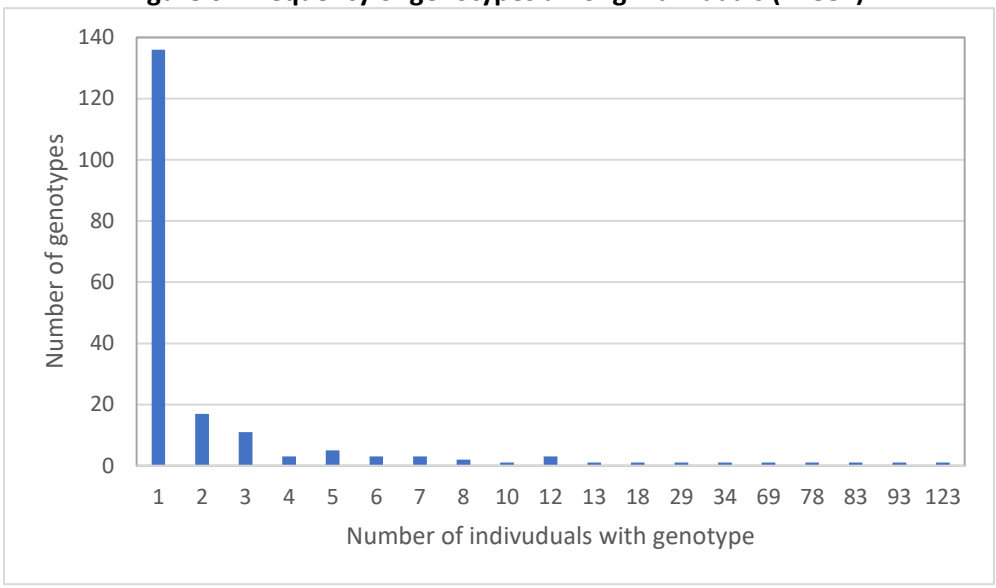


**Figure 6. Frequency of genotypes among individuals (n=881).**




Reserve, Jacalitos Canyon 1 (BLM 3100) and Horse Canyon are also diverse in this respect. The only populations comprised of a single genotype were some of the Panoche and Silver Creek collection sites. A maximum of four were found in the Panoche and Monocline collection sites. At the Panoche site, two of the genotypes were found only in the first sample year and in single individuals; one of the genotypes was found only in the second sample year in a single individual.

**Between Year Variability**. Results suggest that the soil seed bank may buffer some populations against drift, but not others. Across the five sites tested, between year variation accounted for zero to 24% of total variation at a given site. Genetic variation at the Panoche and Kettleman 8500 sites did not differ between years, but between year variation accounted for a portion of total genetic variation at Kettleman 7300 (16%), Lost Hills – Arco (4%), and Traver Ranch (23%) in 2013 and 2015. In addition, 4 to 6 alleles were present in one year but not the other year in each of these populations. At the Panoche site, heterozygosity and the percent of polymorphic alleles were smaller in the second sample year than the first (He = 0.053 versus 0.013, and PLP=57.9% versus 15.8%). Thus, although genetic diversity was not different, heterozygosity did vary between years.

**Hardy-Weinberg Equilibrium and Inbreeding Coefficients.** More than half the sites were comprised entirely of homozygous individuals. The departure from Hardy-Weinberg equilibrium was significant (P<0.001) across all loci at the species level (Table 4). In addition, the departure was significant within all sites and within all loci (P<0.001). The observed heterozygosity ($H_O$) for the species is 0.008 (±0.003) and the expected ($H_e$) is 0.577 (±0.058). The unbiased expected heterozygosity ($uH_e$) equals the expected heterozygosity for populations and loci. Of the 627 site-by-loci combinations, 230 were monomorphic; 391 significantly departed from expected results (P< 0.01 or P<0.001); and only six combinations were in equilibrium. The fixation index

for the species is 0.986 (± 0.005) and for loci is 0.99 (± 0.00). The mean inbreeding coefficient within individuals ($F_{IS}$) is 0.097 (±0.013).

**Population Structure with Microsatellite Markers.** Evaluation of the optimum number of population clusters (K) following the procedure by Evanno et al. (2005) for INSTRUCT and STRUCTURE indicated clear but conflicting maxima for K = 5 and K =2 respectively (Figures 7a and 7b). Clustering using a model that does not assume Hardy-Weinberg equilibrium suggests there are five genetic clusters within the range of *M. congdonii* (Figure 7a and 8). Four of the clusters are sites from 1) Panoche, Silver Creek and Monocline Ridge (sites 1-11); 2) Kettleman Hills and Jacalitos Canyon (21-21); 3) Lost Hills and adjacent San Joaquin Valley (22-26); and 4) Carrizo and Elkhorn Plain and Cuyama Valley. The fifth cluster does not fit a regional or geographic pattern; it includes a subset of individuals from populations in the Kettleman Hills and south (Figure 9).

Clustering assuming Hardy-Weinberg equilibrium suggests the best model for population structure is two genetic clusters (Figure 7b and 8c): 1) all populations from the Kettleman Hills and north (1-21), and 2) all other sampled populations to the south (22-33). The second maxima for INSTRUCT is K=13 (Figures 7a and 7b). For STRUCTURE, the second maxima is for eight clusters (Figure 7b and 8d). STRUCTURE did not appear to overestimate admixture as only 25 % and 37% (for K=2 and K=8) of samples showed significant admixture (<0.95 assignment probability). In comparison, 35% and 69% of samples in the INSTRUCT analysis (for K=5 and K=13) showed significant admixture.

**Figure 7a and 7b. Optimal population clustering using delta K.**

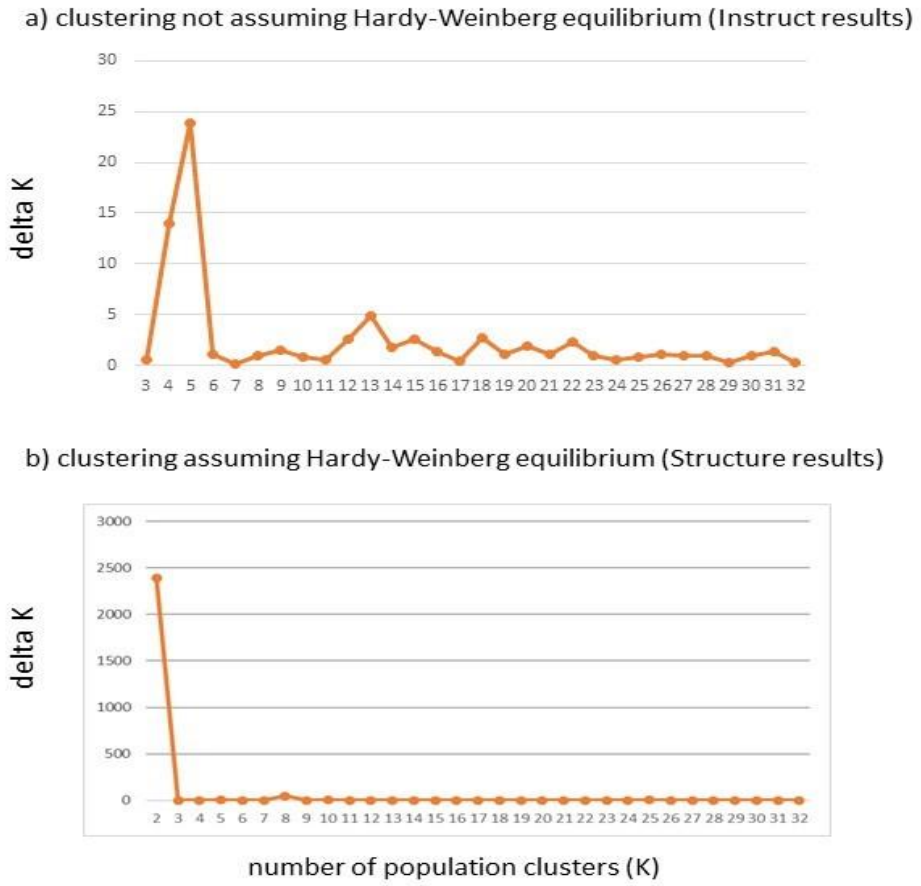
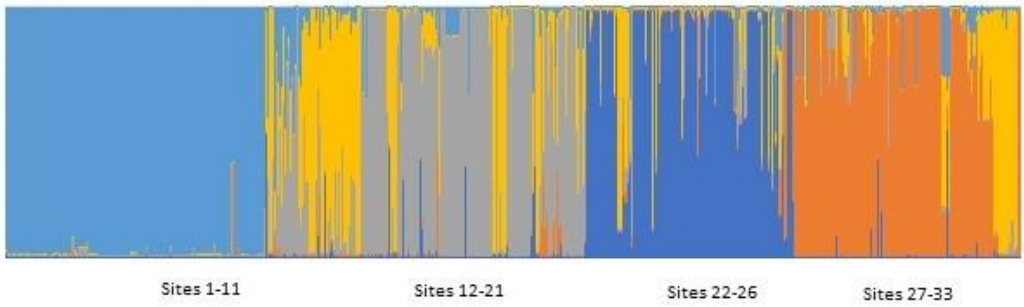
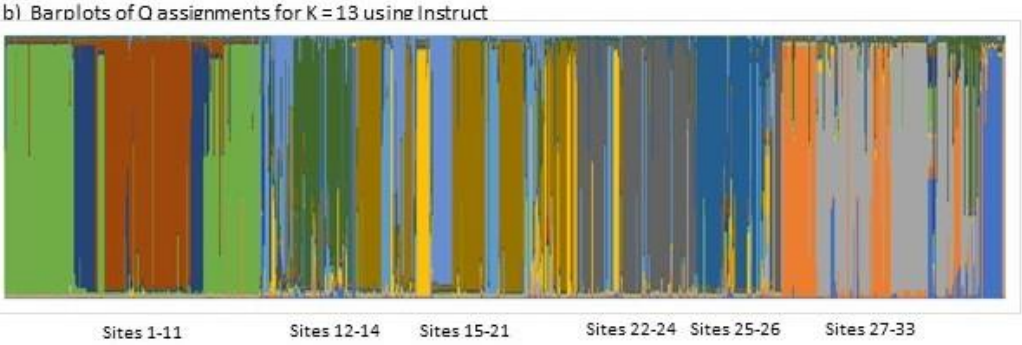


**Figure 8a-d. Results of Population Clustering using INSTRUCT and STRUCTURE for optimum number of clusters (K). Sites arranged from north to south from left to right on the barplots.**

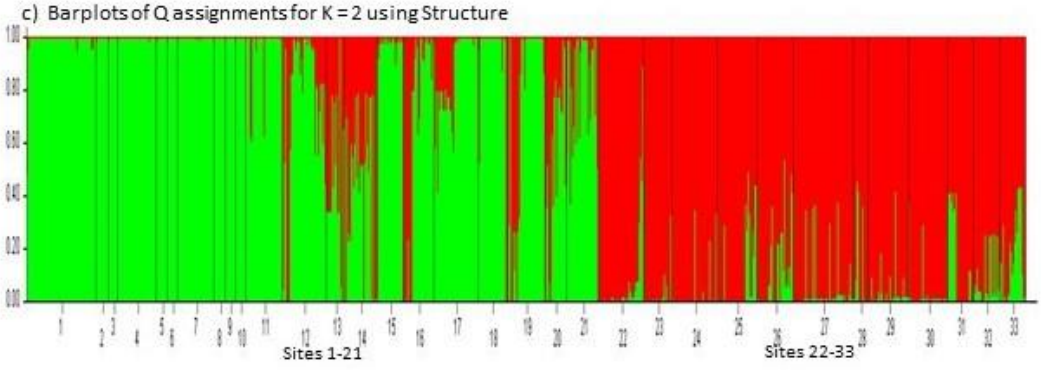
a) Barplots of Q assignments for K=5 using Instruct




b) Barplots of Q assignments for K = 13 using Instruct




c) Barplots of Q assignments for K = 2 using Structure




d) Barplots of Q assignments for K = 8 using Structure

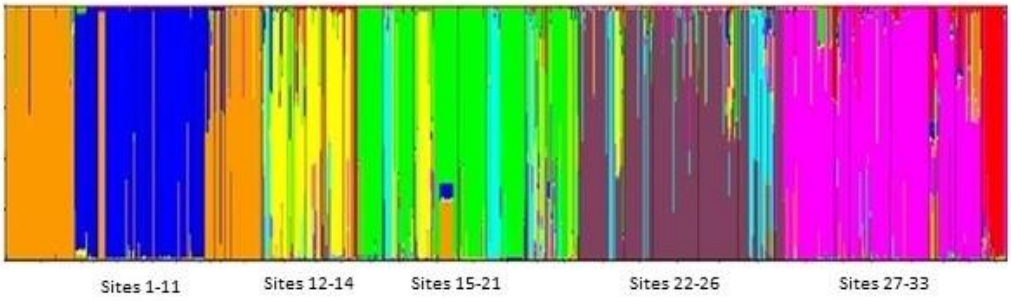
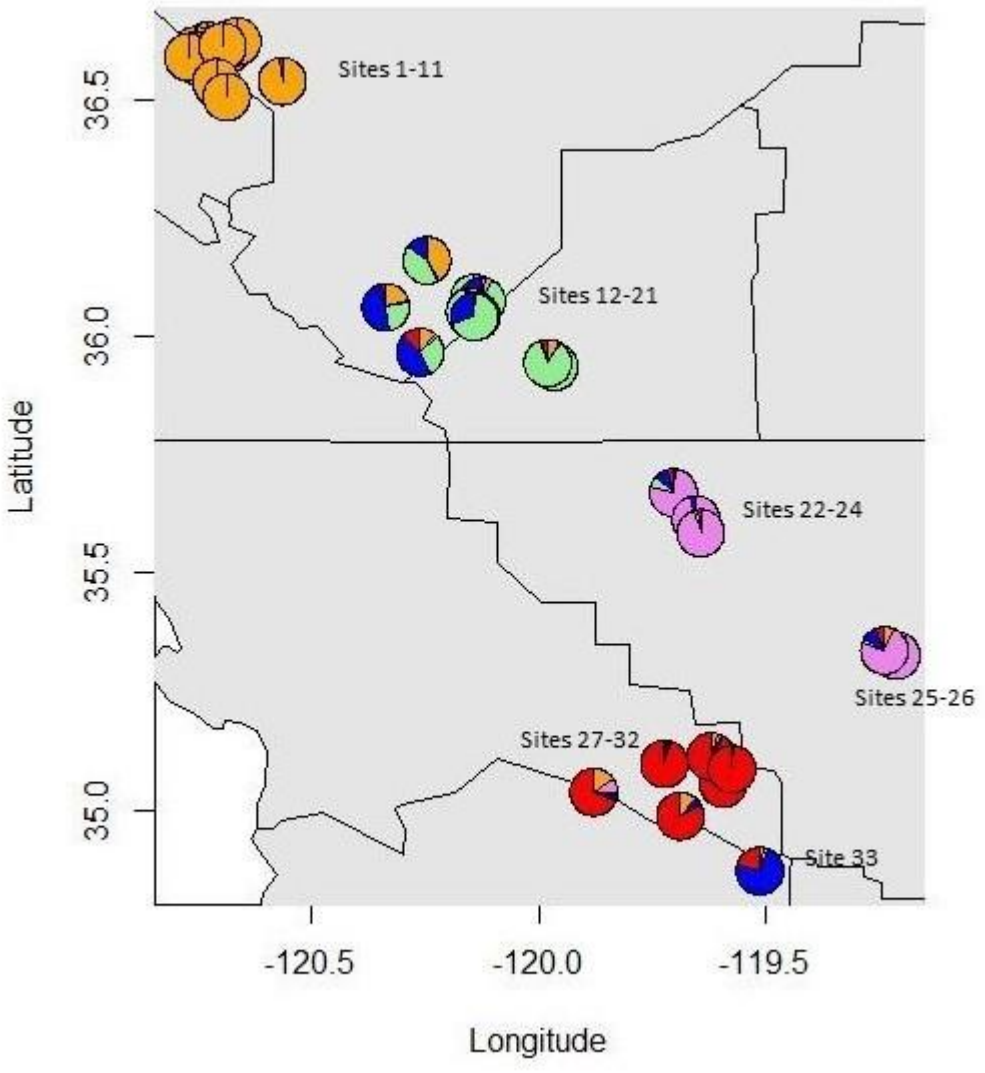


**Figure 9. Distribution of five genetic clusters identified using INSTRUCT.**



DAPC suggests there are three main genetic clusters: 1) the northernmost populations in the Panoche and Silver Creek watersheds and on Monocline Ridge (sites 1-11); 2) Kettleman Hills (Sites 12-21), and 3) all other populations (Sites 22-33). Substructure between the sites in the San Joaquin Valley (Lost Hills and Kern Water Bank, Sites 22 to 26), and populations in the Carrizo and Cuyama Valleys (Sites 27-33) (Figure 10a). DAPC analysis of hierarchical structure within the third set of populations suggests further subdivision between the Lost Hills and eastern San Joaquin Valley populations (Kern Waterbank), and the Carrizo/Elkhorn Plain and Cuyama Valley populations (Figures 10b and 10c). Further structuring within these regions is discussed below.

Bootstrapping of loci using Nei's genetic distance largely supports the above results including also indicates a North-South split between the Kettleman Hills and the Lost Hills (Figure 11). In addition, within the Northern cluster, Nei's genetic distance supports the split between Panoche Creek/Silver Creek/Monocline populations and the Kettleman Hills/Jacilitos Canyon/PVER populations. The analysis also shows strong support for the Jacilitos Canyon populations distinct from the Kettleman Hills and the PVER populations.

In the Southern cluster, the split between the San Joaquin Valley populations and Carrizo and Elkhorn Plains/Cuyama Valley populations is supported. Further subdivision between the Lost Hills populations and Kern Water Bank populations is also strongly supported. In addition, there is also strong support for the distinction of the Santa Barbara Canyon population from the other Cuyama Valley populations and greater similarity of the later with the Carrizo and Elkhorn Plain populations.

**Figure 10a. Scatterplot of the first two principal components of Discriminant Analysis of Principal Components of Microsatellite data from all sample sites. Numbers refer to sites (1-11 = Panoche, Silver Creek, and Monocline; 12 -21 = Pleasant Valley Ecological Reserve; Jacilitos Canyon, Pleasant Valley Ecological Reserve and Kettleman Hills; 22-26 = Lost Hills and Kern Water Bank; and 27-33 Carrizo/Elkhorn Plains and Cuyama Valley).**

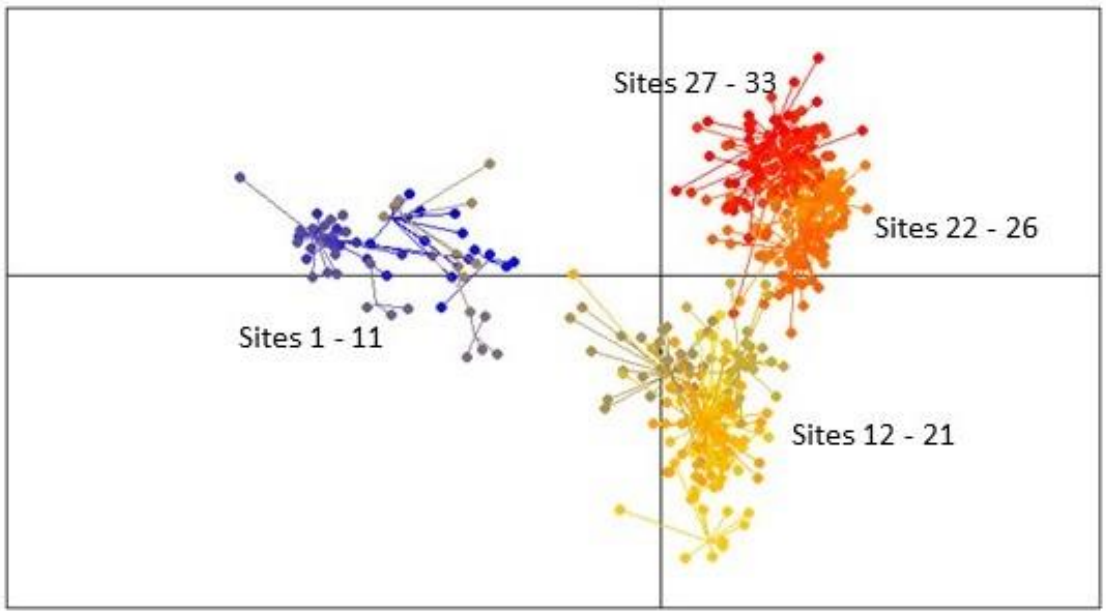



**Figure 10b. Scatterplot of the first two principal components of Discriminant Analysis of Principal Components of SNP data from all sample sites in the Kettleman Hills, Coalinga and Jacilitos Canyon. Numbers refer to sites (12-13 = Jacilitos Canyon; 14 = Pleasant Valley Ecological Reserve; 18 = BLM 8300; 19-21 = all other North Dome and Middle Dome).**

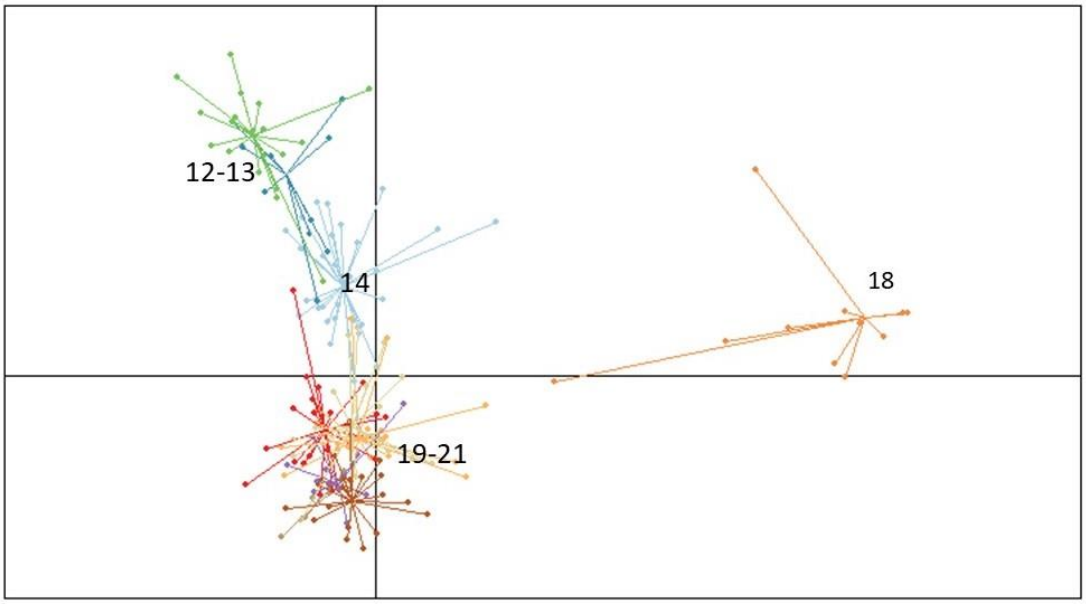


**Figure 10c. Scatterplot of the first two principal components of Discriminant Analysis of Principal Components of microsatellite data from all collection sites south of Kettleman Hills. Numbers refer to sites (22-24 = Lost Hills; 25-26 = Kern water Bank; 27-33 = Carrizo/Elkhorn Plains and Cuyama Valley).**

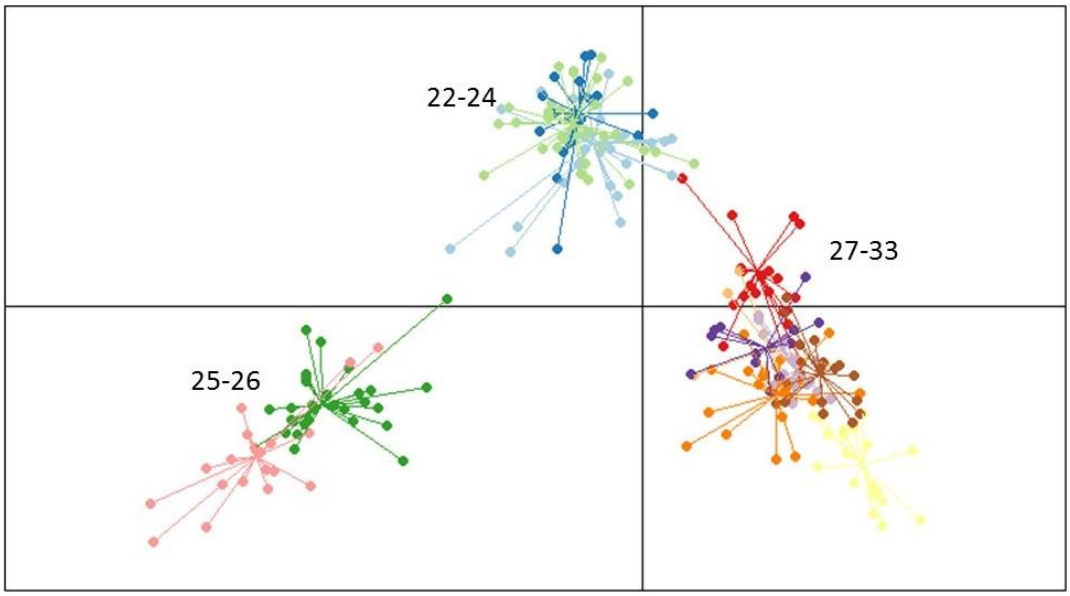
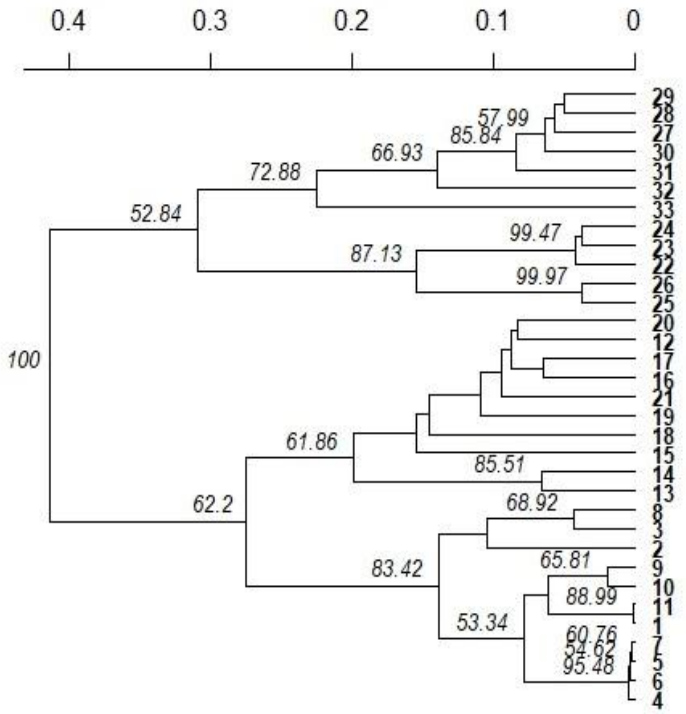



**Figure 11. Dendrogram of *Monolopia congdonii* collection sites using Nei's genetic distance from bootstrapping loci (sample =10,000). Numbers at ends of nodes are collection site numbers.**



**Chloroplast DNA Haplotypes.** Of the eight populations initially screened, 2 genes showed nucleotide variation across populations (3'ndhF and rpl16 intron). The variation in these two markers across 153 samples allowed for definition of six combined region cpDNA haplotypes in *M. congdonii* (see Table 2 and 3 for distribution of cpDNA samples across collection sites and herbarium specimens). Haplotype A appears to be the ancestral type with all other haplotypes derived by a single indel, except Haplotype C, which is derived from 2 indels. In addition, one sample had a haplotype identical to the sympatric species, *M*. *stricta* (SJB532).

Haplotype A was the most frequent in all of the samples (Table 6, Figure 12) but sampling for cpDNA was not proportional throughout the taxon range and highest in the regions it is most frequent. Haplotypes C and D were the next most frequent across the samples. Haplotypes A and C are the most widespread across the sample range and the range of the taxon. Haplotype B was only found on Monocline Ridge, in the Panoche and Silver Creek watersheds, and on the Mendota Plain as far south as 26 miles south of Mendota (SJB1289). Haplotype A is the only other type found in that region; it was found in approximately half the samples (0.53), although it is less frequent (0.30) in the Mendota Plain, than it is in the adjacent foothills. Samples from the Monocline Ridge were all B.

The next region south, Kettleman-Jacilitos-Coalinga, is characterized by Haplotypes A and C (Figure 12). Haplotype A was found in almost two-thirds of the individuals sampled. Haplotype C is rare in this region, although it is the only haplotype in samples from north of Coalinga in the San Joaquin Valley (SJB1283 – SJB 1286). The individual with the *M*. *stricta* Haplotype (Haplotype S) was found in the Kettleman Hills of this region.

The Lost Hills Region and the Eastern Valley Region are characterized by Haplotypes A, C, D and E. They occur in nearly equal frequency in the Lost Hills Region, except Haplotype E, which is very rare in both Regions (0.06 and 0.08). In contrast to the Lost Hills Region, more than half the samples in Eastern Valley were Haplotype C and Haplotypes A and D are infrequent in this Region. Haplotype C is the only type found in the area between the Carrizo/Elkhorn/Cuyama Region and the Lost Hills (Middlewater Plain).

Haplotypes D and E are the predominate haplotypes in Carrizo and Elkhorn Plains and the Cuyama Valley. Haplotype E is more frequent than Haplotype D, especially in the Cuyama Valley (0.92). The ancestral haplotype was not found in any of the samples from the Middlewater Plain, the Carrizo and Elkhorn Plains, or the Cuyama Valley.

**Population Hierarchy and Partitioning of Diversity Among Regions.** These analyses support the same subdivisions but the rank in the hierarchy differs somewhat between analyses (e.g., first order, second, etc.). For example, some analysis suggests the first split is between the North and South, whereas others suggest the first spilt is between three clusters. A hierarchical population structure supporting all subgroups is shown in Table 7 and Figure 14. Analysis of Molecular Variance indicates the four sub regions account for 28% of the total variance, but the largest variance is between samples (41%). Variation between clusters of sites within the

subregions and between sites is lower (Figure 13). Figures 15 and 16 show allelic diversity by region and subregion.

**Table 6. Sample Sizes, Distribution and Proportion of cpDNA Haplotypes Among Collection Sites.**

| Region | Sites | n | A | B | C | D | E | S |
|---|---|---|---|---|---|---|---|---|
| **Panoche-Silver Ck-Monocline** | 1-11 | 30 | 0.53 | 0.47 | 0.00 | 0.00 | 0.00 | 0.00 |
| **Mendota Plain** | n/a | 11 | 0.36 | 0.64 | 0.00 | 0.00 | 0.00 | 0.00 |
| **Kettleman/Coalinga/Jacilitos** | 12-21 | 51 | 0.70 | 0.00 | 0.28 | 0.00 | 0.00 | 0.02 |
| **Lost Hills** | 22-24 | 12 | 0.33 | 0.00 | 0.33 | 0.25 | 0.08 | 0.00 |
| **Eastern Kern** | 25-26, + | 15 | 0.13 | 0.00 | 0.73 | 0.07 | 0.07 | 0.00 |
| **Middlewater Plain** | n/a | 3 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| **Carrizo Elkhorn** | 27-30 | 16 | 0.00 | 0.00 | 0.00 | 0.31 | 0.69 | 0.00 |
| **Cuyama** | 31-33 | 13 | 0.00 | 0.00 | 0.00 | 0.08 | 0.92 | 0.00 |
| **TOTAL** | 1-33 | 151 | 0.41 | 0.14 | 0.21 | 0.07 | 0.17 | 0.00 |

**Figure 12. Distribution and frequency of cpDNA Haplotypes among sample sites. Sample locations of possibly extirpated sites are based on herbarium material.**

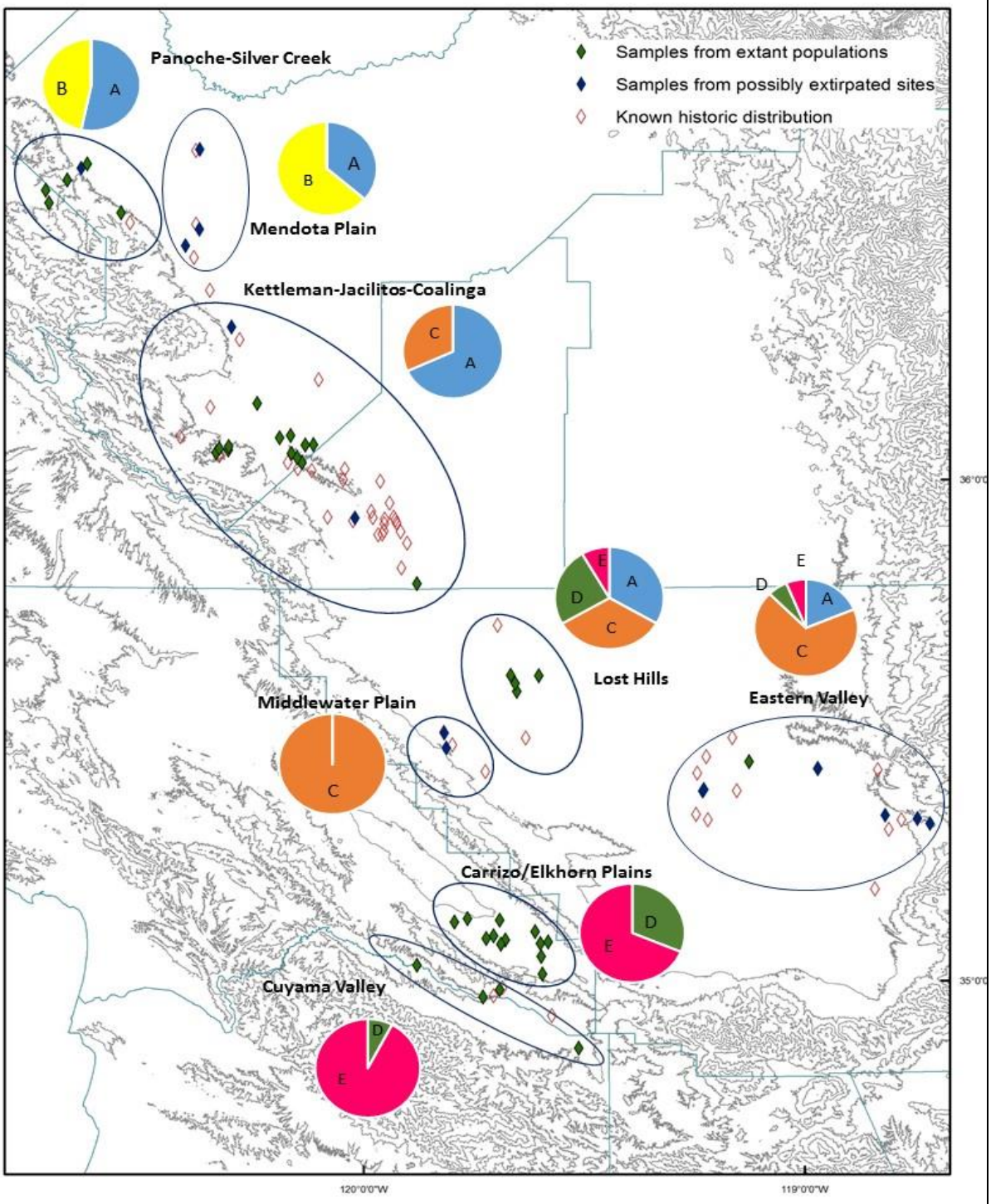


**Table 7. Regional and subregional structure based on consensus of analyses of microsatellite data.**

A. Northern Region
   1. Northern Subregion (Sites 1-11, Mendota Plain cpDNA)
   2. Coalinga Subregion
      a) Jacilitos Canyon (13-14)
      b) Kettleman/PVER (12, 15-21)

B. Southern Region
   1. Southern San Joaquin Valley Subregion
      a) Lost Hills (22-24)
      b) West Kern Water Bank (25-26)
      c) Middlewater Plain (cpDNA only)

   2. Carrizo/Cuyama Subregion
      a) Carrizo Plain/Elkhorn Plain/Cuyama (27-32)
      b) Santa Barbara Canyon (33)

**Figure 13. Percentage of molecular variance based on population structure.**

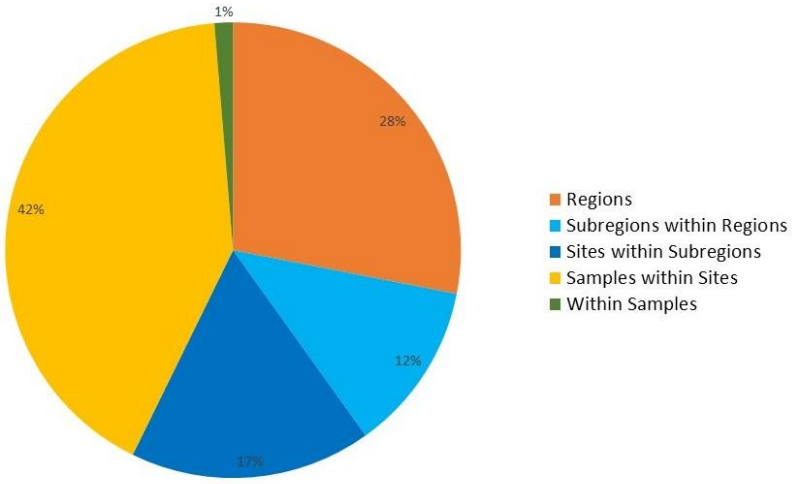


**Figure 14. Map of *Monolopia congdonii* population substructure.**

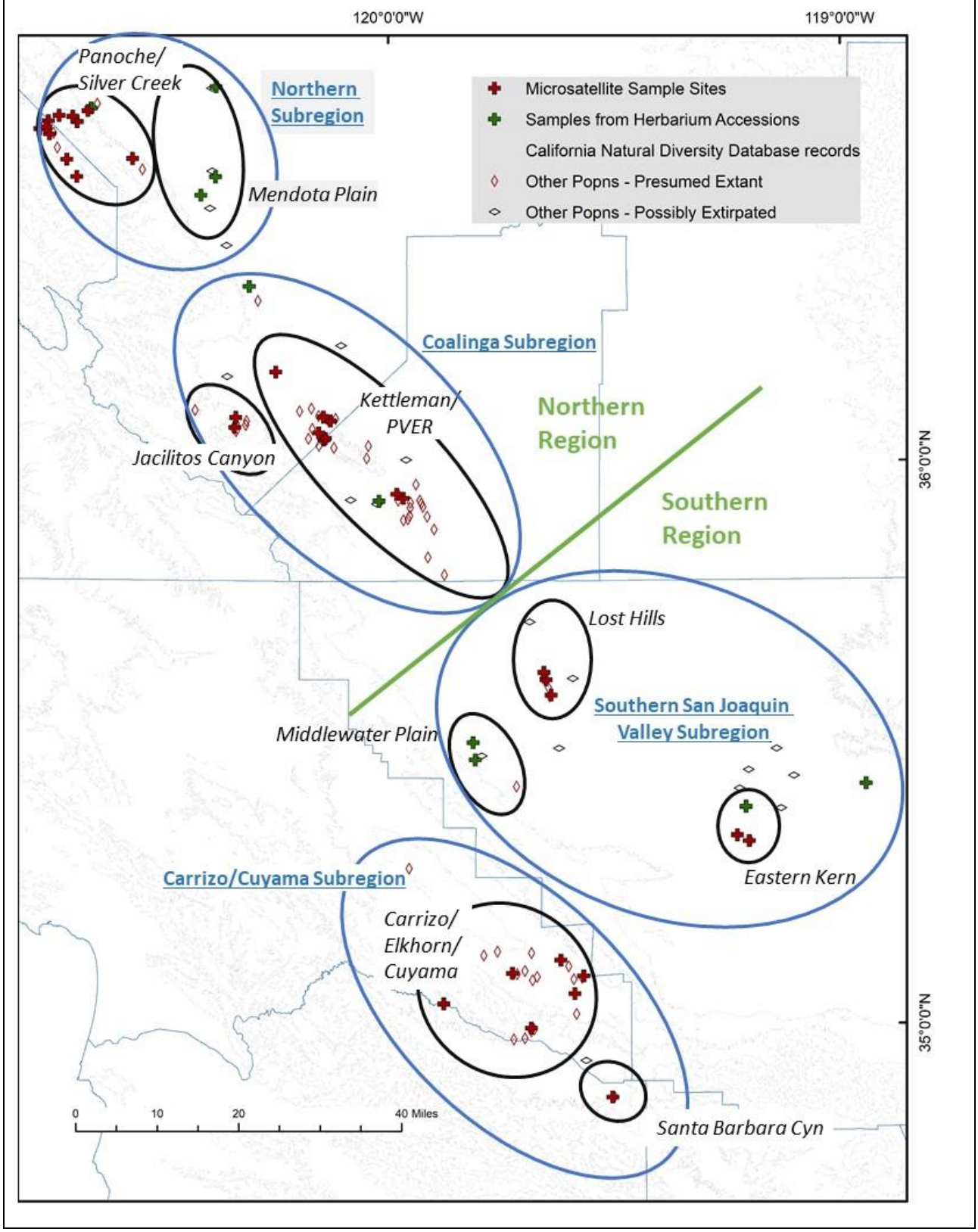
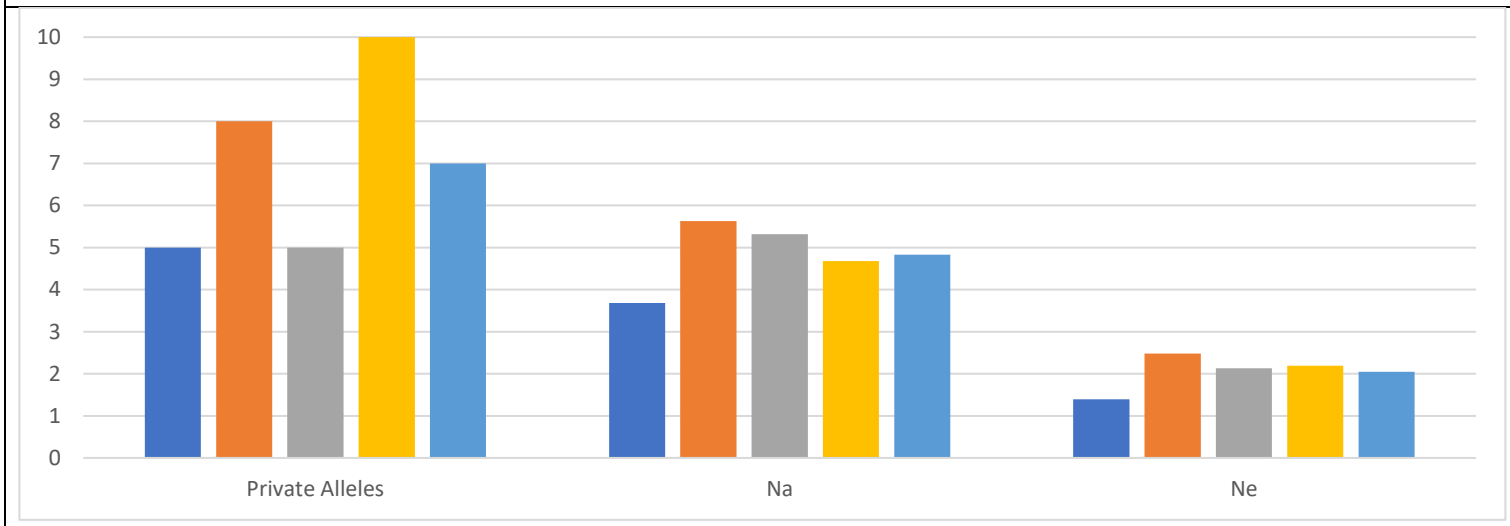


**Figure 15. Allelic patterns in *Monolopia congdonii* by region.**



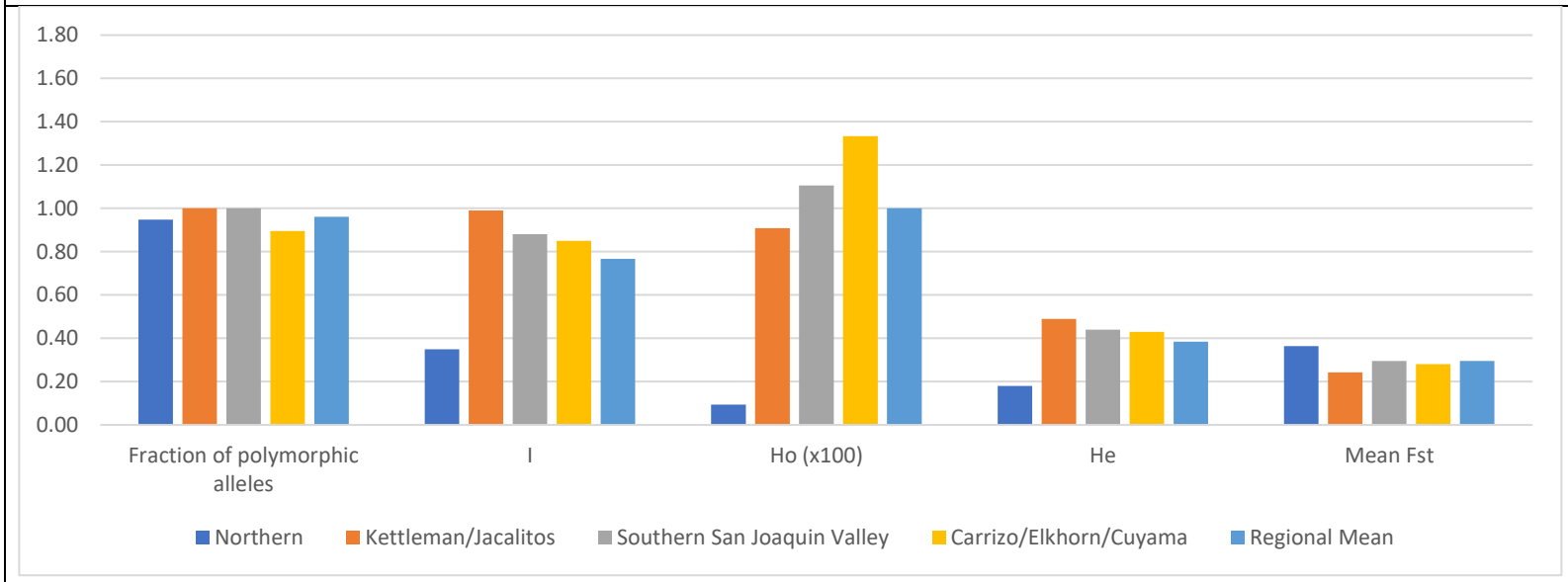
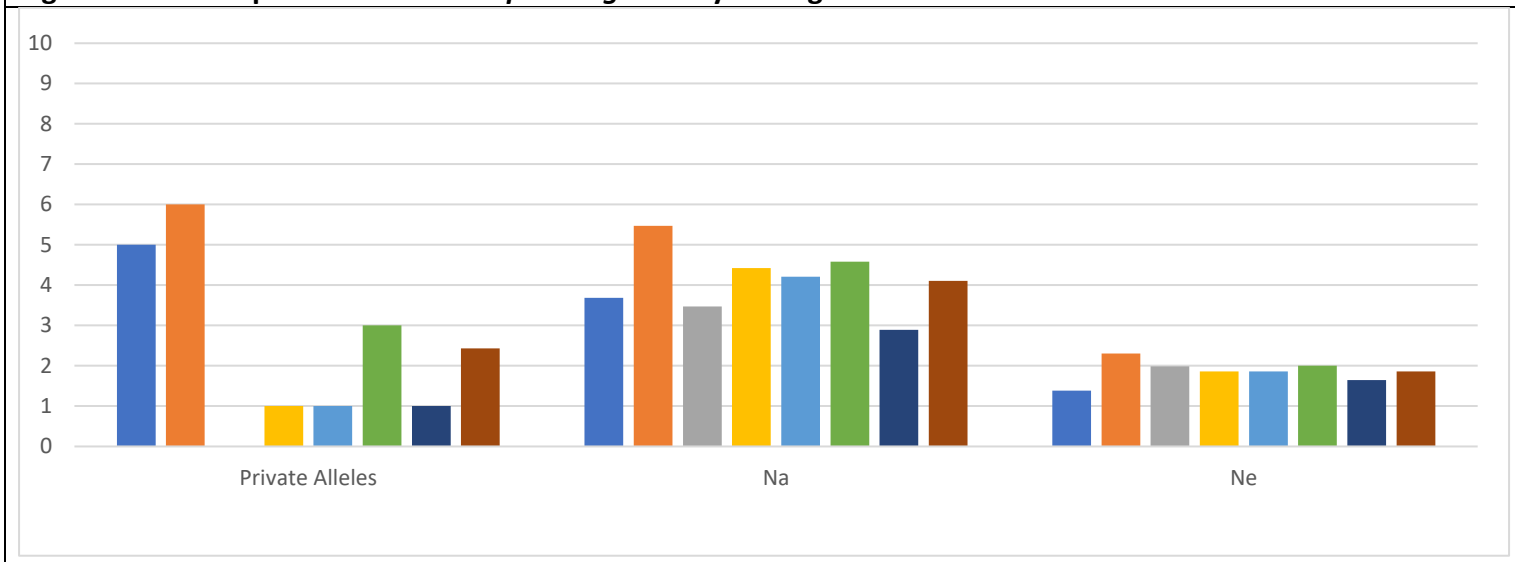
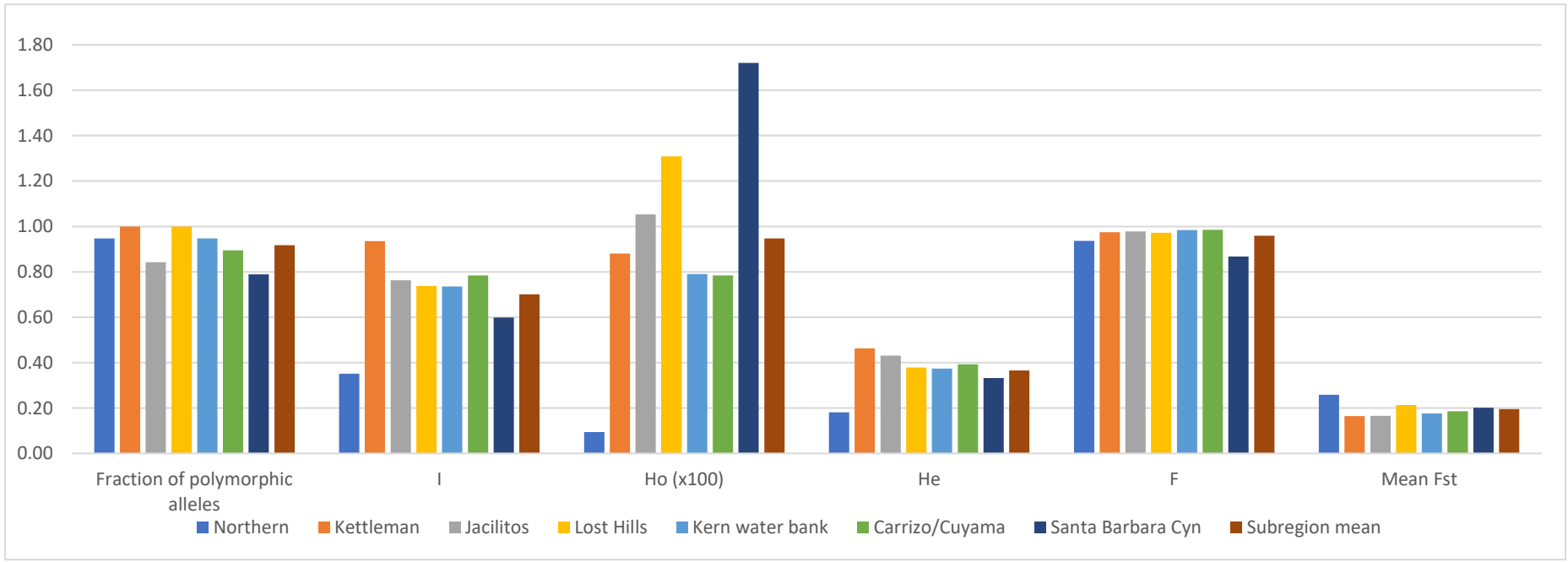



**Figure 16. Allelic patterns in *Monolopia congdonii* by subregion.**



**Population Structure and Migration Rates in the Northern Subregion.** Delta K for both INSTRUCT and STRUCTURE Results support three genetic clusters or gene pools among the eleven collection sites in the Northern Subregion and three areas corresponding to the distribution of these gene pools are delineated (Figure 17). The easternmost of these, Panoche Valley, is located in the Panoche Valley and Silver Creek Ranch (collection sites 4 to 7). The northernmost, Lower Panoche, is comprised of three collection sites in the lower Panoche watershed (collection sites 2-3, 8). The third, Silver Creek-Monocline, is based on four collection sites that are more geographically dispersed. Two are in the upper Silver Creek watershed with the third at the mouth of Silver Creek (aka Panoche, collection site 1). The fourth site is the Monocline Ridge collection site (11). It is important to note that 'collection site' represent collections from patches than are (0.025 mile apart), but may be comprised of gaps <0.25 mi.

Twenty-one of the 225 individuals sampled in the Northern Subregion exhibited admixture ancestry (<0.90 inferred ancestry of any inferred cluster) and more than two-thirds of these individuals were from collection sites 4, 9 and 10. In addition, seven individuals from collection sites assigned to the Panoche Valley gene pool had >0.9 inferred ancestry of the Silver Creek-Monocline gene pool. They included six individuals from Collection Site 4 and one from Collection Site 6. The other 197 individuals exhibited >0.90 inferred ancestry from the cluster assigned to their respective collection site.

The dendrogram based on genetic distance supports these three groups within the Northern Subregion (Figure 18). In the Lower Panoche area, the Panoche T and Right Angle collection sites (3 and 8) are more similar to each other than the third collection site, Lower Panoche (2). Within the Silver Creek – Monocline genepool area, the Panoche and Monocline collection sites (1 and 11) are more similar to each other than the two sites from upper Silver Creek (9 and 10). The three collection sites in the Panoche Valley genepool area are only slight less differentiated from each other. Discriminant Analysis of Principle Components for the Northern Subregion sites (Figure 19) supports the finer structure of the dendrogram. Figure 20 shows distribution and frequency of the two cpDNA haplotypes in the Northern Subregion.

Inferred migration rates as calculated by BayesAss between the three Areas delineated by genetic clusters are very low (Table 8) with mean of less than 0.02% individuals per generation (year). The mean between collection sites within each cluster is only slightly higher (0.03%), although mean rate is lower in Lower Panoche (0.02%) than in the other two (0.04%). As expected, rates within collection sites are much higher; the mean within site migration rate is 0.80%. Migration rates calculated in Migrate and representing less recent migration are similarly low. Isolation by distance was not significant (Mantel test, $R^2 = 0.0039$, $P>0.1$).

AMOVA results indicate 61% of the variation is between the genetic clusters and 39% is between individuals within these genetic clusters. Figure 21 shows the diversity indices for the three genetic clusters. The Silver Creek – Monocline is the most diverse. Lower Panoche is the

least diverse in terms of allelic richness and diversity, but it has the highest divergence among collections sites (Fst=0.807). In contrast, collections sites from Panoche Valley are the most homogenous.

**Figure 17. Distribution and frequency of three *Monolopia congdonii* genetic clusters in the Northern Subregion (blue = Lower Panoche, red = Silver Creek-Monocline, green = Panoche Valley). Pie charts do not represent actual collection site boundaries or distributions.**

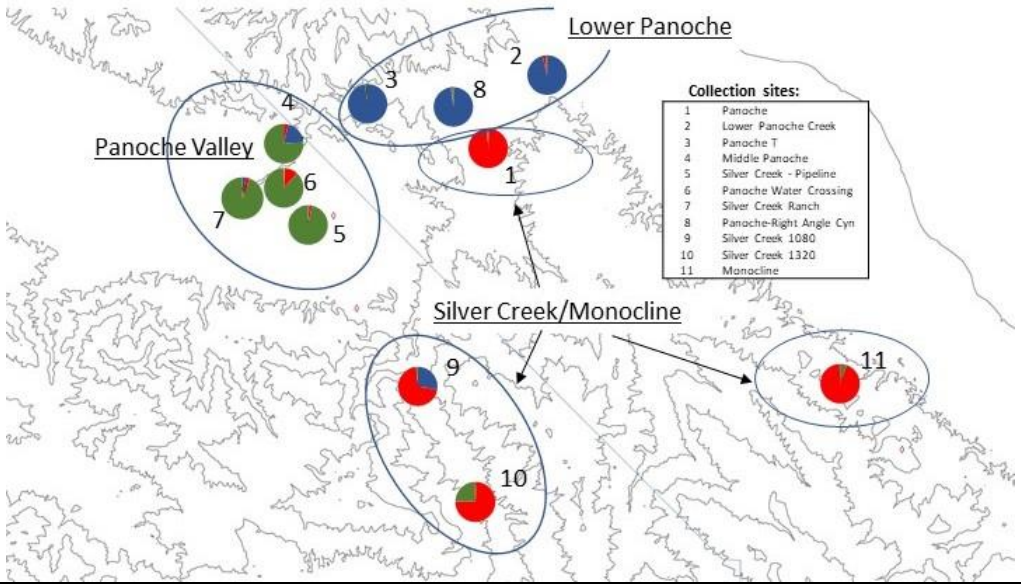



**Figure 18. Dendrogram of *Monolopia congdonii* collection sites in the Northern Subregion using Nei's genetic distance from bootstrapping loci (sample =10,000). Numbers at ends of nodes are collection sites.**

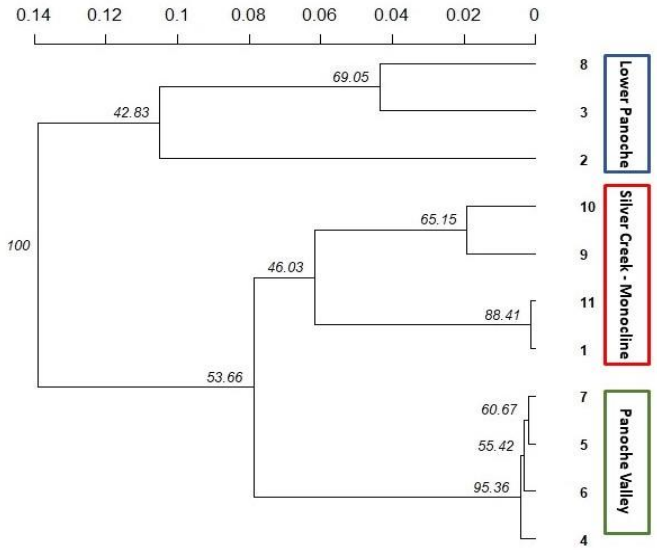


**Figure 19. Scatterplot of the first two principal components of Discriminant Analysis of Principal Components of microsatellite data from *Monolopia congdonii* collection sites from the Northern Subregion.**

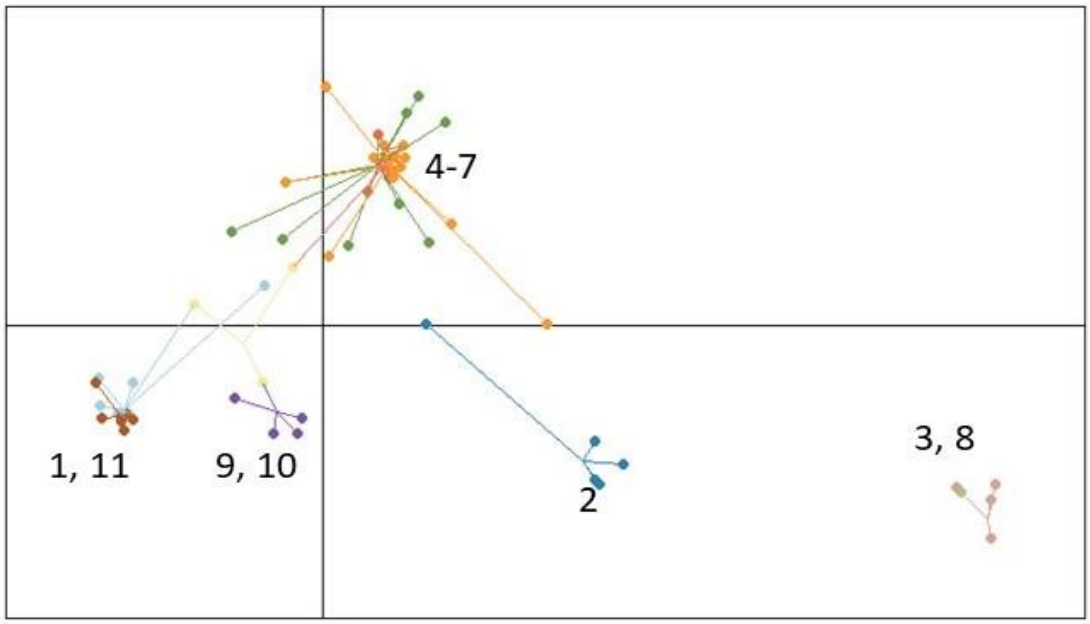


**Figure 20. Distribution and frequency of *Monolopia congdonii* cpDNA haplotypes in the Northern Subregion.**

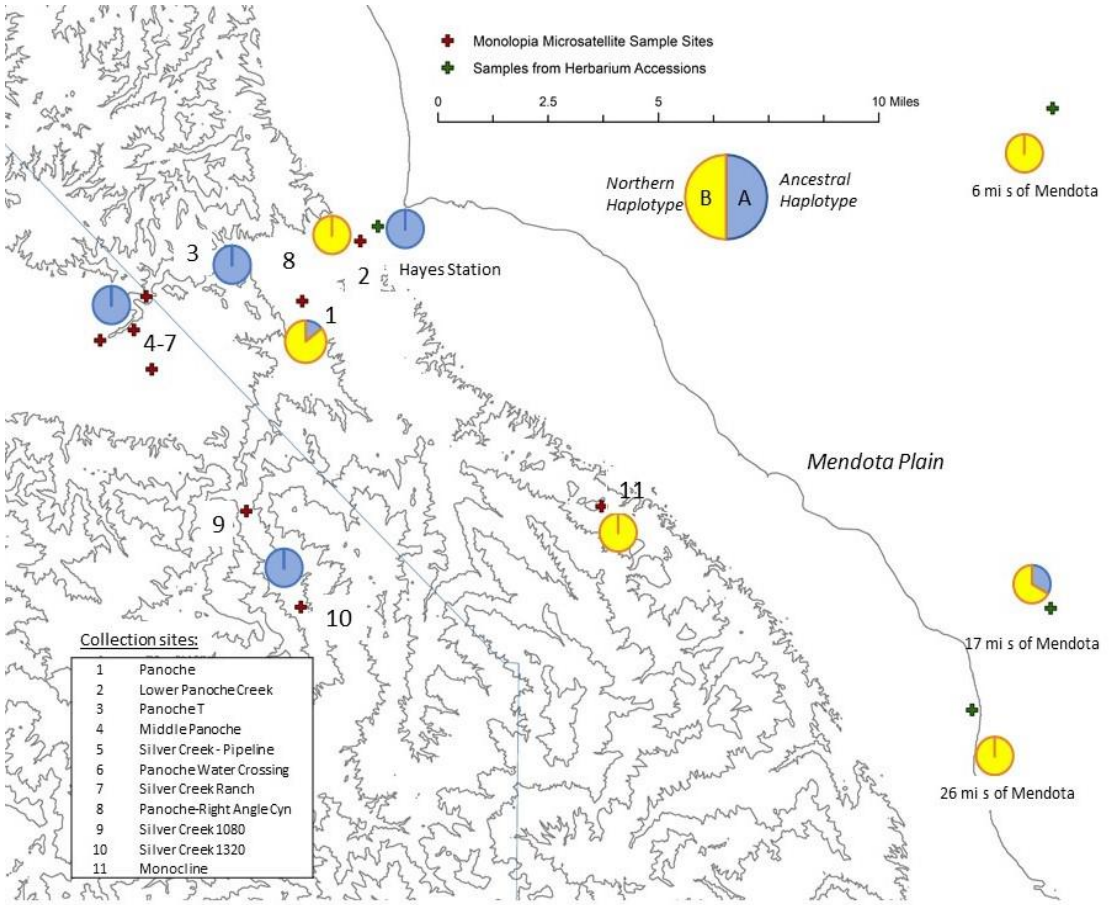
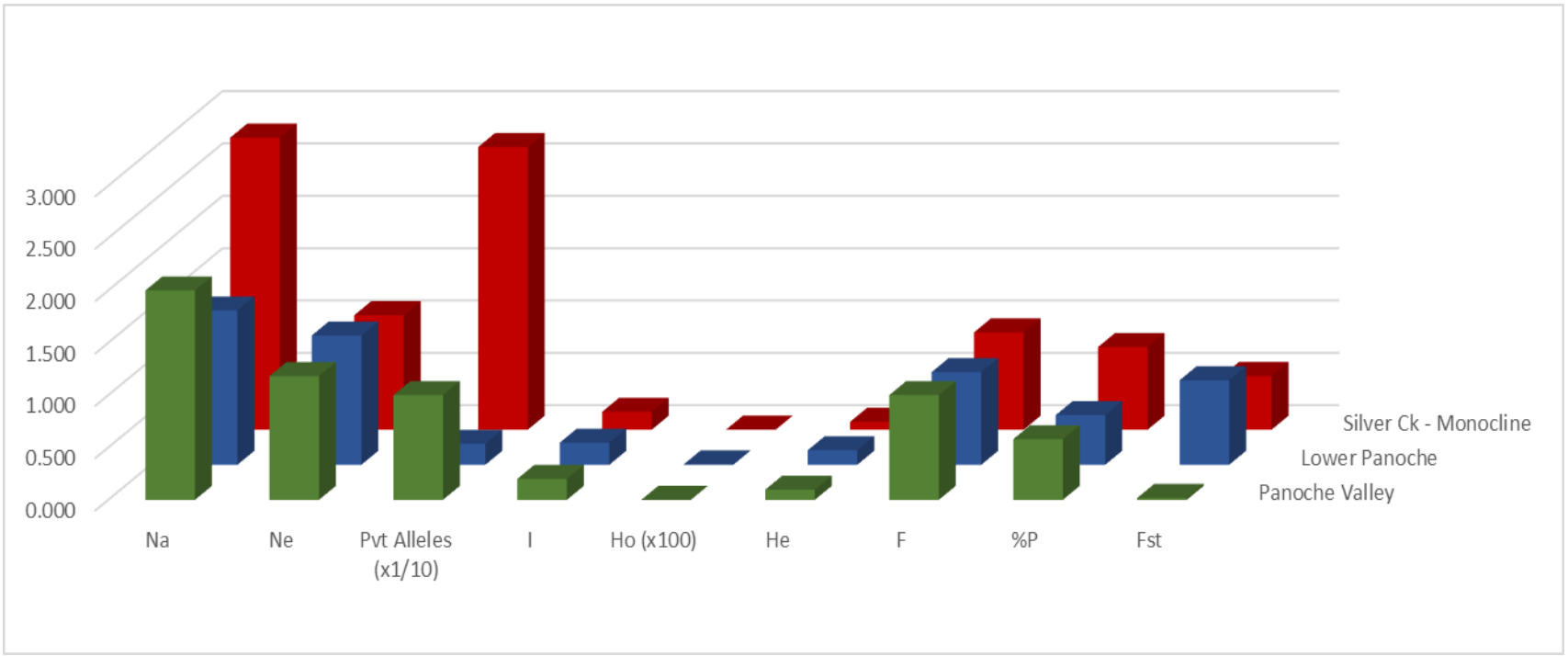


**Table 8. Inferred (posterior mean) migration rates between and within collection sites (underlined), based on microsatellite data, sorted by genetic cluster. Values represent the fraction of individuals in population *i* that are migrants derived from population *j* per generation.**

| | Migrants from collection site (*j*): | 2 | 3 | 8 | 1 | 9 | 10 | 11 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Collection site (*i*): | Gene pool area: | Lower Panoche | | | Silver Creek – Monocline | | | | Panoche Valley | | | |
| 2 | Lower Panoche | 0.8472 | 0.0150 | 0.0154 | 0.0157 | 0.0156 | 0.0151 | 0.0160 | 0.0149 | 0.0149 | 0.0149 | 0.0151 |
| 3 | | 0.0172 | 0.8247 | 0.0175 | 0.0183 | 0.0175 | 0.0175 | 0.0173 | 0.0176 | 0.0173 | 0.0176 | 0.0175 |
| 8 | | 0.0154 | 0.0267 | 0.8297 | 0.0164 | 0.0164 | 0.0161 | 0.0155 | 0.0162 | 0.0153 | 0.0161 | 0.0163 |
| 1 | Silver Creek – Monocline | 0.0046 | 0.0048 | 0.0049 | 0.9470 | 0.0057 | 0.0049 | 0.0054 | 0.0052 | 0.0049 | 0.0079 | 0.0047 |
| 9 | | 0.0166 | 0.0171 | 0.0162 | 0.0171 | 0.8329 | 0.0172 | 0.0169 | 0.0165 | 0.0158 | 0.0167 | 0.0170 |
| 10 | | 0.0172 | 0.0162 | 0.0168 | 0.0174 | 0.1440 | 0.6833 | 0.0162 | 0.0325 | 0.0173 | 0.0173 | 0.0219 |
| 11 | | 0.0081 | 0.0081 | 0.0077 | 0.2294 | 0.0078 | 0.0078 | 0.6998 | 0.0078 | 0.0079 | 0.0078 | 0.0077 |
| 4 | Panoche Valley | 0.0075 | 0.0078 | 0.0074 | 0.0470 | 0.0084 | 0.0071 | 0.0074 | 0.8714 | 0.0073 | 0.0072 | 0.0215 |
| 5 | | 0.0165 | 0.0164 | 0.0156 | 0.0158 | 0.0166 | 0.0155 | 0.0158 | 0.0326 | 0.6822 | 0.0159 | 0.1570 |
| 6 | | 0.0162 | 0.0172 | 0.0176 | 0.0299 | 0.0185 | 0.0166 | 0.0166 | 0.0369 | 0.0170 | 0.6834 | 0.1301 |
| 7 | | 0.0073 | 0.0135 | 0.0082 | 0.0077 | 0.0081 | 0.0080 | 0.0080 | 0.0117 | 0.0076 | 0.0072 | 0.9128 |

**Figure 21. Allelic diversity, fixation index and genetic distance among three genetic populations in the Northern Subregion. (Na=number of alleles per loci, Ne= effective number of alleles; Pvt alleles = number of alleles private among populations x $10^{-1}$; I = Shannon information; Ho = observed heterozygosity (x100); He= expected heterozygosity; F=fixation index; %P=fraction of loci that are polymorphic; and Fst=proportion of the total genetic divergence that separates the populations.**



## DISCUSSION AND RECOMMENDATIONS

**Genetic Diversity and Breeding System.** Overall percent of polymorphic loci, heterozygosity and genetic distance in *Monolopia congdonii* were similar to other selfing taxa (Hamrick and Godt 1998, Nybom 2004). Although these other studies largely used allozymes and therefore the values may not necessarily be comparable, it is interesting to note that these values in *M. congdonii* are closer to the averages for selfing taxa with animal dispersal, which are general higher than gravity dispersed selfing taxa (Hamrick and Godt 1993). While polymorphism and heterozygosity are a magnitude smaller in the Northern Subregion and comparable to selfing taxa that are gravity dispersed.

In addition, heterozygosity in the Arroyo Conejo (0.470) and Kettleman BLM 6800 (0.523) collection sites are as high or greater than rates typical for outcrossing plants (He=0.43 to 0.47). Pleasant Valley Ecological Reserve (He=0.422), Lost Hills-North (0.406), and Jacilitos Canyon 2 (He= 0.419) are close. All other sites were in the range for selfing taxa or lower (He=0.006 to 0.370). Heterozygosity in all collection sites in the Northern Subregion were the lowest, 0.1 or less, but this may be due largely to the low number of polymorphic alleles rather than a difference in selfing rate.

The low levels of within site genetic variation found in *M*. *congdonii* are consistent with its life history traits (annual plant, selfing). Annual selfing taxa and/or early successional taxa allocate most of the genetic variability among populations and typically have the lowest levels of within-population variation (Nybom 2004). However, several populations with higher within site variation and heterozygosity and less genetic distance, suggest there is probably some dispersal to certain sites potentially acting as sinks. Alternatively, other factors such as population size, history and/or habitat heterogeneity may account or contribute to these differences. In most regions, between site differentiation is high but several factors in addition to selfing rate, such as colonization history, low rates of seed or pollen dispersal, and local adaption may also contribute to or help maintain this pattern. Divergence is not correlated to isolation.

Seven of the thirty-three collection sites showed some evidence of bottlenecking. The absence of heterozygote excess is not surprising because of the inherent lack of heterozygosity due to the low number of alleles and the breeding system, but it is also a signature of bottlenecks that would recover faster than the deficiency of rare alleles (Lowe et al. 2004). Alternatively, detection of rare alleles is more sensitive to sample size than other indices (Nei 1987) and most but not all sample sites that did not fit the expected model of allele frequency were small. Rare alleles were detected, however, in several other sites with relatively low sample sizes.

Inbreeding indices were high for all collection sites. This is expected for taxa with a high rate of selfing, but the implications may not be as detrimental as for perennial or outcrossing taxa. Inbreeding costs are different for primarily self-fertilizing taxa due to loss of deleterious recessive alleles expressed in is not expected homozygotes (Husband and Schemske 1996). However, non-lethal and weakly deleterious alleles are not purged and most likely expressed

later in the life history of the taxon rather than early (Husband and Schemske 1996). The effects on an organism with a very short life history, such as *M*. *congdonii*, could be minimal.

**Delineation of Conservation Units.** As expected, genetic structure in *Monolopia congdonii* is high despite the low levels of genetic diversity overall. Distribution of the six cpDNA haplotypes closely resembles the pattern of population structure resulting from microsatellite analysis suggesting more recent barriers to gene flow or selection is similar to historic conditions.

Results of genetic analyses suggest there are four to seven population clusters based on genetic and geographic considerations. Relationship between these different groupings can be considered a hierarchical island model and this pattern best fits the distribution of genetic diversity for the species. Microsatellite data corroborates the cpDNA, but neither genome indicates that individuals within a site or region are monophyletic. Distribution of haplotypes suggests isolation in some regions, but endemic haplotypes occur with more widespread haplotypes. In all these regions or subregions, additional structuring of genetic diversity occurs within these regions and must be considered for management activities.

The Northern Subregion (Monocline-Panoche-Silver Creek) is the most geographically isolated and strongly supported unit. It is the only region with a unique cpDNA haplotype (B) and microsatellite genotypes that are the most divergent from populations in the rest of its range including the highest mean Fst values among regions and subregions. These results elevate its conservation value. The cpDNA results, in combination with the microsatellite data, corroborate a long history of genetic isolation in this region but it is interesting that haplotype B was found in the few samples from the Mendota Plain and south towards the Coalinga Subregion.

The Coalinga Subregion is characterized by haplotypes A and C. Although neither of these are unique to the region, the ancestral haplotype (A) is the most frequent in this region suggesting it could be the location where the taxon evolved and dispersed to the north and south. Microsatellites also indicate it may have the highest diversity among all Subregions in terms of number of alleles, effective number of alleles and the highest number of private alleles.

Habitat loss and fragmentation for *M*. *congdonii* has probably been the highest in the Southern San Joaquin Valley Subregion. It is possible that populations occurred more or less across the southern valley and that there were historic populations between the Lost Hills and Eastern Valley sites in the Southern San Joaquin Valley Subregion. These two areas share the same cpDNA haplotypes but differ in haplotype frequencies. Microsatellite results corroborate the similarity but distinction between the two areas. These results are not conclusive as to the presence of historic populations.

Similarly, the Cuyama Valley and Carrizo/Elkhorn Plain share the same suite of haplotypes but frequencies differ. Microsatellites from each valley are more similar to each other but show some divergences especially in Santa Barbara Canyon. Recognition of both the Lost Hills and Eastern Valley as separate units, but combination of the Carrizo and Elkhorn Plains region and the Cuyama Valley region into one is recommended until further evidence suggests otherwise. Further investigation into the Santa Barbara Canyon site and microsatellite differences with other

Cuyama Valley and Carrizo/Elkhorn Plain populations is needed, including analyzing more samples from the Carrizo/Elkhorn Plains.

Samples from the Middlewater Plain and the valley floor northeast of Coalinga were comprised solely of Haplotype C, but occur in different areas, and these areas may warrant conservation status. Microsatellite data from these regions would be useful to determine if unique alleles of genotypes were or are in these populations. Populations were recently rediscovered in the Middlewater Plain, but there are no known extant populations on the floor of the San Joaquin Valley northeast of Coalinga.

There is a sharp transition in haplotypes northeast of Coalinga, possibly within 10 miles of the southernmost haplotype B sample. Likewise, haplotype C drops out west of the Middlewater Plain. These results suggest a strong historical barrier to gene flow between these areas, or stronger selection for genotypes associated with Haplotype C in the San Joaquin Valley floor, south of Cantua Creek and east of the South Coast Range foothills.

**Genetic Diversity and Recommendations for Restoration in the Northern Subregion.** Although the genetic diversity in the Northern Subregion is lower than all other Subregions, it is the most genetically unique. Diversity indices at collection sites in the Northern Subregion on average were significantly lower ($P<0.01$ to P0.001) than sites in other Subregions including percent of polymorphic loci, observed and expected heterozygosity, and Shannon Information. For example, the mean percentage of polymorphic loci across all sites in the Northern Subregion was 28% in contrast to a mean of 81% for all other populations. Sample size cannot account for this as the number of samples from the Northern Subregion was higher than for any other except the Coalinga Subregion. In contrast, the Northern Subregion has the highest genetic differentiation (Fst) among Subregions, an above average number of private alleles for the subregion and a unique cpDNA haplotype. Evidence for recent bottlenecks is weak and the low diversity and high divergence is potentially due to founder effects and subsequent isolation.

Four important characteristics of genetic diversity in the Northern Subregion should be taken into account when determining restoration activities in this area. First, it is probably the most genetically distinct Region or Subregion in the extant range of *Monolopia congdonii* based on the combined results of microsatellite and cpDNA data. This evidence suggests distinction is due to a history of isolation rather than isolation due to recent fragmentation of the range and/or drift. Second, genetic diversity within the Subregion is low with an average of one to two alleles per locus at a site and the highest level of homozygosity among all Subregions. This lack of diversity is most likely due to founder effects and lack of genetic exchange between Subregions.

Third, patterns of genetic diversity within the Subregion suggest there is limited gene flow between gene pools. There are three distinction gene pools within the region that show some geographic pattern, but not they are not geographically isolated from each other. The most distinct of these genepools, Lower Panoche, is adjacent to and close proximity to sites that

support the other two gene pools. The 'collection sites' are more than 0.25 miles apart and collections within these sites are <0.25 miles apart. The distance between collection site 1 and site 8 are just 0.75 miles apart and maintain some of the greatest genetic distance within the subregion (Figure 17). Conversely, sites 1 and 11 are over eight miles apart but minimally divergent from each other (Figure 17). However, it is not sufficient to understand if these genetic populations among sites are maintained by dispersal limitation (pollen and seeds) or if there is dispersal (most likely by fruit) but strong selection for genotypes that prevents establishment of these other genotypes.

Fourth, between year genetic differences in the recruited populations include heterozygosity and allelic diversity. Thus, loss of genetic diversity in this Region would be significant for the region and the species. Restoration and management should take these characteristics into consideration and the following recommendations:

1) Movement of genetic material in or out of the Northern Subregion should not occur in order to conserve the major gene pools.

2) Conservation and protection of gene pools and populations that support them by appropriate management and restoration of existing habitat is preferable and more likely to succeed than establishing new populations.

3) Movement of genetic material between the three gene pools/Areas within the Northern Subregion should be avoided, especially where small numbers of individuals occur to prevent swamping of genetic diversity. In addition, the Monocline population is the only population sampled that potentially only supports haplotype B and transfer of germplasm from other populations in this region is strongly discouraged.

4) Any new populations in the Northern Subregion should be located as far away as possible from existing sites and surveys should be done to confirm there are not existing unsampled populations that may represent additional gene pools in these areas.

5) Material for these new populations should take into consideration the geographic context and use material from the most appropriate gene pool. Sites that are outside of these geographic contexts could use material from all three gene pools.

6) Sampling material for seed storage or new sites, should occur over several years to maximize diversity.

**Other Recommendations for Restoration and Management.** All of the proposed Conservation Units or Subregions and Areas are important components of the genetic diversity in *Monolopia congdonii*. Potential for loss of genetic diversity is high due to the low frequency of many private alleles and large proportion of rare genotypes range wide or in regions. Movement of genes between these areas is not recommended in order to minimize risk of losing genetic diversity within *M. congdonii*. If there comes a time where assisted relocation is justified,

moving or mixing of genepools should be limited to areas where the probability of contaminating natural populations is very low.

In addition, only a few collection sites could be said to be rich in allelic diversity (i.e., Na> 2) or exhibited relatively high heterozygosity for the taxon (He>0.4), but those sites were among the least distinct (low Fst). They may represent sinks where or patters and they may receive migrants from several populations. Conversely, many sites with low allelic diversity, were among the most distinct. Those populations with low genetic diversity, may support the most unique diversity within a region. Thus, a large number of populations within a region, are necessary to protect the full, range of genetic diversity within a region. In addition, these genetically distinctive sites are often small. Therefore, they are also more vulnerable to swamping by large introductions of different genotypes. These patterns should be taken into account to minimize probability of introducing nonlocal genes.

**Acknowledgments.** Thanks to the Molecular Phylogenetics Lab (MPL) and the Evolutionary Genetics Lab (EGL) at UCB for use of facilities; to Lydia Smith for training on the ABI 3730; Bridget Wessa for help with sequencing of cpDNA regions; and Mark Wright for help running INSTRUCT.

## REFERENCES CITED


Aguilar, R., Quesada, M., Ashworth, L., Herrerias-Diego, Y. and J. Lobo. 2008. Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. Molecular Ecology 17:5177-5188.

Alsos, I.G., Ehrich, D., Thuiller, W., Eidesen, P.B., Tribsch, A., Schonswetter, P., Lagaye, C., Taberlet, P. and C. Brochmann. Genetic consequences of climate change for northern plants. Proceedings of the Royal Society B:Biological Sciences 279:2042-2051.

Anderson, J.T., Panetta, A.M. and T. Mitchell-Olds. 2012. Evolutionary and ecological responses to anthropogenic climate change. Plant Physiology 160(4):1728-1740.

Beerli, P. and J. Felsenstein. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in subpopulations by using a coalescent approach. Proceedings of the National Academy of Sciences 98(8):4563-4568.

Blacket, M.J., Robin, C., Good, R.T., Lee, S.F., and A.D. Miller. 2013. Universal primers for fluorescent labelling of PCR fragments—an efficient and cost-effective approach to genotyping by fluorescence. Molecular Ecology Resources 12(3):456-63.

Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134:1289–1303.

Charlesworth, D., and J. R. Pannell. 2001. Mating system and population structure in the light of coalescence theory. Pp. 73–95 *In* J. Silvertown and J. Antonovics, eds. Integrating ecology and evolution in a spatial context. Blackwell Scientific, London.

Cole, C.T. 2003. Genetic variation in rare and common plants. Annual Review of Ecology, Evolution and Systematics 34:213-237.

Cristescu, R., Sherwin W.B., Handasyde, K., Cahill, V. and D.W. Cooper. 2010. Detecting bottlenecks using BOTTLENECK 1.2.02 in wild populations: the importance of the microsatellite structure. Conservation Genetics 11:1043–1049.

Consortium of California Herbaria. 2016. Online database of vascular plant specimens in California herbaria. Online. Available: http://ucjeps.berkeley.edu/consortium/ (accessed October 2016).

Culley, T.M., Stamper, T.I., Stokes, R.L., Brzyski, J.R., Hardiman, N.A., Klooster, M.R., and B.J. Merritt. 2013. An Efficient Technique for Primer Development and Application that Integrates Fluorescent Labeling and Multiplex PCR. Applications in Plant Sciences 1(10): doi:10.3732/apps.1300027.

Doi, H., Takahashi, M. and I. Katano. 2010. Genetic diversity increases regional variation in phenological dates in response to climate change. Global Change Biology 16:373-379.

Evanno, G., Regnaut, S., J. Goudet. 2005. Detecting the number of clusters of individuals using the software Structure: a simulation study. Molecular Ecology 14(8): 2611–2620.

Gao, H., Williamson, S., and C.D. Bustamante. 2007. A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. Genetics 176:1635–1651.

Glemin, S., Bazin, E. and D. Charlesworth. 2006. Impact of mating systems on patterns of sequence polymorphism in flowering plants. Proceedings of the Royal Society of Biology 273:3011–3019.

Hamrick, J.L. and M.J.W. Godt. 1989. Allozyme diversity in plant species. In Plant Population Genetics, Breeding, and Genetic Resources, eds. Brown, A.H.D., Clegg, M.T., Kahler, A.L. and B.S. Weir, pp. 43-63. Sunderland, MA: Sinauer.

Hayden, M.J., Nguyen, T.M., Waterman, A. and K.J. Chalmers. 2008. Multiplex-Ready PCR: A new method for multiplexed SSR and SNP Genotyping. BMC Genomics 9(80) http://www.biomedcentral.com/1471-2164/9/80.

Holleley, C.E. and P.G. Geerts. 2009. Multiplex Manager 1.0: a cross-platform computer program that plans and optimizes multiplex PCR. BioTechniques 46: 511–517.

Husband, B.C. and D.W. Schemske. 1996. Evolution of the magnitude and timing of inbreeding depression in plants. Evolution 50:54–70.

Ingvarsson, P.K. 2002. A metapopulation perspective on genetic diversity and differentiation in partially self-fertilizing plants. Evolution 56(12):2368–2373.

Johnson, l.A., and D.E. Soltis. 1994. matK DNA sequences and phylogenetic reconstruction in Saxifragaceae sensu stricto. Systematic Botany 19: 143–156.

Jombart, T. 2008. Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 2008, 24:1403-1405.

Jombart, T., Devillard, S., Dufour, A.-B. and D. Pontier. 2008. Revealing cryptic spatial patterns in genetic variability by a new multivariate method. Heredity 101:92-103.

Jombart, T., Devillard, S. and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics 11(94) doi:10.1186/1471-2156-11-94

Jordan, W.C., Courtney, M.W. & Neigel, J.E. 1996. Low levels of intraspecific genetic variation at a rapidly evolving chloroplast DNA locus in North American duckweeds (Lemnaceae). American Journal of Botany 83: 430--439.

Jump, A.S., Marchant, R. and J. Penuelas. 2009. Environmental change and the option value of genetic diversity. Trends in Plant Science 14:51-58.

Kamvar, Z.N., Tabima, and N. J. Grünwald. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281 http://dx.doi.org/10.7717/peerj.281

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., and A. Drummond. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12):1647-1649.

Kelchner, S.A. and L.G. Clark. 1997. Molecular evolution and phylogenetic utility of the chloroplast *rpl16* intron in *Chusquea* and the Bambusoideae (Poaceae). Molecular Phylogenetics and Evolution 8: 385--397. http://dx.doi.org/10.1006/mpev.1997.0432.

Leimu, R. and M. Fischer. 2008. A Meta-Analysis of Local Adaptation in Plants. PLoS ONE 3:e4010.

Linhart, Y.B. and M.C. Grant. 1996. Evolutionary significance of local genetic differentiation in plants. Annual Review of Ecology and Systematics 27: 237–277.

Mazer, S. and B. Henrickson. 1993. Demography and Reproductive Biology of San Joaquin Woolly Threads (*Lembertia congdonii*). Final Report to the State of California, Département of Fish and Game. Contract #FG1460.

McGlaughlin, M.E. and E.A. Friar. 2011. Evolutionary diversification and geographical isolation in *Dubautia laxa* (Asteraceae), a widespread member of the Hawaiian silversword alliance. Annales of Botany 107: 357-370. http://dx.doi.org/10.1093/aob/mcq252

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Nybom, H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Molecular Ecology 13(5): 1143-1155.

Lowe, A., Harris, S. and P. Ashton. 2004. Ecological Genetics. Blackwell Publishing, Malden, Massachusetts.

Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J. And L. De Meester* 2013. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Molecular Ecology doi: 10.1111/mec.12561.

Panero, J.L. & Crozier, B.S. 2003. Primers for PCR amplification of Asteraceae chloroplast DNA. Lundellia 6: 1--9.

Parnell, J.R. and B. Charlesworth. 1999. Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. Evolution 53(3): 664-676.

Peakall R. and P.E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6: 288-95.

Piry, S., Luikart, G., J.M. Cornuet. 1999. BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. Journal of Heredity 90(4):502–503

Pollak, E. 1987 On the theory of partially inbreeding finite populations. I. Partial selfing. Genetics 117:353–360.

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

Reusch, T. B., Ehlers, A., Hammerli, A. and B. Worm. 2005. Ecosystem recovery after climatic extremes enhanced by genotypic diversity. Proceedings of the National Academy of Sciences USA 102:2826–2831.

Rice, W.R. 1989. Analyzing tables of statistical tests. Evolution 43(1):223–225.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. Genetics 145:1219-1228.

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

Sgro, C.M., Lowe, A.J. and A.A. Hoffmann. 2010. Building evolutionary resilience for conserving biodiversity under climate change. Evolutionary Applications 4: 326–337.

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

Swofford, D.L. 2002. *PAUP**. Phylogenetic Analysis using parsimony (*and other methods) v4.0b10. Sinauer Associates.

Vallone, P.M. and J.M. Butler. 2004. AutoDimer: a screening tool for primer-dimer and hairpin structures. Biotechniques. 37(2):226-231.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M. and P. Shipley. 2004. MICROCHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4:535–538.

Wilson G.A. and B. Rannala. 2003. Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163:1177–1191.