UNIVERSITY OF CALIFORNIA
RIVERSIDE

Patterns of Genetic and Morphological Variation in Manzanitas

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

Doctor of Philosophy

in

Plant Biology

by

Glen R. Morrison

September 2023

Dissertation Committee:
Dr. Amy Litt, Chairperson
Dr. Jason Stajich
Dr. Alan Brelsford

The Dissertation of Glen R. Morrison is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

I thank my loving wife, Jaz. Without her encouragement, I could not have done this.

The text of this dissertation, in part, is a reprint of the material as it appears in "*Subspecies Differentiation in an Enigmatic Chaparral Shrub*", published in American Journal of Botany, in its 6th edition of its 107th volume, in June 2020. The co-author Amy Litt directed and supervised the research which forms the basis for the third chapter of this dissertation. Contributions of other co-authors on that publication are listed in the Author Contributions section within the third chapter.

ABSTRACT OF THE DISSERTATION

Patterns of Genetic and Morphological Variation in Manzanitas

by

Glen R. Morrison

Doctor of Philosophy, Graduate Program in Plant Biology
University of California, Riverside, September 2023
Dr. Amy Litt, Chairperson

Manzanitas (*Arctostaphylos*, Ericaceae) are the most taxonomically diverse genus of woody plants in western North America, with an epicenter of diversity in the California Floristic Province (CFP). For more than a century, botanists have recognized the high diversity of the genus, and have worked to come to a satisfactory understanding of how best to circumscribe taxa in the genus. Additionally, there has been much curiosity about how the genus became so diverse in the CFP. However, addressing these questions using the tools of investigation that have been historically available has proven difficult.

I undertook three studies, utilizing updated and modern methods, aimed at addressing these questions. In the first chapter, I present the results of a phylogeographic study using reduced representation genome sequence data to reconstruct the biogeographic history of three widespread

species in western North America (*A. patula*, *A. pungens*, and *A. pringlei*). For each species, I tested whether they originated in the CFP, or outside of the CFP. I found a likely origin for *A. patula* to be within the CFP, a likely origin for *A. pungens* outside of the CFP, and I found an unclear geographic origin for *A. pringlei* with regard to the CFP - leaving the question of the origins of diversity in the genus unanswered..

In the second chapter, I present the results of a morphological study aimed at testing distinguishability and distinctiveness of manzanita taxa. I found that taxa can largely be distinguished, but also found overlap among many taxa in overall morphological variation. Together, this indicates that distinguishability is accounted for by specific, taxonomically prescribed, diagnostic traits. I also found evidence that morphological variation in the genus is not composed of naturally discrete, separated units, but rather, is relatively continuous.

In the third chapter, I present the results of a study investigating genetic differentiation among subspecies of a widespread species, *Arctostaphylos glandulosa,* with 10 described subspecies that are not completely separable by either morphology or geography. I found that the pattern of genetic differentiation within *A. glandulosa* corresponds to geography, rather than subspecies identity, with the exception of one subspecies, *A. glandulosa* subsp. *gabrielensis*, which was genetically distinct.

Taken together, this dissertation research shows the potential for improving our understanding of manzanitas with continued investigation utilizing modern methods, but also indicates that much more work remains to be done to better understand this challenging genus.

Table of Contents

List of Figures

Addendum 2: Classroom morphology activities.

List of Tables

List of Appendices

INTRODUCTION:

Manzanitas (*Arctostaphylos*, Ericaceae) are a diverse group of shrubs and small trees with an epicenter of diversity in the California Floristic Province (CFP), the region of western North America with a Mediterranean-climate. The CFP is a global biodiversity hotspot, supporting a large number of endemic plant species that are found only within the CFP. Manzanitas are emblematic of this endemism, as they are the most species-diverse woody plant genus native to the CFP. The diversity of the genus has long been apparent to those studying the flora of the CFP, and over the course of the last century many individual authors have described new taxa, leading to the current taxonomy which recognizes 60 species and 107 taxa, including subspecies.

While many manzanita taxa have been described, there has been a tendency for taxonomists to describe taxa without putting the delimiting traits in the context of the overall pattern of morphology in the genus. As a consequence, the morphological basis of considering some taxa species, and others subspecies, is not clear. Additional confusion regarding the morphological basis of the taxonomy comes from hybridization, which has been demonstrated to occur between many species, and has been inferred to be common based on observations that many species overlap in morphological traits and variation. Of course, the latter may be due to the former, but it may also be the case that some taxa that are now recognized, that are morphologically similar and/or hybridize readily, may be better considered subspecies, or conspecific populations of no taxonomic rank. As might be expected, the distinctiveness of manzanita subspecies is even less clear.

The traits that are used to describe subspecies are often difficult to assess, and there is often morphological overlap among taxa to a confusing degree. Additionally, while some manzanita subspecies have fairly distinct and non-overlapping ranges, many do not, with some polytypic species having subspecies whose geographic ranges are fully encompassed within the ranges of other subspecies. In such cases, some individual populations may be found in which all plants are identifiable to a single subspecies, but in other populations, plants may be identifiable to several subspecies, or none. While there is hardly any botanical agreement on what exactly a subspecies is, differences in geographic range or ecological preferences are usually considered necessary. Many subspecies of manzanitas do not meet either criterion. This leaves the biological meaning of manzanita subspecies an open question.

While there is considerable uncertainty around how to best divide and describe diversity in manzanitas, there is not much doubt that the genus is quite diverse. Understanding how manzanitas became so diverse has thus long been a goal of manzanita researchers. One part of this is to elucidate the historical biogeography of manzanita species, i.e. the context of where and when species evolved, and how they came to be where they are today. Because all extant species diversity is represented in the CFP, and there are few species with ranges that extend beyond the CFP borders, this has caused many to assume that the current pattern of diversity is the result of extensive speciation within the CFP. This is a reasonable hypothesis, of course, but it is not the only possible explanation. We know manzanitas were once more widespread across western North America, and it is plausible that some or many species may have arisen elsewhere, expanding into the CFP before going extinct elsewhere. However, these competing hypotheses have never been tested.

In this dissertation, I address questions stemming from these complexities and uncertainties. Chapter one is a phylogeographic study aimed at elucidating the biogeographic

histories of three widespread western North American manzanita species. I bring new, empirical evidence to the discussion of the origins of manzanita diversity in the CFP, as well as the overall floristic endemism of the region. My other two chapters are aimed at assessing how well taxonomy corresponds to the structure of natural variation. Chapter two is a study of morphological variation in the genus, asking whether taxa are morphologically distinguishable and/or distinct, including specific analyses on species, subspecies, narrow endemics taxa, and widespread taxa. Chapter three is a study of genetic variation in subspecies of a particularly diverse polytypic species, in which I compare the structure of genetic variation with subspecific identity and other non-taxonomic factors, such as geography and environment.

The individual studies I conducted as part of this dissertation deal with related, but distinct, aspects of diversity and evolution. However, these studies are united in their application of modern techniques to generate and analyze data on natural variation, at scales that have not been achieved in previous studies of manzanitas. This approach allowed me to gain new insights where previous studies aimed at addressing the same questions and hypotheses may not have succeeded. With continued research undertaken at similar scales, there is much potential for future discovery in this historically challenging genus.

Chapter 1: Biogeographic History of Three Western North American Manzanita Species.

ABSTRACT:

Premise: The California Floristic Province (CFP) is a region of western North America with a Mediterranean-type climate that is known for its remarkably high endemic plant diversity. Understanding the origin of this endemic diversity is a long-standing goal of North American biogeography, and two main hypotheses have been put forward. The first hypothesis is that the high endemic diversity of the CFP is the result of diversification within the CFP, following the onset of the Mediterranean climate. The second hypothesis is that lineages and species arose elsewhere in North America, and migrated into the CFP before disappearing elsewhere in North America. Manzanitas (genus *Arctostaphylos*, Ericaceae) are emblematic of CFP plant lineages in their endemic diversity, with all 60 species of manzanitas represented in the CFP, but only a handful extending outside of the CFP. The manzanita species with ranges that include areas outside of the CFP occur in areas implicated in the second hypothesis, making manzanitas a good system for testing the two hypotheses. In this study we reconstruct the most likely biogeographic history of three widespread species to determine whether they likely originated within the CFP, consistent with hypothesis 1, or outside of the CFP, consistent with hypothesis 2.

Methods: We sampled *A. patula, A. pungens*, and *A. pringlei,* three species that occur both within and outside of the CFP, across their ranges. I generated double-digest restriction site-associated (ddRADseq) data for a set of 128 samples, and estimated relationships among the samples using phylogenetic methods, before applying models of historical biogeographic reconstruction for each species.

Key results: We found that *A. patula* likely arose in the southern Sierra Nevada, within the CFP, while *A. pungens* likely arose outside of the CFP, in the Basin and Range Province or northern Sierra Madre Occidental of mainland Mexico. Our analyses regarding *A. pringlei* were inconclusive and consistent with the current two-region distribution being a result of vicariance that resulted from the formation of the Sonoran and Mojave Deserts. Additionally, we found evidence of widespread interspecific gene flow between *A. pungens* and both *A. pringlei* and *A. patula*, but not between *A. pringlei* and *A. patula.*

Conclusion: We found different patterns of geographic origin and dispersal history for each species, with support for both hypotheses. Overall, the results suggest that speciation in areas outside of the CFP may have contributed to the modern endemic diversity of manzanitas in the CFP. By extension, these results show the possible importance of historical immigration of plant lineages into the CFP in forming the endemic diversity we see there today.

INTRODUCTION:

On the Pacific coast of North America lies a region with exceptionally high botanical diversity, known as the California Floristic Province (CFP). The CFP ranges from southern Oregon down through California into northern Baja California, Mexico, and is defined by its Mediterranean-type climate, with dry, hot summers, and cool wet winters. Of the approximately 7,000 native plant species that occur in the California Floristic Province, around 60 percent are endemic to the CFP (IUCN 2016; Burge et al. 2016). Understanding how and why this endemic diversity came into being has long been a goal of biogeographers seeking to understand patterns of organismal diversity and distributions in North America.

Over the last century two major hypotheses have been proposed to explain the origin of endemic diversity in the CFP. The first is that the diversity we see today is the result of diversification and speciation *within* the CFP, starting from around the time of the onset of the Mediterranean-type climate regime around 15 million years ago (Millar and Woolfenden 2016; Bruce G. Baldwin et al. 2012; Kauffmann, Parker, and Vasey 2021). Various mechanisms have been proposed as drivers of this within-CFP diversification, ranging from the establishment of a relatively stable climate that allowed for increased retention of newly evolved species (Raven and Axelrod 1978), to the unique features of the CFP's geology and topography (e.g. highly varied soil substrate types, varied elevations and aspects, etc.) that created a diversity of habitats leading to adaptive diversification of lineages present in California (Thornhill et al. 2017; Lancaster and Kay 2013; Burge et al. 2016; B. G. Baldwin 2014). Of course, these mechanisms are not mutually exclusive, and may have worked in tandem, producing a situation in which rapid diversification

inside the CFP existed alongside conditions for relatively high long-term survival of newly evolved species.

The second hypothesis that has been proposed to account for the current endemic diversity of the CFP is that species may have evolved outside of the CFP and migrated in in the past, but have since gone extinct outside of the CFP. This hypothesis is supported by the presence of plant communities inland from the CFP that have similar plant lineages and overall structure to plant communities in the CFP. In parts of the southwestern US, a shrubland biome is found that is often called “interior chaparral”, which supports a number of genera that also include species found in the CFP (Brown 1978). These floristic ties may be the result of dispersal from the chaparral biome of the CFP, but may just as likely be the result of dispersal in the other direction, consistent with the second hypothesis.

Additionally, fossil evidence suggests that in the past, a more mesic climate existed in much of the desert basins of the southwestern US, meaning that shrubland plant communities in the interior and those in the area of the modern CFP may have been better connected in the past (Hastings 1965; Brown 1978). Some regions like the modern Great Basin are thought to have been a high elevation, relatively less arid plateau (the Nevadaplano), between the middle Miocene and the early Pliocene (roughly 12 to 5.3 million years ago) (Wolfe 1964). These have a fossil flora of plant communities more like those in the modern CFP than those of the Great Basin today (Millar and Woolfenden 2016; Kauffmann, Parker, and Vasey 2021; Bruce G. Baldwin et al. 2012). Likewise, fossilized packrat middens in parts of the modern Sonoran and Mojave Deserts show that chaparral-like shrubland plant communities were present in areas that are currently desert. It has been suggested that many elements of these paleofloras went extinct locally as

climates became more arid in much of western North America (Raven and Axelrod 1978). Before going extinct, however, such lineages may have spread to the CFP where they persisted and/or diversified (Raven and Axelrod 1978).

While this potential paleo-migration has long been of interest to biogeographers of North America and the CFP, most studies attempting to elucidate the origins of CFP endemism have focused on the first hypothesis; these often only include sampling from the CFP, and are thus unable to evaluate evidence for historical immigration from elsewhere. Many of these studies do not even include sampling from the portions of the CFP outside of the political borders of the state of California (Burge et al. 2016; Thornhill et al. 2017).

Of the many plant lineages with disproportionately high endemic diversity within the CFP, perhaps none is more emblematic of this pattern than manzanitas, *Arctostaphylos* Adans. (Ericaceae, subfamily Arbutoideae). Manzanitas are shrubs and small trees known for their striking red bark and twisting branches, and with at least 60 currently recognized species, they are the most species-diverse woody genus in all of western North America (Kauffmann, Parker, and Vasey 2021; Keeley, Parker, and Vasey 2017). All manzanita species occur within the CFP, and 55 of them occur *only* in the CFP. This center of diversity in the CFP may appear to lend support to the first hypothesis, that manzanitas arose in the CFP and only a few species were able to disperse beyond it. However, there is fossil evidence of manzanitas outside of the CFP going back as far as the onset of the Mediterranean-type climate in the CFP. The oldest verified fossil of a manzanita is material of *A. masoni* Wolfe (Wolfe 1964), dated to around 15.8 million years ago during the time of the Nevadaplano (Wolfe 1964, Axelrod 1958). This species, identified from multiple fossil leaves recovered from beds in western Nevada, may have resembled the extant,

prostrate species *A. nevadensis,* which occurs today to the west of the fossil site, in the Sierra Nevada north through the Klamath and Cascade Ranges, as well as in the Elkhorn Range of interior/eastern Oregon. The early presence of manzanitas in the western areas of what was once the Nevadaplano might suggest an ancestral range for this genus, or at least some of its species, outside of the CFP, in other regions of western North America, consistent with hypothesis 2.

I used three widespread manzanita species, all of which occur within and outside of the CFP, to test these hypotheses. *Arctostaphylos pungens* Kunth (Mexican manzanita, pointleaf manzanita, or pingüica) has the largest range of the three species, occurring throughout the mountainous regions of mainland Mexico and northern Baja California, and in the United States in several mountain ranges of Southern and Central California, as well as in the southern Basin and Range Province (Fig 1.1). The range of *A. pungens* is particularly intriguing as the vast majority of its range is actually outside of the CFP, and is shifted considerably farther to the south than other manzanitas. *Arctostaphylos patula* Greene (greenleaf manzanita) has the second largest range, with a more northerly distribution, inhabiting the mountains that form the western, southern and eastern flanks of the Great Basin, with a few scattered populations on higher mountains within the Great Basin, as well as populations in the higher mountains of Southern California and northern Baja California (Fig 1.1). The range of *Arctostaphylos pringlei* Parry (pinkbract manzanita) is composed of two discontiguous regions, one in the mountains of Southern California and northern Baja California, and one further inland, spanning the mountains of Arizona, and extending slightly into very southern edges of Nevada and Utah (Bruce G. Baldwin et al. 2012; Kauffmann, Parker, and Vasey 2021; Parker, Vasey, and Keeley 2020).

I leveraged genome-scale DNA sequencing and biogeographic reconstruction to identify the most likely hypotheses for the biogeographic history of each species and evaluated evidence for an origin of each species within or outside of the CFP, as well as finer scale signals of dispersal or vicariance events that may have produced the modern ranges of these species. The results suggest a different history for each species, with support for both hypotheses. This suggests a complex history for this genus, and ultimately the CFP, with a combination of both hypotheses underlying the current endemic diversity.

METHODS:

Plant material—

To assemble our sample set I first determined the overall ranges of our focal species using the range maps in the *Field Guide to Manzanitas* (Kauffmann, Parker, and Vasey 2021), supplemented by public data (notes, specimen imagery, etc.) from georeferenced herbarium collections available through the California Consortium of Herbaria (cch2.org), SEINet (https://swbiodiversity.org/), and Tropicos databases (tropicos.org). I then identified target collecting localities across the range of each species, including localities at far extremes of the ranges (e.g. southernmost, northernmost ends of a range etc.), as well as localities well spaced out within the main body of the range. I targeted localities where recent collections indicated that a given species was likely currently present. For areas where few herbarium collections have been made, or collections are too old for the exact location to be known (i.e. pre-GPS), I searched the iNaturalist database of photographic observations (inaturalist.org) of that species, checking for

photographs that show diagnostic features, to identify specific localities with current populations. When feasible, we collected young floral bud material or young leaf material for DNA extraction, and when such fresh growth was unavailable, we collected mature leaf material. If possible, we kept material on ice in the field before storing it at -20° C. Otherwise we collected and stored material on silica gel. Samples from Mexico were collected by Laura Trejo-Hernández (Universidad Nacional Autónoma de México), Maria Socorro González-Elizondo (Instituto Politecnico Nacional, CIIDIR Herbarium), and Lluvia Flores-Rentería (California State University, San Diego), and were collected on silica gel, and shipped to UCR for extraction. We collected a voucher specimen for each individual we sampled. In total we sampled ten localities of *A. patula*, 28 localities of *A. pungens*, and 8 localities of *A. pringlei*, sampling 5 samples per species per locality when possible, and at least three otherwise.

Herbarium specimens for collections made in the United States were deposited in the University of California, Riverside Herbarium (UCR), and those collected in Mexico were deposited at the Herbarium of Instituto Politécnico Nacional, CIIDIR Unidad Durango (CIIDIR). See Tables 1.1-1.3 lists of collections (of *A. patula, A. pringlei,* and *A. pungens*, respectively) included in this study, with associated dates, collection numbers, and geographic coordinates. All collections were made with advance permission or official permits for all localities.

DNA extraction—

We performed all lab work with the exception of DNA sequencing in our laboratory at UC Riverside. After grinding samples of frozen fresh tissue or silica-dried tissue by hand for 10-15 minutes in liquid nitrogen, I extracted DNA using the Qiagen DNeasy kit (catalog number:

69104) (Hilden, Germany), and then estimated DNA concentration using a spectrophotometer (Eppendorf BioSpectrometer, Hamburg, Germany). We checked for the presence of high molecular weight DNA by loading at least 4 uL of DNA solution on a 1% agarose gel, alongside 1 uL of a 1kb+ DNA ladder (catalog number N3200S) (New England Biolabs, Ipswitch, Massachusetts, USA), and running at approximately 80 V for 40-50 minutes. I imaged these gels using a BioRad GelDoc xr+ gel documentation imaging system (Bio-Rad, Hercules, California, USA), and inspected the resulting DNA for brightness, and distribution of molecular weight of each sample.

For the 238 samples that passed these quality control steps, we prepared ddRADseq libraries using a modified version of the protocol published in Brelsford et al. (2016), which is itself a modification of the protocols used in Parchman et al. (2012) and Peterson et al. (2012). The full protocol is provided in Appendix 1.1, and primarily differs from the protocol in Brelsford et al. (2016) by using restriction enzymes PstI and MspI for digestion, and by pooling individual samples into batches after the ligation of barcodes, rather than after PCR amplification. Libraries were sequenced by Novogene (Beijing, China) at their facility in Davis, CA. The total of 238 samples were prepared simultaneously as three batches of 79 or 80 samples each, and each batch of prepared libraries was sequenced in its own lane on an Illumina (San Diego, California, USA) HiSeq platform, using a 150 bp, paired-end sequencing chemistry, to a target depth of approximately 5 million reads per sample..

We processed raw Illumina reads using the iPyRAD pipeline v.0.9.85 (Eaton and Overcast 2020), with the default specifications, but using a reference genome and setting the minimum number of samples per recovered locus to 52 (40% of samples being processed). After

initially demultiplexing samples, we aligned reads to a chromosome-scale genome assembly for the congeneric bigberry manzanita, *Arctostaphylos glauca* Lindl. (Huang et al. 2022). After aligning reads, I found that a proportion of samples were not sufficiently productive in terms of mapped reads for downstream analyses. The iPyRAD statistics for the alignment showed a steep drop-off in mapped/aligned reads after 128 samples. I therefore retained these top 128 for subsequent analysis. The samples in this subset had a minimum of 768 loci represented, a maximum of 2,262 loci, and a mean of 1,476 loci (with a standard deviation of 429 loci). This sample set provided between one and five samples from thirty localities (Fig 1.1).

Assessment of species-level differentiation and identification of admixed individuals—

I first analyzed this dataset using fastStructure (Raj, Stephens, and Pritchard 2014) on the entire dataset on the entire data set to identify putatively admixed individuals, as hybridization is thought to be common in *Arctostaphylos* (Dobzhansky 1953; Gankin 1967; Schmid, Mallory, and Tucker 1968; Gottlieb 1968; Keeley 1976; Ellstrand et al. 1987; Nason, Ellstrand, and Arnold 1992; Schierenbeck, Stebbins, and Patterson 1992; Parker et al. 2020; Serkanic, Parker, and Schierenbeck 2021; Kauffmann, Parker, and Vasey 2021). This analysis revealed a number of individual admixed samples, and several localities composed of admixed individuals. Cluster results at k = 2 and k = 3 were used for subsequent analysis; higher k values were found to be effectively identical to the result at k = 3, with additional clusters receiving zero, or miniscule assignment proportions for any sample (Fig. 1.2). Because some samples appeared more admixed at either k = 2 or k = 3, I retained the fastStructure results at both k values to identify putative admixed individuals.

The first step in performing a historical biogeographical reconstruction is to estimate the relationships among samples from different areas, and to root that topology using an outgroup (Matzke 2014). I therefore added into the dataset two samples from different genera in the same subfamily of the Ericaceae (Arbutoideae), *Arbutus unedo* L. (strawberry tree) and *Arctous alpina* (L.) Nied. (alpine bearberry), to serve as outgroup taxa in rooting our tree. The data for these samples were prepared and sequenced using the same ddRADseq method as for the *Arctostaphylos* samples. To build a phylogenetic hypothesis we used IQ-TREE (Nguyen et al. 2015), to find the best supported model of nucleotide evolution (TVM+F+I+R5), and estimate a topology. Given the computational challenge of computing traditional bootstraps for the number of tips and loci included in the analysis, we used the ultra-fast bootstrap (UBS) method, computing 1,000 UBS to evaluate support for the topology. We then visualized the resulting tree using FigTree (tree.bio.ed.ac.uk/software/figtree/).

For some subsequent analyses in which admixed individuals would confound phylogeographic signals, we prepared a second dataset excluding any sample that had less than 95% assignment to a single fastStructure cluster at k = 2 or k = 3 in the analysis performed on the initial 128-sample dataset of all three species. Applying this threshold left us with 80 samples of relatively clear species assignment. Using this 80-sample subset, we computed a new phylogenetic topology for phylogeographic interpretation and inference using the same method described above. We computed a time-calibrated topology using the chronos function, with the relaxed model (lambda parameter set to 1), in the ape, v. 5.0 (Paradis and Schliep 2019), in the R statistical language and environment v. 4.3.0 (R Core Team 2023). I specified the divergence date of *Arctostaphylos* from *Arbutus unedo* at a minimum of 21.2 million years ago, and a maximum of 39.2 millions years ago, based on molecular clock divergence dates estimated from the

sequences of the ITS region and nuclear large subunit ribosomal DNA previously published in Hileman et al. (Hileman, Vasey, and Parker 2001).

We used this time-calibrated topology for historical biogeographic inference using the BioGeoBEARS package (Matzke 2014) in R. In order to perform this analysis, I first grouped localities into areas for biogeographic inference. Some of these areas corresponded to physiographic regions (e.g. the Basin and Range Province) and some to subdivisions, for instance northern and southern portions of major mountain ranges (Fig. 1.3). These subdivisions allowed for the detection of dispersal within those mountain ranges. I then extracted three single-species subtrees from the main time-calibrated tree, and from each of those trees, I produced a simplified topology by reducing single-locality clades to a single tip. To better inform model inference in the BioGeoBEARS analysis, I constructed geographic distance matrices for each species, by calculating the pairwise geographic distances between the approximate centers (centroids) of each polygon shown on Fig.1.3. I also constructed adjacency matrices for each species, encoding which areas are adjacent to which others. The BioGeoBEARS analysis uses these matrices in calculating the likelihood of dispersal from a given area to another. Where an area within a species range is very remote from any other area within the range (e.g. the Colorado Plateau portion of the range of *A. patula*, Fig. 1.3), that area was noted as not adjacent to any other area. Finally, I evaluated likelihood statistics for BioGeoBEARS model formulations to select a best supported model for each species.

RESULTS:

The sequencing batches produced a total of 756 million raw reads, with 99.5% of individual raw reads passing the read-level quality control filters implemented in iPyRAD. 74.8% of individual raw reads that passed quality control filters were able to be aligned to the reference genome and were processed into 2,525 unique loci. Of the 128 manzanita samples that made it through all quality control, fastStructure identified 48 (37.5%) as admixed. *A. patula* had the highest percentage of putatively admixed samples (50%, twelve of 24 samples), followed by *A. pungens*, with 33 of 85 samples (approximately 39%). *A. pringlei* had the smallest percentage, with three of nineteen samples (approximately 16%).

Phylogenetic relationships and monophyly of focal species—

A. pringlei is monophyletic in this analysis and is found to be sister to a clade containing all samples of *A. patula* and *A. pungens* (Fig. 1.4). Within this clade, both *A. patula* and *A. pungens* are found to be polyphyletic. In the case of *A. patula,* the species is polyphyletic due to the placement of a single, relatively highly admixed sample identified as *A. patula* amongst samples of *A. pungens*. Without that sample*, A. patula* is monophyletic, and *A. pungens* paraphyletic. A main clade of *A. pungens* is dominated by samples that show little or no signal of admixture. All other samples of *A. pungens,* placed elsewhere in the tree, show relatively high admixture as inferred in the fastStructure analyses (Fig. 1.2).

When all putatively admixed samples are omitted, phylogenetic analyses recover monophyletic clades for all three species (Fig. 1.5). In this version of the tree, *A. pringlei* is sister

to a clade containing both *A. patula* and *A. pungens*, which are in turn sister to each other. It was this topology that I used as a starting point in preparing simplified within-species, population trees that were used as an input for historical biogeographic reconstructions.

***Best supported hypothesis of biogeographical history for* A. patula**—

The results of the BioGeoBEARS analysis of *A. patula* identifies the southern half of the Sierra Nevada, in the CFP, as the likely ancestral area for the species (Figs.1. 6-1.7). The lineage from the northern Sierra Nevada is inferred to be the result of dispersal northward along the Sierra Nevada approximately 8.5 million years ago (Figs. 1.6-1.7). The populations outside of the Sierra Nevada are inferred to be the result of dispersal out of the Sierra Nevada, approximately 10 million years ago, first southwards into the Southern California area, and subsequently northeast into the Colorado Plateau, approximately four million years ago (Figs. 1.6-1.7).

***Best supported hypothesis of biogeographical history for* A. pringlei**—

The topology recovered for *A. pringlei* divides our samples into two clades, estimated to have diverged approximately 10 million years ago, one composed of samples from the CFP, and one composed of those from the Basin and Range Province (Fig. 1.8). This bifurcated topology and bifurcated geographic range precluded a meaningful inference of biogeographical history for *A. pringlei*. The best biogeographic hypothesis produced by BioGeoBEARS is uninformative, with the species most likely arising somewhere within Southern California or the Basin and Range Province, i.e. somewhere within its current range. This kind of result in BioGeoBEARS can be consistent with a vicariance event (Nicholas J. Matke, personal communication).

***Best supported hypothesis of biogeographical history for* A. pungens**—

Analysis suggests the ancestral area for *A. pungens* is in the area of the Basin and Range Province or northern Sierra Madre Occidental. The results suggest that populations in Southern California are the result of dispersal into that region approximately 10 million years ago (Figs. 1.9-1.10). The populations further south and east in Mexico are the result of a series of dispersal events beginning with southward dispersal to the Trans-Mexican Volcanic Belt, approximately nine million years ago (Figs.1. 9-1.10). Subsequently, from the Trans-Mexican Volcanic Belt, one descendant lineage dispersed northward into the Sierra Madre Oriental, approximately seven million years ago, and another dispersed further southward from the Trans-Mexican Volcanic Belt to the Sierra Madre del Sur, approximately four million years ago (Figs. 1.9-1.10).

***Admixed* A. patula *and* A. pungens**—

While our fastStructure analyses detected some populations that appear relatively genetically pure, in some populations we detected individuals with mixed assignment (Fig. 1.2). These ranged from apparent F1 hybrids (based on approximately 50/50% assignment to two clusters), to those that were largely assigned to one genotype with a small amount of a second one. Collections of *A. patula* and *A. pungens* from Navajo Mountain, in the Navajo Nation, appear consistently admixed (Fig. 1.2). We collected plants which we morphologically determined as *A. pungens* at a location ~1,550 m above sea level, in the middle slopes of the mountain, and collected plants which we determined as as *A. patula* at ~2,650 m above sea level These elevations are consistent with the general habitat preferences of those species. In spite of the plants of each population showing species-diagnostic characters (e.g. presence of compressed

inflorescence axis diagnostic of *A. pungens*, and the deeply green and round leaves that are diagnostic of *A. patula*), they also appeared to share some morphological similarity. For example, all the plants on Navajo Mountain had the mounded habit, which is typical of *A. patula,* but not typical of *A. pungens*. Our genetic data showed that all sampled individuals from Navajo Mountain are likely admixed between the two species, with the individuals from the lower elevation *A. pungens*-like locality averaging 83% assignment to the *A. pungens* cluster, and the higher elevation *A. patula*-like locality averaging 60% assignment to the same *A. pungens*-dominated cluster.

DISCUSSION:

Phylogenetic relationships and monophyly of focal species—

In the topology of the tree estimated with all samples, including the putatively admixed samples, those samples are consistently placed as early diverging lineages near the base of the *A. patula-A. pungens* clade, or near the base of the two clades that contain most samples of *A. patula* and *A. pungens* (Fig. 1.4). The bifurcating structure of a phylogenetic tree cannot accurately represent samples that are likely the result of reticulating ancestry (hybridization), and the presence of variability from both genomes may account for their placement nearer the base of clades. The monophyly of all three focal species, when these samples are removed, is also consistent with this hypothesis. Taken all together, and considering the identification of three species-correlated clusters in the fastStructure results, the results confirm that *A. patula* and *A. pungens* are distinct , but with some ongoing interspecific gene flow.

Interestingly, the inferred closer relationship of *A. patula* and *A. pungens* found in the phylogenetic analysis is at odds with the pattern of clustering found in the fastStructure analysis (Fig. 1.2). At k = 2, the results show one cluster dominated by *A. pungens*, and one dominated by *A. patula* and *A. pringlei*, suggesting that *A. patula* and *A. pringlei* are more genetically similar to each other than they are to *A. pungens*. This difference in results may be due to differences in how the these two methods (phylogenetic analysis vs. fastStructure) handle the input data, with fastStructure weighting all characters equally and maximum likelihood weighting differentially, and how the results produced by the two methods are affected by sample size (Meirmans 2018). Although this complicates estimates of relationships, it does not interfere with the goal of the fastStructure analysis, which was to identify admixed individuals.

Previous phylogenetic studies utilizing the nuclear ribosomal ITS region have found evidence of two clades (the "large" and "small" clades), representing a relatively ancient divergence event in *Arctostaphylos* (Boykin et al. 2005; Wahlert, Parker, and Vasey 2009). *A. pringlei* has been placed in the large clade, and *A. patula* in the small clade. Interestingly, whereas the analysis of Boykin et al. (2005) put *A. pungens* and *A. pringlei* together in the large clade, the analysis of Wahlert et al. (2009) was ambiguous: one sample of *A. pungens* from the CFP (Southern California) was placed in the large clade, whereas a second sample from the Basin and Range Province (Arizona) was placed in the small clade. The results of the current phylogenetic analysis suggest a closer relationship between *A. pungens* and *A. patula* than between *A. pungens* and *A. pringlei* (Fig 1.5), which lends support to the placement of *A. pungens* in the small clade with *A. patula*, in contrast to the results of Boykin et al. (2005). If that placement is correct, the fastStructure results (k = 3, 4) showing that admixture between *A. pringlei* and *A. pungens* (introgression between clades according to the phylogenetic results of

this study) is as common as admixture between *A. patula* and *A. pungens* (introgression within clade) (Fig 1.2); this is inconsistent with hypotheses suggesting hybridization is unlikely or infrequent between the two ITS clades (Parker et al. 2020). These results therefore suggest a potentially weaker barrier to hybridization across the clades than was previously thought, but also clearly indicate that further analysis of the phylogenetic placement of *A. pungens* is needed, including the impact of admixed individuals in providing literally mixed signals.

Historical biogeography of A. patula—

The best supported hypothesis for the biogeographic history of *A. patula* suggests the species originated in the southern Sierra Nevada, with one lineage dispersing northward within the mountain range, and another lineage dispersing to Southern California and subsequently to the northeast and inland to the Colorado Plateau (Figs. 1.6-1.7). While the Sierra Nevada is relatively contiguous with the mountains of Southern California, making dispersal feasible, the dispersal of *A. patula* from Southern California to the Colorado Plateau, some 300 miles (480 km) away, is a much more substantial event, which is made more intriguing when the habitat preferences of *A. patula* are taken into account. *A. patula* occurs today at relatively high altitudes, where summer temperatures are relatively mild, and winter temperatures are low with typically snowy conditions. These conditions do not occur on the higher peaks of the mountains of the deserts between Southern California and the Colorado Plateau. However, the estimated timing of the divergence of plants in the Colorado plateau from those in Southern California (just over four million years ago) is close to the upper estimated date range for the formation of the higher peaks of the Sierra Nevada (5-10 million years ago), which brought about a rain shadow that contributed to the formation of the dry climate in the Basin and Range Province (Millar and

Woolfenden 2016). It is conceivable that the higher elevation peaks of the early Great Basin may have furnished climates more suitable for *A. patula* to spread gradually across the area of the Great Basin from Southern California to the Colorado Plateau.

Dispersal by bird is another possible mechanism to explain the movement of *A. patula* across the range gap of the Great Basin. The range of another manzanita species, *Arctostaphylos uva-ursi* (L.) Spreng., supports the plausibility of long-range dispersal of manzanitas seeds. *A. uva-ursi* has a large circumboreal distribution, in both North America and Eurasia, but also occurs in relatively isolated populations hundreds of miles from the next closest population. One population of *A. uva-ursi* is even known from Guatemala, separated from the nearest conspecific populations (in high mountains of the Basin and Range Province) by approximately 1,800 miles (2,900 km), and the entire length of Mexico. Researchers have hypothesized that the seeds of *A. uva-ursi* were carried to such remote locations by frugivorous birds (Kauffmann, Parker, and Vasey 2021). Such a mechanism could explain the long-distance dispersal of *A. patula* suggested by this analysis.

Historical biogeography of A. pringlei—

The topology we recovered for *A. pringlei* is consistent with a historical vicariance event (Fig. 1.8). Fossil and paleoclimate reconstructions suggest that the truly arid regions of North America, like the Mojave and Sonoran Deserts that separate the two halves of the range of *A. pringlei*, may have only developed in the late Neogene period, likely sometime in the last 10 millions years (Millar and Woolfenden 2016). This timing is consistent with the time of divergence between the inland clade (*A. pringlei* subsp. *pringlei*) and the coastal clade (*A.*

pringlei subsp. *drupacea*). It is possible that *A. pringlei* once occurred over a more contiguous range, covering the area of Southern California inland to the mountains of the modern Basin and Range Province, and was subsequently divided by increasing aridification that came with the climatic development of the Mojave Desert. These two populations subsequently diverged to form two subspecies.

Historical biogeography of A. pungens—

The best supported hypothesis for the biogeographic history of *A. pungens* is that the species originated somewhere in the area of the Basin and Range Province, or northern Sierra Madre Occidental, with subsequent dispersals into the CFP in Southern California and southwards through the Sierra Madre Occidental to reach the rest of its range in Mexico by way of the Trans-Mexican Volcanic Belt (Figs. 1.9-1.10). The analysis suggests that the species had reached all of these regions by about 4 million years ago (Figs. 1.9-1.10). The inferred history involves sequential dispersal from one area southward to a geographically adjacent one, with two exceptions: (1) the dispersal from the Basin and Range to Southern California, and (2) the dispersal from the Trans-Mexican Volcanic Belt to the northern Sierra Madre Oriental (Fig. 1.9).

The lineage of *A. pungens* that dispersed from the Basin and Range Province to the CFP is estimated to have diverged from other lineages in the interior around 10 million years ago (Fig. 1.9). This puts the divergence at around the same time that the Sonoran and Mojave Deserts are thought to have begun to form (Millar and Woolfenden 2016). This suggests that at the time of dispersal into Southern California, the climates of the intervening areas of the modern Mojave and Sonoran Deserts may have been somewhat more mesic, allowing more contiguous habitat in

between the Basin and Range Province and the Southern California, and a lower barrier to dispersal.

The dispersal of *A. pungens* from the Trans-Mexican Volcanic Belt to the northern Sierra Madre Oriental would have occurred across the Mexican Plateau, the high elevation arid plateau that lies between the Sierra Madres Occidental and Oriental, and north of the Trans-Mexican Volcanic Belt. This dispersal event is estimated to have occurred around seven million years ago. The prevailing hypothesis on the climatic history of the Mexican Plateau is that the climate has been relatively stable there since the middle Miocene (15-11 million years ago) (Fig. 1.9). Taken together, these dates imply the Mexican Plateau would have posed some barrier to dispersal even at the estimated time of dispersal. However, it is worth noting that the estimated probability of dispersal from the Trans-Mexican Volcanic Belt to the northern Sierra Madre Oriental is approximately equal to the probability of dispersal from the Trans-Mexican Volcanic Belt to the southern Sierra Madre Oriental (Fig. 1.10), which is a more plausible geographical route for dispersal, due to the relative contiguity of the two areas.

CONCLUSION:

I found a different best supported hypotheses of biogeographic history for each of the three species, with an origin in the CFP for *A. patula*, an origin outside of the CFP for *A. pungens*, and an unclear area of origin for *A. pringlei*. Taken together, these results demonstrate that speciation both within and outside of the CFP may have contributed to the current pattern of endemic diversity we see in manzanitas within the CFP. Furthermore, if manzanitas are representative of the kind of biogeographic histories that gave rise to modern CFP flora, these results likewise suggest that the high endemism of the CFP flora may be the result of multiple historical biogeographic processes.

ACKNOWLEDGEMENTS:

I thank all past and present members of Team Manzanita, at UC Riverside, whose contributions to field work, lab work, analysis, and general support made this work possible. I thank Dinusha Maheepala, Yi Huang, and Alex Rajewski for their instruction in lab methods, and their patience. I thank the UC-MEXUS CONACYT program for providing the funding that supported this research. I thank Kathryn Kennedy, Jeremy Palmer, Ben Roe for their help in obtaining collection permits. I thank Angela Buehlman in particular for her help with collections trips in the interior Southwestern US, and Tito Abbo, Andrew C. Sanders, and Elliot Levine for making certain collections in California on my behalf. I thank Arnold Clifford for his invaluable help in my field work in Navajo Nation. I thank Stephanie Ickert-Bond for collecting samples of *Arctous alpina*. I thank undergraduate researchers Stephanie Tyler, Jenni Kao, Vijay Sim, and Lucie Nguyen for their assistance in the grinding samples and DNA extraction. I thank Tito Abbo for his immense help in sequencing library preparation. I thank Alan Brelsford and German Lagunas-Robles for sharing their expertise, and support, in troubleshooting lab work.

FIGURES AND TABLES:

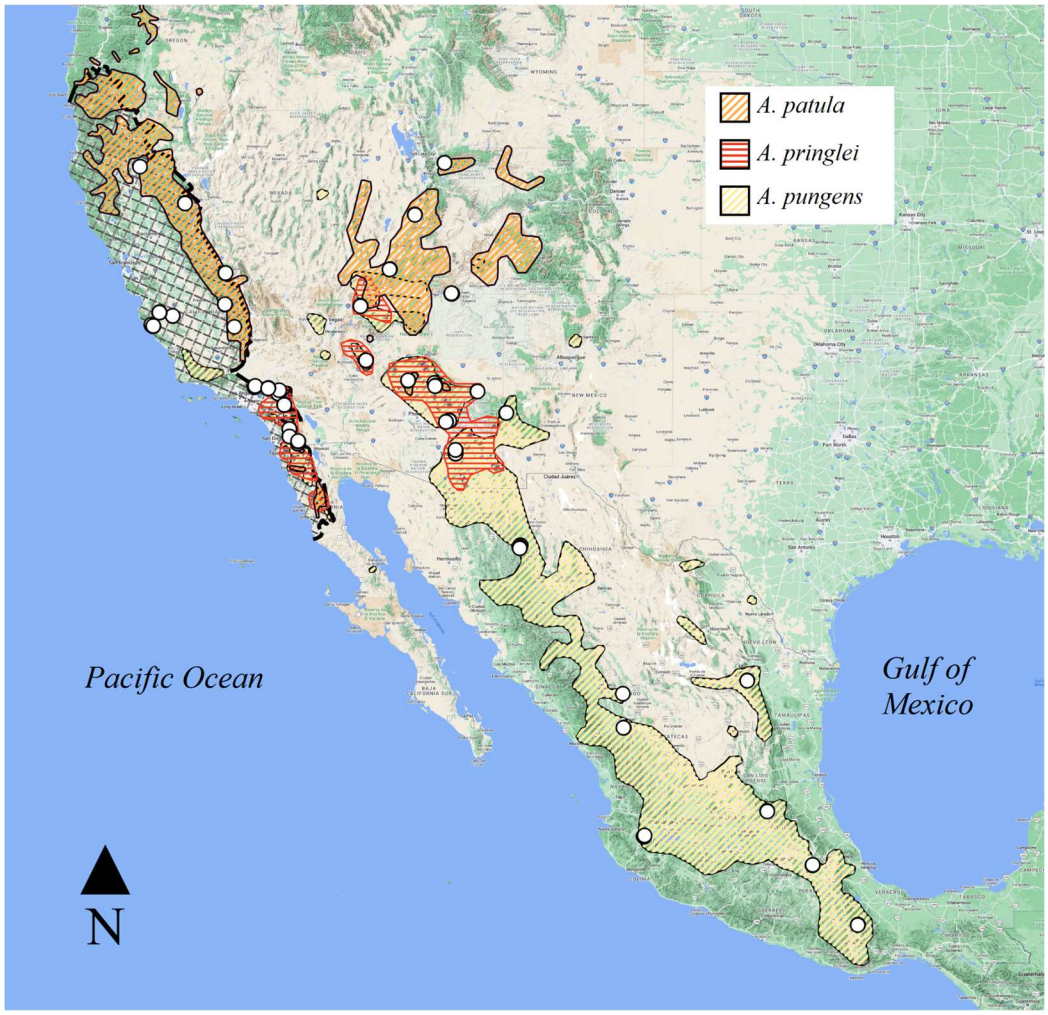


Fig. 1.1. Ranges of the three focal species. Range polygons were hand drawn following the overall shape of the species ranges as determined by examining georeferenced herbarium collection evidence. Map created using QGIS software, using a Google Maps basemap layer.

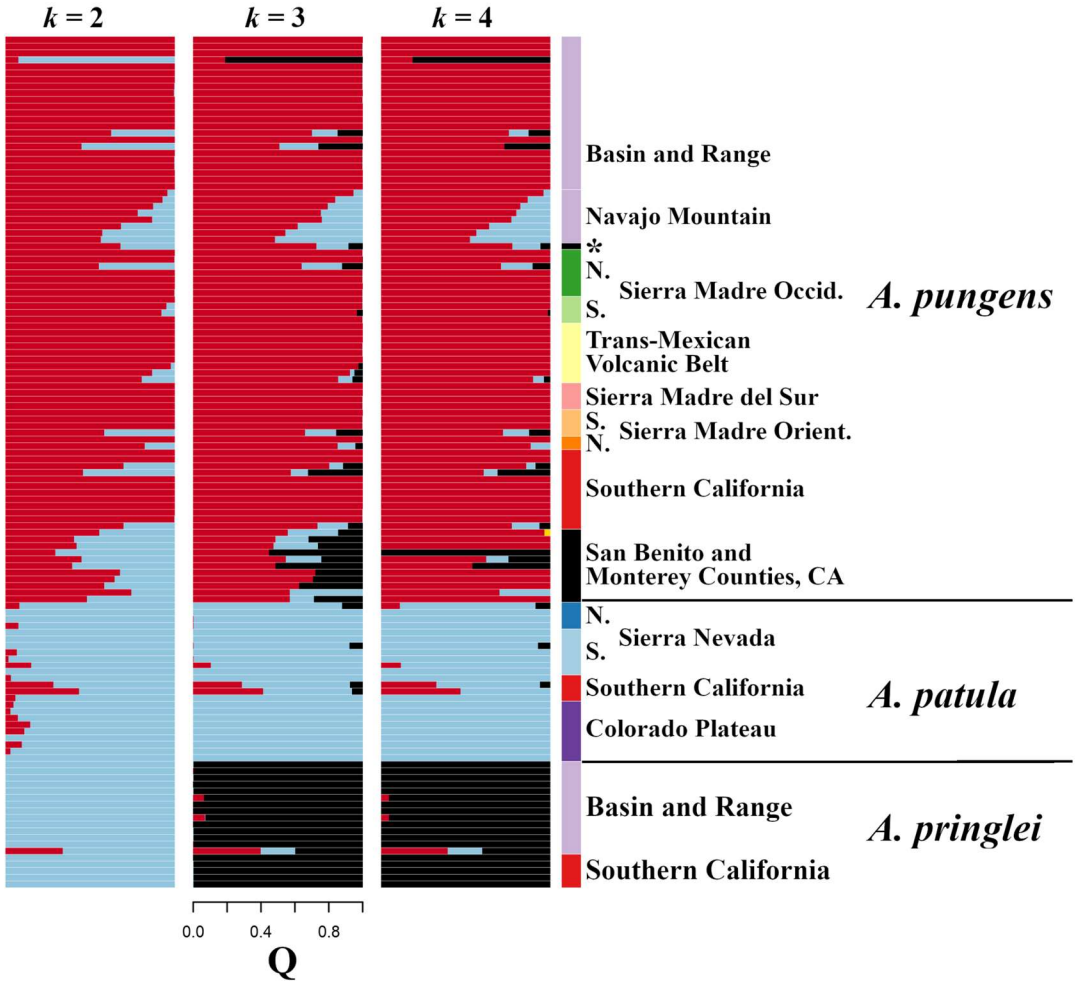


Fig.1.2. Results of fastStructure analysis for $k = 2$ to $k = 4$, with the same order of individual samples (rows in each plot). Colors used in the fastStructure plots are arbitrary, and chosen to show different cluster results, and are not related to the colors used in geographic notations at right, which use the same color-coding as used to show areas in Fig. 1.3, with the exception of the two sets of samples marked with black on the right hand side. These samples/localities were not included in the biogeographic reconstruction, due to putative admixture. The sample marked with the asterisk is a single sample from northern Baja California, Mexico.

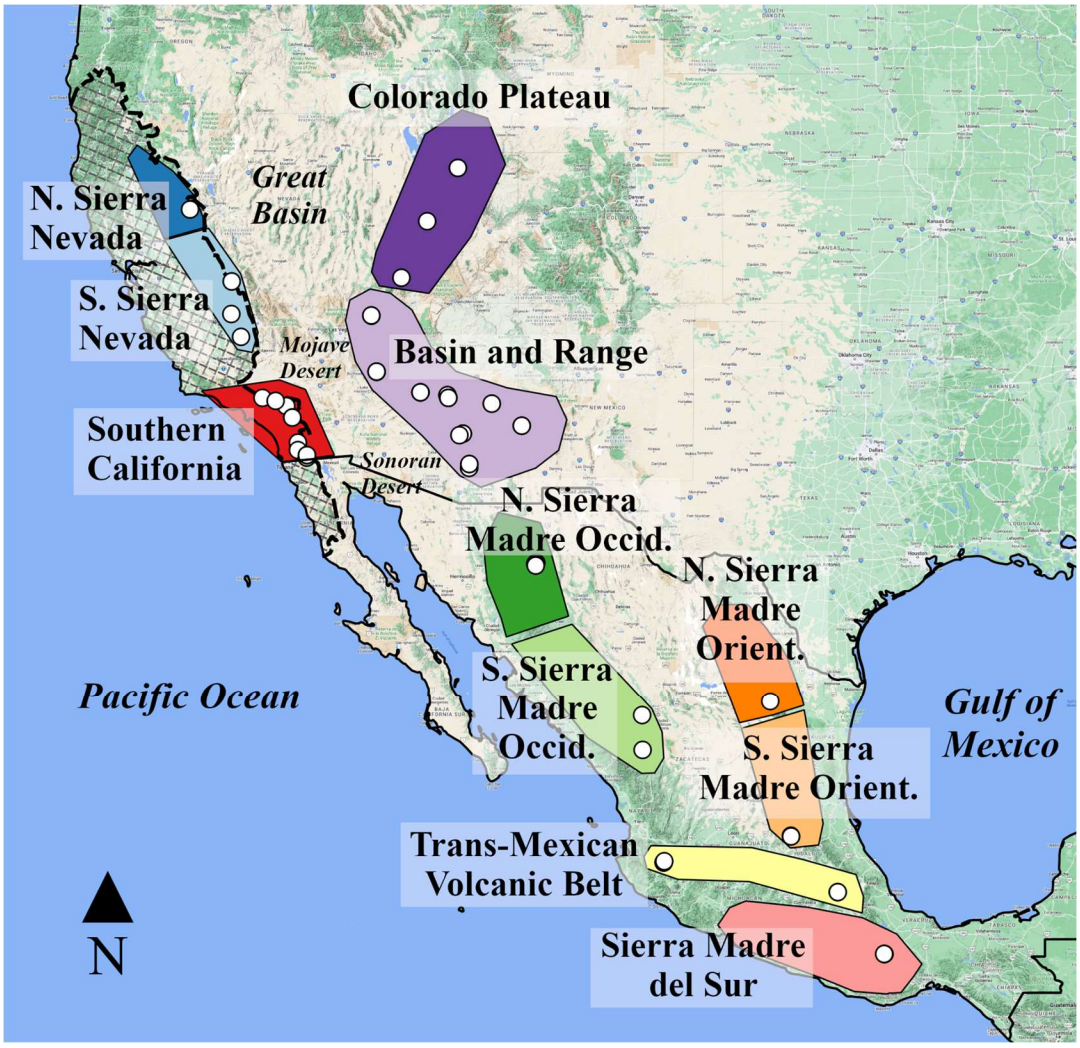


Fig. 1.3. Map of localities represented in the subset of the dataset used for phylogeographic inference (excluding putative admixed individuals). Polygons drawn onto the map show the areas that were used to divide species ranges for use in the BioGeoBEARS analysis performed on each species.

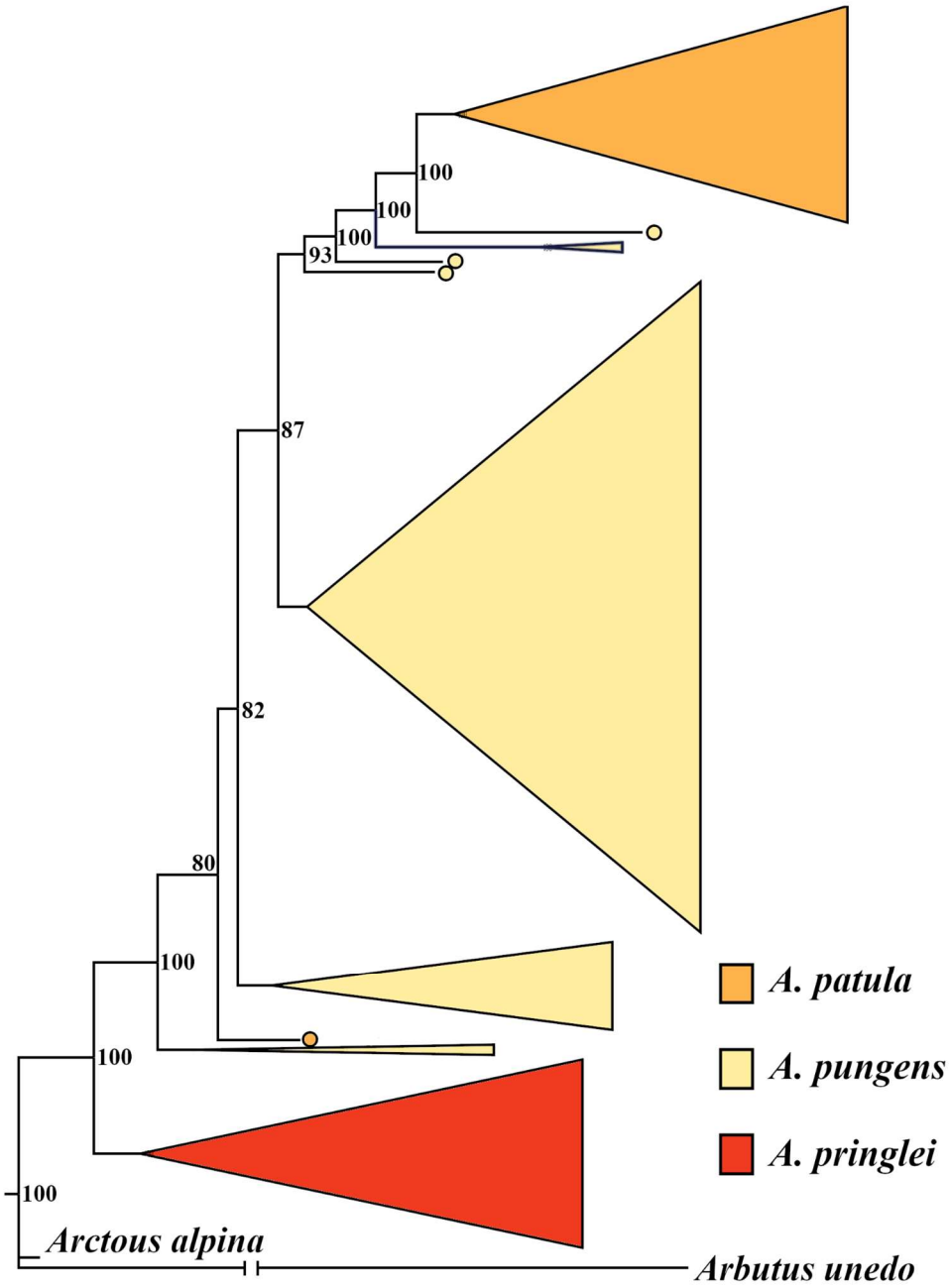


Fig. 1.4. Phylogeny built using all samples. Clades composed of samples of a single species are collapsed into triangles to highlight species relationships and monophyly or lack thereof. Colors match those used on the species range maps in Fig. 1.1.

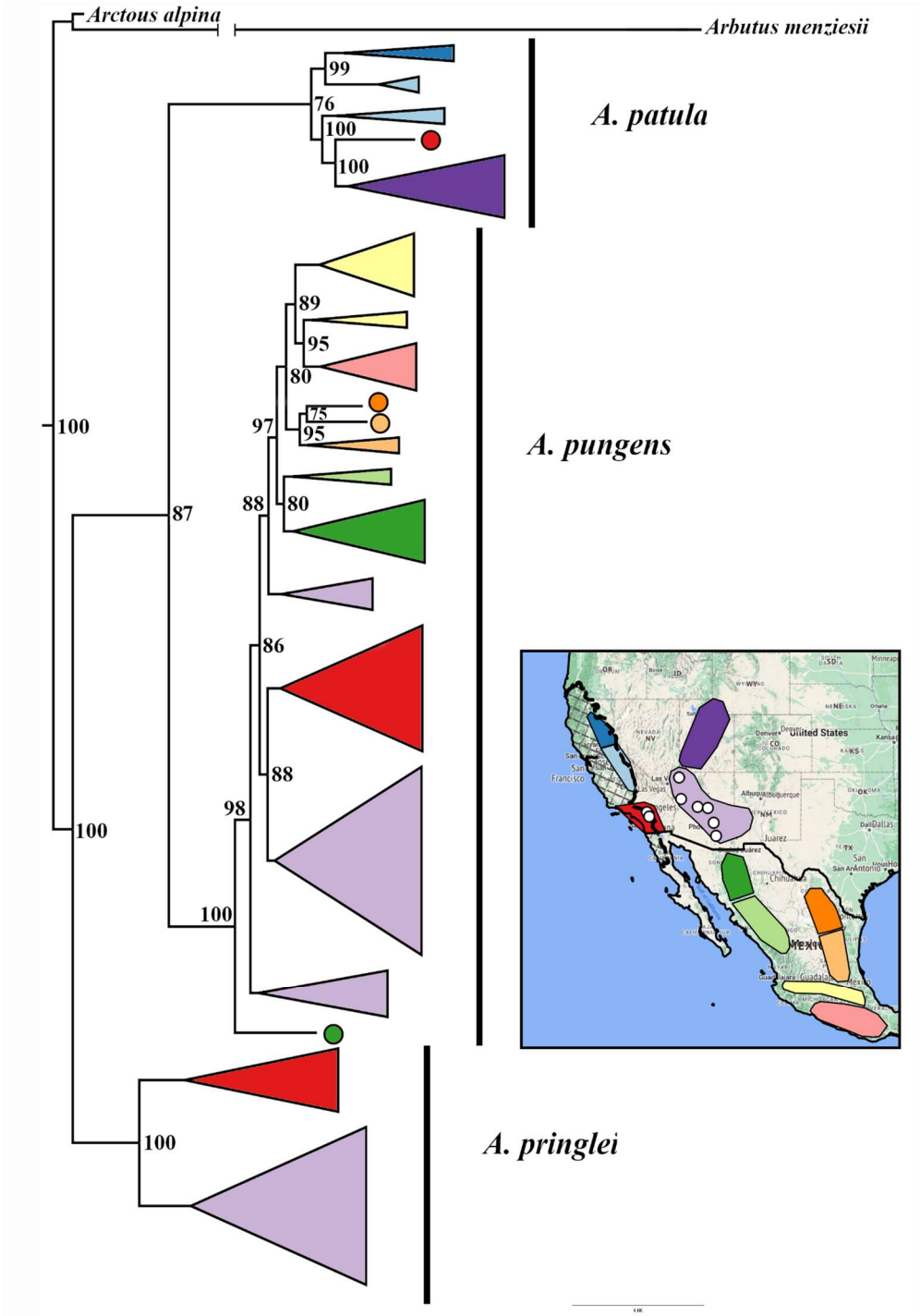

Arctous alpina
Arbutus menziesii
99
76
100
100
A. patula
89
95
80
75
95
97
100
A. pungens
88
80
87
86
88
98
100
100
100
A. pringlei
United States

Fig. 1.5. Relationships estimated among samples in the dataset excluding putatively admixed individuals. Topology is a consensus tree from IQTree, with some clades composed of samples of a single species from the same region collapsed into simplified triangles. Tips are colored by geographic region, using the same color-coding as used in the map in Fig. 1.3. The inset map is the same map as shown in Fig. 1.3, provided here for ease of cross-referencing with colors/areas marked on the tree..

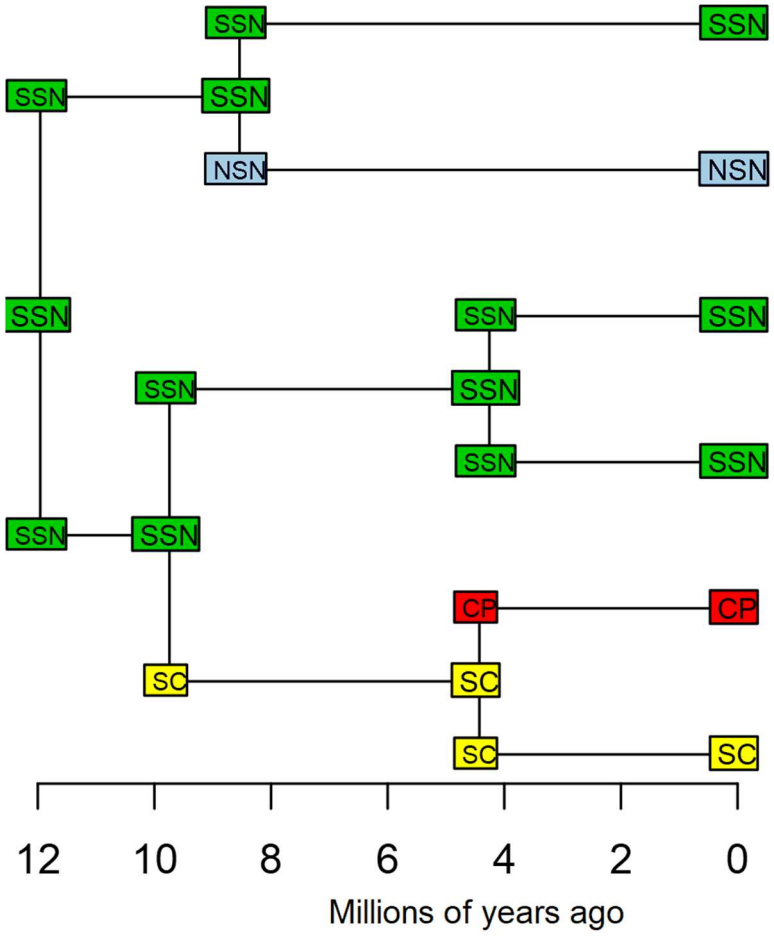


Fig. 1.6. Diagram showing the best supported hypothesis for the historical biogeography of *A. patula*. Text labels at nodes give the most likely range at that point in evolutionary history: "NSN" = northern Sierra Nevada; "SSN" = southern Sierra Nevada; "SC" = Southern California; "CP" = Colorado Plateau. Note that the colors used on this tree are a different set than those used on Fig. 1.3.

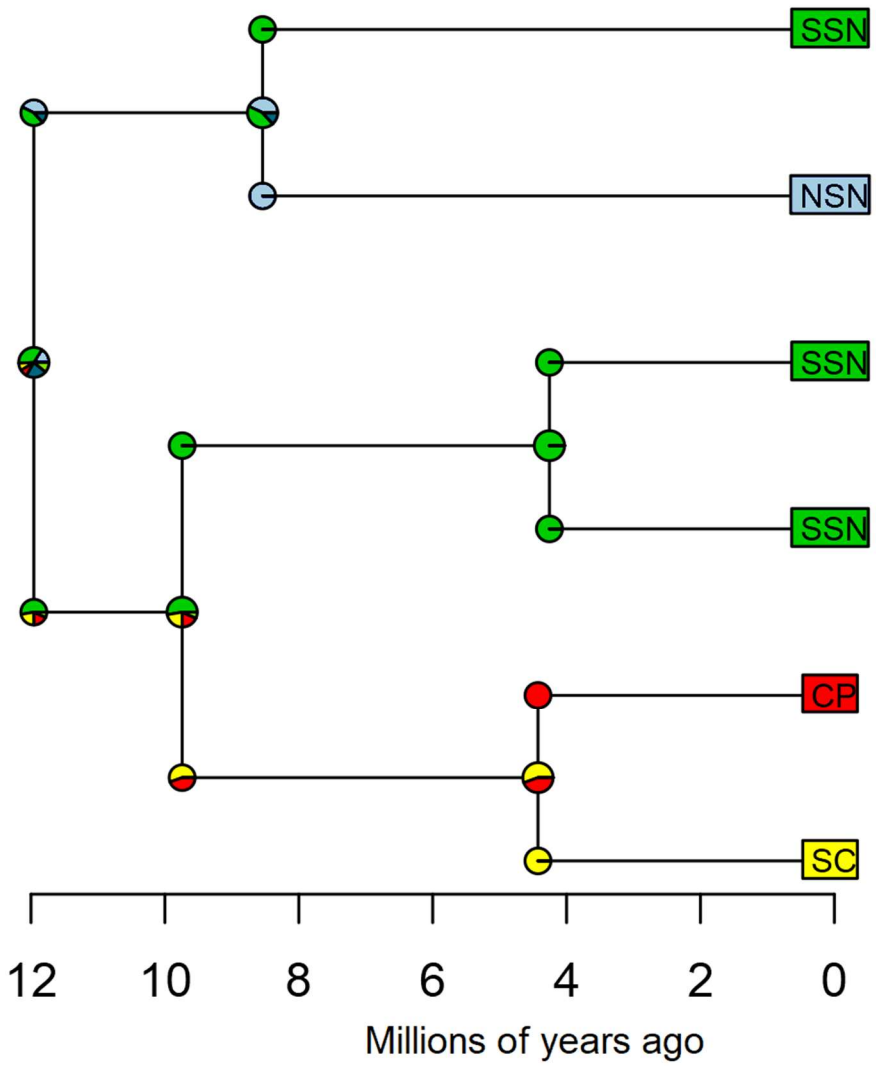


Fig. 1.7. Diagram showing BioGeoBEARS model results for competing hypotheses of historical biogeography for *A. patula*. Pie charts at nodes give the relative probability of occurrence within each area at that point in evolutionary time. "NSN" = northern Sierra Nevada; "SSN" = southern Sierra Nevada; "SC" = Southern California; "CP" = Colorado Plateau. Note that the colors used on this tree are a different set than those used on Figs. 1.3 and 1.5.

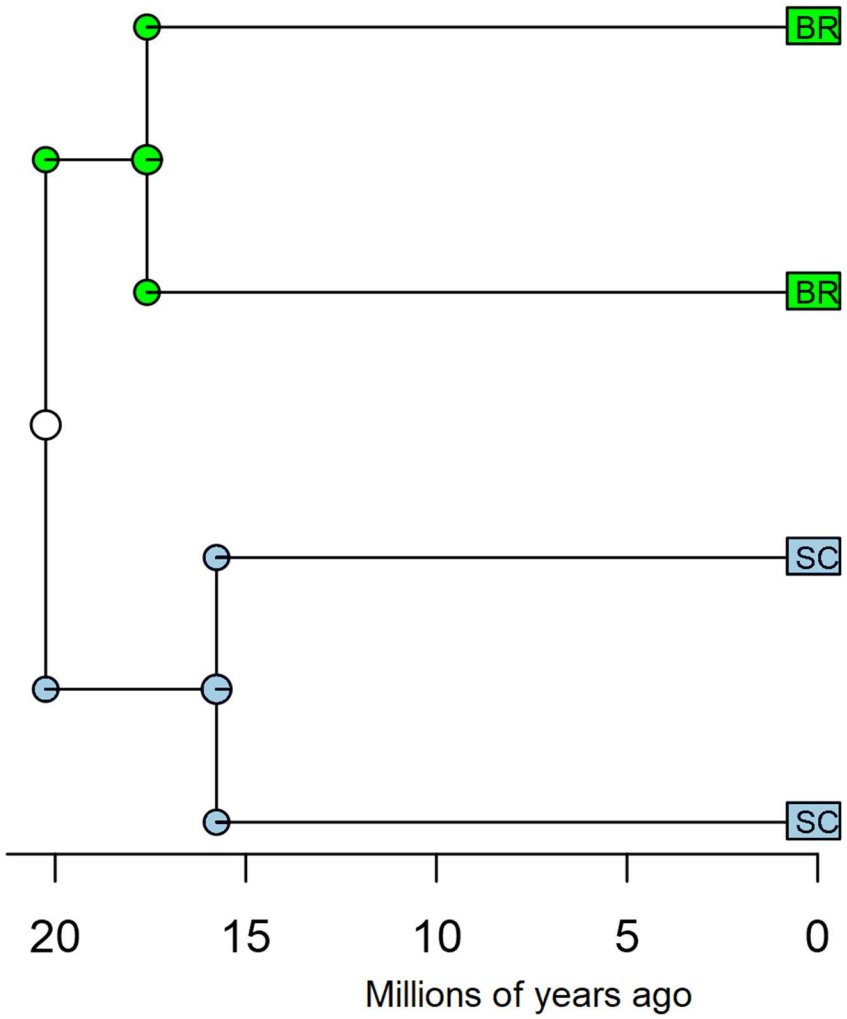


Fig. 1.8. Diagram showing BioGeoBEARS model results for competing hypotheses of historical biogeography for *A. pringlei*. Pie charts at nodes give the relative probability of occurrence within each area at that point in evolutionary time: "BR" = Basin and Range province; and "SC" = Southern California. The white pie chart at the root indicates that the BioGeoBEARS analysis was unable to make inferences on the probability of occurrence in one of the other area based on the topology and geographic information provided to the analysis.

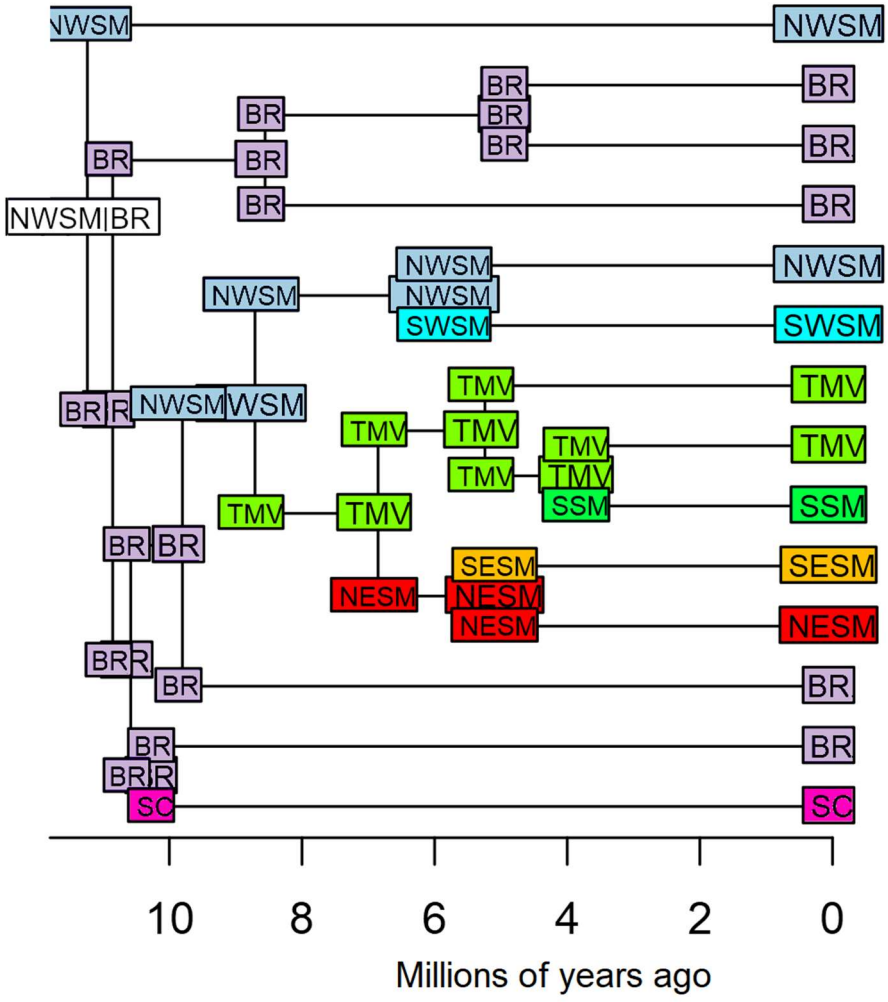


Fig. 1.9. Diagram showing the best supported hypothesis for the historical biogeography of *A. pungens*. Text labels at nodes give the most likely range at that point in evolutionary history: "NSWM" = northern Sierra Madre Occidental; "BR" = Basin and Range province; "SWSM" = southern Sierra Madre Occidental; "TMV" = Trans-Mexican Volcanic Belt; "SSM" = Sierra Madre del Sur; "NESM" = northern Sierra Madre Oriental; "SESM" = southern Sierra Madre Oriental; and "SC" = Southern California.

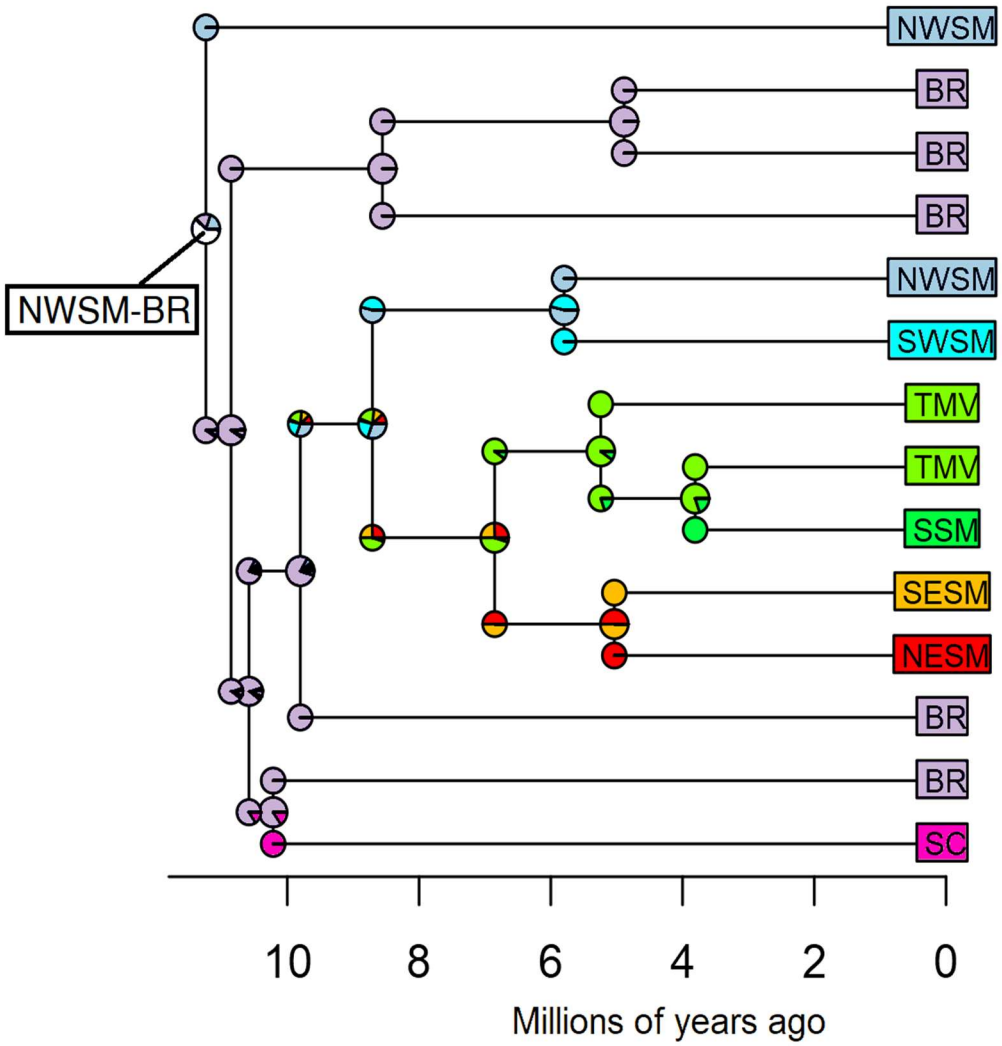


Fig. 1.10. Diagram showing BioGeoBEARS model results for competing hypotheses of historical biogeography for *A. pungens*. Pie charts at nodes give the relative probability of occurrence within each area at that point in evolutionary time: "NSWM" = northern Sierra Madre Occidental; "BR" = Basin and Range province; "SWSM" = southern Sierra Madre Occidental; "TMV" = Trans-Mexican Volcanic Belt; "SSM" = Sierra Madre del Sur; "NESM" = northern Sierra Madre Oriental; "SESM" = southern Sierra Madre Oriental; and "SC" = Southern California.

Table 1.1. List of *Arctostaphylos patula* specimens, sorted by state, and collection number, with geographic coordinates provided for each collection (WGS1984 projection).

State	Collection Number	Latitude (° N)	Longitude (° W)
California	Y. Huang 618-17	34.32697	117.83264
	E. Levine 23	36.07071	118.53626
	E. Levine 25	36.69947	118.86771
	E. Levine 26	36.69947	118.86771
	E. Levine 27	36.69947	118.86771
	E. Levine 48	40.49816	121.87987
	G. Morrison 120	34.36694	117.80024
	G. Morrison 123	34.36704	117.80018
	G. Morrison 172	39.51286	120.28217
	G. Morrison 173	39.51287	120.28217
	G. Morrison 174	39.51283	120.28218
	G. Morrison 181	37.58804	118.85513
	G. Morrison 183	37.58798	118.85519
	G. Morrison 185	37.58795	118.85519
	G. Morrison 189	34.23507	116.95962
Utah	G. Morrison 486	37.68787	113.03867
	G. Morrison 487	37.68787	113.03867
	G. Morrison 488	37.68787	113.03867
	G. Morrison 489	37.68787	113.03867
	G. Morrison 493	39.20102	112.15603
	G. Morrison 494	39.20102	112.15603
	G. Morrison 495	39.20102	112.15603
	G. Morrison 497	40.59977	111.10318
	G. Morrison 498	40.59977	111.10318

Table 1.2. List of *Arctostaphylos pringlei* specimens, sorted by state, and collection number, with geographic coordinates provided for each collection (WGS1984 projection).

State	Collection Number	Latitude (° N)	Longitude (° W)
California	G. Morrison 192	34.13636	116.98484
	G. Morrison 196	34.13605	116.98487
	G. Morrison 405	33.80972	116.77786
	G. Morrison 406	33.80951	116.77791
	G. Morrison 407	33.80967	116.77800
Nevada	G. Morrison 476	36.66020	114.07015
	G. Morrison 477	36.66020	114.07015
	G. Morrison 478	36.65251	114.06626
Arizona	G. Morrison 421	33.35106	110.93237
	G. Morrison 425	33.34979	110.93204
	G. Morrison 431	32.36852	110.71777
	G. Morrison 432	32.36845	110.71785
	G. Morrison 433	32.36843	110.71793
	G. Morrison 452	34.44261	111.46064
	G. Morrison 453	34.44292	111.46149
	G. Morrison 454	34.44297	111.46146
	G. Morrison 457	34.52234	112.38673
	G. Morrison 460	34.52202	112.38744
	G. Morrison 469	35.09879	113.88795

Table 1.3. List of *Arctostaphylos pungens* specimens, sorted by country, state, and collection number, with geographic coordinates provided for each collection (WGS1984 projection).

Country	State or Reservation	Collection Number	Latitude (° N)	Longitude (° W)
MX	Baja California	L. Flores-Rentería SJAp2	31.98450	115.96884
	Chihuahua	S. González-Elizondo 2021-16-a	29.58563	108.41655
		S. González-Elizondo 2021-16-b	29.58563	108.41655
		S. González-Elizondo 2021-16-d	29.58563	108.41655
		S. González-Elizondo 2021-32-b	29.50278	108.43583
		S. González-Elizondo 2021-32-c	29.50278	108.43583
		S. González-Elizondo J66 2583	29.55183	108.45219
		S. González-Elizondo J66 2584	29.55183	108.45219
	Durango	S. González-Elizondo 2021-42-c	23.84817	104.77218
		S. González-Elizondo 2021-42-d	23.84817	104.77218
		S. González-Elizondo 2021-42-e	23.84817	104.77218
		S. González-Elizondo 2021-52-c	24.93132	104.78869
	Jalisco	L. Trejo-Hernández 2003	20.29289	104.05092
		L. Trejo-Hernández 2004	20.29342	104.05092
		L. Trejo-Hernández 2008	20.29864	104.04886
		L. Trejo-Hernández 2009	20.30119	104.04792
		L. Trejo-Hernández 2011	20.33308	104.02981
	Nuevo León	S. González-Elizondo 2021-11-c	25.36304	100.39710
		S. González-Elizondo 2021-11-e	25.36304	100.39710
	Oaxaca	L. Trejo-Hernández 1977	17.34044	96.48231
		L. Trejo-Hernández 1978	17.33981	96.48611
		L. Trejo-Hernández 1980	17.34186	96.49064
		L. Trejo-Hernández 1985	17.34208	96.49392
	Querétaro	L. Trejo-Hernández 1970	21.12386	99.68181
		L. Trejo-Hernández 1971	21.12453	99.68431
		L. Trejo-Hernández 1972b	21.12772	99.68636
		L. Trejo-Hernández 1976	21.14539	99.69153
	Tlaxcala	L. Trejo-Hernández 1965	19.34378	98.08672
		L. Trejo-Hernández 1966	19.34306	98.08664
		L. Trejo-Hernández 1967	19.34222	98.08358
		L. Trejo-Hernández 2016	19.34222	98.08358
US	Arizona	G. Morrison 417	33.30659	111.05074
		G. Morrison 418	33.30663	111.05070
		G. Morrison 426	32.47797	110.71104
		G. Morrison 430	32.47797	110.71083

	G. Morrison 442	34.20310	109.94303
	G. Morrison 444	34.20296	109.94232
	G. Morrison 446	34.36709	111.43060
	G. Morrison 447	34.36707	111.43057
	G. Morrison 448	34.36713	111.43059
	G. Morrison 449	34.36706	111.43069
	G. Morrison 450	34.36704	111.43076
	G. Morrison 461	34.52329	112.38657
	G. Morrison 462	34.52322	112.38652
	G. Morrison 464	34.52304	112.38649
	G. Morrison 465	34.52290	112.38656
	G. Morrison 471	35.09879	113.88606
	G. Morrison 472	35.10343	113.89011
	G. Morrison 473	35.10402	113.89010
	G. Morrison 474	35.10408	113.89000
California	Tito Abbo 17	33.09427	116.58572
	Tito Abbo 18	33.09434	116.58583
	Tito Abbo 19	33.09435	116.58586
	Tito Abbo 21	33.09441	116.58601
	Tito Abbo 22	32.66569	116.28201
	Tito Abbo 25	32.66554	116.28197
	Y. Huang 218-07	32.88002	116.58475
	G. Morrison 250	36.47337	121.18559
	G. Morrison 251	36.47345	121.18570
	G. Morrison 252	36.47363	121.18604
	G. Morrison 254	36.47692	121.18716
	G. Morrison 268	36.11946	121.46420
	G. Morrison 270	36.11945	121.46439
	G. Morrison 271	36.08688	121.40485
	G. Morrison 272	36.08697	121.40487
	G. Morrison 273	36.08643	121.40464
	G. Morrison 339	34.30067	117.33708
	G. Morrison 341	34.28246	117.34960
	G. Morrison 363	36.38036	120.71306
	G. Morrison 364	36.38000	120.71139
	Andrew Sanders 43396	32.73626	116.27582
	Andrew Sanders 43399	32.73626	116.27582
	Andrew Sanders 43401	32.73626	116.27582
Navajo Nation	G. Morrison 500	37.01205	110.84028
	G. Morrison 501	37.01205	110.84028
	G. Morrison 503	37.01205	110.84028

	G. Morrison 504	37.01205	110.84028
	G. Morrison 506	37.01314	110.86646
	G. Morrison 507	37.01314	110.86646
	G. Morrison 508	37.01314	110.86646
	G. Morrison 509	37.01314	110.86646
Nevada	G. Morrison 481	36.66020	114.07015
New Mexico	G. Morrison 437	33.57627	108.91326
	G. Morrison 439	33.57692	108.91458
	G. Morrison 440	33.57703	108.91450

LITERATURE CITED:

Baldwin, B. G. 2014. "Origins of Plant Diversity in the California Floristic Province." *Annual Reviews* 45: 347–69.

Baldwin, Bruce G., Goldman H. Douglas, David J. Keil, Robert Patterson, Thomas J. Rosatti, and Dieter H. Wilken, eds. 2012. *The Jepson Manual: Vascular Plants of California, 2nd Edition*. University of California Press.

Boykin, Laura M., Michael C. Vasey, V. Thomas Parker, and Robert Patterson. 2005. "Two Lineages of *Arctostaphylos* (Ericaceae) Identified Using the Internal Transcribed Spacer (ITS) Region of the Nuclear Genome." *Madrono* 52 (3): 139–47.

Brown, David E. 1978. "The Vegetation and Occurrence of Chaparral and Woodland Flora on Isolated Mountains within the Sonoran and Mojave Deserts in Arizona." *Journal of the Arizona-Nevada Academy of Science* 13 (1): 7–12.

Burge, Dylan O., James H. Thorne, Susan P. Harrison, Bart C. O'Brien, Jon P. Rebman, James R. Shevock, Edward R. Alverson, et al. 2016. "Plant Diversity and Endemism in the California Floristic Province." *Madroño* 63 (2): 3–206.

Dobzhansky, Theodosius. 1953. "Natural Hybrids of Two Species of *Arctostaphylos* in the Yosemite Region of California." *Heredity* 7 (1): 73–80.

Eaton, Deren A. R., and Isaac Overcast. 2020. "ipyrad: Interactive assembly and analysis of RADseq datasets." *Bioinformatics* 36 (8): 2592–94.

Ellstrand, Norman C., Janet M. Lee, Jon E. Keeley, and Sterling C. Keeley. 1987. "Ecological Isolation and Introgression: Biochemical Confirmation of Introgression in an *Arctostaphylos* (Ericaceae) Population." *Acta Oecologica, Oecologica Plantarum.* 8 (4): 299–308.

Gankin, R. 1967. "A New Species of *Arctostaphylos* from Santa Barbara County, California." *The Four Seasons*.

Gottlieb, Leslie D. 1968. "Hybridization Between *Arctostaphylos viscida* and *A. canescens* in Oregon." *Brittonia* 20 (1): 83.

Hastings, James Rodney. 1965. *The Changing Mile: An Ecological Study of Vegetation Change with Time in the Lower Mile of an Arid and Semiarid Region*. University of Arizona Press.

Hileman, Lena C., Michael C. Vasey, and V. Thomas Parker. 2001. "Phylogeny and Biogeography of the Arbutoideae (Ericaceae): Implications for the Madrean-Tethyan

Hypothesis." *Systematic Botany* 26 (1): 131–43.

Huang, Yi, Merly Escalona, Glen Morrison, Mohan P. A. Marimuthu, Oanh Nguyen, Erin Toffelmier, H. Bradley Shaffer, and Amy Litt. 2022. "Reference Genome Assembly of the Big Berry Manzanita (*Arctostaphylos glauca*)." *The Journal of Heredity* 113 (2): 188–96.

Kauffmann, Michael Edward, Tom Parker, and Michael Vasey. 2021. *Field Guide to Manzanitas: California, North America, and Mexico, Second Edition*. Backcountry Press.

Keeley, Jon E. 1976. "Morphological Evidence of Hybridization between *Arctostaphylos glauca* and *Arctostaphylos pungens* (Ericaceae)" *Madroño* 23 (8): 427–34.

Keeley, Jon E., V. Thomas Parker, and Michael C. Vasey. 2017. "Characters in *Arctostaphylos* Taxonomy." *Madroño* 64 (4): 138–53.

Lancaster, Lesley T., and Kathleen M. Kay. 2013. "Origin and Diversification of the California Flora: Re-Examining Classic Hypotheses with Molecular Phylogenies." *Evolution; International Journal of Organic Evolution* 67 (4): 1041–54.

Matzke, N. J. 2014. "Model Selection in Historical Biogeography Reveals That Founder-Event Speciation Is a Crucial Process in Island Clades." *Systematic Biology* 63: 951–70.

Meirmans, Patrick G. 2018. "Subsampling Reveals That Unbalanced Sampling Affects STRUCTURE Results in a Multi-Species Dataset." *Heredity* 122 (3): 276–87.

Millar, Constance I., and Wallace B. Woolfenden. 2016. "Ecosystems Past: Vegetation Prehistory." In *Ecosystems of California*, edited by Harold Mooney, Erika Zavaleta, and Melissa Chapin. University of California Press.

Nason, John D., Norman C. Ellstrand, and Michael L. Arnold. 1992. "Patterns of Hybridization and Introgression in Populations of Oaks, Manzanitas, and Irises." *American Journal of Botany* 79 (1): 101–11.

Parker, V. Thomas, Christina Y. Rodriguez, Gail Wechsler, and Michael C. Vasey. 2020. "Allopatry, Hybridization, and Reproductive Isolation in *Arctostaphylos*." *American Journal of Botany* 107 (12): 1798–1814.

Parker, V. Thomas, Michael C. Vasey, and Jon E. Keeley. 2020. "*Arctostaphylos* in: Flora of North America Vol. 8." Flora of North America [Online]. 2020. http://floranorthamerica.org/Arctostaphylos.

Raj, Anil, Matthew Stephens, and Jonathan K. Pritchard. 2014. "fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets." *Genetics* 197 (2): 573–89.

Raven, P. H., and D. I. Axelrod. 1978. *Origin and Relationships of the California Flora*. Berkeley.: University of California Press.

R Core Team. 2023. “R: A Language and Environment for Statistical Computing.” Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Schierenbeck, Kristina A., G. Ledyard Stebbins, and Robert W. Patterson. 1992. “Morphological and Cytological Evidence for Polyphyletic Allopolyploidy in *Arctostaphylos mewukka*.” *Syst. Evol.* 179: 187–205.

Schmid, Rudolf, Thomas E. Mallory, and John M. Tucker. 1968. “Biosystematic Evidence for Hybridization Between *Arctostaphylos nissenana* and *A. viscida*.” *Brittonia* 20 (1): 34.

Serkanic, Steven, V. Thomas Parker, and Kristina Schierenbeck. 2021. “Complexity in Polyploid Species Origin and Establishment: *Arctostaphylos mewukka* (Ericaceae).” *Systematic Botany* 46 (3): 666–77.

Thornhill, Andrew H., Bruce G. Baldwin, William A. Freyman, Sonia Nosratinia, Matthew M. Kling, Naia Morueta-Holme, Thomas P. Madsen, David D. Ackerly, and Brent D. Mishler. 2017. “Spatial Phylogenetics of the Native California Flora.” *BMC Biology* 15 (1): 96.

Wahlert, Gregory A., V. Thomas Parker, and Michael C. Vasey. 2009. “A Phylogeny of *Arctostaphylos* (Ericaceae) Iinferred from Nuclear Risomal ITS Sequences.” *Journal of the Botanical Research Institute of Texas* 3 (2): 673–82.

Wolfe, J. A. 1964. “Miocene Floras from Fingerrock Wash Southwestern Nevada.” 454-N. US Geological Survey.

Appendix 1.1. Protocol for preparation of Double-Digest Restriction Site-Associated DNA Sequencing (ddRADseq) libraries.

1. Prepare a 96 well plate, loading 200 ng, or as close as possible, of gDNA into each well. Total volume should be 17.8 uL, so load extra water for higher concentration samples (> ~11.2 ng/ uL) to reach a total volume 17.8 uL. Note: the order of DNA samples going into the plate should be randomized.
2. To each active well in the 96 well plate, add 2.2 uL of restriction master mix, mixing by pipetting 10-30 times when loading each well. Seal the plate and incubate at 37°C for 6 hours.

Restriction Master Mix:

Reagents	Volume (µL) per sample/well
Cutsmart buffer (10x)	2
PstI-HF	0.1 (= 2 units)
MspI	0.1 (= 2 units)

3. On a new plate, load 9 uL of each restriction digest product from the previous step, and then add 1.6 uL of ligation master mix, and 1 uL of barcoded PstI adaptor to each well. Seal the plate with a cover seal and centrifuge the plate, then incubate in a thermocycler according to the following program:16°C for 3 hours (cover temp: 70°C, reaction volume: 12 µL); 12°C indefinitely.

Ligation Master Mix:

Reagents	Volume (µL) per sample/well
10x Cutsmart	0.26
100 mM ATP	0.12
MspI adaptor (10pM/uL)	1
Water	0.05
T4 DNA Ligase	0.17

4. Pool the now barcoded samples into microcentrifuge tubes, and label accordingly.
5. Prepare a 1% agarose gel with TAE buffer, using a well comb that forms wells of a capacity of at least 40 uL. Consider increasing overall gel volume to form a deeper well to increase capacity of each well, as needed.
6. Load 40 uL of ligation pool into an individual well. Each well will go through a separate gel excision and purification, so the number of wells you will need to fill will depend on concentrations of ligation products going into this reaction, and the productivity anticipated for the following (PCR) step. If performing this protocol for the first time, with no existing expectations, it is suggested to load at least 4 wells for processing. Leave every third well empty; these will be loaded with DNA ladder.
7. Load a DNA ladder in the empty wells between sample wells. Make sure the ladder gives resolution at 200 bp and 500 bp, the bounds of size selection for this protocol.
8. Run the gel at 80 V for 80 minutes. Place the gel on a UV light source, while wearing UV-protective PPE. Before turning on the UV light source, have a sterilized safety razor blade handy. Darken the room, and turn on the UV, and working quickly to avoid UV degradation of DNA, cut the 200-500 bp region from each sample well, and place in a microcentrifuge tube.
9. Using a commercially available gel DNA extraction kit, isolate the DNA from the gel excisions.
10. Clean up the resulting DNA extractions from the gel excisions using magnetic beads (Ampure or similar), using a 1:1 ratio of bead solution to sample.

11. Perform a PCR using 10.06 µL of sample (purified, size-selected ligations, from previous step) to 9.94 µL of PCR master mix, doing as many reactions as anticipated will be needed to produce target library DNA mass. Include one water control each time you set up this PCR step, also of 20 µL reaction volume.

PCR Master Mix:

Reagent	Volume (µL) per reaction
Q5 Buffer	4
dNTP (10mM)	0.4
PCR Primer Mix [IllPCR1+IllPCR2-6]	1.34
Q5 High GC Enhancer	4
Q5 Hot Start DNA Polymerase	0.2

12. Perform a final bead cleanup with light size-selecting activity (ratio = ~0.8) on the pools of PCR products, eluting in a lower volume of water if increased concentration is necessary.

Chapter 2: Morphological Analysis of Manzanita Species and Subspecies.

ABSTRACT:

Introduction: There are ~60 species and 109 taxa (including subspecies) in the genus *Arctostaphylos* (Ericaceae). For over a century there has been debate on how many species should be recognized, and how to best circumscribe taxa. Overall, most manzanitas are morphologically similar, sharing many traits, and often being differentiated by traits that are a matter of degree, rather than kind, or that can be difficult for non-experts to assess. Identification can also be challenging as fruits and inflorescences are often both needed for identification, despite not predictably being available on the plants at the same time. In addition, frequent hybridization has led to populations of individuals that do not neatly conform to the descriptions of any one species or subspecies. As a consequence, it is not clear that current taxonomy best represents patterns of morphological diversity. In this study I evaluate the morphological distinguishability and distinctiveness of manzanita species and subspecies.

Methods: I measured 32 vegetative morphological traits from a sample set of 326 herbarium specimens, representing 33 species (22 narrow endemic species, 3 widespread monotypic species, and six widespread polytypic species) and 43 taxa (including subspecies). I used only vegetative traits, since they are always available and not dependent on season. I used classification-based and ordination-based methods to evaluate (1) whether taxa are able to be distinguished on the basis of vegetative morphology, and (2) whether the current taxonomy corresponds to the natural structure of variability.

Key results: Classification-based methods (linear discriminant analysis) found that species can be distinguished with a high degree of accuracy using the vegetative traits in the dataset. Most

subspecies can also be distinguished, although at a lower rate of accuracy. However, ordination-based methods (NMDS) showed that the current taxonomy may not correspond well to the overall pattern of variation.

Conclusion: Most manzanita taxa are distinguishable on the basis of vegetative morphology. However, that distinguishability relies on individual delimiting traits, and many taxa overlap morphologically. This suggests that there may be other ways to circumscribe taxa that capture the distribution and pattern of morphological variation more accurately.

INTRODUCTION:

Manzanitas (genus *Arctostaphylos* Adans., Ericaceae) are a taxonomically diverse group of woody plants, almost all of which are found only in the California Floristic Province (CFP), the region of North America with a Mediterranean-type climate (Kauffmann et al., 2021). With 60 recognized species, the genus *Arctostaphylos* is the most species-diverse genus of woody plants in the CFP, by approximately a factor of two compared with the second most diverse (Baldwin et al., 2012). Currently, 19 species are divided into subspecies as well. This produces a total of at least 107 currently recognized taxa (Kauffmann et al., 2021). The apparent diversity of this genus has long attracted the attention of North American botanists, and for around a century many botanists have taken up studies of manzanitas, describing or revising taxa. For the last approximately twenty years the taxonomy has been relatively stable, with only a handful of new taxa being described or existinging taxa being merged (Keeley and Massihi, 1994; Keeley et al., 1997; Parker and Vasey, 2004; Parker et al., 2007). However, in spite of this stability, suggesting agreement among experts, the practical work of identifying manzanita taxa is very difficult for a number of reasons.

Manzanita taxa are described, and identified, based on a distinctive set of morphological characters. For example, botanists unfamiliar with making identifications in the genus often collect material in flower, which is typically the most valuable material for identification in angiosperm genera, only to find out when keying specimens that flowers are hardly ever used as a basis for circumscribing or identifying manzanita taxa because they vary little among species. Characters like the presence or absence of a woody swelling where the stem(s) and roots meet (basal burl), hair length, density and glandularity, as well as characters of the immature ("nascent") inflorescences and fruits, all figure heavily into the descriptions of taxa and their

identification. Often, one has to assess aspects of the nascent inflorescences as well as of the fruits to successfully key out a specimen, despite those two phenological stages rarely overlapping on an individual plant.

In addition to the difficulty of evaluating both young inflorescences and mature fruits at the same time, a person identifying a manzanita often needs to differentiate between morphological conditions that appear to be subjective. For example, descriptions of hair densities and lengths (e.g. "hairs … dense and long") (Kauffmann et al., 2021), can be perplexing for someone who hasn't previously seen a large panel of manzanita diversity. How long is "long", or how sparse is "sparse", outside of the context of the full range of variability? Learning to recognize these length and density categories is challenging without significant experience with a wide range of manzanita species and subspecies. Even reproductive traits can be challenging: a taxon may be described as having "scale-like" inflorescence bracts, which can be hard for the uninitiated to interpret without having the alternative condition ("leaf-like") available for comparison.

Beyond the idiosyncrasies of the morphological characters used to describe and identify manzanita taxa, most manzanita taxa are very similar overall. All but a few are shrubs or tree-like shrubs, usually at least a meter tall and no more than several meters tall, with smooth red bark, simple and entire, generally elliptical, leaves, always in an alternate arrangement, and similar fruits with seeds that are either not fused or partly fused. Taxa that depart from this set of features are usually seen as most distinct, or easiest to identify. For example, several taxa are markedly prostrate, or have different fruit morphology (e.g. with complete seed fusion). Such taxa are the minority though, and for most manzanita taxa, morphological differences usually appear to be a

matter of degree rather than kind. This combination of overall similarity, and the unusual delimiting characters, along with the multiple, often non-overlapping phenological stages that may need to be observed, makes manzanita identification famously difficult and confusing.

The confusion that many face in trying to identify manzanitas has led to some to question the high level of species and subspecies diversity of the genus. As far back as 1969, predating many later descriptions of new taxa, Peter Raven expressed skepticism at the extent of described taxonomic diversity. Raven believed that authors had been describing taxa based on traits of possibly little importance, and not viewing diagnostic traits in the context of character variation across the genus, stating that "the overall pattern of variation tends to become more and more obscure as the new taxa are proliferated" (Raven, 1969).

Even if the traits used to identify manzanita species were less problematic, ease of identification is not equivalent to support for the underlying taxonomic treatment. For instance, in many other plant species, subspecies are easy to distinguish, but are still considered subspecies, therefore morphological distinctiveness is not always seen as sufficient for species recognition. *A. pringlei* is an example of such a species in manzanitas. There are two recognized subspecies of *A. pringlei*, which can be diagnosed morphologically based on whether the seeds are fused or not, and which occur over relatively large, separated ranges (overlapping only in one region of Baja California). Nonetheless, these remain classified as subspecies. Conversely, there are numerous species on the Central Coast of California with geographically close and narrow ranges that are very similar morphologically, and share traits that make them distinctive as a set from other manzanita species such as "earlobe-shaped" ("auriculate") leaf bases, but remain classified as a separate species.

More than half (60%) of manzanita taxa are “narrow endemics”, species and subspecies that occur over a small geographic range and are often represented by only one or two populations (Kauffmann et al., 2021). At many localities, this reduces the number of potential identifications down to just one or a few. This geographic pattern has led to a reliance on geography in identification of many manzanita taxa. The introduction to the *Field Guide to Manzanitas* (Kauffmann et al., 2021) states that “knowing where you are can be the best step toward proper identification” The prominent role of geographic location in identification can reinforce the understanding that these taxa are geographically discrete, by adding identified collections or observational records to the range map without going through the keying process to make a determination based on morphology. Sometimes the application of keys in nature can serve as a test of the supposed morphological distinctions among taxa, and a reliance on geographic information in identification, rather than inspection of delimiting characters, can circumvent this opportunity.

A further complication in the taxonomy of *Arctostaphylos* is interspecific gene flow. Hybridization and introgression have been detected in multiple sets of species (Dobzhansky, 1953; Gankin, 1967; Gottlieb, 1968; Schmid et al., 1968; Keeley, 1976; Ellstrand et al., 1987; Nason et al., 1992; Schierenbeck et al., 1992; Parker et al., 2020; Kauffmann et al., 2021; Serkanic et al., 2021), and are suspected to be frequent overall. The presence of hybrid and introgressed individuals contributes to the difficulty of describing species with unique traits, as trait states that evolved in one species become shared and increase the morphological overlap among species, Several manzanita taxa are currently hypothesized to be the result of historical or ongoing interspecific gene flow, due to having a combination of traits associated with other

species, for example *A. refugioensis* (Gankin, 1967) and *A. gabilanensis* (Parker and Vasey, 2004).

Perhaps most emblematic of all the complications discussed above are the tetraploid manzanita species. While the majority of species are diploid, 11 are tetraploid (two of these are reported as having both diploid and tetraploid individuals) (Parker et al., 2012), and these tetraploid species are in general more morphologically variable than diploid species. This is evident in the fact that nearly all have been subdivided into subspecies (Parker et al., 2012), with as many as ten currently recognized (in *A. glandulosa*). The differences among the subspecies of these tetraploids are generally less clear than the differences among species, and it is common to find populations of these species with individuals that key out to multiple different subspecies, as well as those that do not key well to any subspecies (Huang et al., 2020).

While phylogenetic analyses have been carried out in the genus (Hileman et al., 2001; Boykin et al., 2005; Wahlert et al., 2009), those studies have not been conducted in a manner that allowed for the boundaries among taxa to be evaluated. Two studies that have used molecular data to assess boundaries among subspecies of *A. glandulosa*, a widespread highly variable species (Burge et al., 2018; Huang et al., 2020), found mixed support for taxonomic boundaries as currently described. Consequently, as is the case for most groups, the taxonomy of *Arctostaphylos* is still based on morphology that has not been tested.

While the descriptions of new taxa over the past few decades have included morphological analyses distinguishing the new taxon from a few closely similar taxa (Keeley and Massihi, 1994; Keeley et al., 1997), to date there has not been any analysis testing whether taxa

across the genus are morphologically discrete. In this chapter, I undertake a systematic study aimed at answering the following questions.

1. Are currently recognized taxa distinguishable on a morphological basis alone, without information on geographic origin?
2. Does the structure of morphological variation correspond to the current taxonomy?

These questions are similar, but are not equivalent. It is possible for the answer to question 1 to be yes, while the answer to question 2 is no. A taxonomic treatment does not have to reflect the actual structure of morphological differentiation to allow for taxa to be consistently identified correctly; that is, there may be other more optimal methods of grouping individuals or populations based on morphological similarity than the currently recognized taxonomy. However, for a morphologically based taxonomic treatment to reflect existing patterns of variation the taxa described should be diagnosable/identifiable by morphology. In an ideal scenario, a morphologically based taxonomic treatment would circumscribe taxa that are not only diagnosable by morphology, but that also define taxa that correspond to the existing structure of morphological variation, i.e. answer yes to both questions posed here. In this study, I evaluated 32 morphological traits in 43 manzanita taxa (33 species and 18 subspecies) to address these two questions.

METHODS:

Selection of taxa and specimens for measurement—

I selected taxa that provide a range of variation in terms of the geographic regions where they are found, the size of ranges (i.e. widespread and narrow endemic taxa), and monotypic as well as polytypic species. I also focused on taxa that had a minimum of five specimens in the herbarium at University of California, Riverside (UCR) and that are part of ongoing studies at UC Riverside using genomic data for phylogenetics and species delimitation. The final list of taxa included 33 species and 43 taxa, including subspecies. Of the 33 species, 24 were monotypic, and nine were polytypic. For four of the nine polytypic species (*A. manzanita, A. parryana, A. patula* and *A. nevadensis*), I sampled from only one subspecies. In the case of *A. patula*, which was relatively recently divided into two subspecies with the description of *A. patula* subsp. *gankinii* (Parker and Vasey, 2016; Kauffmann et al., 2021), I had extensive and widespread sampling, but did not find any plants that keyed out to *A. patula* subsp. *gankinii*. For the remaining five polytypic species, I sampled between two and eight subspecies.

Of the 33 species included, 11 were widespread, while 22 were narrow endemic species. None of the subspecies were narrow endemics. The relative range size (widespread versus narrow endemic) of each species is noted in Table 2.1. For narrow endemics species I included a minimum of five specimens from a single locality. For monotypic widespread species, I included a minimum of five specimens from a minimum of three localities, making an effort to choose populations that covered the geographic range of each species. For the widespread species that have subspecies, I targeted a minimum of five samples per subspecies, taking as wide a

geographic sampling as possible within each subspecies. In total, the 33 species, and 44 taxa including subspecies, represent approximately 55% of currently recognized species, and 41% of currently recognized taxa, respectively, and were represented by a total of 326 specimens in the final dataset (Appendix 2.1).

Selection of variables for data acquisition—

I first identified a list of characters that are used in dichotomous keys and species descriptions, based on a review of Jepson Herbarium eFlora and the *Field Guide to Manzanitas* (Baldwin et al., 2012; Kauffmann et al., 2015). Some variables were simple dimensional measures that were measured with a ruler, while other variables were categorical or ordinal. Several of these latter variables were binary (e.g. presence/absence of a basal burl), while others included several ranked states (e.g. hair density "glabrous", "sparse", "moderately dense", or "dense").

Reproductive characters were excluded from our variable list for several reasons. Reproductive characters are linked to seasonal phenology, meaning specimens often don't have both nascent inflorescences and fruits, the traits of which feature prominently in identification keys and species descriptions. In addition to this logistical issue for scoring traits for analyses, using vegetative characters, which are always present, allows the results to be translated into field-applicable guidance that will be useful no matter the phenological state of an individual plant.

Data collection and pre-analysis processing—

The majority of data were recorded by myself, with additional data recorded by undergraduate researchers whom I had trained myself to ensure they were logging each variable equivalently to myself. Quantitative data was collected either from herbarium specimen labels (for variables like plant height), or measured on specimens using a ruler.

Pilot study on quantitative measurement of hair lengths and densities—

Two of the more difficult traits to evaluate in identifying manzanita taxa are the lengths and densities of hairs. Hairs are divided into four length categories: pubescence, short hairs, medium hairs, and long hairs. Hair densities are divided into three categories: sparse, moderately dense, and dense. In an effort to provide a quantitative framework for these categories, I conducted two pilot studies measuring densities and lengths of stem hairs. To quantify hair densities, I took a photograph on camera-equipped dissecting scope, at 8X magnification, of the stem surface of three stems on each of 22 samples, representing five taxa that differ in the descriptions of hair length and density. I then used imageJ (Schneider et al., 2012) to measure hair density on these images by (1) calibrating the scale of the image, using an in-image scale bar provided by the microscope, (2) drawing a rectangle on the surface of the stem of approximately

2 mm^2, (3) recording the exact actual area of that rectangle, and (4) counting all the individual hairs within that rectangle. I then divided the number of hairs by the area of the rectangle in which they were counted.

To quantify hair lengths I made three cross-sections from each of three different stems per individual, for 24 individuals representing six taxa (3-6 individuals per taxon), and recorded a micrograph on a TM4000 scanning electron microscope (Hitachi Inc., Tokyo, Japan). I then measured the lengths of individual hairs along the perimeter of the cross-section, using imageJ. To measure the length of individual hairs shown in a micrograph, I (1) calibrated the measurement tool in imageJ using the in-image scale bar provided by the microscope, (2) manually drew a line down the center of each individual hair along its entire length, and (3) applied the measurement tool to that line to get the length of the hair in millimeters.

Using these hair length and hair density data I then plotted a histogram to determine if there were multiple peaks indicating multiple distributions corresponding to the different hair length and density classes that could be used to define quantitative categories. For both hair lengths and densities, I found a continuous and unimodal distribution and therefore could not find an empirical basis for translating the commonly used descriptive categories into quantitative ranges (Fig. 2.1).

Final variables used in analyses—

Because attempts to quantify hair measurements failed to produce repeatable objective categories, I used a ranked system for hair length and density, based on the classes of hair lengths

and densities used in taxonomy and keys. I scored/logged these hair variables by viewing a mature stem segment or leaf under magnification of a hand lens or dissecting scope and recording the density class for hairs of each length class present on the leaves or stems of the specimen. The number of stem segments or leaves viewed per specimen varied depending on the amount of material on a given sheet, but was typically between two and five stems, and at least ten leaves. Because manzanita stem hairs are usually only found on stems for one or two seasons of growth, until the development of mature bark, I scored stem segments that were in the most recent mature growth that had not yet developed bark.

A single individual can have multiple lengths of hair with different densities. I therefore logged data in a descriptive, textual format using a consistent set of terms to describe different states, and later translated these textual descriptions into ordinal variables. Separately for stems and leaves, I divided hair length into four variables (pubescence, short, medium, and long), and for each variable I recorded the density of hairs of that length. I categorized hair densities as “no hairs”, “sparse”, “moderately dense”, and “dense” which I encoded as 0,1, 2, and 3, respectively. For example the textual description “dense pubescence, sparse long hairs”, indicating that there were two types of hairs (pubescence and long) of different densities (dense and sparse respectively) would be recorded as 0 (no hairs in that length category) for the short hair and medium hair variables, 3 for the pubescence variable, and 1 for the long hair variable. A textual description like “moderately dense pubescence, dense medium hairs” would be recorded as 2 for the pubescence variable, 3 for the medium hairs variable, and 0 for everything else. When a stem or leaf was completely glabrous, with no hairs of any lengthI entered 0 for each variable. The inclusion of eight total variables encoding hair length and density, four each for leaves and stems, presents the possibility that hair variables will constitute a large percentage of overall variation

recovered. Although this may impact inference and results, it parallels how hairs are handled in taxonomy and keys, since hairs are often described in combinations of different densities for different length classes, thus the number of variables used reflects that complexity.

Leaf base shape was also recorded as a set of variables, rather than a single variable, because taxonomic descriptions and keys commonly use ranges of shapes (e.g. “tapered to truncate”), rather than a single shape description. This reflects the variation in leaf shape that can be found on a single individual in some species. I logged binary variables (presence/absence) for four leaf base shape categories (tapered, rounded, truncate, and auriculate) (See Fig. 2.2 for examples of these shapes). I employed this strategy because using a single value for leaf base shape would not allow me to record the range of shapes represented. For example, if the leaf bases on a specimen ranged in shape from tapered to truncate, and a central value was recorded, this would record the leaf base shape for that specimen as the same as a specimen that had strictly rounded leaves, and no other shape.

For quantitative measures of leaf dimensions, in order to minimize the possible impact of developmental time point differences on leaf dimension measurements, and to provide an objective mechanism for scoring leaf traits, I recorded two sets of leaf measurements (blade length, midrib length, blade width, petiole length, length of midrib from petiole to widest point of blade) for each specimen (Fig. 2.3). One set was recorded for the single largest leaf on a distal, mature segment of growth, avoiding leaves on either new or immature growth, or those on stem segments that were multiple growth seasons old. We recorded the dimensions of the largest leaf, rather than estimating average dimensions, as attempts to apply random sampling to obtain average leaf dimensions were prevented by differences in the amount, condition, and age of material available for each specimen. In contrast, measures of maximum dimensions proved

repeatable and objective and provided a measure of leaf size comparable to the upper end of the size range that would be reported in a taxonomic description or in a dichotomous key. Additionally, I wanted to prioritize the number of specimens and taxa included and this approach provided an objective way to obtain measurements from as many specimens as possible. The second set of leaf measurements was recorded in an identical manner, but from the largest leaf from older stem segments of growth. On some occasions, the largest leaf overall was also in the mature distal stem segment, and in such cases, the same measurements were recorded for both sets of leaf measurements. Both sets of lead dimension measurements were retained in the final dataset.

I translated the several binary variables (e.g. absence/presence of basal burl) in the raw dataset into a 0/1 numerical format. In specimens for which it was unclear whether the morphology shown conformed to the "present" state of a binary character, such as a weakly auriculate leaf base, the variables were coded as 0.5. See Table 2.2 for a list of the 32 final format variables used for analyses.

Additionally, to explore the possible effect of applying the commonly used categories of length and density on the results of our main analyses, I ran separate analyses in which the hair variables were simplified to binary absence/presence for stem or leaf hairs. If the use of the commonly used but perhaps not clearly defined categories for length and density biases results, then the removal of that information from the dataset should produce different results. This also provided an opportunity test for a possible effect of heavy weighting of hair variables, due to the large number of those variables that were included.

Before formatting the raw data variables into a final, standardized format, I removed three raw data variables that had proved difficult to score consistently: leaf greenness, longest internode, and whole plant width. Leaf greenness could not be logged consistently as the leaves on many specimens lose their color over time. Longest internode values were found to be difficult to record, as leaves or leaf scars could be obscured on the underside of glued material. Whole plant width was frequently missing from specimen data.

For leaf dimension measurements, I retained the length of the leaf blade as a single absolute scale variable, to represent differences in general leaf "size" among specimens and taxa, but used the other raw leaf dimension variables to calculate proportional variables reflecting shape. The formulae for these calculated variables are given in Table 2.3, alongside visual depictions of the raw dimensional variables originally recorded in Fig. 2.3. The computation of these proportional variables was intended to provide numerical representations of shapes like "auriculate", "ovate", and "obovate" that are commonly used to describe leaf shapes in botany (examples of these leaf base shapes can be found in Fig. 2.2).

This final curated dataset was a matrix of 32 variables, recorded without missing values, for 326 samples, representing 33 species and 50 taxa, including subspecies.

Data analysis—

Taxonomically-informed analysis

In order to address our two main questions, I took two paths of analysis. The first path, aimed at answering question 1, relied on a linear discriminant analysis (LDA) (Fisher, 1936). The LDA

model was tasked with identifying taxa based on the morphological data, which could then be evaluated for accuracy against human determinations on a sample-by-sample basis, using a leave-one-out cross-validation method. This provided a test of the morphological distinctiveness of the taxa in the dataset.

Before running the LDA models, I transformed the dataset into continuous and uncorrelated variables using principal components analysis (PCA) (Pearson, 1901). I retained all resulting components of the PCA, rather than selecting a set of higher-variance components, as is done in some applications of PCA, to avoid potentially removing taxonomically informative, but low variance, components. Such components could differentiate one taxon from all others, and would thus have small total variance, despite being informative.

Finally, because the sample sizes among taxa in the dataset were variable, with several taxa (widespread species) having multiple times as many individuals sampled as narrow endemic species, I performed random sampling within taxa to achieve equal sample sizes. This procedure was repeated 5,000 times, with each step consisting of (1) randomly subsampling individual taxa to a maximum of five samples per taxon, (2) performing a PCA on the resulting subsampled dataset, (3) computing an LDA, and (4) calculating a "confusion matrix" (a matrix that represents the number of samples in the analysis that were *accurately* classified by the LDA model) for that LDA. After repeating this procedure 5,000 times, I averaged the confusion matrices for all repetitions. I implemented this procedure in the R statistical computing environment (R Core Team, 2023).

Because polyploid species are often more morphologically variable than diploid species, and as a result, are more often subdivided into subspecies that may themselves be variable, I ran the above procedure for two configurations of the dataset. One configuration included all samples, and one included only diploid species. The ploidy of each taxon is noted in Table 2.1. For each of these, I subsequently ran separate LDAs trained on full taxonomic determinations (including subspecies where available), or trained simply on species-level determinations. The results of the models trained using different ploidy-based sample sets, and different taxonomic ranks for training, were compared to determine if patterns exist, e.g. taxa of one rank or ploidy being more distinguishable or distinct than those of the other rank of ploidy. Additionally, because I wanted to investigate whether these analyses could detect a signal of morphological differences among populations within species, I ran a separate LDA that was trained on all the data (not subsampled) for three widespread species (*A. patula*, *A. pungens* and *A. pringlei*). For these species we selected discrete populations for each species that were geographically distant from one another. This LDA was trained and its accuracy assessed at the population level, rather than species level.

Taxonomically–naïve analysis

In order to address our second goal, assessing whether the structure or pattern of morphological variation reflects current taxonomic circumscriptions, I employed a second analytical path. Starting from the same curated, final dataset as above, I computed a non-metric multidimensional scaling (NMDS) analysis using the vegan package in R (Oksanen et al., 2022). This method of analysis projects high-dimension distances among samples into a lower-dimension space which can be more readily interpreted and subsequently analyzed (Kruskal,

1964). To compute an NMDS I first had to compute a distance matrix representing the pairwise morphological distance between each sample and every other sample.

Because I had a mixture of continuous numerical variables and ordinal variables, I calculated the distance matrix using Gower distances, as that particular distance method is well suited for such mixed datasets (Gower, 1971). I then used the metaMDS function in the R vegan package to compute the NMDS (Oksanen et al., 2022). The computation of an NMDS requires the specification of a set number of new dimensions upon which to project the distances in the original distance matrix (Kruskal, 1964; Minchin, 1987). I ran several numbers of dimensions, starting at two, and increasing the dimensionality until a projection that had a good fit to the original distance matrix. I assessed goodness of fit at each number of dimensions using the "stress" statistic reported by metaMDS, increasing the number of dimensions until achieving a stress value that was less than 0.1, which is generally considered a good approximation of the higher dimensional distances (Minchin 1987). After producing the NMDS, which resulted a 5-dimensional space, I computed a matrix of Pearson correlation coefficients (Freedman et al., 2007) between the input variables and each NMDS dimension. This correlation matrix was constructed to allow for interpretation of morphological meaning of each NMDS dimension.

After computing the NMDS, I performed a k-means clustering analysis on the resulting latent variables (MacQueen, 1967). In order to run a k-means analysis, a specific value for k must be specified, which represents the number of clusters among which the samples are to be sorted. I tested k values ranging from 2 to 60, as I wanted to make sure to include a clustering result at the number of species or taxa included in each particular analysis, as well as including a range of values below and above. This range allowed me to find an optimal number of clusters for each k-

means result. To find optimal k values, I used the "elbow method", calculating within-cluster sums of squares (WCSS) for each k value, and plotting WCSS as a function of k, to find the "elbow" in the graph, where the rate of decrease in WCSS slows, and the curve flattens, which corresponds to the optimal number of clusters for that analysis. It should be noted that "optimal" means the largest value for k at which the clusters are still informative. Finally, I evaluated the k-means results by comparing cluster assignments of individual samples to their taxonomic determination, to assess whether individual taxa were being assigned to species-exclusive clusters, or were being co-clustered with other taxa.

The analytical path described in this section was performed separately for the full dataset, and a diploid-only subset of the dataset, each of which were subsequently analyzed with species or all taxa as the basis of interpretation.

Because k-means clustering is theoretically able to produce clusters with differing numbers of samples (MacQueen, 1967), I ran the NMDS and k-means analyses on the full dataset without subsampling, in contrast to the LDAs which require equal sample numbers. However, in practice, since larger taxonomic samples naturally capture more variation, I also ran the NMDS and k-means procedures on randomly subsampled versions of the dataset, in which the sample size of individual taxa was balanced and equal to five.

Hypervolume overlap analysis

In order to further explore morphological distinction among taxa in the high-dimensional space of the NMDS, I also computed pairwise hypervolume overlap statistics using the hypervolume package in R (Blonder et al., 2023). These statistics are derived by pairwise

comparisons, calculating the proportion of the hypervolume occupied by the samples of one species or subspecies that overlaps with the hypervolume of another species or subspecies. I calculated these hypervolume overlaps for all pairs of species/subspecies, and calculated the degree of overlap among them. Results from the overlap analysis were used to explore patterns of shared variation including similarities among geographically close taxa or specific trait conditions/states shared among taxa.

RESULTS:

Taxonomically-informed analysis—

Because we wanted to test the morphological distinctiveness of the taxa included in this study, we trained LDAs to test for taxonomic signal in the dataset. The LDAs show a high degree of accuracy overall. The LDA models trained only on species-level determinations showed a minimum accuracy of approximately 97%, with many species being classified to the correct determination 100% of the time (Fig. 2.4). However, in the case of the LDA that included both subspecies (sixteen) and species (twenty-six) determinations, while most taxa were still classified with high accuracy (>90%), all but one subspecies (subsp. *adamsii*) of the polyploid species *A. glandulosa* were frequently misclassified as another subspecies of *A. glandulosa*. One subspecies (*A. glandulosa* subsp. *mollis*) was classified correctly only approximately 69% of the time (Fig. 2.5). Another subspecies, *A. glandulosa* subsp. *cushingiana* was classified with relatively high accuracy (approximately 92%), but received a large share of misclassifications from other subspecies; approximately 38% of misclassified *A. glandulosa* were grouped into *A. glandulosa* subsp. *cushingiana*. An LDA model trained on only diploid taxa showed high accuracy, with 20

out of 30 diploid taxa (67%) being classified by the LDA to the correct taxon, only two taxa (7%) being classified at lower than 98% accuracy, and the singler lowest classification accuracy being 84% (Fig. 2.6).

Because I wanted to test for morphological differentiation among populations of widespread species, I ran an LDA trained on the population-level for three widespread species, *A. patula*, *A. pungens*, and *A. pringlei*. This LDA showed very high classification accuracy, with only one sample being incorrectly assigned, indicating that there is meaningful morphological differentiation even among populations of some species (Fig. 2.7).

Finally, to address the possible effect of my use of the qualitative and possibly subjective categories commonly used to describe hair lengths and densities, I ran another pair of LDAs of the same structure as the first LDA described above, one with all species, and one with all taxa including subspecies, but with the stem and leaf hair variables each collapsed to a single, binary variable: the presence/absence of hairs, of any length or density, on each organ (stem hairs present/absent, and leaf hairs present/absent). In the LDA trained on species, with simplified hair variables, 17 of 33 species were predicted with at least 99% accuracy, while the other 16 species were predicted with between 80% and 97% accuracy. In the LDA trained on all taxa including subspecies (Fig. 2.8) 17 of forty-four taxa were predicted with at least 99% accuracy, while the other 27 taxa were predicted with between 55% and 97% accuracy. Species were predicted more accurately than subspecies, representing 15 of the 17 taxa predicted with at least 99% accuracy. Subspecies were predicted with less accuracy (mean accuracy of 93% for species, versus a mean accuracy of 84% for subspecies), which was a weakly significant difference (Welch's two-side test, $t = 2.43$, $df = 28.2$, p-value = 0.0216). These LDAs, taken together show that accuracy of

prediction is lower, but still generally high, when the commonly used categories of hair length and density are simplified to binary absence/presence of hairs on stems or leaves.

Taxonomically-naïve analysis—

In order to evaluate whether dominant patterns of morphological variation in the dataset corresponded to currently recognized taxa, I used a combination of NMDS and k-means clustering to sort samples into taxonomically-uninformed clusters, and then compared composition of those clusters to the species determinations. I performed four comparisons of this type: (1) with all species and subspecies; (2) with all species, but not subspecies; (3) with only diploid species, and (4) with only narrow endemic species.

The matrix showing correlations between individual input variables and each NMDS dimension (Fig. 2.9) showed that the first NMDS dimension most strongly correlates with the range of variation from sessile, auriculate and/or overlapping (imbricate) leaves, to those with longer petioles, tapered leaf bases, and non-overlapping leaves. Consistent with these associations, the first NMDS dimension separates a set of species from California's Central Coast region that are sometimes termed the "auriculate" species, having mostly sessile leaves, with auriculate leaf bases, and overlapping, imbricate foliage (Figs. 2.9-2.10). The second NMDS dimension appears to reflect the spectrum of leaf base shapes that are not auriculate, with a strong negative correlation with tapered leaf bases, a weak positive correlation with rounded leaf bases, and a strong positive correlation with truncate leaf bases (Fig. 2.9). The third NMDS dimension is correlated with several hair density variables, and the presence or absence of stem and leaf

glands, and this dimension appears to separate samples with more strongly glandular hair morphology from those lacking such morphology (Fig. 2.9). The fourth and fifth NMDS variables show less clear patterns of association with individual input variables (Fig. 2.9).

The clusters produced by the k-means analysis did not correspond to taxa. Additionally, the results of k-means analyses were found to be unstable for all four of the datasets tested, with repeated runs rendering assignments of individual samples into different clusters. Therefore in some instances a taxon might be co-clustered mostly with a second species, but in the next instance co-clustered mostly with a third species, but not the second. Such unstable results were found when running the k-means algorithm on the full sample set, as well as when running it on subsampled versions of the dataset that adjusted for more balanced sample sizes within taxa or species, and when making comparisons of species or subspecies to various cluster assignments. This instability was also found in the results of the k-means clustering performed on all other subsets of taxa (diploid-only, species-only, and narrow endemics-only). Additionally, this instability was observed across the investigated ranges of possible values from k, i.e. whether I split samples into a small or large number of clusters. This instability in the k-means result suggests that there are no well supported clusters in the dataset. This unstable pattern of clustering provides an answer to question 2, that the structure of morphological variation does not correspond to the current taxonomy.

Hypervolume overlap—

The hypervolume overlap analysis (Fig. 2.11) showed that only one species (*A. nissenana*) had zero morphological overlap in the NMDS, but many taxa did occupy relatively distinct morphological space. Diploid species overlapped each other to a lesser degree than did polyploid species (Fig. 2.11). Narrow endemic species (all of which included here are also diploid) overlapped with other narrow endemic species to a lesser degree than they overlapped with widespread species (Figs. 2.11-2.12). One noticeable exception was the substantial overlap of two narrow endemic "auriculate" species, *A. auriculata* from the San Francisco Bay area and *A. andersonii* from the Monterey Bay area of California, with a third auriculate species from the Monterey Bay area, *A. pajaroensis* (Figs. 2.11-2.12, and see map in Fig. 2.13). In most other cases, narrow endemic species had less than 1% overlap with any other narrow endemic species. However, these species did often show considerable overlap (sometimes greater than 50%) with various widespread species (Figs. 2.11-2.12).

Many species, both narrow endemic and widespread, overlapped substantially with *A. manzanita* (represented here by the single subspecies *A. manzanita* subsp. *manzanita*), in particular, which is a relatively widespread, tetraploid, subspecies. Two widespread taxa, *A. pungens* and *A. parryana* subsp. *parryana*, overlapped almost entirely with *A. manzanita* subsp. *manzanita*, having less than 10% unique variation when compared with *A. manzanita* subsp. *manzanita* (Fig. 2.11). Conversely, *A. manzanita* subsp. *manzanita* only overlaps other taxa by a small percentage, because the morphological hypervolume occupied by *A. manzanita* subsp. *manzanita* is much larger than the hypervolumes occupied by other species that overlap with it. Thus from the perspective of *A. manzanita* subsp. *manzanita*, the percentage overlap is small.

The hypervolume overlap analysis also highlights that in some polytypic species, several of the subspecies share many morphological trait states. *A. glandulosa* subspecies, which are very well sampled in this study, show this most clearly, but the same pattern can be seen between subspecies of *A. pringlei*, as well. This might be expected, since subspecies are not expected to be as diverged as species (Persoon, 1805; Clausen, 1941). When overlap is calculated between various taxa and *A. glandulosa* as a whole (not divided by its subspecies), overlap is noticeably increased, reflecting the larger hypervolume/range of variation encompassed by the species as a whole.

Overall, the analysis of hypervolume overlap showed a mixed pattern of morphological distinctiveness among taxa, with some taxa overlapping relatively little with others, and some taxa sharing much variation with multiple other taxa. This result suggests that at least some sets of taxa are not morphologically distinct in their vegetative morphology.

DISCUSSION:

In this study we addressed two questions, one asking whether currently recognized manzanita taxa are morphologically distinct, and one asking whether the current taxonomy is the best way of describing variability in the genus. These questions are closely related, but are not the same, and our analyses produced different answers to each. The LDA results show that manzanita taxa are morphologically distinguishable, even when considering only vegetative traits, and that some subspecies, perhaps as expected, are less distinct from each other than species are. In

contrast, the results of the k-means clustering analyses showed little to no correspondence between taxa and clusters, indicating that current taxonomic boundaries may not accurately reflect the natural structure of vegetative morphological differentiation. Comparisons of morphological overlap among taxa showed only one taxon being entirely non-overlapping with other taxa. This suggests that the distinguishability of currently recognized taxa, as shown by the LDA results, is due to differences in specific delimiting traits. It only requires one trait to distinguish species in these analyses, and even a difference that seems subtle to someone with an identification key can produce this result (dense hairs vs moderately dense hairs). However, in species such as *A. pringlei* or species in other genera, subspecies are equally distinguishable, often by a single discrete trait state (seeds fused in *A. pringlei* subsp. *drupacea* and seeds free in *A. pringlei* subsp. *pringlei*), which raises the question of why such differences are considered diagnostic of species in some cases and subspecies in other cases.

I also found that while only one taxon was detected as occupying a unique morphological space, a number of other taxa were mostly non-overlapping with other taxa, with this being more so for narrow endemic species and more so for species rather than subspecies. I also found that certain widespread taxa (e.g. *A. glauca* and *A. patula*) had more morphological variation (larger hypervolume) overall than narrow endemic species, producing an asymmetric pattern wherein the morphological hypervolume of some narrow endemic species substantially overlaps that of one or more widespread taxa, but the percentage overlap of the widespread taxa with the narrow endemic is small.

While the LDA results showed that the taxa included in this analysis are distinguishable even when only vegetative morphology is considered (question 1), the instability of the results of

the k-means analyses suggest that the current taxonomy does not necessarily best reflect the existing structure of morphological variation (question 2). The results of the hypervolume overlap analyses show that most taxa have a percentage of overlap with other taxa, with some showing extensive overlap; this also can be seen as suggesting that current species boundaries do not correspond to the structure of morphological differentiation in the genus.

One reason the results of these analyses differ may be underlying differences in how variables are treated/used by the different methods. The classification-based method (LDA) used to address the first question is able to prioritize some variation over other variation in training the model, by linking variation in a given input variable, or combinations of variables, to one of the input categories/labels (species or subspecies). These input categories are known to the LDA in model construction, but are not known to the NMDS or k-means clustering analyses, although they are necessary for the interpretation of the NMDS using the hypervolume overlap analysis. In this way the LDA is performing a procedure similar to the process of building a dichotomous key, and may find taxon-specific combinations of trait states that may account for a small proportion of overall variation, but are strongly linked to individual taxa. This lends support to *distinguishability* (question 1), but does not necessarily lend support to the taxonomy reflecting natural structure of morphological variation (question 2).

A taxonomically-naïve analysis might be able to separate taxa as neatly as the LDAs have in this study, if more variables were to be included. However, given the amount of morphological overlap among taxa, it would likely require a very large number of distinguishing variables. For example, reproductive traits are important in manzanita taxonomy and were not included here. Nonetheless, the LDA was able to classify most specimens to the correct species without those

reproductive characters. This is intriguing, as a common complaint of those attempting to identify manzanitas is that they need characters from multiple phenological stages, e.g. nascent inflorescences and fruits, which often are not found on a single individual at the same time. The strong accuracy of the LDA in these analyses shows the potential for developing dichotomous keys based entirely on vegetative traits that do not rely on reproductive traits at all. Other smaller findings relating to characters used in keys, may likewise be figured into the design of future keys, for example, that specimens can have unclear or intermediate conditions for several characters thought to be binary (e.g. presence of a basal burl, or presence of certain leaf base shapes).

Some of this overlap shown in the NMDS-based analyses appears geographically-linked, and this is not surprising, as it is well understood that some species with shared morphological characters occur in relatively close geographic proximity, for example the three auriculate species of the Monterey Bay and San Francisco Bay area (*A. auriculata, A. andersonii*, and *A. pajaroensis*, Figs. 2.11-2.12 and Fig. 2.13). In other cases, there is no clear geographic pattern to the morphological overlap, for example the overlap found between the widespread taxa *A. pungens* and *A. parryana* subsp. *parryana*, which are found largely in Southern California or outside of the CFP, and *A. manzanita* subsp. *manzanita*, which is found in the Sierra Nevada and Coastal Range north of the Bay Area. In particular, widespread and polytypic species, which are most variable, like *A. glandulosa*, complicate matters, as their range of morphology often overlaps other species considerably. By contrast, diploid species, as a subset, overlap each other less frequently and less substantially, although most do have some overlap with other diploids (mean of approximately 20%).

The larger amount of morphological variability in widespread taxa, as compared with narrow endemic species, may be due to increased variability in habitat and climate which can lead to local adaptation or plasticity, or to founder effects in partially isolated populations spread across a wide geographic area. It is also intriguing that so much of the variation of many species falls within the hypervolume representing variation in *A. manzanita* subsp. *manzanita*, as this is only one of six subspecies of *A. manzanita*. The inclusion of more subspecies of *A. manzanita*, as well as missing subspecies from other polytypic species, would likely lead to even greater overlap.

It should be noted that the meaning of overlap here is difficult to interpret taxonomically, with regard to question 2. Whether or not two species or subspecies are found to be non-overlapping in terms of morphological variation, we cannot say how different is different enough to mean that two species are truly distinct species, or how similar is similar enough for two species to be considered one. In the case of subspecies, attempts at such judgments are even more fraught, as there exists little consensus in botany on what a subspecies actually is (Huang et al., 2020). Such questions are ultimately subjective, and of a philosophical nature, depending on how species, subspecies, and even populations, are defined. However, studies of this kind form the empirical context by which to make these philosophical judgements in a way that is grounded in the structure of natural variation.

Studies are currently being undertaken currently using next generation sequencing methods, and similar sampling as was included in this study, that are seeking to assess the genetic distinctiveness of manzanita species. This morphological analysis can help inform the

interpretation of those coming studies, as the pairing of morphological and genetic data is often recommended as a best practice when conducting delimitation studies using sequencing data.

CONCLUSION:

Our results show that the manzanita taxa we sampled are for the most part morphologically distinguishable, but simultaneously show that the current taxonomy may not be the most natural way of partitioning the overall pattern of morphological variation in this genus. The results show promise for increased emphasis on vegetative characters in the construction of keys for taxonomic identification, which would make the identification process more accessible to people making actual identifications in the field or from herbarium specimens lacking reproductive material.

ACKNOWLEDGEMENT:

I thank all past and present members of Team Manzanita, at UC Riverside, whose contributions to field work, lab work, analysis, and general support made this work possible. I thank Tito Abbo in particular for his work in processing collections into herbarium specimens, which facilitated this work, and I thank him also for contributing to data collection. I thank undergraduate researchers Makayla Drew, Vijay Sim, Lucie Nguyen and Anna Rothfusz for contributing to data collection. I thank all those who helped with permitting for field collections, and those who helped me with collections, in particular, Yi Huang, Tito Abbo, Angela Buehlman, and Amy Litt. I thank Andy Sanders and Teresa Salvato of the UC Riverside

Herbarium for their work in processing specimens that facilitated this work. I thank Jon E. Keeley and Tom Parker for giving me feedback in the selection of morphological traits to measure.

FIGURES AND TABLES:

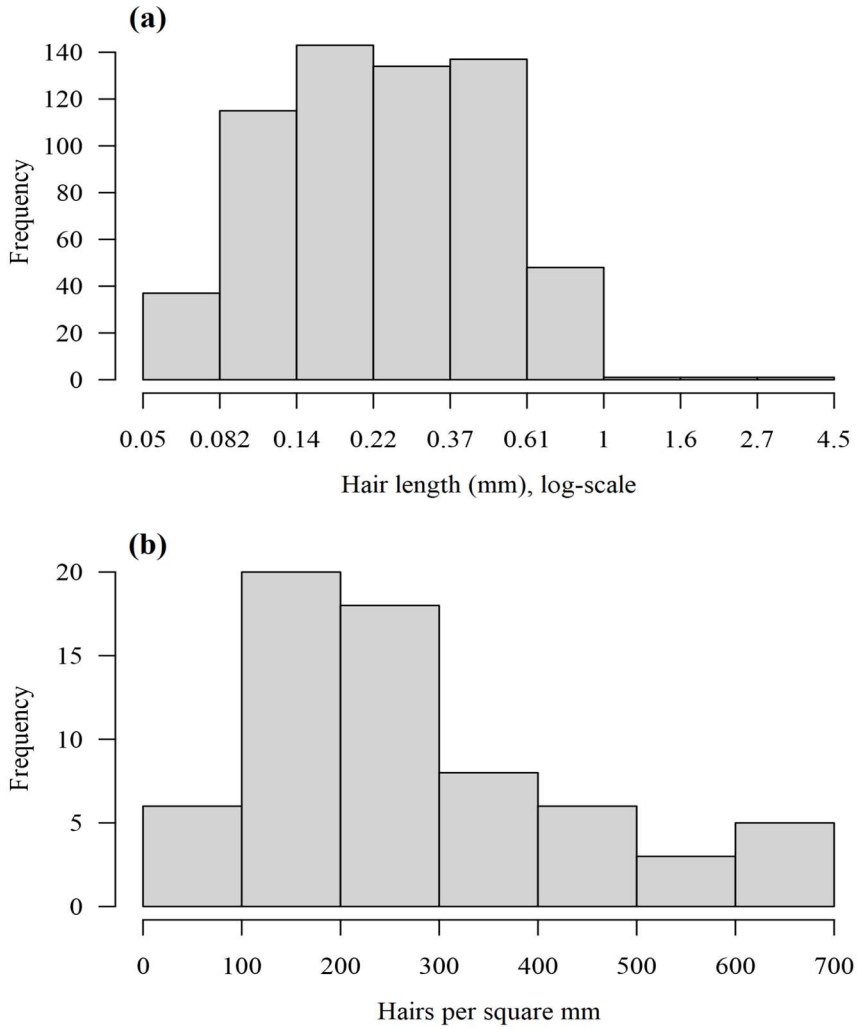


Fig. 2.1. Histograms showing the distribution of hair length (a) and density (b) measurements from the pilot studies. The x-axis for the histogram of hair lengths (a) is plotted on a log-scale, due to the presence of a small number of very high values (very long hairs) that otherwise compress the rest of the distribution (lower values, shorter hairs).

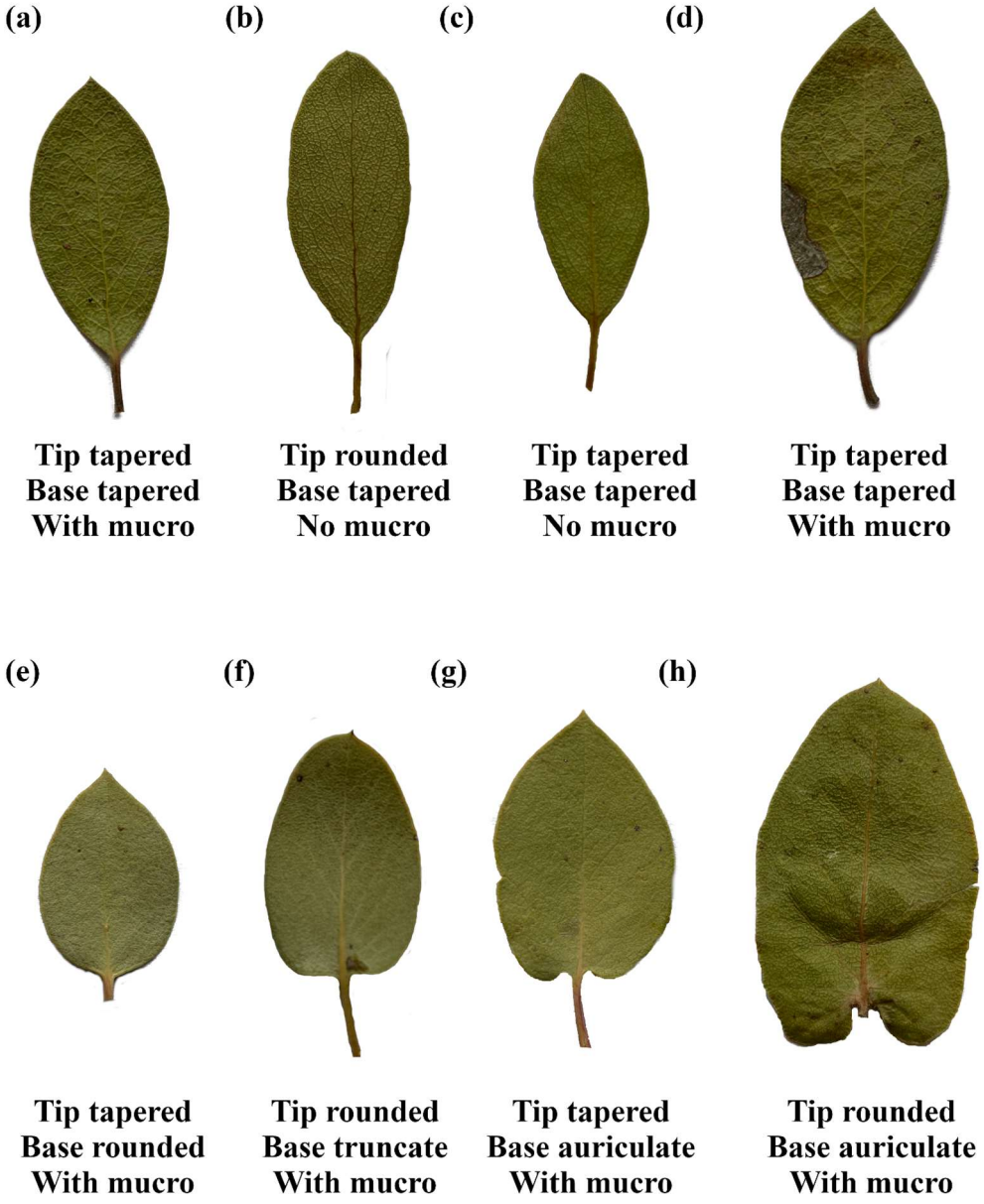


Fig. 2.2. Photographic examples of leaves with different base and tip shape, and showing leaves with and without mucros, the small pointed structure at the tip, which was a trait included in the analysis.

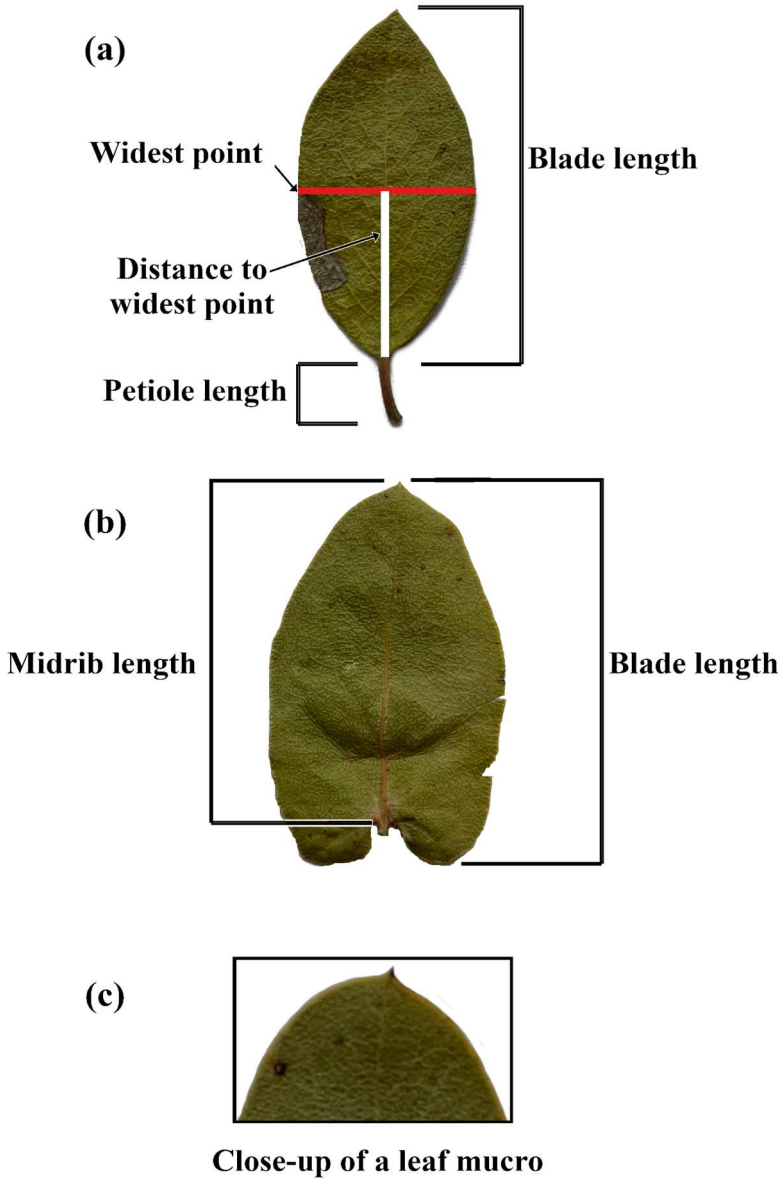


Fig. 2.3. Photos of manzanita leaves, labeled to show the various quantitative dimensions recorded (a-b), as well as a close-up view of a leaf tip mucro (c), a trait that was also included in the dataset.

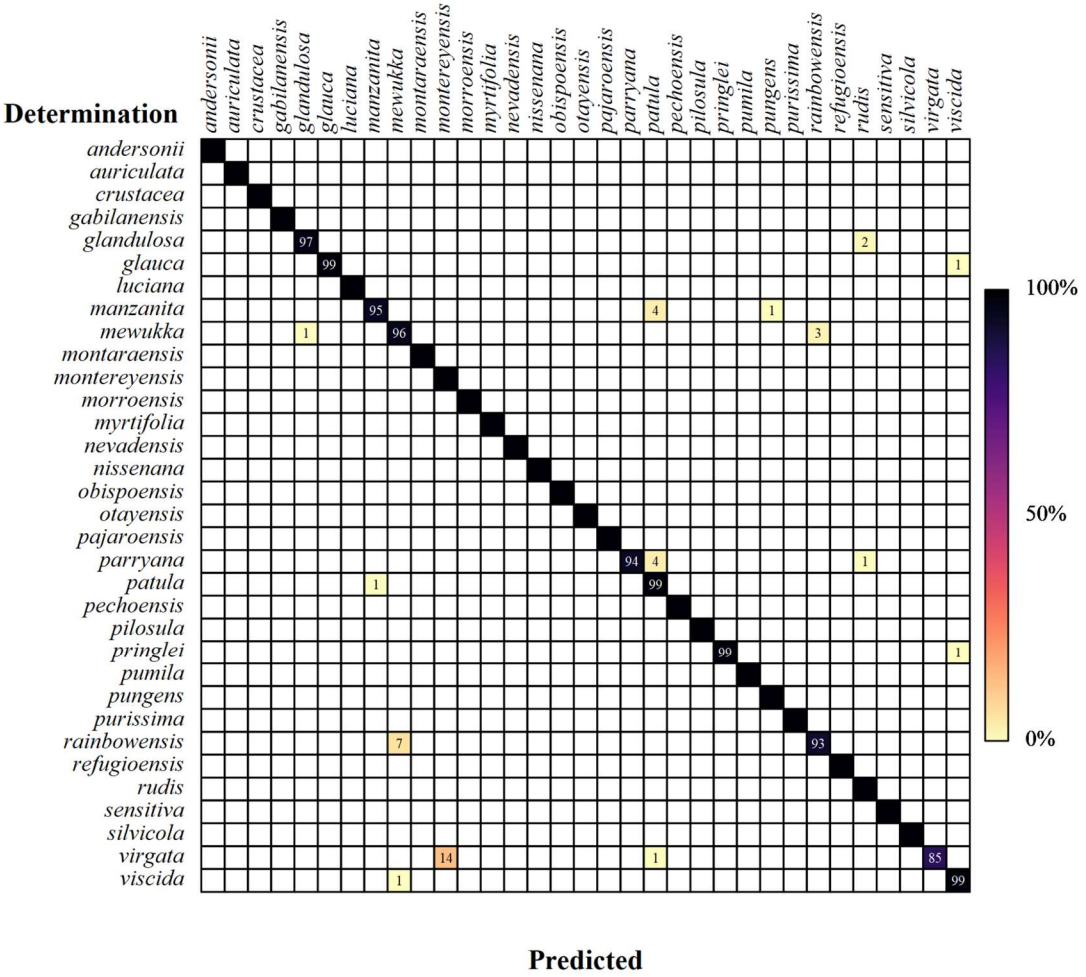


Fig. 2.4. Results of the linear discriminant analysis trained on species-levels determinations. Colors are scaled to show accuracy of LDA assignments for a given species, compared against any human determination. Percentages (rounded to nearest whole number) greater than 0% and less than 100% are plotted in each square.

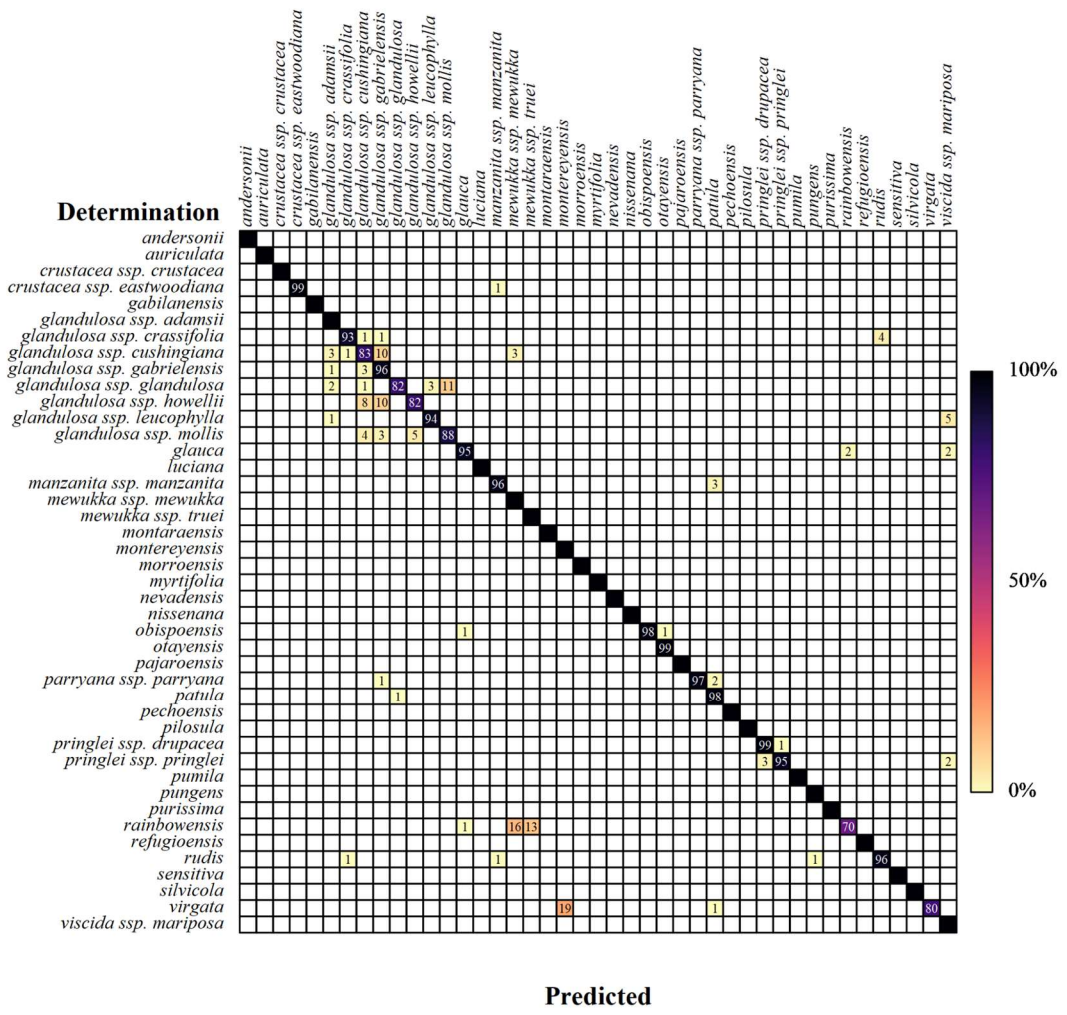


Fig. 2.5. Results of the linear discriminant analysis trained on both species and subspecies determinations. Colors are scaled to show accuracy of LDA assignments for a given species or subspecies, compared against the human determination. Percentages (rounded to nearest whole number) greater than 0% and less than 100% are plotted in each square.

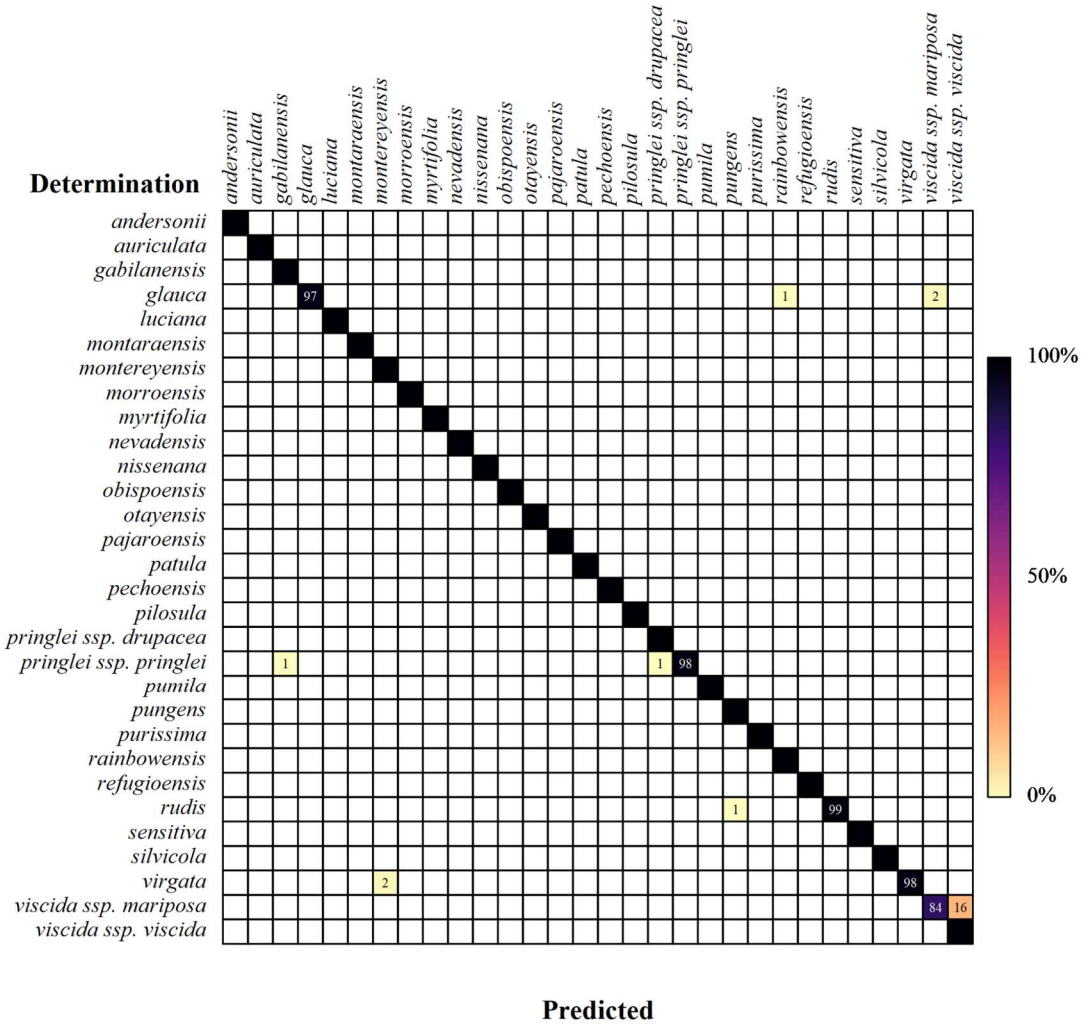


Fig. 2.6. Results of the linear discriminant analysis trained on only diploid species and subspecies determinations. Colors are scaled to show accuracy of LDA assignments for a given species, compared against the human determination. Percentages (rounded to nearest whole number) greater than 0% and less than 100% are plotted in each square.

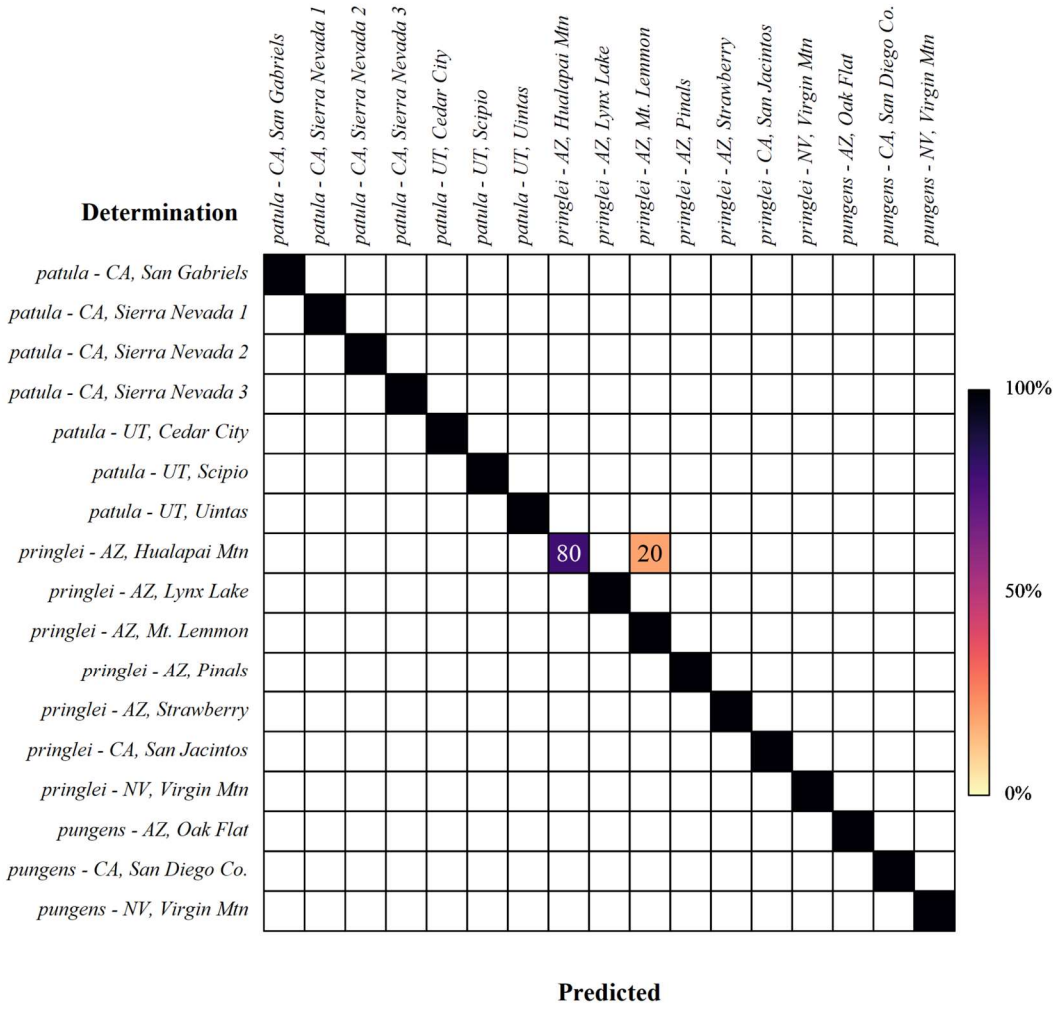


Fig. 2.7. Results of linear discriminant analysis trained on the locality/population level for three species in which we tested distinguishability at the population scale. Colors are scaled to show accuracy of LDA assignments for a given species at a given locality/population, compared against the human determination, and actual locality. Percentages (rounded to nearest whole number) greater than 0% and less than 100% are plotted in each square.

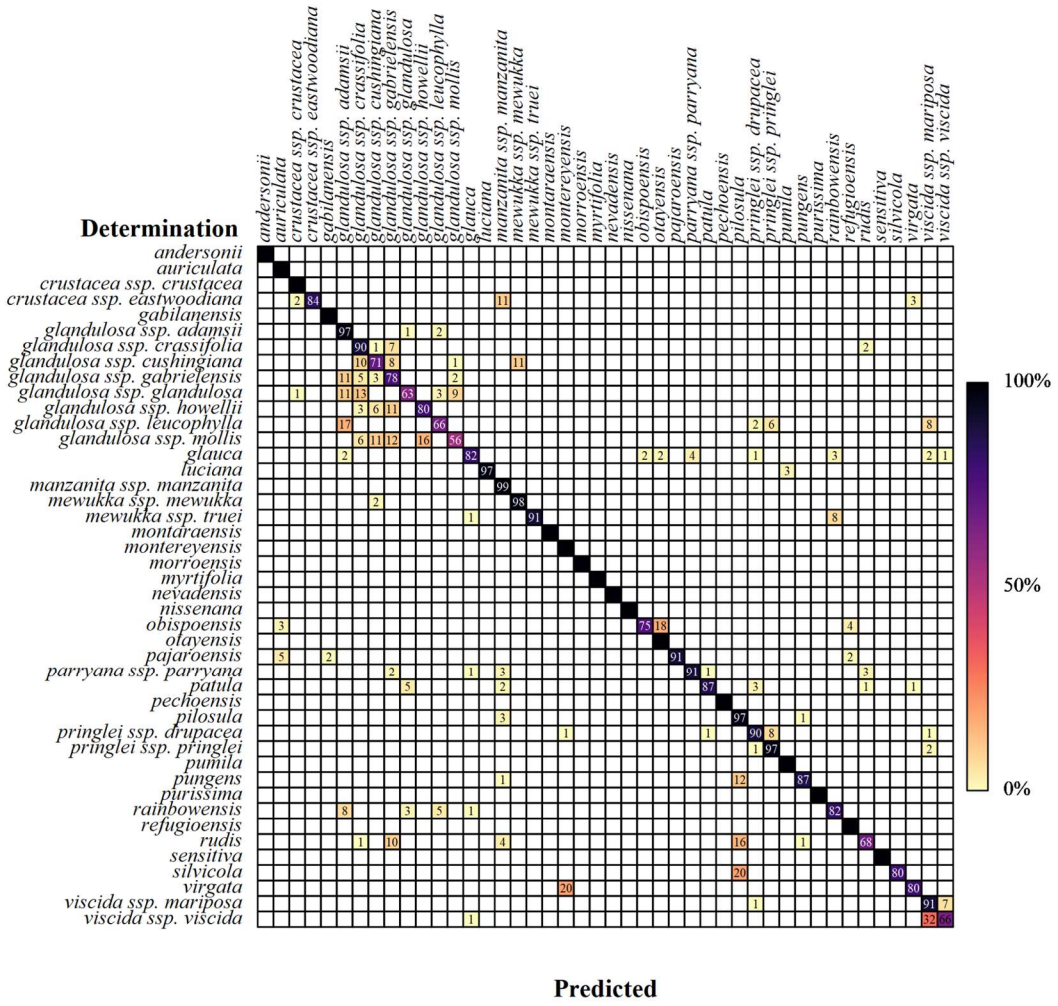


Fig. 2.8. Results of the linear discriminant analysis trained on species and subspecies determinations, using a version of the dataset where the stem and leaf hair variables are simplified to a binary presence/absence variable for each organ (leaf or stem). Colors are scaled to show accuracy of LDA assignments for a given species or subspecies, compared against the human determination. Percentages (rounded to nearest whole number) greater than 0% and less than 100% are plotted in each square.

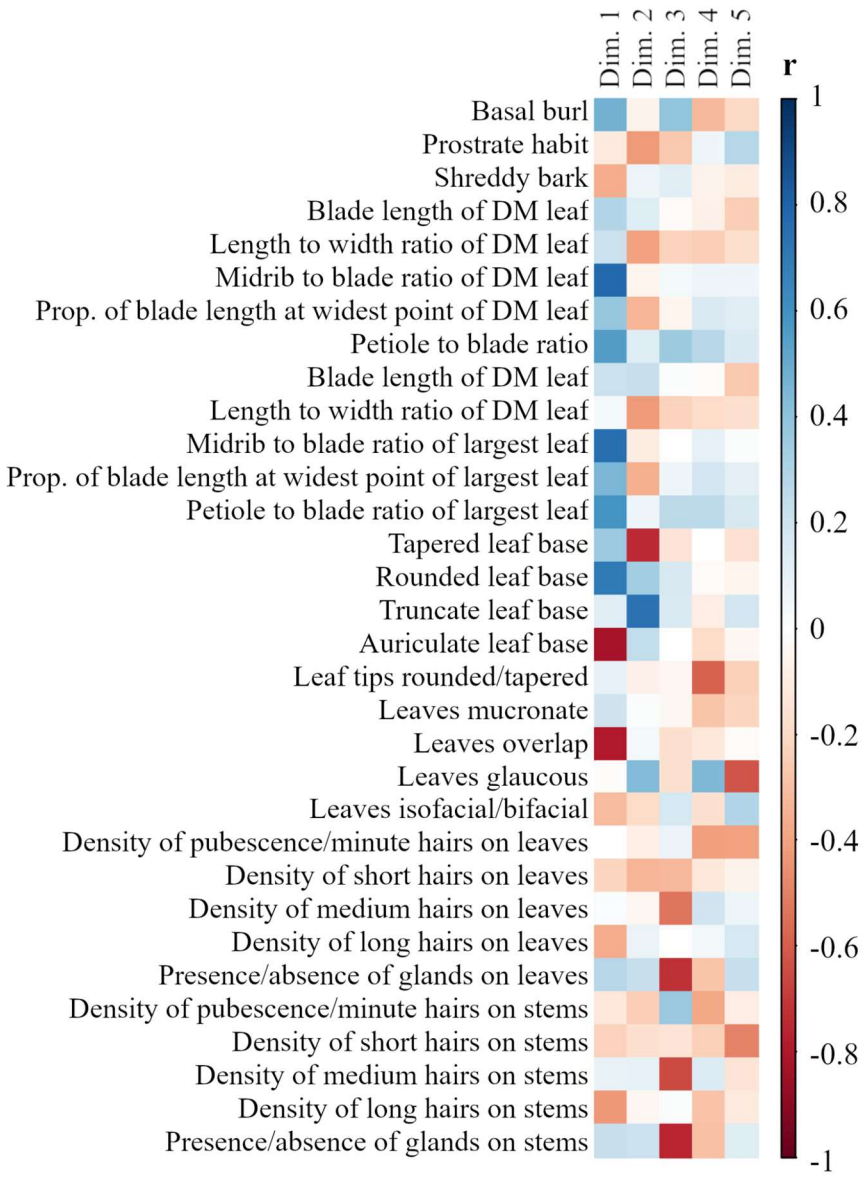


Fig. 2.9. Correlation between individual variables and the five variables used in the NMDS calculated on the entire dataset. Correlation values represent Pearson correlation coefficients. Scale at right shows colors that correspond to negative and positive values of the correlation coefficient. “DM” is an abbreviation for “distal mature”.

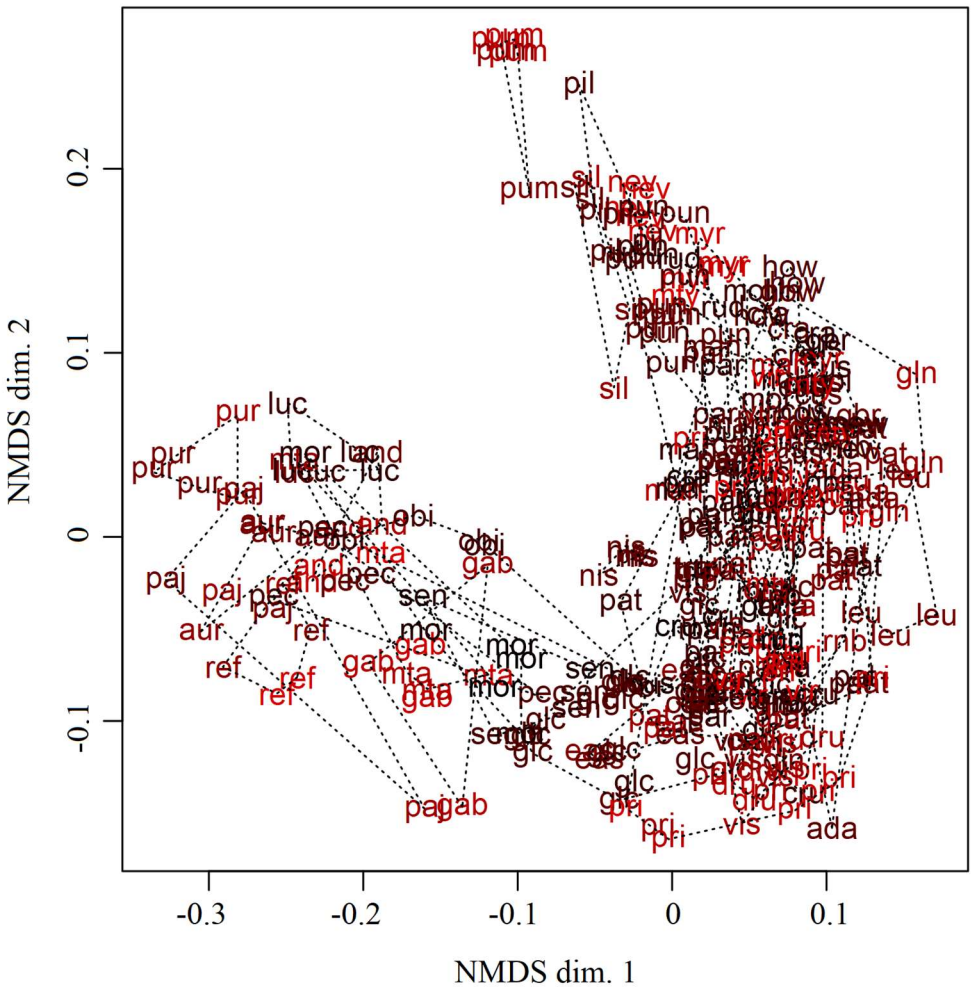


Fig. 2.10. NMDS biplot of all samples, and all taxa included. Dimension 3 is encoded as color of text points. Points/samples plotted as three-letter abbreviations for species (see Table 2.1 for taxa abbreviated). Polygons drawn around samples are convex hulls, which are the smallest area enclosing all samples from a given taxon.

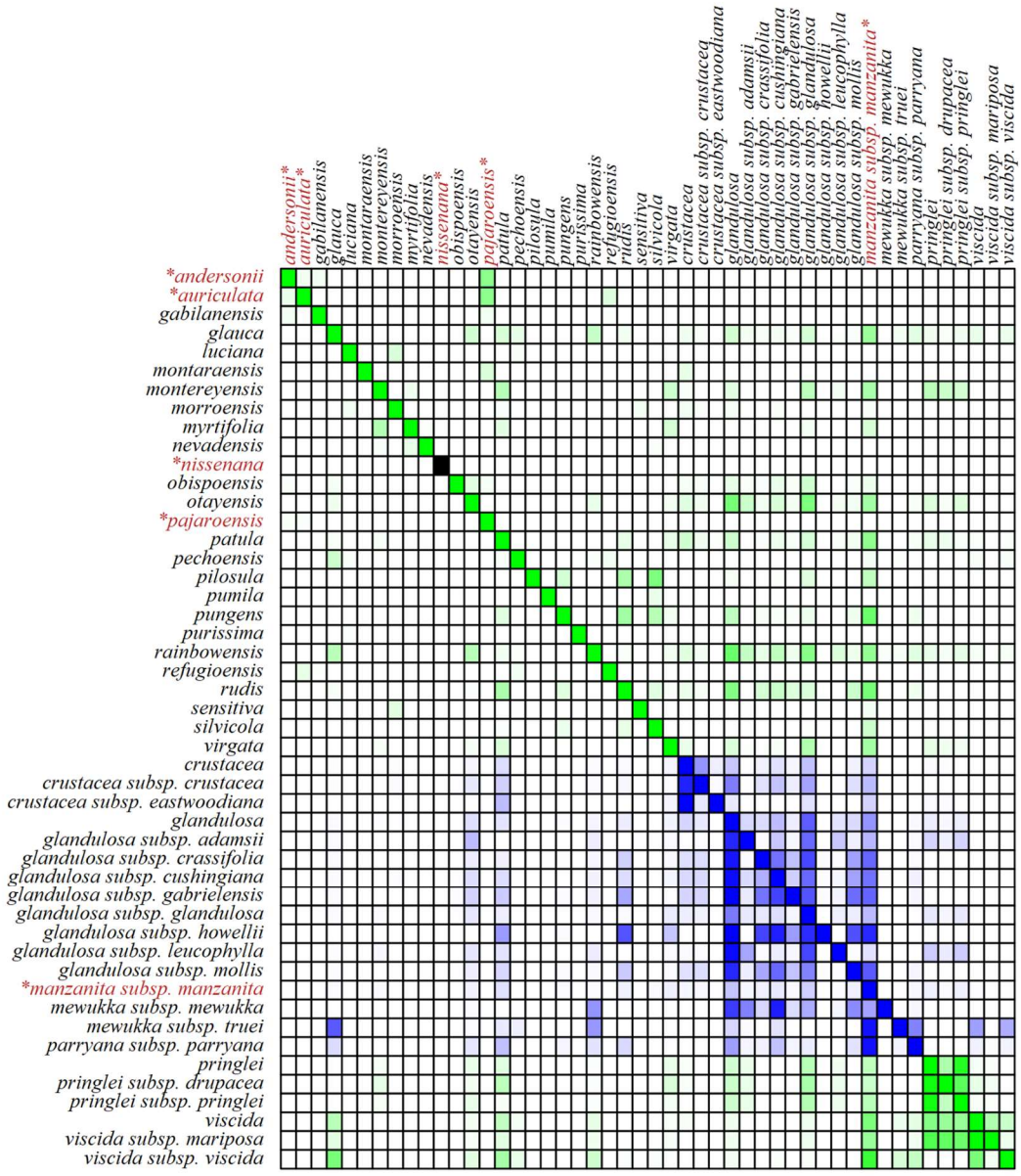

*andersonii
*auriculata
gabilanensis
glauca
luciana
montaraensis
montereyensis
morroensis
myrtifolia
nevadensis
*nissenana
obispoensis
otayensis
*pajaroensis
patula
pechoensis
pilosula
pumila
pungens
purissima
rainbowensis
refugioensis
rudis
sensitiva
silvicola
virgata
crustacea
crustacea subsp. crustacea
crustacea subsp. eastwoodiana
glandulosa
glandulosa subsp. adamsii
glandulosa subsp. crassifolia
glandulosa subsp. cushingiana
glandulosa subsp. gabrielensis
glandulosa subsp. glandulosa
glandulosa subsp. howellii
glandulosa subsp. leucophylla
glandulosa subsp. mollis
*manzanita subsp. manzanita
mewukka subsp. mewukka
mewukka subsp. truei
parryana subsp. parryana
pringlei
pringlei subsp. drupacea
pringlei subsp. pringlei
viscida
viscida subsp. mariposa
viscida subsp. viscida

Fig. 2.11. Matrix showing the hypervolume overlap of each taxon (rows) with every other taxon (columns). Taxa are sorted to show monotypic species in the higher rows, and polytypic species and individual subspecies in the lower rows, sorted alphabetically within each of those categories. Colors are scaled by proportion of overlap, with pure white being zero overlap, and deepest blue or green (e.g. on the diagonal) being 100% overlap. Taxa plotted in green are diploid, and taxa plotted in blue are tetraploid. One species (*A. nissenana*), which was the only taxon with zero calculated overlap with any other taxon, is plotted in black, and is a diploid. The names of several taxa mentioned as examples in the text are marked with asterisks and colored red for ease of reference with the text.

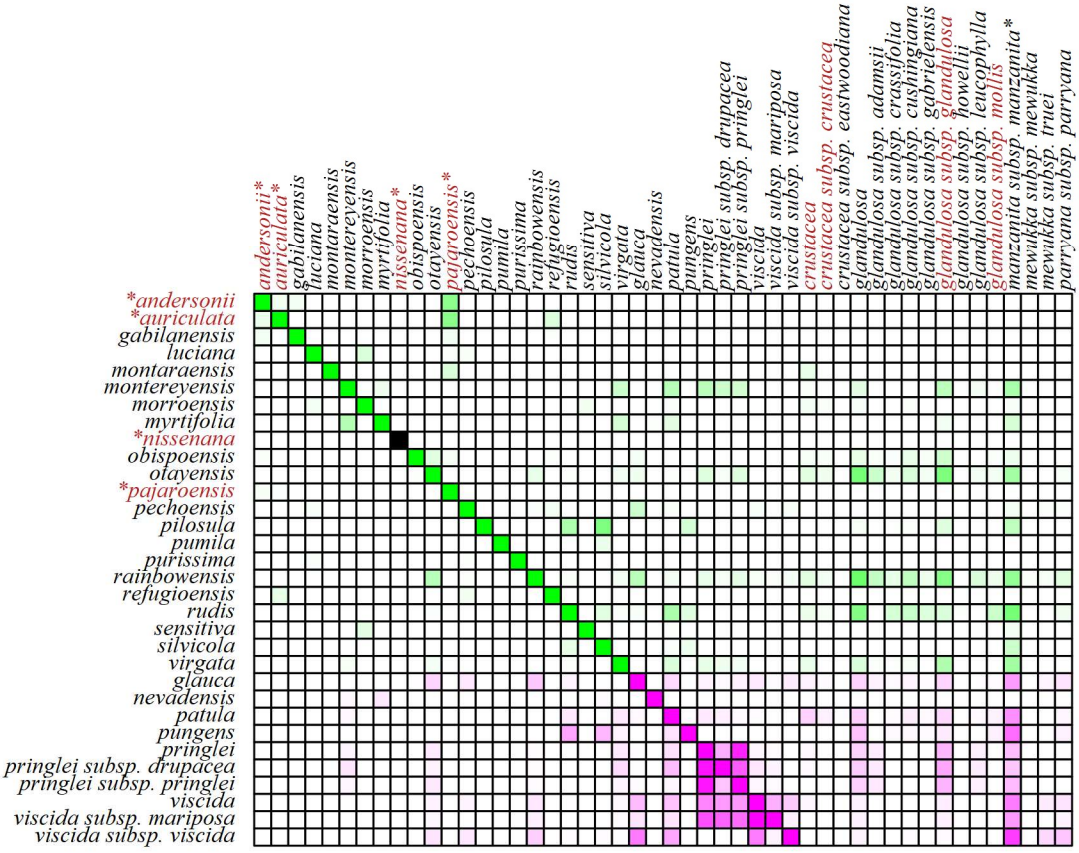

*andersonii
*auriculata
gabilanensis
luciana
montaraensis
montereyensis
morroensis
myrtifolia
*nissenana
obispoensis
otayensis
*pajaroensis
pechoensis
pilosula
pumila
purissima
rainbowensis
refugioensis
rudis
sensitiva
silvicola
virgata
glauca
nevadensis
patula
pungens
pringlei
pringlei subsp. drupacea
pringlei subsp. pringlei
viscida
viscida subsp. mariposa
viscida subsp. viscida
crustacea
crustacea subsp. crustacea
crustacea subsp. eastwoodiana
glandulosa
glandulosa subsp. adamsii
glandulosa subsp. crassifolia
glandulosa subsp. cushingiana
glandulosa subsp. gabrielensis
glandulosa subsp. glandulosa
glandulosa subsp. howellii
glandulosa subsp. leucophylla
glandulosa subsp. mollis
manzanita subsp. manzanita*
mewukka subsp. mewukka
mewukka subsp. truei
parryana subsp. parryana

Fig. 2.12. Matrix showing the hypervolume overlap of each diploid taxon (rows), with all other taxa (diploid and tetraploid), displayed as columns. Taxa are sorted in order of (1) narrow endemic species (which are all diploid), (2) widespread diploid and monotypic species, and (3) polytypic species and individual subspecies, sorted alphabetically within each of those categories. Columns are sorted in the same way, but with tetraploid taxa included at the right end. Colors are scaled by proportion of overlap, with pure white being zero overlap, and deepest blue or green (e.g. on the diagonal) being 100% overlap. Taxa plotted in green are narrow endemic species, and taxa plotted in magenta are widespread diploid taxa. One species (*A. nissenana*), which was the only taxon with zero calculated overlap with any other taxon, is plotted in black at its diagonal position. The names of several taxa mentioned as examples in the text are marked with asterisks and colored red for ease of reference with the text.

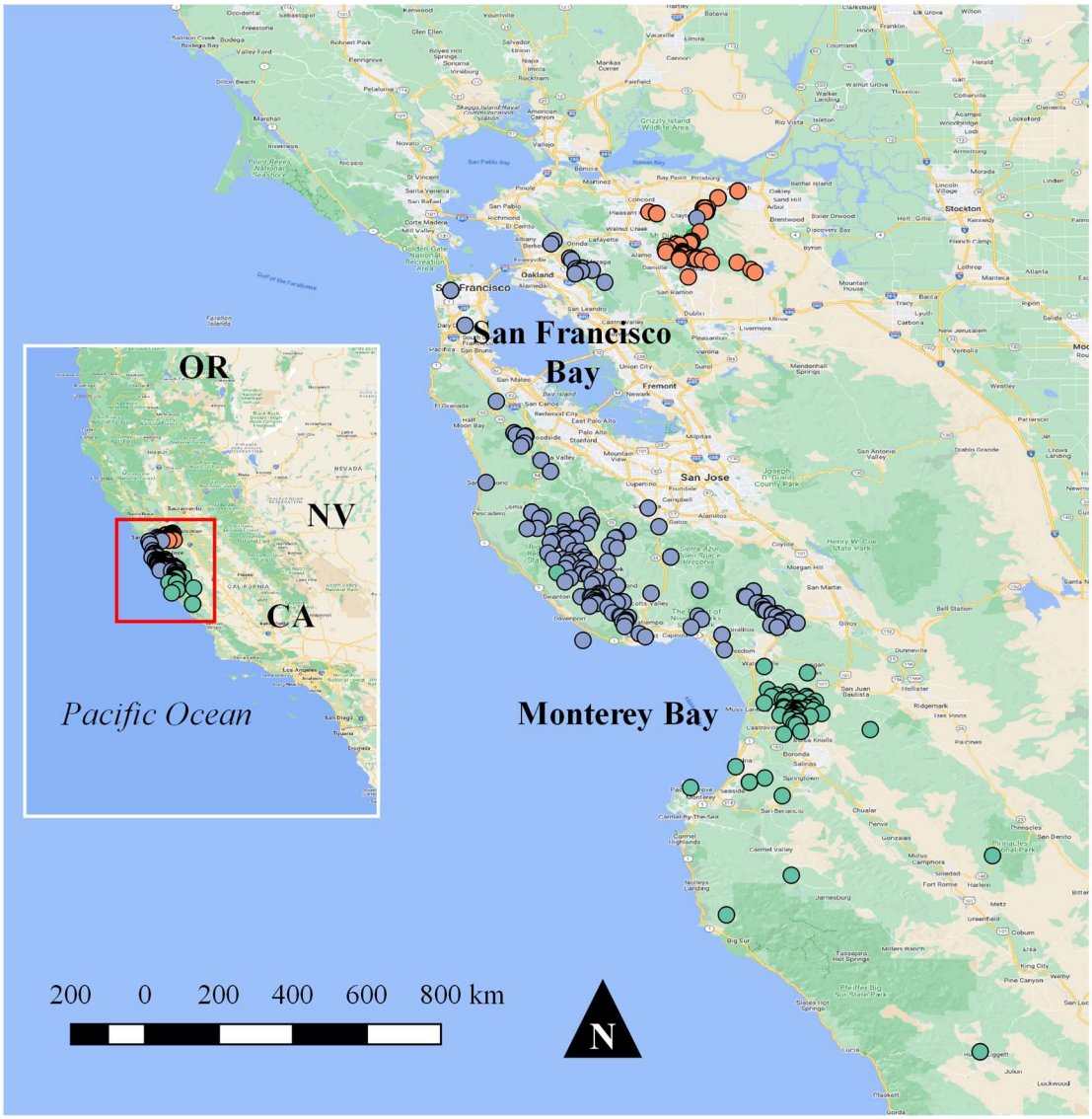


Fig. 2.13. Map of all locations for geo-referenced herbarium collections of *A. pajaroensis* (green points), *A. auriculata* (orange points), and *A. andersonii* (purple points), showing the approximate ranges of each species. Data downloaded from California Consortium of Herbaria database (www.cch2.org). Smaller map in the inset shows the area of the larger map. Basemap from Google. Map made in QGIS (https://qgis.org/).

Table 2.1. Taxa included in this study, in alphabetical order, listed with their ploidy, range classification used here in analyses (narrow endemic, or widespread), and the three-letter code used in Fig. 2.10.

Taxon	Ploidy	Range	Three-letter code
A. andersonii A. Gray	Diploid	Narrow	And
A. auriculata Eastw.	Diploid	Narrow	Aur
A. crustacea Eastw. subsp. *eastwoodiana* (P.V. Wells)	Tetraploid	Widespread	Eas
A. *crustacea* Eastw. subsp. *crustacea*	Tetraploid	Widespread	Cru
*A. gabilanensi*s V.T. Parker & M.C. Vasey	Diploid	Narrow	Gab
A. glandulosa Eastw. subsp. *leucophylla* J.E. Keeley, M.C. Vasey & V.T. Parker	Tetraploid	Widespread	Leu
A. glandulosa Eastw. subsp. *crassifolia* (Jeps.) P.V. Wells	Tetraploid	Widespread	Cra
A. glandulosa Eastw. subsp. *cushingiana* (Eastw.) J.E. Keeley, M.C. Vasey & V.T. Parker	Tetraploid	Widespread	Cus
A. glandulosa Eastw. subsp. *glandulosa*	Tetraploid	Widespread	Gln
A. glandulosa Eastw. subsp. *mollis* (J.E. Adams) P.V. Wells	Tetraploid	Widespread	Mol
A. glandulosa Eastw. subsp. *adamsii* (Munz) Munz	Tetraploid	Widespread	Ada
A. glandulosa Eastw. subsp. *gabrielensis* (P.V. Wells) J.E. Keeley, M.C. Vasey & V.T. Parker	Tetraploid	Widespread	Gbr
A. glandulosa Eastw. subsp. *howellii* (Eastw.) P.V. Wells	Tetraploid	Widespread	How
A. glauca Lindl.	Diploid	Widespread	Glc
A. luciana P.V. Wells	Diploid	Narrow	Luc
A. manzanita Parry subsp. *manzanita*	Tetraploid	Widespread	Man
A. mewukka Merriam subsp. *mewukka*	Tetraploid	Widespread	Mew
A. mewukka Merriam subsp. *truei* (W. Knight) P.V. Wells	Tetraploid	Widespread	Tru
A. montaraensis Roof	Diploid	Narrow	Mta
A. montereyensis Hoover	Diploid	Narrow	Mty
A. morroensis Wiesl. & B. Schreib.	Diploid	Narrow	Mor

A. myrtifolia Parry	Diploid	Narrow	Myr
A. nevadensis A. Gray	Diploid	Widespread	Nev
A. nissenana Merriam	Diploid	Narrow	Nis
A. obispoensis Eastw.	Diploid	Narrow	Obi
A. otayensis Wiesl. & B. Schreib.	Diploid	Narrow	Ota
A. pajaroensis (J.E. Adams) J.E. Adams	Diploid	Narrow	Paj
A. parryana Lemmon subsp. *parryana*	Tetraploid	Widespread	Par
A. patula Greene	Diploid	Widespread	Pat
A. pechoensis (Abrams) Dudley ex Eastw.	Diploid	Narrow	Pec
A. pilosula Jeps. & Wiesl.	Diploid	Narrow	Pil
A. pringlei Parry subsp. *drupacea* (Parry) P.V. Wells	Diploid	Widespread	Dru
A. pringlei Parry subsp. *pringlei*	Diploid	Widespread	Pri
A. pumila Nutt.	Diploid	Narrow	Pum
A. pungens Kunth	Diploid	Widespread	Pun
A. purissima P.V. Wells	Diploid	Narrow	Pur
A. rainbowensis J.E. Keeley & Massihi	Diploid	Narrow	Rnb
A. refugioensis Gankin	Diploid	Narrow	Ref
A. rudis Jeps. & Wiesl.	Diploid	Narrow	Rud
A. sensitiva Jeps.	Diploid	Narrow	Sen
A. silvicola Jeps. & Wiesl.	Diploid	Narrow	Sil
A. virgata Eastw.	Diploid	Narrow	Vir
A. viscida Parry subsp. *mariposa* (Dudley) P.V. Wells	Diploid	Widespread	Mar
A. viscida Parry subsp. *viscida*	Diploid	Widespread	Vis

Table 2.2. List of the 32 morphological variables used in the analyses, divided into categories based on organs involved, or characters concerned, and providing information on the type of data variable each is (quantitative, binary, or ordinal).

Categories	Variable name	Variable type
Growth form	Basal burl	Binary
	Prostrate habit	Binary
Bark	Shreddy bark	Binary
Various leaf variables relating to dimension, shape, color, etc.	Blade length of distal mature leaf	Quantitative
	Length to width ratio of distal mature leaf	Quantitative
	Midrib to blade ratio of distal mature leaf	Quantitative
	Proportion of blade length at widest point on distal mature leaf	Quantitative
	Petiole to blade ratio of distal mature length	Quantitative
	Blade length of largest leaf	Quantitative
	Length to width ratio of largest leaf	Quantitative
	Midrib to blade ratio of largest leaf	Quantitative
	Proportion of blade length at widest point on largest leaf	Quantitative
	Petiole to blade ratio of largest length	Quantitative
	Tapered leaf bases	Binary
	Rounded leaf bases	Binary
	Truncate leaf bases	Binary
	Auriculate leaf bases	Binary
	Leaf tips rounded/tapered	Binary

	Leaves mucronate	Binary
	Leaves overlap	Binary
	Leaves glaucous	Binary
	Leaves isofacial/bifacial	Binary
Hair variables	Density of pubescence on leaves	Ordinal
	Density of short hairs on leaves	Ordinal
	Density of medium hairs on leaves	Ordinal
	Density of long hairs on leaves	Ordinal
	Presence/absence of glands on leaves	Binary
	Density of pubescence on stems	Ordinal
	Density of short hairs on stems	Ordinal
	Density of medium hairs on stems	Ordinal
	Density of long hairs on stems	Ordinal
	Stem glands	Binary

Table 2.3. Formulae used to calculate several leaf shape variables. See Fig. 2.3 for visual illustration of the raw data variables used in these formulae.

Calculated variable	**Formula**
Length to width ratio	= (Blade length) / (Blade width)
Midrib to blade ratio	= (Midrib length) / (Blade length)
Proportion of blade length at widest point	= (Distance to widest point) / (Blade length)
Petiole to blade ratio	= (Petiole length) / (Blade length)

LITERATURE CITED:

Baldwin, B. G., G. H. Douglas, D. J. Keil, R. Patterson, T. J. Rosatti, and D. H. Wilken eds. . 2012. The Jepson Manual: Vascular Plants of California, 2nd Edition. University of California Press.

Blonder, B., W. C. from Cecina Babich Morrow, D. J. Harris, S. Brown, G. Butruille, A. Laini, and D. Chen. 2023. hypervolume: High Dimensional Geometry, Set Operations, Projection, and Inference Using Kernel Density Estimation, Support Vector Machines, and Convex Hulls.

Boykin, L. M., M. C. Vasey, V. Thomas Parker, and R. Patterson. 2005. Two lineages of *Arctostaphylos* (Ericaceae) identified using the internal transcribed spacer (ITS) region of the nuclear genome. *Madrono* 52: 139–147.

Burge, D. O., V. T. Parker, M. Mulligan, and C. García Valderamma. 2018. Conservation Genetics of the Endangered Del Mar Manzanita (*Arctostaphylos glandulosa* subsp. *crassifolia*) Based On Rad Sequencing Data. *Madroño* 65: 117–130.

Clausen, R. T. 1941. On the use of the terms 'subspecies' and 'variety'. *Rhodora* 43: 157–167.

Dobzhansky, T. 1953. Natural hybrids of two species of *Arctostaphylos* in the Yosemite region of California. *Heredity* 7: 73–80.

Ellstrand, N. C., J. M. Lee, J. E. Keeley, and S. C. Keeley. 1987. Ecological isolation and introgression: biochemical confirmation of introgression in an *Arctostaphylos* (Ericaceae) population. *Acta Oecologica, Oecologica Plantarum.* 8: 299–308.

Fisher, R. A. 1936. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics* 7: 179–188.

Freedman, D., R. Pisani, and R. Purves. 2007. Statistics (international student edition).

Gankin, R. 1967. A new species of *Arctostaphylos* from Santa Barbara County, California. *The Four Seasons*.

Gottlieb, L. D. 1968. Hybridization between *Arctostaphylos viscida* and *A. canescens* in Oregon. *Brittonia* 20: 83.

Gower, J. C. 1971. A General Coefficient of Similarity and Some of Its Properties. *Biometrics* 27: 857–871.

Hileman, L. C., M. C. Vasey, and V. T. Parker. 2001. Phylogeny and Biogeography of the Arbutoideae (Ericaceae): Implications for the Madrean-Tethyan Hypothesis. *Systematic botany* 26: 131–143.

Huang, Y., G. R. Morrison, A. Brelsford, J. Franklin, D. D. Jolles, J. E. Keeley, V. T. Parker, et al. 2020. Subspecies differentiation in an enigmatic chaparral shrub species. *American journal of botany* 107: 923–940.

Kauffmann, M. E., T. Parker, and M. Vasey. 2015. Field Guide to Manzanitas: California, North America, and Mexico. Backcountry Press.

Kauffmann, M. E., T. Parker, and M. Vasey. 2021. Field Guide to Manzanitas: California, North America, and Mexico, Second Edition. Backcountry Press.

Keeley, J. E. 1976. Morphological Evidence of Hybridization between *Arctostaphylos glauca* and *Arctostaphylos pungens* (Ericaceae). *Madroño* 23: 427–434.

Keeley, J. E., and A. Massihi. 1994. A*rctostaphylos glauca*, a New Burl-Forming Manzanita from Northern San Diego County, California. *Madroño* 41: 1–12.

Keeley, J. E., A. Massihi, J. D. Rodríguez, and S. A. Hirales. 1997. *Arctostaphylos incognita*, a New Species and its Phenetic Relationship to Other Manzanitas of Baja California. *Madroño* 44: 137–150.

Kruskal, J. B. 1964. Nonmetric multidimensional scaling: A numerical method. *Psychometrika* 29: 115–129.

MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. L. M. L. Cam, and J. Neyman [eds.],. University of California Press, Berkeley.

Minchin, P. R. 1987. An Evaluation of the Relative Robustness of Techniques for Ecological Ordination. *Vegetatio* 69: 89–107.

Nason, J. D., N. C. Ellstrand, and M. L. Arnold. 1992. Patterns of Hybridization and Introgression in Populations of Oaks, Manzanitas, and Irises. *American journal of botany* 79: 101–111.

Oksanen, J., G. L. Simpson, F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, et al. 2022. vegan: Community Ecology Package.

Parker, V. T., C. Y. Rodriguez, G. Wechsler, and M. C. Vasey. 2020. Allopatry, hybridization, and reproductive isolation in *Arctostaphylos*. *American journal of botany* 107: 1798–1814.

Parker, V. T., and M. C. Vasey. 2004. *Arctostaphylos gabilanensi*s (Ericaceae), a Newly Described Auriculate-Leaved Manzanita from the Gabilan Mountains, California. *Madroño* 51: 322–325.

Parker, V. T., and M. C. Vasey. 2016. Two New Subspecies of *Arctostaphylos* (Ericaceae) From California and Implications For Understanding Diversification In This Genus. *Madrono; a West American journal of botany* 63: 283–291.

Parker, V. T., M. C. Vasey, and J. E. Keeley. 2012. *Arctostaphylos*, in Jepson Flora Project (eds.). *Jepson eFlora*. Website https://ucjeps.berkeley.edu/eflora/ [accessed 5 September 2023].

Parker, V. T., M. C. Vasey, and J. E. Keeley. 2007. Taxonomic Revisions in the Genus *Arctostaphylos* (Ericaceae). *Madrono; a West American journal of botany* 54: 148–155.

Pearson, K. 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2: 559–572.

Persoon, C. H. 1805. Synopsis Plantarum, seu Enchiridium Botanicum, complectens enumerationem systematicam Specierum hucusque cognitarum. Lutetiorum.

Raven, P. H. 1969. Supplement to A California Flora by Philip A. Munz. *Madrono; a West American journal of botany* 20: 239–240.

R Core Team. 2023. R: A Language and Environment for Statistical Computing.

Schierenbeck, K. A., G. L. Stebbins, and R. W. Patterson. 1992. Morphological and cytological evidence for polyphyletic allopolyploidy in *Arctostaphylos mewukka*. *Syst. Evol.* 179: 187–205.

Schmid, R., T. E. Mallory, and J. M. Tucker. 1968. Biosystematic Evidence for Hybridization Between *Arctostaphylos nissenana* and *A. viscida*. *Brittonia* 20: 34.

Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* 9: 671–675.

Serkanic, S., V. T. Parker, and K. Schierenbeck. 2021. Complexity in Polyploid Species Origin and Establishment: *Arctostaphylos mewukka* (Ericaceae). *Systematic botany* 46: 666–677.

Wahlert, G. A., V. T. Parker, and M. C. Vasey. 2009. A Phylogeny of *Arctostaphylos* (Ericaceae) Inferred from Nuclear Risomal ITS Sequences. *Journal of the Botanical Research Institute of Texas* 3: 673–682.

Appendix 2.1: Herbarium specimens included in morphological dataset.

Determination,County,Country,Voucher deposited,Collection number,

Arctostaphylos andersonii A. Gray, Morrison 107, Morrison 108, Morrison 109, Morrison 111, Morrison 112, ***Arctostaphylos auriculata*** **Eastw.**, Abbo 061, Abbo 062, Abbo 063, Abbo 065, Abbo 066, ***Arctostaphylos crustacea*** **Eastw. subsp.** ***eastwoodiana*** **(P.V. Wells)**, Huang 718-26, Huang 718-27, Huang 718-28, Huang 718-29, Huang 718-30, ***Arctostaphylos crustacea*** **Eastw. subsp.** ***crustacea***, Morrison 018, Morrison 019, Morrison 093, Morrison 094, Morrison 207, Morrison 232, ***Arctostaphylos gabilanensis*** **V.T. Parker & M.C. Vasey**, Morrison 245, Morrison 246, Morrison 247, Morrison 248, Morrison 249, ***Arctostaphylos glandulosa*** **Eastw. subsp.** ***leucophylla*** **J.E. Keeley, M.C. Vasey & V.T. Parker**, Huang 317-07, Huang 317-27, Huang 317-34, Huang 317-37, Huang 318-14, Litt 318-17, ***Arctostaphylos glandulosa*** **Eastw. subsp.** ***crassifolia*** **(Jeps.) P.V. Wells**, Huang 618-02, Huang 618-04, Huang 618-05, Sanders 42680, Sanders 42682, Sanders 42683, ***Arctostaphylos glandulosa*** **Eastw. subsp.** ***cushingiana*** **(Eastw.) J.E. Keeley, M.C. Vasey & V.T. Parker**, Huang 718-07, Huang 718-08, Huang 718-09, Huang 718-36, Huang 718-37, Huang 718-39, Huang 818-01, Huang 818-02, Huang 818-03, Huang 818-04, Huang 818-05, Morrison 216, Morrison 217, Morrison 218, Morrison 225, ***Arctostaphylos glandulosa*** **Eastw. subsp.** ***glandulosa***, Huang 718-33, Huang 818-14, Litt 317-09, Litt 317-19, Litt 317-22, Morrison 003, Morrison 226, Stickrod 070, ***Arctostaphylos glandulosa*** **Eastw. subsp. mollis (J.E. Adams) P.V. Wells**, Huang 318-08, Huang 318-10, Litt 317-02, Morrison 005, Morrison 011, Morrison 012, ***Arctostaphylos glandulosa*** **Eastw. subsp.** ***adamsii*** **(Munz) Munz** , Litt 317-04, Litt 317-05, Litt 317-29, Litt 317-31, Litt 317-35, Litt 317-36, ***Arctostaphylos glandulosa*** **Eastw. subsp.** ***gabrielensis*** **(P.V. Wells) J.E. Keeley, M.C. Vasey & V.T. Parker**, Litt 417-01, Litt 417-02, Litt 417-03, Morrison 126, Morrison 128, Sanders 43384, Sanders 43386, Sanders 43387, ***Arctostaphylos glandulosa*** **Eastw. subsp.** ***howellii*** **(Eastw.) P.V. Wells**, Morrison 258, Morrison 261, Morrison 262, Morrison 263, Morrison 266, ***Arctostaphylos glauca*** **Lindl.**, Abbo 057, Elvin 2899, Huang 318-0402, Levine 77, Morrison 013, Morrison 253, Morrison 256, Morrison 259, Morrison 260, Morrison 264, Morrison 265, Morrison 267, Morrison 275, Morrison 278, Morrison 280, Morrison 282, Morrison 283, Morrison 296, Morrison 297, Morrison 298, Morrison 301, Morrison 336, Morrison 338, Morrison 356, Morrison 359, Morrison 380, Morrison 381, Morrison 382, Morrison 398, Morrison 400, Morrison 413a, Provance 411-29, Sanders 43388, Swinney 13253, Swinney 17558, ***Arctostaphylos luciana*** **P.V. Wells**, Morrison 029, Morrison 030, Morrison 031, Morrison 032, Morrison 033, Morrison 034, ***Arctostaphylos manzanita*** **Parry subsp.** ***manzanita***, Morrison 143, Morrison 144, Morrison 158, Morrison 159, Morrison 244, Stickrod 243, ***Arctostaphylos mewukka*** **Merriam subsp.** ***mewukka***, Morrison 138, Morrison 139, Morrison 164, Morrison 165, Morrison 169, ***Arctostaphylos mewukka*** **Merriam subsp.** ***truei*** **(W. Knight)**

P.V., Morrison 163, Morrison 166, Morrison 167, Morrison 170, ***Arctostaphylos montaraensis* Roof**, Morrison 227, Morrison 228, Morrison 229, Morrison 230, Morrison 231, ***Arctostaphylos montereyensis* Hoover**, Morrison 072, Morrison 073, Morrison 074, Morrison 075, Morrison 076, ***Arctostaphylos morroensis* Wiesl. & B. Schreib.**, Morrison 041, Morrison 042, Morrison 043, Morrison 044, Morrison 045, Morrison 046, ***Arctostaphylos myrtifolia* Parry**, Morrison 146, Morrison 147, Morrison 148, Morrison 149, Morrison 150, ***Arctostaphylos nevadensis* A. Gray**, Morrison 176, Morrison 177, Morrison 178, Morrison 179, Morrison 180, ***Arctostaphylos nissenana* Merriam**, Morrison 151, Morrison 152, Morrison 153, Morrison 154, Morrison 155, ***Arctostaphylos obispoensis* Eastw.**, Morrison 035, Morrison 036, Morrison 037, Morrison 038, Morrison 039, ***Arctostaphylos otayensis* Wiesl. & B. Schreib.**, Buehlman 035, Buehlman 036, Huang 818-09, Huang 818-10, Huang 818-11, Huang 818-12, ***Arctostaphylos pajaroensis* (J.E. Adams) J.E. Adams**, Morrison 369, Morrison 371, Morrison 372, Morrison 373, Morrison 379, ***Arctostaphylos parryana* Lemmon subsp. *parryana***, Huang 718-44, Huang 718-45, Huang 718-46, Huang 718-47, Huang 718-48, Morrison 114, Morrison 115, Morrison 116, Morrison 117, Morrison 118, Morrison 292, Morrison 294, ***Arctostaphylos patula* Greene**, Levine 20, Levine 21, Levine 22, Levine 23, Levine 24, Morrison 119, Morrison 120, Morrison 121, Morrison 122, Morrison 123, Morrison 124, Morrison 172, Morrison 173, Morrison 174, Morrison 175, Morrison 181, Morrison 182, Morrison 183, Morrison 184, Morrison 185, Morrison 486, Morrison 487, Morrison 488, Morrison 489, Morrison 490, Morrison 491, Morrison 492, Morrison 493, Morrison 494, Morrison 495, Morrison 496, Morrison 497, Morrison 498, Morrison 499, ***Arctostaphylos pechoensis* (Abrams) Dudley ex Eastw.**, Morrison 020, Morrison 021, Morrison 022, Morrison 388, Morrison 390, ***Arctostaphylos pilosula* Jeps. & Wiesl.**, Morrison 023, Morrison 024, Morrison 025, Morrison 026, Morrison 027, ***Arctostaphylos pringlei* Parry subsp. *drupacea* (Parry) P.V. Wells**, Bodden 006, Sanders 33186, Kelly 12B, Morrison 404, Morrison 405, Morrison 406, Morrison 407, Morrison 408, Morrison 412, ***Arctostaphylos pringlei* Parry subsp. *pringlei***, Morrison 421, Morrison 422, Morrison 423, Morrison 424, Morrison 425, Morrison 431, Morrison 432, Morrison 433, Morrison 434, Morrison 435, Morrison 451, Morrison 452, Morrison 453, Morrison 454, Morrison 455, Morrison 456, Morrison 457, Morrison 458, Morrison 459, Morrison 460, Morrison 466, Morrison 467, Morrison 468, Morrison 469, Morrison 470, Morrison 476, Morrison 477, Morrison 478, Morrison 479, Morrison 480, ***Arctostaphylos pumila* Nutt.**, Morrison 081, Morrison 082, Morrison 083, Morrison 084, Morrison 085, ***Arctostaphylos pungens* Kunth**, Abbo 017, Abbo 018, Abbo 019, Abbo 020, Abbo 021, Morrison 416, Morrison 417, Morrison 418, Morrison 419, Morrison 420, Morrison 481, Morrison 482, Morrison 483, Morrison 484, Morrison 485, ***Arctostaphylos purissima* P.V. Wells**, Huang 718-10, Huang 718-11, Huang 718-12, Huang 718-13, Huang 718-14, ***Arctostaphylos rainbowensis* J.E. Keeley & Massihi**, Abbo 002, Abbo 005, Abbo 006, Abbo 010, Abbo 011, Abbo 015, ***Arctostaphylos refugioensis* Gankin**, Huang 718-01, Huang 718-02, Huang 718-03, Huang 718-04, Huang 718-05, ***Arctostaphylos rudis* Jeps. & Wiesl.**, Huang 718-15, Huang 718-16, Huang 718-17, Huang 718-18, Huang 718-19, Huang 718-20, ***Arctostaphylos sensitiva* Jeps.**, Morrison 211, Morrison 212, Morrison 213, Morrison 214, Morrison 215, ***Arctostaphylos silvicola* Jeps. & Wiesl.**, Morrison 101, Morrison 102, Morrison 103, Morrison 104, Morrison 105, ***Arctostaphylos virgata* Eastw.**,

Morrison 234, Morrison 235, Morrison 236, Morrison 237, Morrison 238, ***Arctostaphylos viscida* Parry subsp. *mariposa* (Dudley) P.V. Wells**, Morrison 130, Morrison 132, Morrison 133, Morrison 141, Morrison 156, Morrison 157, ***Arctostaphylos viscida* Parry subsp. *viscida***, Morrison 168, Morrison 171,

Chapter 3: Subspecies differentiation in an enigmatic chaparral shrub species.

ABSTRACT:

Premise: Delimiting biodiversity units is difficult in organisms in which differentiation is obscured by hybridization, plasticity, and other factors that blur phenotypic boundaries. Such work is more complicated when the focal units are subspecies, the definition of which has not been broadly explored in the era of modern genetic methods. Eastwood manzanita (*Arctostaphylos glandulosa* Eastw.), is a widely distributed and morphologically complex chaparral shrub species with much subspecific variation that has proven challenging to categorize. Currently ten subspecies are recognized, however, many of them are not geographically segregated, and morphological intermediates are common. Subspecies delimitation is of particular importance in this species, as two of the subspecies are rare. The goal of this study was to apply an evolutionary definition of subspecies to characterize structure within Eastwood manzanita.

Methods: We used publicly-available geospatial environmental data and reduced-representation genome sequencing to characterize environmental and genetic differentiation among subspecies. In addition, we tested whether subspecies could be differentiated by environmentally-associated genetic variation.

Key results: Our analyses do not show genetic differentiation among subspecies of Eastwood manzanita, with the exception of one of the two rare subspecies. In addition, our environmental analyses did not show ecological differentiation, though limitations of the analysis prevent strong conclusions.

Conclusion: Genetic structure within Eastwood manzanita does not correspond to current subspecies circumscriptions but rather reflects geographic distribution. Our study suggests that

subspecies concepts need to be reconsidered in long-lived plant species, especially in the age of next generation sequencing.

INTRODUCTION:

A comprehensive understanding of biodiversity is crucial for ecologists, conservationists, land managers, policy makers, and others whose work depends on the accurate recognition of biodiversity units (Regan, Colyvan, and Burgman 2002; Mace 2004; Renwick et al. 2017; Keller et al. 2011). Given current rates of extinction in plants, discovering, identifying, and delimiting plant biodiversity units is more critical than ever (Thomas et al. 2004; Pimm et al. 2014; Grooten and Almond 2018). However, drawing boundaries around taxonomic units is difficult in groups in which populations are poorly differentiated and/or vary along a continuous cline (Bradburd, Coop, and Ralph 2018; Razkin et al. 2017; Carstens et al. 2013; Jörger and Schrödl 2013). A number of factors can blur boundaries, including hybridization, introgression, local adaptation, and phenotypic plasticity (Grant 1972; R. G. Harrison and Larson 2014). These factors can result in highly variable populations and species complexes that include a range of phenotypes that cannot be easily divided into distinct groups using standard genetic, morphological, or ecological criteria.

Eastwood manzanita (*Arctostaphylos glandulosa* Eastw., Ericaceae), a widespread tetraploid shrub found in the chaparral of southern Oregon, California, and the Baja California (MX) peninsula, is a phenotypically complex system in which delimiting subspecies has been challenging (Keeley, Vasey, and Parker 2007; Baldwin et al. 2012; Kauffmann, Parker, and Vasey 2015). Typical of manzanitas, Eastwood manzanita has twisting branches covered with red bark, small simple drought-adapted leaves, and clusters of white-to-pink urn-shaped flowers (Fig. 3.1). Plants of this species produce a burl, a large woody structure that develops where roots and stem meet and that contains dormant buds (Wieslander and Schreiber 1939; Jepson 1916). This allows the plant to resprout and persist through numerous wildfires. Seed germination is fire-

dependent. Thus, a population of Eastwood manzanita may include some individuals that are potentially hundreds or thousands of years old, having resprouted after multiple fires, as well as younger individuals that have grown from seed produced by individuals from the original population or brought in by dispersers from other populations (Keeley and Hays 1976; Parker 2015; Moore and Vander Wall 2015). Currently, ten subspecies are recognized based on traits related to hair density and morphology, leaf color, inflorescence characters, and seed fusion, defined in previous morphometric studies (Fig. 3.1) (Keeley, Vasey, and Parker 2007; Baldwin et al. 2012). However, the morphological boundaries among subspecies can be indistinct, with intermediate phenotypes and individuals that don't conform to any one subspecies. Moreover, most subspecies overlap in their geographic range (Kauffmann, Parker, and Vasey 2015), and multiple subspecies may be found in the same population, raising the need for a clearer understanding of the relationship between genetic, morphological,and ecological patterns among these subspecies.

Characterizing subspecies differentiation in Eastwood manzanita is particularly critical because two of the currently recognized subspecies are narrow endemics of conservation concern (Fig. 3.2): *A. glandulosa* subsp. *crassifolia* (Del Mar manzanita) is federally listed as endangered (ecos.fws.gov/), and this subspecies, along with *A. glandulosa* subsp. *gabrielensis* (San Gabriel manzanita), is listed as rare in the California Native Plant Society Inventory of Rare Plants (http://www.rareplants.cnps.org/). Therefore, the ability to distinguish these two subspecies is required for conservation management.

There is currently no agreed-upon definition for a subspecies and relatively few authors have directly considered the concept of subspecies. Most authors define subspecies as conspecific

groups of one or more populations that have evolutionary meaning (Patten 2015; Patten, Unitt, and Sheldon 2002). Though various authors define "evolutionary meaning" differently, the phrase implies some level of genetic differentiation among subspecies as evolution requires the inheritance of allelic differences. However, previous methods for evaluating genetic variation have not always been capable of detecting differentiation, due to the limited nature of such data (Martien et al. 2017). Modern genomic techniques provide greater power to estimate genetic differentiation than previous methods, allowing us to better investigate evolutionary units (N. Harrison and Kidner 2011; Andrews et al. 2016).

We define subspecies as genetically differentiated populations within a species that have unique morphology or demonstrate a difference in adaptation to the local environment (Haig et al. 2006). Because previous work on Eastwood manzanita (Keeley, Vasey, and Parker 2007) indicated that in some populations morphological diagnosis of subspecies can be difficult to apply, we used next-generation sequencing data and online map-based resources to ask whether currently recognized subspecies are (1) genetically differentiated and/or (2) environmentally differentiated, or if other genetic or environmental structure can be detected within this enigmatic species.

METHODS:

Sampling—

We collected 137 accessions from 7 subspecies of Eastwood manzanita in Southern California (Fig. 3.3, Appendix 3.1). An additional three samples from coastal San Diego County,

California and two samples of an eighth subspecies collected in northern Baja California (Burge et al. 2018) were included in this study. In addition to Eastwood manzanita, we included seventeen samples of five diploid manzanita species collected from this same sampling area (Appendix 3.2). This multispecies sample set was used for comparison of genetic differentiation among species-level taxa. Because a previous study that performed chromosome counts for many manzanita taxa reported no variation in ploidy within *A. glandulosa* (Philip V. Wells 1968), we did not perform independent checks for ploidy on our *A. glandulosa* samples.

Identification of samples—

We identified samples to species and subspecies using the dichotomous key in the *Field Guide to Manzanitas: California, North America and Mexico* (Kauffmann, Parker, and Vasey 2015), and, when identification was still unclear, we cross referenced with the dichotomous key in Keeley et al. (Keeley, Vasey, and Parker 2007). Identifications were confirmed by VTP and JEK. Subspecies identifications were thus in line with the most up-to-date taxonomy based on morphometric study by Keeley et al. (Keeley, Vasey, and Parker 2007). In the case of two Eastwood manzanita DNA samples from Burge et al. (2018), subspecies identification was not recorded; two others were identified as subspecies *erecta*. Voucher specimens were not available for those four collections and for three samples of the diploid species.

DNA extraction and quality control—

We ground 150-200mg frozen floral bud, young leaf, or flower tissue in liquid nitrogen, and used the Qiagen DNEasy Plant Mini Kit (Qiagen: Hilden, Germany) to extract DNA,

modifying the protocol as described in Appendix S1. We used a Qubit® 2.0 fluorometer (Invitrogen: Carlsbad, CA) to quantify the concentration, and checked for the presence of high molecular weight DNA by running extracts on a 1% agarose TAE gel (Invitrogen: Carlsbad, CA) and imaging on a Bio-Rad Gel Doc (Bio-Rad Laboratories: Hercules, California).

Double digest restriction-site associated DNA sequencing (ddRAD-seq) library preparation and sequencing—

One set of libraries (96 samples) was prepared and sequenced at LGC Genomics (Berlin, Germany) (Appendix S2). We prepared another set containing the remaining 58 samples at the University of California, Riverside (UCR), and sequenced them at the UCR Genomics Core Facility. The protocol used for library preparation at UCR (Appendix S3) was based on the protocol developed by Brelsford et al. (Brelsford, Dufresnes, and Perrin 2016), which was based on the protocol outlined in Parchman et al. (Parchman et al. 2012) and Peterson et al. (Peterson et al. 2012). We modified this protocol to use the same enzymes as LGC and to incorporate the normalization step in the LGC protocol. Both libraries were sequenced using paired 150 bp reads on an Illumina NextSeq 500 V2 that was configured to provide ~1.5 million reads per sample. Tetraploid and diploid samples were sequenced for the same target number of reads and in the same sequencing lanes.

Sequence data processing—

All data processing was done on the High Performance Computer Cluster at UCR. We removed adapter sequences, and verified that cut sites for both PstI and MspI were present for

each read pair. To demultiplex sequences, we separated samples based on their indexed Illumina adapter sequence, and then by their inline barcode sequence using the process_radtags program in Stacks V. 2.1 (Catchen et al. 2013). We used the resulting sequence files to construct two sets of sequence data: one containing the 137 Eastwood manzanita samples ("Eastwood manzanita data set"), and one containing the five other species ("multispecies data set").

We called Single Nucleotide Polymorphisms (SNPs) using freebayes V. 1.3.1 (Garrison and Marth 2012), a software tool that can call diploid or polyploid genotypes. Because freebayes requires a reference genome, and no assembly is currently available for *Arctostaphylos*, we constructed a ddRAD reference file by (1) merging read pairs with PEAR V. 0.9.10 (Zhang et al. 2014), (2) sampling 200,000 reads from each merged sequence file, and (3) clustering these reads by 95% sequence similarity to form contigs using CD-HIT-EST V. 3.1.1 (W. Li and Godzik 2006). We constructed separate reference files for the Eastwood manzanita and multispecies data sets, and then aligned the individual sequence files to each respective reference file using BWA-MEM in bwa V. 0.7.12 using default parameters (H. Li 2013). We then called variants from the aligned reads using the freebayes-parallel script provided in freebayes (Garrison and Marth 2012). We ran freebayes three separate times: (1) on the Eastwood manzanita data set assuming diploidy, (2) on the Eastwood manzanita data set assuming tetraploidy, and (3) on the multispecies data set assuming diploidy. Because the diploid and tetraploid samples were sequenced to the same target read number, coverage per genome copy is greater for the diploid samples than for the tetraploid samples. However, the method we used to make our reference files is based on simple clustering of pooled sequences by similarity, so differences in coverage per genome copy should not impact the reference construction.

We filtered the resulting VCF (variant call format) files to impose quality and variant type controls using custom bash and R scripts. This filtering removed indels, multi-nucleotide polymorphisms (SNPs at successive nucleotides that may be linked), and combinations of different types of variants, and left only simple SNPs. We eliminated the other variants, as they could not be analyzed using the methods we used. Additionally, we removed loci with greater than 20% missing data across samples, and loci that had a minor allele recorded for fewer than 3 samples. We refer to this tetraploid SNP data set as the "4N" data set. We also generated two data sets in which genotypes were called as diploid. For the "2N" SNP data set, we allowed up to two alleles per individual at any given locus, but allowed more than two alleles across all samples. Additionally, as many analyses require SNPs to be biallelic (having no more than two alleles across all samples at a given locus), we created a third SNP data set ("2N-biallelic"), by removing loci from the 2N SNP data set that had more than two alleles total across all samples. The multispecies data set was processed the same as the 2N data set. After quality filtering, the 4N, 2N, 2N-biallelic, and multispecies SNP data sets contained 4,018, 3,395, 3,337, and 21,660 SNPs, respectively. We removed eleven Eastwood manzanita individuals that had ≥ 50% missing data, leaving 126 individuals for analyses.

SNP data processing—

For downstream analyses, we converted the filtered VCF files to nexus format alignment files with custom scripts in R V. 3.6.0 (R Core Team 2018), using standard IUPAC ambiguity codes to represent heterozygosity while retaining information for each allele. For STRUCTURE analyses (Pritchard, Stephens, and Donnelly 2000), which require unlinked loci, we used custom R scripts to select only the first SNP from each unique RAD fragment.

Genetic distance analyses—

We analyzed all data sets in SplitsTree4 V. 4.15.1 (Huson and Bryant 2006), using the uncorrected *p* measure of genetic distance, defined as the number of nucleotide differences between two sequences divided by the length of the sequences (Nei and Kumar 2000). We retained information for heterozygous genotypes by setting SplitsTree4 to average ambiguous states in the calculation of uncorrected *p*. We computed a network visualization for each data set using the NeighborNetwork method (Bryant and Moulton 2002). To visualize patterns across these networks, we used the phangorn package (Schliep et al. 2016) in R. Additionally, we used multidimensional scaling analysis (MDS) (Gower 1966), calculated using the cmdscale function in R, to visualize genetic distances. We investigated the results of MDS models calculated with as many as four dimensions, but the results yielded no additional geographic or taxonomic pattern beyond those of the two-dimensional MDS. We therefore calculated our MDS analyses with two dimensions. To search for patterns of clustering within the MDS results we performed k-means clustering (Forgy 1965; MacQueen 1967) using the kmeans function in R (R Core Team 2018). We used three statistical methods, implemented in the factoextra R package (Kassambara and Mundt 2016), to determine the optimal number of clusters: the within-group sum-of-squares, average silhouette width, and gap statistic methods (Rousseeuw 1987; Forgy 1965; Tibshirani, Walther, and Hastie 2001). To test whether genetic differentiation among samples is associated with geographic distance, we performed a simple Mantel test (Sokal 1979) in the R package ade4 (Dray, Dufour, and Others 2007).

In addition to MDS, we performed non-metric multidimensional scaling (NMDS) using the prabclust function in the R package prablus (Hennig and Hausdorf 2019). This package calculates pairwise shared allele distances, defined as one minus the proportion of alleles shared between two samples (ignoring loci with missing data for one or both samples) (Bowcock et al. 1994), and performs NMDS on this distance matrix. We selected nine dimensions for the NMDS, as this was the least number of dimensions that yielded a stress value of less than 0.1 (Clarke 1993). Because of the high dimensionality of the NMDS analysis, we could not visually inspect the result so we instead performed Gaussian clustering implemented in the prabclust function, to sort samples into clusters. We tested clustering results ranging from two to four clusters ($k = 2$ to 4), and used leave-one-out cross-validation (Friedman, Hastie, and Tibshirani 2001) to evaluate whether the clusters were distinct at each value of k. We then compared clustering results with taxonomic determination and geographic pattern. Because the clustering results at $k = 4$ provided no additional geographic or taxonomic insight when compared with the result at $k = 3$, we did not perform clustering for higher values of k. As the NMDS implemented in prabclust can only use diploid, biallelic SNPs (Hennig and Hausdorf 2019), we used the 2N-biallelic SNP data for this analysis, and could not compare results of Gaussian clustering among the three Eastwood manzanita SNP data sets.

Structure analysis—

For each Eastwood manzanita data set, we used the ParallelStructure package (Besnier and Glover 2013) in R to implement computation of STRUCTURE analyses (Pritchard, Stephens, and Donnelly 2000) for fifteen separate calculations of each value of k (the number of genetic clusters) ranging from $k = 1$ to $k = 9$. We ran each independent STRUCTURE calculation for

1,100,000 MCMC generations, discarding the first 100,000 generations as burn-in. To infer the best supported value of *k*, we used the Evanno et al. Δ*k* method (Evanno, Regnaut, and Goudet 2005), implemented on the STRUCTURE Harvester website (Dent and vonHoldt 2013). We created STRUCTURE-style bar charts using a custom R plotting function.

Principal Components Analysis—

We conducted principal component analyses (PCA) to estimate genetic differentiation, as an alternative method of ordination (based on similarities) to MDS (based on distances). For the 2N-biallelic data set, we used VCFtools v0.1.13 (Danecek et al. 2011) to convert the VCF file into a numeric genotype matrix. We used the scale2 function in R package flashpcaR (Abraham, Qiu, and Inouye 2017) to scale the numeric genotypes and conducted a PCA in R using factoextra (Kassambara and Mundt 2016).

For the multispecies data set, doing a similar analysis in R is computationally heavy and prohibitively slow, due to the large number of SNPs. Thus, we used the option "--pca" in the PLINK (Purcell et al. 2007) package to perform principal component analysis to estimate genetic differentiation in the multispecies dataset.

Ecological differentiation analysis—

To obtain a sufficient number of samples to test whether Eastwood manzanita subspecies are differentiated by habitat, we used geo-referenced herbarium collection records (n = 1648) of the seven California Eastwood manzanita subspecies included in this study from the Consortium

of California Herbaria (http://ucjeps.berkeley.edu/consortium/), and cleaned the data by removing duplicates and updating taxonomic names (Appendix S4). We used environmental variables, from publicly available sources, that have been suggested to be correlated with the distribution of manzanitas and other chaparral shrub species (Franklin 1998), including soil pH, downloaded from SoilGrid (https://soilgrids.org/, ~0.25 km^2 resolution), and the 19 Bioclimatic variables, along with solar radiation, sourced from Worldclim (http://worldclim.org/version2, ~1 km^2 resolution) (Appendix S5). We used ArcGIS v10.2.2, to extract the environmental values for the coordinates of the specimens and performed principal component analysis using the R package factoextra (Kassambara and Mundt 2016).

Environment-genotype association analysis—

To determine if Eastwood manzanita subspecies are genetically differentiated at loci that are potentially linked to local environmental adaptation, we generated the "environment-associated SNP data set" dataset. We used the environmental data and Pearson's correlation coefficient (r) to calculate the pairwise correlation between environmental variables and eliminated those that were highly correlated ($|r| > 0.7$) (Appendix S6), leaving seven variables: BIO3 Isothermality, BIO5 Max Temperature of Warmest Month, BIO9 Mean Temperature of Driest Quarter, BIO12 Annual Precipitation, BIO14 Precipitation of Driest Month, Solar Radiation and Soil pH (Appendix S7). We also used the scale2 function in R package flashpcaR to scale the 2N-biallelic data set and then used the latent factor mixed model implemented in R package LFMM (Frichot et al. 2013) to find SNPs that are highly associated with the environmental variables ($P < 1 * 10^{-5}$ for a z-test). We set the number of latent factors (K) to two

in accordance with the results of the STRUCTURE analysis and p-value histogram (Appendix S8) as recommended by Frichot et al. (2013).

To determine which genes contain the environment-associated SNPs, we identified the contigs containing the environment-associated SNPs and BLASTed (Altschul et al. 1997) them against GenBank (https://blast.ncbi.nlm.nih.gov/). We used the default setting and chose megablast. If no significantly similar genes were recovered using megablast, we repeated the search using discontiguous megablast and blastn.

RESULTS:

San Gabriel manzanita subspecies alone is supported as genetically distinct in some analyses

To evaluate whether subspecies of Eastwood manzanita are genetically distinct, we analyzed the 4N SNP data set using MDS, NeighborNetwork, and STRUCTURE analyses. The results of all three analyses suggest that there is no correspondence between the structure of genetic variation within Eastwood manzanita and the subspecies taxonomy (Figs. 3.4, 3.5). In the MDS analysis (Fig. 3.4a), samples of most subspecies overlap widely. Subspecies *gabrielensis* (San Gabriel manzanita) is an exception to this pattern, as it has almost no overlap with other subspecies.

To determine if there are genetic clusters that do not correspond to currently recognized subspecies, we performed k-means clustering on all three genetic data sets (Appendix S9). We used three statistical methods, the within-cluster sum of squares, silhouette, and gap statistic methods (Rousseeuw 1987; Forgy 1965; Tibshirani, Walther, and Hastie 2001), but they did not

provide a consistent optimal number of clusters (*k*) (Appendix S10), suggesting there is no single optimal number of clusters for these MDS results. At all values of *k*, the clusters do not show any pattern correlated with subspecies identity, but rather, roughly correspond to geographic areas within the range of our sampling (Appendix S11-S13). One exception is that at $k = 3$, the five samples of San Gabriel manzanita from the type locality (Mill Creek Summit in the San Gabriel Mountains) are identified as a distinct cluster (Fig. 3.6, Appendix S12).

In addition to MDS, we assessed whether there were any clusters that do not correspond to subspecies by performing an NMDS analysis paired with Gaussian clustering. NMDS has the same analytical goal as MDS, but uses different underlying mathematics and can thus produce different patterns than MDS. A linear discriminant analysis with leave-one-out cross validation showed no clear separation of subspecies in the NMDS clusters (Appendix S14-S16). The results of Gaussian clustering on the NMDS show a geographic pattern that is similar to that of the k-means results on the MDS (Appendix S11-S13). The results of the Gaussian clustering, however, differ in identifying three of the five Mill Creek Summit samples as a distinct cluster at $k = 2$ (Appendix S14).

In the NeighborNetwork analysis (Fig. 3.4b), samples from most subspecies are intermingled across the network. There are samples from individual subspecies that group together closely, but other samples from the same subspecies fall elsewhere in the network. Two subspecies, *gabrielensis* and *erecta*, form exclusive groups in the network, however we only have two samples of the latter. Most tips in the network are very long, and show little shared edge length with adjacent tips, however, some groups of tips show more shared edge length, indicating shared genetic variation among these samples. Groups of samples showing these longer shared edges include subsets of subsp. *gabrielensis, cushingiana*,and *glandulosa*, as well as some

clusters of mixed subspecies identity. The five *A. glandulosa* subsp. *gabrielensis* from the type locality, which were identified as a distinct cluster in both the k-means and NMDS analyses, show considerable shared edge length, supporting their distinctness from other *A. glandulosa* (Appendix S13, S15-S17).

The STRUCTURE analysis (Fig. 3.5) shows strongest support for $k = 2$ (Appendix S18). Most samples of subsp. *leucophylla, adamsii*, and *erecta* show assignment to a single cluster (Fig. 3.5a). Individuals of other subspecies show assignment to one of the two clusters or to a combination of the two. At $k = 2$, subsp. *gabrielensis* does not appear to be distinct from other subspecies. At $k = 3$ *or* 4 (Figs. 3.5b, 3.5c), there is increased variation within subspecies, but no difference among any subspecies. STRUCTURE analyses of the 4N data set do not support the distinct identity of subspecies *gabrielensis*, in contrast to the MDS and NeighborNetwork analyses.

Results of analyses of genetic structure are similar when assuming diploidy—

Because some methods of analysis are only available for diploid genetic data, many authors have analyzed sequence data from polyploid species as diploid (Stobie et al. 2018; Burge et al. 2018; Lachmuth, Durka, and Schurr 2010; Rodzen, Famula, and May 2004). To evaluate the effect this assumption may have on results, we compared the results of analyses based on the 4N data set to the 2N and 2N-biallelic data sets. The MDS and NeighborNetwork analyses based on the 2N data set (Fig. 3.7) yielded a similar pattern of clustering, with almost no correspondence between subspecies and genetic structure. As with the 4N data set, subsp. *gabrielensis* and the two subsp. *erecta* samples form groups in the NeighborNetwork analysis (Fig. 3.7b). In addition, all samples of subsp. *crassifolia* also group together in this analysis, in contrast to the 4N

analysis. All other subspecies are highly interspersed with each other. Results of the STRUCTURE analyses based on the 2N data set (Fig. 3.8) are also largely consistent with those of the 4N analysis. At $k = 2$, the most strongly supported value, there is no differentiation among subspecies. However, subsp. *leucophylla* and *adamsii*, but not *erecta* in this case, share a genotype that is found less frequently in other subspecies, similar to the results of the analysis with the 4N data set. In contrast to the 4N analysis, though, at $k = 3$ and $k = 4$, the analysis based on the 2N data set shows most individuals of subsp. *gabrielensis* sharing a genotype that is largely genetically distinct from other subspecies.

The results of the MDS and NeighborNetwork analyses based on the 2N-biallelic (Appendix S19, S20) data set are nearly identical to those based on the 2N data set. The results of the STRUCTURE analysis on the 2N-biallelic data set (Appendix S21) differ slightly from the analysis based on the 2N data set, with more individuals assigned to a single cluster and fewer assigned to multiple clusters.

Because principal components analysis (PCA) is commonly used to analyze large-scale genetic data sets, we performed PCA using the 2N-biallelic data set (Appendix S22). Although the percent of variation explained by PC1 and PC2 is very low (<2%), this analysis reveals a similar pattern of genetic differentiation as the MDS analysis. The combined results of our analyses suggest distinctness of only the San Gabriel manzanita subspecies.

Genetic variation in Eastwood manzanita corresponds to a north-south gradient

A previous study of genetic structure in Eastwood manzanita based on a smaller sample set suggested genetic variation along a north-south transect (Burge et al. 2018), therefore we tested this hypothesis with our data set. Using the 4N data set, we sorted the results of the STRUCTURE analysis by latitude and found a gradient of genotype change from north to south (Fig. 3.9, Appendix S23). Plotting samples coded by genotype on a map of southern California shows this north-south gradient (Fig. 3.10). Coding the samples in the MDS and NeighborNetwork analyses as north (Transverse Ranges and north) vs. south (south of the Transverse Ranges) rather than by subspecies shows non-overlap of the two groups, although they do not form separated clusters (Appendix S24). Moreover, using the Mantel test, we detected a significant correlation between genetic distance and geographic distance among pairwise samples (Mantel $r = 0.27$, $P < 0.0001$) (Appendix S25). Analyses using the 2N and 2N-biallelic data sets gave indistinguishable results (Appendix S26-S27).

Sorting the samples in the STRUCTURE analysis by latitude (Fig. 3.9) exposes a pattern in subsp. *cushingiana* in which the samples of this subspecies fall largely into two discrete groups, one consisting of a northern genotype and one a southern genotype. This is consistent with the localities of our collections, which come from two discrete locations, one northern and one southern (Fig. 3.3). This pattern can also be seen in MDS plots, in which the samples of subsp. *cushingiana* form two clusters, one overlapping with southern samples, and the other forming a cluster close to subsp. *gabrielensis* and other northern samples (Figs. 3.5a, 3.7a). Subspecies *glandulosa* shows a similar, but not as clear-cut, pattern in the STRUCTURE results

(Fig. 3.9). These subspecies both have wide ranges, and span the latitudinal distribution of our samples, thus it is not surprising to see north-south variation within each of them.

Broad-scale environmental data fail to distinguish Eastwood manzanita subspecies—

To test whether Eastwood manzanita subspecies can be distinguished by broad-scale topoclimatic and edaphic factors, we performed PCA with environmental data extracted from online mapping resources. We used data from herbarium specimens to provide statistical power. The range of environmental variability captured in this data set differs among subspecies, but samples of most subspecies overlap (Fig. 3.11). PC1 and PC2, which respectively explain 73% and 25% of the variation, are most heavily weighted by four environmental variables: solar radiation, BIO 4 Temperature Seasonality (standard deviation * 100), BIO 12 Annual Precipitation and BIO16 Precipitation of Wettest Quarter (Appendix S28). Subspecies *glandulosa* has the widest range of variation along the niche dimensions considered, and occupies an area on the plot that encompasses the ranges of all other subspecies. Subspecies *cushingiana* occupies the second largest space, although the number of samples classified as subsp. *cushingiana* (n = 221) is far fewer than subsp. *glandulosa* (n = 830).

Analyses using only environment-associated SNPs suggests subspecies* cushingiana *is also, in part, genetically distinct—

To determine if subspecies might be distinguished by differences in genes that potentially play a role in local adaptation, we identified SNPs correlated with variation in environmental variables, and then identified likely genes containing those SNPs. Using the same environmental

variables as in the previous analysis, and the 2N-biallelic data set**,** we identified 73 SNPs that are highly associated with seven of the environmental variables ($P < 1 * 10^{-5}$) (Fig. 3.12, Appendix S29). As some SNPs were correlated with more than one environmental variable, we identified 50 unique SNPs that we included in the environment-associated SNP data set.

We performed PCA, MDS and STRUCTURE analyses to evaluate whether subspecies of Eastwood manzanita are genetically distinguishable at loci potentially important in environmental adaptation (Figs. 3.13, 3.14). The percent of variation explained by PC1 and PC2 increased greatly compared to the result using the full 2N-biallelic data set (Fig. 3.13a, Appendix S22), presumably the result of a considerably smaller data set. Plots of PC3 and greater showed no additional pattern of subspecies differentiation. Samples of most subspecies show a higher degree of overlap, although subsp. *gabrielensis* still forms a largely distinct group. However, a subset of the subsp. *cushingiana* samples also form a discrete group in this analysis. The MDS analysis suggests a similar result (Fig. 3.13b).

The STRUCTURE analysis using the environment-associated SNP data set shows the greatest support for $k = 6$ (Fig. 3.14a). Although the value of k is large, the percentage of assignment to each cluster in most samples is similar, suggesting low levels of differentiation across the subspecies. However, a subset of individuals from subsp. *cushingiana* and all samples of subsp. *gabrielensis* share a genotype that is rare in the remaining samples, consistent with results from the 2N-biallelic data set (Appendix S21). Because $k = 2$ was the most highly supported value for the 2N-biallelic data set, we analyzed the environment-associated SNP data set using $k = 2$ as well (Fig. 3.14b, 3.14c). The results are similar to those with $k = 6$. The assignment to clusters is similar among most samples, with subsp. *gabrielensis* and a subset of

subsp. *cushingiana* samples sharing a genotype which is found scattered in a few samples of other subspecies (Fig. 3.14a, 3.14b). However, using the reduced data set, we found fewer individuals from other subspecies sharing this genotype (Fig. 3.14b, 3.14c). Thus the environment-associated SNP data set does not differentiate most subspecies, but the signal differentiating some samples of subsp. *cushingiana* and subsp. *gabrielensis* is still detected. In contrast to the analyses with the full data set, however, we do not see a north-south gradient of genetic differentiation with this data set (Appendix S30). Although subsp. *gabrielensis* and the northern samples of subsp. *cushingiana* still show a distinct genotype, many of the individuals from other subspecies that shared this "northern" genotype in the analyses of the full data set do not share it in the reduced data set, thereby obscuring the north-south gradient.

In the LFMM analysis, we found that the SNPs identified assuming K=1, 3, and 4 almost entirely overlap with the SNPs based on the assumption of K=2 (Appendix S31). Therefore, we used K=2 in the analysis and traced the 50 environment-associated SNPs to 44 unique contigs, the length of which varied from 130 bp to 260 bp. A BLAST search (Altschul et al. 1997) found a match for 41 of these contigs (Table 3.1). The majority of these genes are predicted to play a role in functions such as cell division, protein elongation, cytoskeleton-related processes, and transcription, and therefore relationships with adaptation to local environmental conditions are difficult to establish.

DISCUSSION

Most Eastwood manzanita subspecies are not differentiated by reduced-representation genomic sequence data or broad-scale environmental data—

Our analyses were unable to detect differentiation among most Eastwood manzanita subspecies on the basis of genomic data, coarse-scale environmental variables, or environment-associated genetic variants. These results are consistent with the morphological variability seen across the species. Instead of a correspondence to taxonomy, we see across our genetic analyses that genetic structure within *A. glandulosa* shows a geographic pattern.

As in the genetic analyses, our analyses of environmental variables found overlap among all subspecies. However, although the geographic distribution of the herbarium specimens we used largely matches the described ranges of subspecies, subspecies taxonomy has changed over the decades, and older specimens may have been classified using a currently outdated system. Furthermore, we are limited in the conclusions we can draw by the coarse resolution of the environmental data and the limited number of ecological factors considered. As is always the case in such analyses, taxa may be separated on niche axes not captured by the environmental data considered, at the scale of the analysis (Fletcher et al. 2013). Therefore, these results should be seen as preliminary and a source of hypotheses for additional testing with updated specimen identification, additional habitat information such as soil water potential, mineral differences, and vapor pressure deficit, as well as finer-scaled data describing these factors. Although it is not surprising to find genetic and habitat similarity among members of the same species, some level

of divergence in these factors would be expected among subspecies as evolutionary units, which may be identified with further study.

Our results did show that genetic structure within Eastwood manzanita reflects geographic distribution, confirming the results of Burge et al. (2018). MDS and NeighborNetwork analyses show that the samples can be divided into northern and southern groups, however, these groups are not clearly separated, and the overall pattern is thus better described as a gradient of genetic variation (Appendix S24). The STRUCTURE results show a transition from a predominantly northern genotype to predominantly southern at the Transverse Ranges of southern California (Fig. 3.9). This suggests a pattern of genetic divergence by geographic distance, confirmed by the significant correlation between genetic and geographic distance (Appendix S25-S27). Such a north-south pattern of genetic variation has also been observed in other studies of the California biota (Sork et al. 2016; Schierenbeck 2017; Zink, Lott, and Anderson 1987; Burge et al. 2011) and the Transverse Ranges have been suggested as a barrier to genetic continuity (Calsbeek, Thompson, and Richardson 2003; Sgariglia and Burns 2003; Forister, Fordyce, and Shapiro 2004); (Chatzimanolis and Caterino 2007). Our analyses do not indicate a barrier at the Transverse Ranges, but rather suggest a continuum (Figs. 3.9, 3.10).

Our analyses using the 4N, 2N and 2N-biallelic data sets generate largely similar results (Figs. 3.4-3.5, 3.7-3.8, S1-3), suggesting that the loss of information caused by assuming diploidy in a tetraploid sample set may not prevent detection of genetic structure. In fact the genetic patterns appears most clear in the 2N-biallelic and least clear in the 4N analyses, suggesting that the biallelic SNPs may hold the strongest evolutionary signal. This is supported by the observation that most of the loci (~98%) in the data set are biallelic. Overall our results suggest

that the easier and more rapid analyses based on the use of biallelic genetic data can give a good approximation of the genetic structure. Testing in additional systems is needed to determine if this is broadly true.

Analyses support distinction of San Gabriel manzanita , but not Del Mar manzanita—

Two subspecies of Eastwood manzanita are considered rare or threatened, San Gabriel manzanita (*A. glandulosa* subsp. *gabrielensis*), found in the San Gabriel and Sierra Madre mountains of California, and Del Mar manzanita (*A. glandulosa* subsp. *crassifolia*), found along the coast of San Diego County (Fig. 3.2). A previous study of Del Mar manzanita, based on RAD-Seq data and morphometric analyses, concluded that circumscription of this subspecies based on vegetative morphology was ineffective (Burge et al., 2018). An earlier study found that fruit shape was distinctive in subsp. *crassifolia* (Keeley, Vasey, and Parker 2007), but fruits were not included in the study by Burge et al. The Burge et al. study was unable to reach a conclusion regarding genetic boundaries of the subspecies, largely due to insufficient sampling of other subspecies. Our results, with broader sampling across the species, suggest that Del Mar manzanita is not genetically distinct from other subspecies (Figs. 3.4, 3.5). The only exception is the 2N and 2N-biallelic NeighborNetwork analyses, in which the Del Mar manzanita samples cluster together (Figs. 3.7, S1-3). However, neither the MDS nor the STRUCTURE analyses of the diploid data sets show any differentiation of subsp. *crassifolia*. These results, along with inconclusive morphological distinction and lack of environmentally-associated genetic differentiation along the niche dimensions considered, suggest that the recognition of Del Mar manzanita as a distinct subspecies should be reconsidered.

San Gabriel manzanita (*A. glandulosa* subsp. *gabrielensis*) was historically treated as a separate species, *A. gabrielensis* (P. V. Wells 1992, 2000) but later was transferred into *A. glandulosa* as a subspecies (Keeley, Vasey, and Parker 2007). PCA, MDS and NeighborNetwork analyses using all three genetic data sets indicate that this subspecies is genetically distinct from the others (Figs. 3.4, 3.7, S1-4). To determine if San Gabriel manzanita is as distinct as other species are from each other, we analyzed genetic differentiation of five diploid species that are morphologically both distinct and consistent (Appendix S32-S34). The results showed that all species formed discrete clusters, and occupied well-separated regions of the MDS and PCA plots. In contrast, samples of San Gabriel manzanita fall relatively close to the other subspecies (Figs. 3.4, 3.6, Appendix S19, Appendix S22). The results of the NeighborNetwork analysis of the multi-species data set show genetic clusters completely coincide with species identity, with samples from each species grouping together, and separated from other species (Appendix S34). In contrast to this, NeighborNetwork analyses of Eastwood manzanita subspecies show that although San Gabriel manzanita samples form a unique cluster, they are not separated from the rest of the network (Figs.3. 4, 3.6, Appendix S20). Although these results cannot be compared directly, the lack of discrete separation of the San Gabriel manzanita samples from the other subspecies suggests closer relationship than seen in the multispecies analyses. However, we found that samples of San Gabriel manzanita from the type locality (Mill Creek Summit in the San Gabriel Mountains) showed greater differentiation from other samples of *A. glandulosa* than did the samples of San Gabriel manzanita from other localities (Appendix S35). The Mill Creek samples were also identified as a distinct cluster in the k-means and Gaussian clustering analyses (Fig. 3.6, Appendix S12-S13, S15-S16).

Our results may support the original status of San Gabriel manzanita as a distinct species, though it would possibly need to be circumscribed more narrowly as just the plants from Mill Creek Summit. It has been hypothesized that San Gabriel manzanita may be a hybrid of *A. glandulosa* and *A. parryana* (Keeley et al. 2007; Kauffmann et al. 2015). The latter is a broadly sympatric species that has noticeable morphological similarities with San Gabriel manzanita, including shiny bright green leaves and fusion of the nutlets in the fruits (Kauffmann, Parker, and Vasey 2015). If this hypothesis is correct, the variability that we found among the Mill Creek Summit samples may be the result of independent hybridization events and introgression. The samples of San Gabriel manzanita from other localities, which have typical morphology for San Gabriel manzanita and are genetically intermediate between the Mill Creek Summit cluster and other samples, may be hybrids between *A. glandulosa* and the putatively distinct species, *A. gabrielensis*. A definitive evaluation of this hypotheses, however, requires further sampling, including additional San Gabriel manzanita populations and *A. parryana*.

Unlike PCA, MDS and NeighborNetwork analyses, STRUCTURE analyses, with all three genetic data sets at most values of *k,* do not indicate that the genotype of San Gabriel manzanita is unique, although it is not common in other subspecies (Figs. 3.5, 3.7, Appendix S21). PCA, MDS, and NeighborNetwork analyses use either allele frequencies (PCA) or distance measures (MDS and NeighborNetwork) to characterize genetic variability or relatedness in a given population of samples (Bryant and Moulton 2004; Jombart, Pontier, and Dufour 2009). These analyses invoke few assumptions regarding the structure of genetic diversity in the sample population. In contrast, STRUCTURE is based on the assumption of Hardy-Weinberg (HW) equilibrium, and constructs a genetic model that assigns samples to clusters that minimize HW disequilibrium (Falush, Stephens, and Pritchard 2003; Pritchard, Stephens, and Donnelly 2000).

The evolutionary assumptions implemented to construct complex models in STRUCTURE analyses likely explain the differences in results obtained in these analyses. Moreover, the formula STRUCTURE uses to calculate expected genotype frequencies differs for diploid and polyploid samples, which may explain the differences in the results obtained using the 4N vs the 2N or 2N-biallelic data sets (Dufresne et al. 2014).

Taken together, our results indicate that San Gabriel manzanita shows a degree of genetic differentiation from the other subspecies and suggests it should remain of conservation concern, however a final determination regarding its taxonomic rank cannot be made without further sampling, including a disjunct population reported from Santa Barbara county.

Using genetic loci potentially related to environmental adaptation produces a similar result to the full SNP data set—

Subspecies are often considered to arise through local adaptation (Grant 1972; Haig et al. 2006; Walsh et al. 2017; Patten, Unitt, and Sheldon 2002). This suggests that subspecies should differ genetically at loci related to responses to environmental factors. Although we found that Eastwood manzanita subspecies are not differentiated on the basis of genome-wide genetic data, we investigated whether an analysis of environmentally-linked genetic loci might uncover differences among subspecies. Both PCA and MDS using the 50 2N-biallelic SNPs that varied in correlation with environmental variables showed a largely similar pattern to the analyses of the full SNP data set (Fig. 3.13, Appendix S19, S22), with perhaps even more pronounced results showing tight clustering of most samples and divergence of samples of subsp. *gabrielensis*. However, with this strongly reduced genetic data set, subsp. *cushingiana* also is found to be

partly distinct. Some samples of this subspecies cluster with the other subspecies, but five samples, corresponding to those from the northern part of our sampling range (Fig. 3.3), are divergent. This suggests that local adaptation may play a role in genetic differentiation of subsp. *gabrielensis* and *cushingiana*, which is consistent with the habitat differentiation found in a previous study (Keeley, Vasey, and Parker 2007). The partial genetic distinction within subsp. *cushingiana* is also supported by other analyses that suggest that the subspecies can be subdivided into a north and a south component (e.g., Fig. 3.9). This reflects our sampling from two geographically separated regions (Figs. 3.2, 3.3). Further investigation, including sampling across the range, is needed to elucidate the patterns of genetic variation of subsp. *cushingiana*.

A goal of molecular analyses is to draw direct lines from genetic variants to phenotypes. We evaluated the genes containing environmentally-correlated SNPs to determine if we could identify such connections between genotype and environmental adaptation. However, identification of genes containing environmentally-associated SNPs did not reveal any genes that can readily be related to adaptation to habitat differences. Many function in multiple biological processes, making connection with specific environmental adaptations impossible. Establishing such a link would take careful analysis of gene function, not possible at this time due to limits posed by the plants themselves, including difficult culturing and long generation times.

Subspecies recognition in Eastwood manzanita—

Although Eastwood manzanita has been divided into multiple subspecies, our analyses suggest that with the exception of San Gabriel manzanita, the eight subspecies we sampled are not well differentiated genetically. This is consistent with the overlapping morphological

boundaries among the subspecies. Within the ranges of many subspecies there are populations that are fairly uniform phenotypically, and that represent the archetype for that subspecies. However, in between those populations are heterogeneous populations that obscure much of those distinctions. Recognition of the individual phenotypically uniform populations as subspecies leaves the plants found in much of the range of the species as unclassifiable. Furthermore, such population-level variation may allow recognition of different phenotypes, but currently we do not have data to suggest any degree of genetic differentiation, which would be required for subspecies to be evolutionarily significant units, according to our definition. Given that some populations of *A. glandulosa* show consistent morphologies that can be identified as a single subspecies, but our analyses detected no genetic differentiation, the complex may be something akin to the syngameon concept sometimes applied to groups of woody plants species that show extensive interspecific gene flow (Cavender-Bares 2019; Grant 1972).

One hypothesis that could explain the existence of populations with mixed phenotypes in these long-lived perennial shrubs is that populations that had uniform phenotypes when they were established may have become more variable over time due to subsequent gene flow from other populations. This is consistent with the predicted very long lifespan of these plants and the fact that recruitment is tied to infrequent fires. These factors may create populations composed of plants of varying ages with genetic admixture from multiple genetic lineages within the species. Studies have shown that the seeds are dispersed by mammals, including large mammals, which may allow them to be moved over large distances (Keeley and Hays 1976; Parker 2015). Although nothing is known about the distance that pollen travels in this species, solitary bees, honey bees and bumblebees, which have been documented pollinating manzanitas including *A. glandulosa* subsp. *mollis*, can travel up to several miles (Hagler et al. 2011; Zurbuchen et al.

2010; Fulton and Carpenter 1979; Osborne et al. 2008). Such mixing of genetic material, leading to diverse phenotypes, is therefore possible, however little is known about the demographic structure of Eastwood manzanita populations, an area that needs further investigation.

Another hypothesis to explain the morphological variability in this species is that Eastwood manzanita may be a heterogeneous assemblage of tetraploid hybrids that arose multiple times, possibly from multiple different progenitor species pairs. An additional factor contributing to the variability might be introgression into these tetraploid populations from diploid species via unreduced gametes (Ramsey and Schemske 1998). This is consistent with the regionally localized ranges of many of the subspecies (Fig. 3.2). Much more in-depth sampling across the ranges of these subspecies, and inclusion of potential progenitor species, will be needed to evaluate this hypothesis. This would include more northern populations, including *A. glandulosa* subsp. *howellii*, an unsampled subspecies from the central coast of California, as well as expanded sampling of Mexican subspecies. These analyses will benefit from the sequencing of a manzanita genome, which is underway.

Subspecies concepts have not historically focused on plants, particularly not on long-lived perennial plants. It is possible that due to the differences in population structure that result from the long life-span, dynamics of gene flow, and immobility of individuals, subspecific differentiation in these species may not be well described by available concepts. Further consideration of subspecies concepts, and aspects of practical application, are needed to encompass a wider variety of organisms with diverse life histories. As we now have greater power than ever to assess genetic structure within a species, we have an opportunity to evaluate

diversity in an ever-greater array of groups, and must ensure our conceptual framework keeps pace accordingly.

ACKNOWLEDGEMENTS:

The authors thank Michael Vasey for sharing observations on Eastwood manzanitas, Dawn Nagel for technical support, Jacob Landis and Alex Rajewski for guidance on library preparation, Sarah Wilson for assisting in analyses, Matt Guilliams for helping with field work, and Dylan O. Burge for contributing samples and expertise. This research was funded by the University of California, Riverside, and the California Native Plant Society.

AUTHOR CONTRIBUTIONS:

Research was designed by YH, GM and AL. Field collections were made by YH, GM, AL, AS and NS. Collection and identification were assisted by VTM, JEK and GW. Data were generated by YH, GM and NS, with the guidance of AL and AB. Analyses were performed by YH and GM, with guidance from TS and DJ. The manuscript was written by YH and GM, with contributions from all other authors.

DATA AVAILABILITY:

The ddRADseq data are available in fastq format as a NCBI SRA data set at BioProject ID PRJNA611781. Environmental data, reference tables used in data processing, ddRADseq reference fasta files, VCF files, and SNP data sets in nexus and STRUCTURE formats, are available from the Dryad Digital Repository: https://doi.org/10.6086/D1NQ3B (Huang et al., 2020). The custom scripts we used are available at github/team-manzanita/glandulosa_2020.

SUPPORTING INFORMATION:

Additional Supporting Information may be found online in the supporting information tab for this article (https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1496).

Appendix S1. Modification of DNA extraction using the Qiagen DNEasy Plant Mini Kit (Qiagen: Hilden, Germany).

Appendix S2. LGC ddRAD-seq protocol: ddRAD library construction on 96 samples Arctostaphylos; PstI-MspI (provided by LGC, Berlin, Germany).

Appendix S3. UCR ddRAD-seq protocol.

Appendix S4. Number of herbarium records for each Eastwood manzanita subspecies included in the analyses of subspecies habitat differentiation.

Appendix S5. URL, units and resolution of environmental variables.

Appendix S6. Correlation of environmental variables evaluated for the analysis of subspecies habitat differentiation.

Appendix S7. Subspecies identification and environmental data for Eastwood manzanita herbarium records included in analysis of subspecies habitat differentiation. **APPENDIX S8.** Histograms of adjusted p-values of LFMM analyses for seven climatic variables when K = 1, 2, 3, and 4.

Appendix S9. K-means clustering results for k = 2 to k = 4 on a two-dimensional MDS of the 4N data set.

Appendix S10. Statistical evaluations for the best supported number of clusters in the k-means clustering analyses of MDS dimensions.

Appendix S11. Results of k-means clustering, for k = 2, on the MDS of the 4N SNP data set.

Appendix S12. Results of k-means clustering, for k = 3, on the MDS of the 4N SNP data set.

Appendix S13. Results of k-means clustering, for k = 4, on the MDS of the 4N SNP data set.

Appendix S14. Results of Gaussian clustering with prabclust, for k = 2, on the NMDS of the 4N SNP data set.

Appendix S15. Results of Gaussian clustering with prabclust, for k = 3, on the NMDS of the 4N SNP data set.

TABLES AND FIGURES:

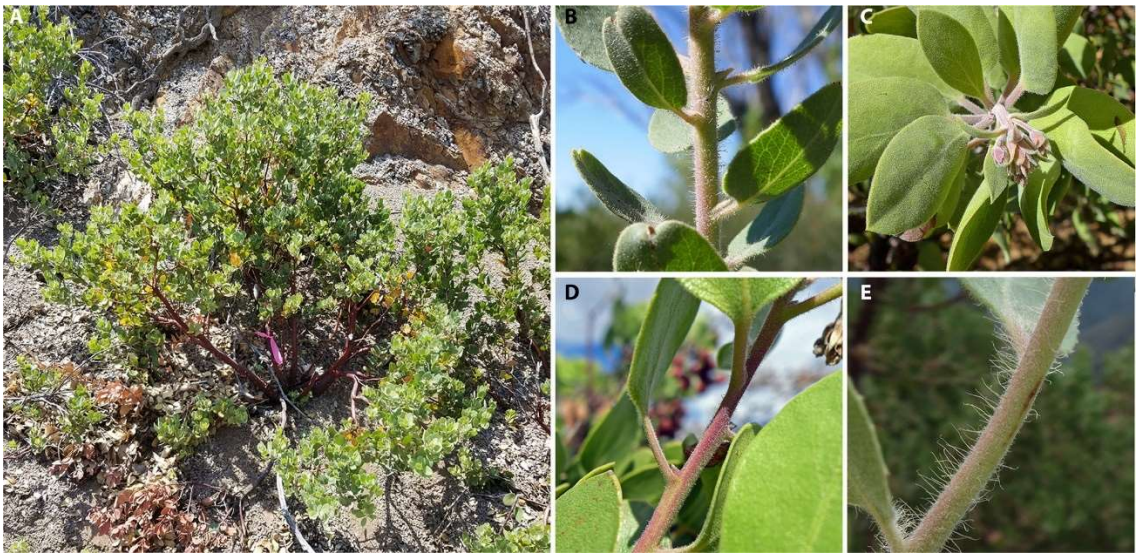


Fig. 3.1. Variation in hair traits in subspecies of Eastwood manzanita. a. Eastwood manzanita, Santa Barbara County. b. subsp. *glandulosa*, short- and medium-length hairs, long hairs with terminal glands. c. subsp. *cushingiana*, dense short hairs lacking glands. c. subsp. *gabrielensis*, relatively sparse, short hairs lacking glands. e. subsp. *mollis*, short and very long wavy non-glandular hairs. Photo credits: a. A. Litt. b-d Neil Kramer. e. Michael Charters, CalFlora.net.

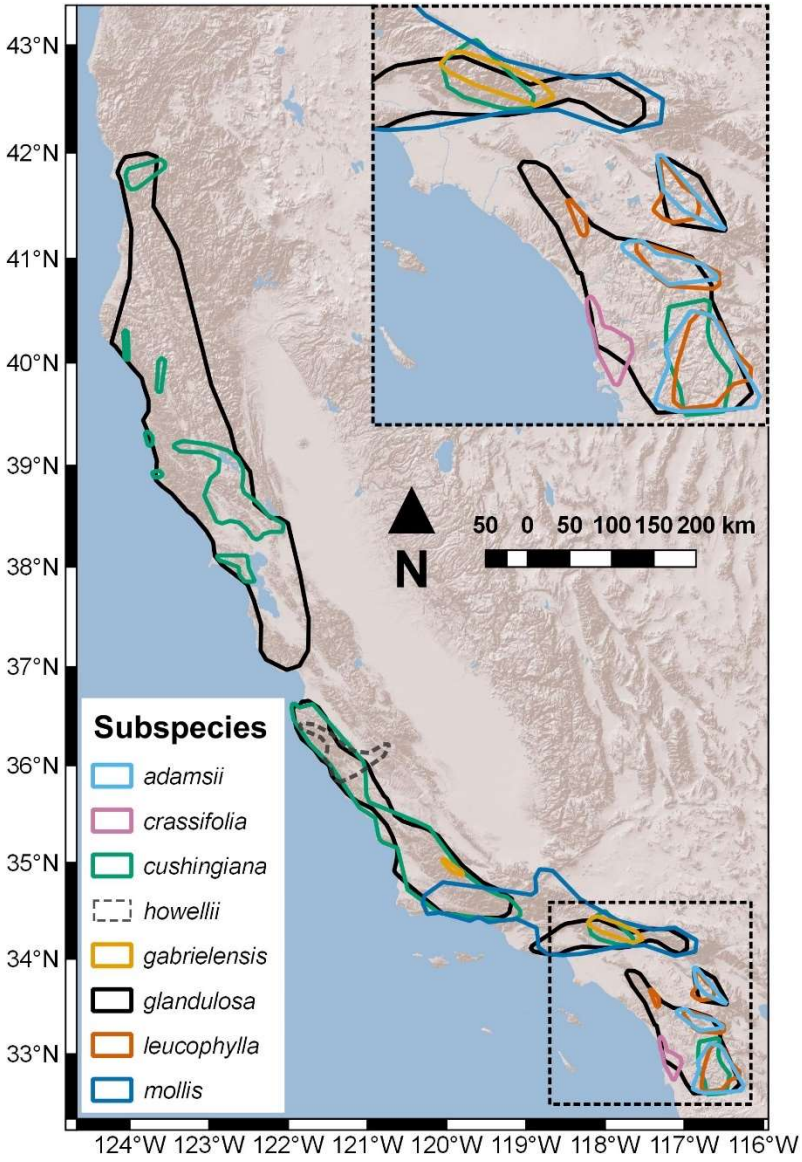


Fig. 3.2. Map of California with the ranges of the 8 subspecies of Eastwood manzanita found in California. The two Mexican subspecies are not well-databased and therefore not included. Inset map at top-right shows southern California region, marked with a dotted box on the main map.

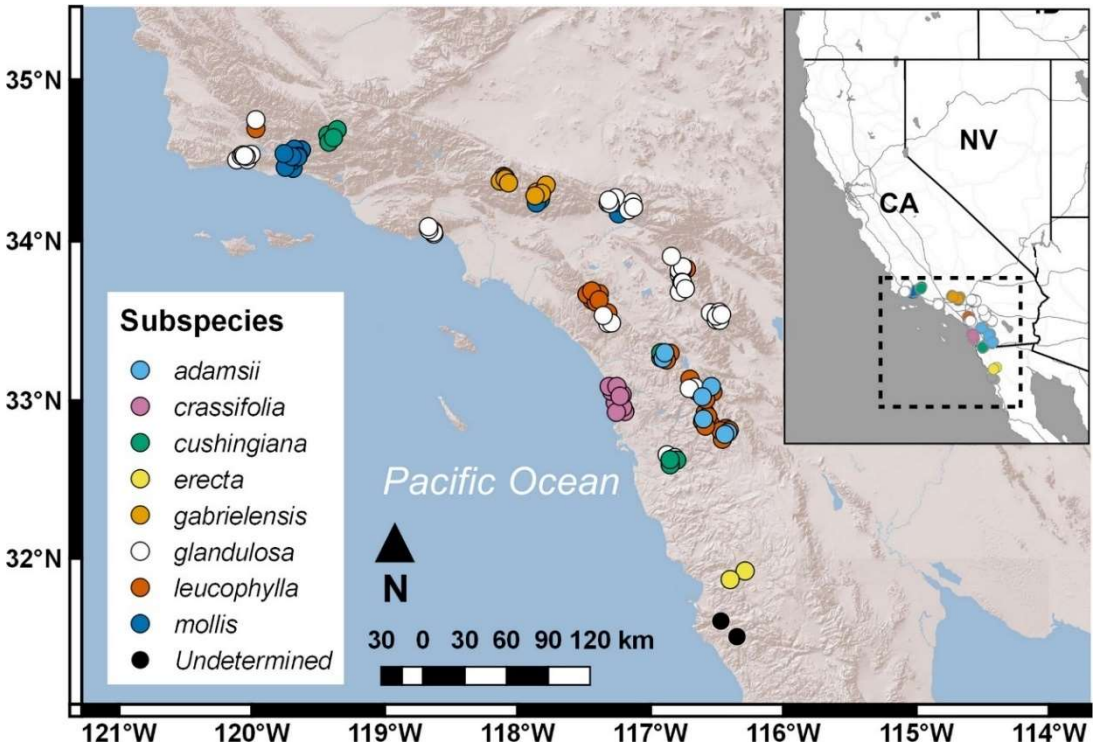
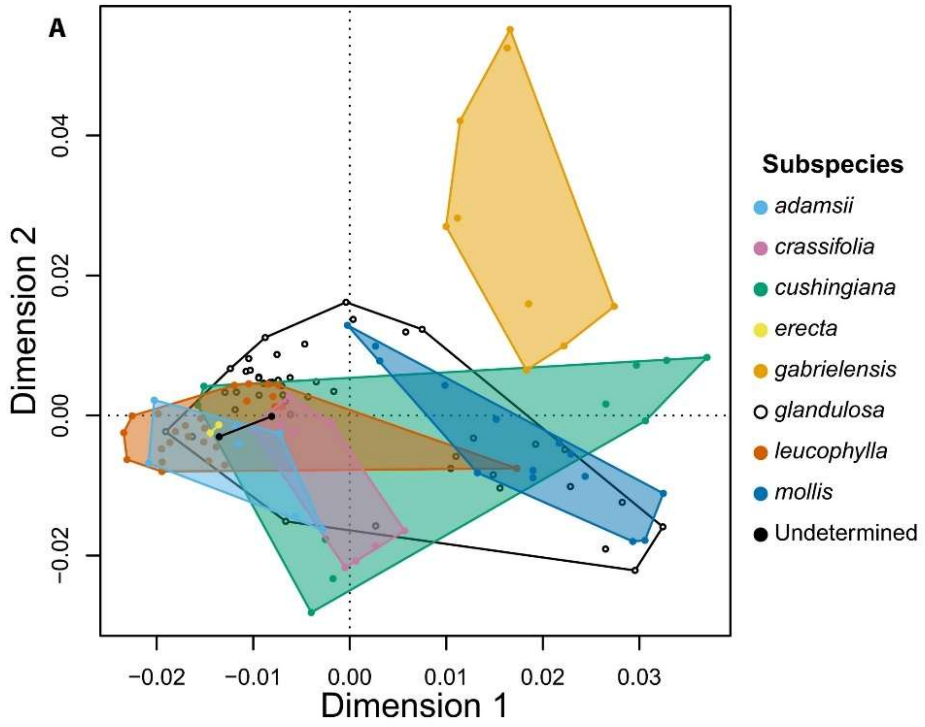


Fig. 3.3. Map of collection localities for samples included in genetic analyses. Colors indicate subspecies identification. Because some samples would overlap on the map, a random "jitter" value between -0.15 and 0.15 degrees longitude and latitude was applied to each point. Inset shows the area of focus relative to the state of California.


A
Dimension 2
Dimension 1
0.04
0.02
0.00
−0.02
−0.02
−0.01
0.00
0.01
0.02
0.03
Subspecies
adamsii
crassifolia
cushingiana
erecta
gabrielensis
glandulosa
leucophylla
mollis
Undetermined

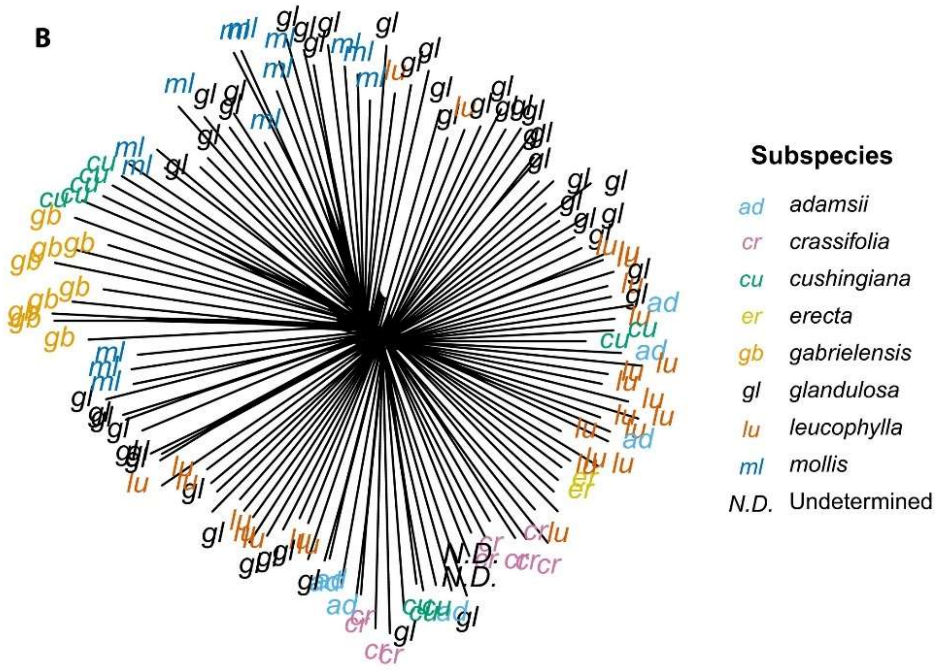

B
Subspecies
ad adamsii
cr crassifolia
cu cushingiana
er erecta
gb gabrielensis
gl glandulosa
lu leucophylla
ml mollis
N.D. Undetermined

Fig. 3.4. MDS analysis (a) and NeighborNetwork (b) for 4N data set. (a) Two dimensional representation of genetic distance among Eastwood manzanita samples. Points and polygons are colored by subspecies identification. Polygons are minimum areas that enclose all samples of each subspecies. (b) Tips are labelled and colored by subspecies identification.

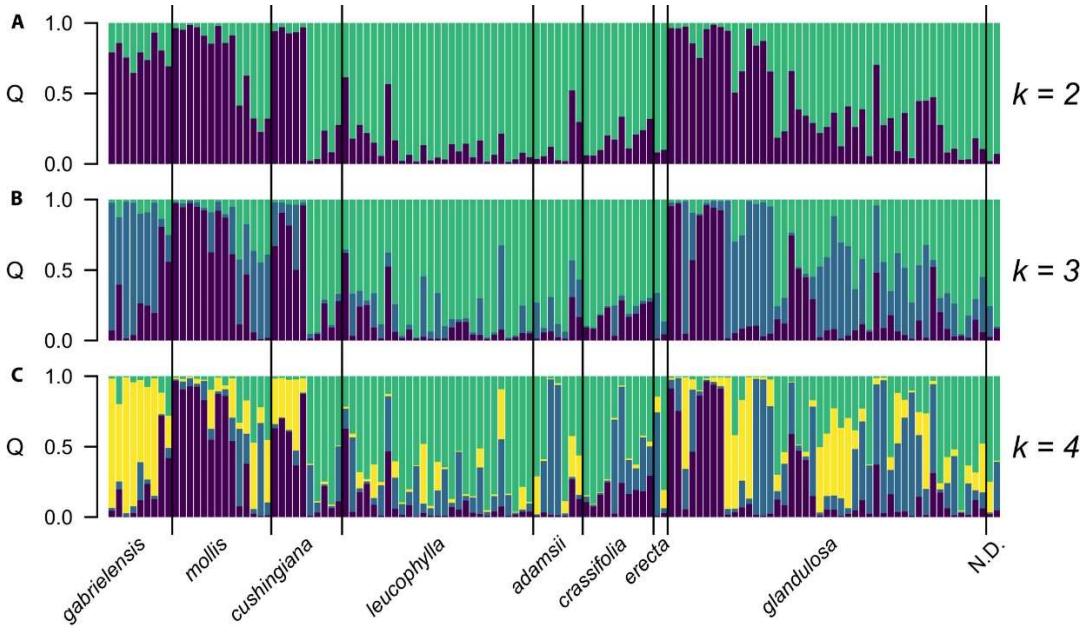


Fig.3.5. STRUCTURE results for $k = 2$ to $k = 4$ (top to bottom) for the 4N data set. Vertical bars represent individuals. Colors within bars represent ancestral clusters of differing genotypes. The proportion of each color in each bar represents the probability of assignment (Q) to each cluster. Individuals are sorted along the x-axis by subspecies, then by latitude of collection.

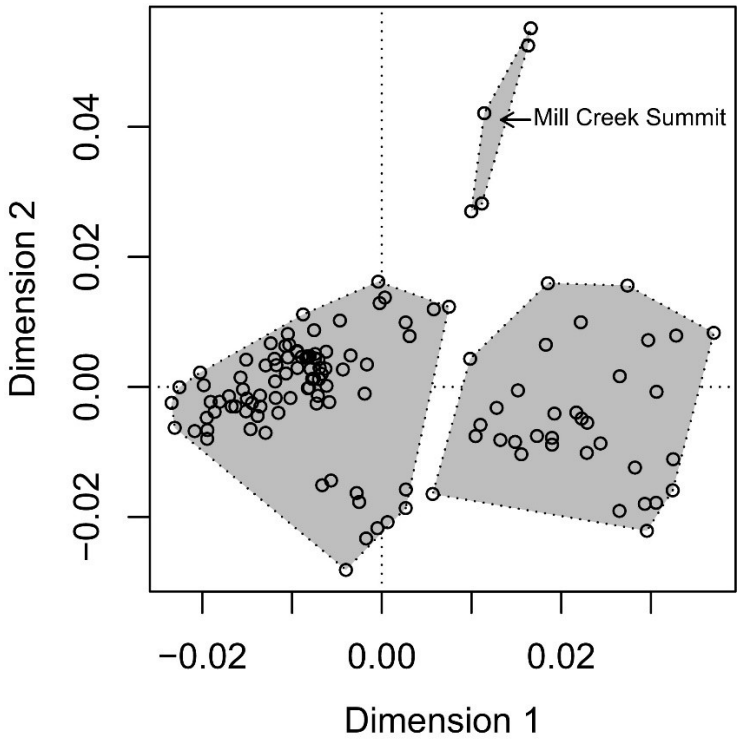


Fig. 3.6. Results of k-means clustering, for k = 3, on the MDS of the 4N SNP data set. Points represent individual samples and polygons mark the boundaries of the k-means clusters. The labelled cluster is composed of the five samples from the type locality of *A. glandulosa* subsp. *gabrielensis* at Mill Creek Summit in the San Gabriel Mountains.

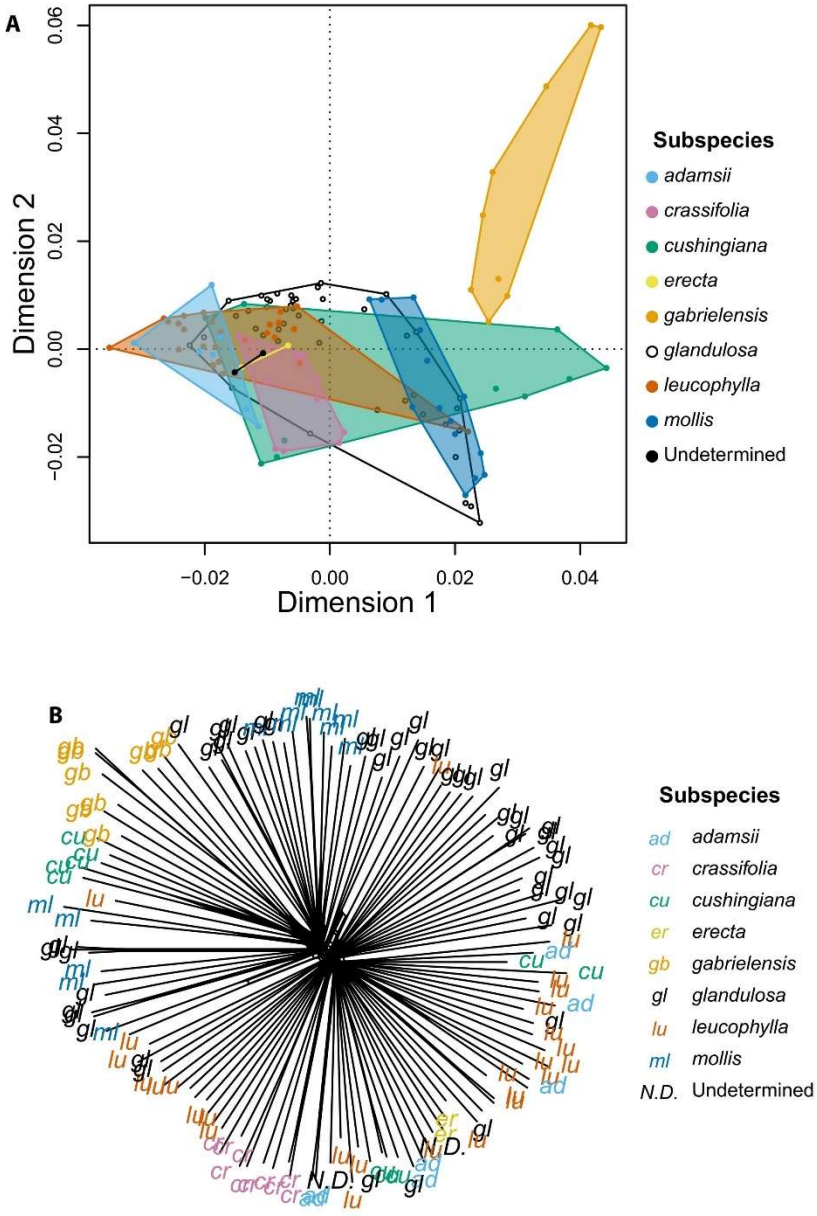


Fig. 3.7. MDS analysis (a) and NeighborNetwork (b) for 2N data set. Graphics and colors as in Fig. 3.4.

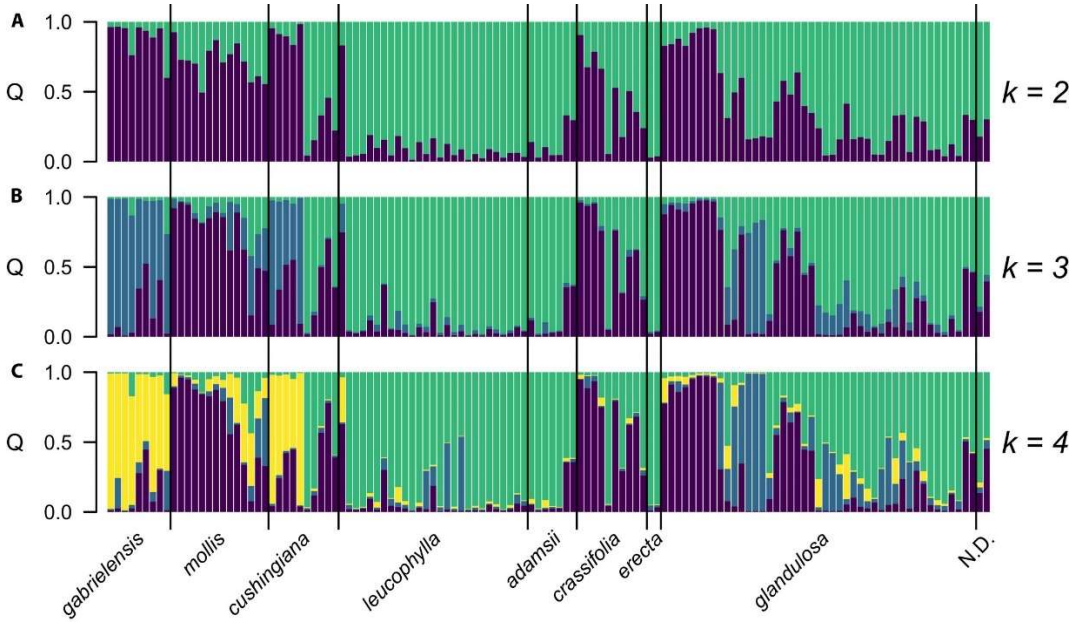


Fig. 3.8. STRUCTURE results for $k = 2$ to $k = 4$, for the 2N data set. Graphics and colors as in Fig. 3.5.

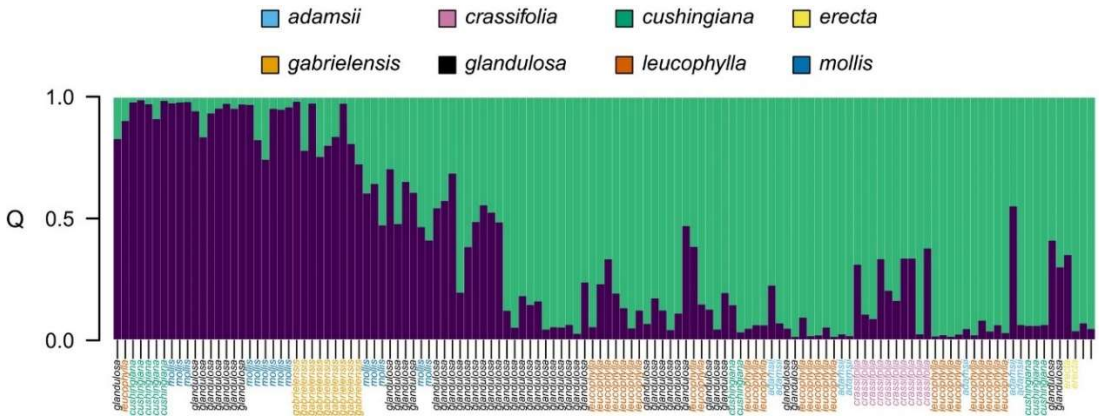


Fig. 3.9. STRUCTURE results for $k = 2$, for the 4N data set, sorted by latitude. Graphics and colors as in Fig. 3.5, except samples sorted by latitude first, and then by subspecies. Two samples undetermined to subspecies (farthest right in graph) are not labeled.

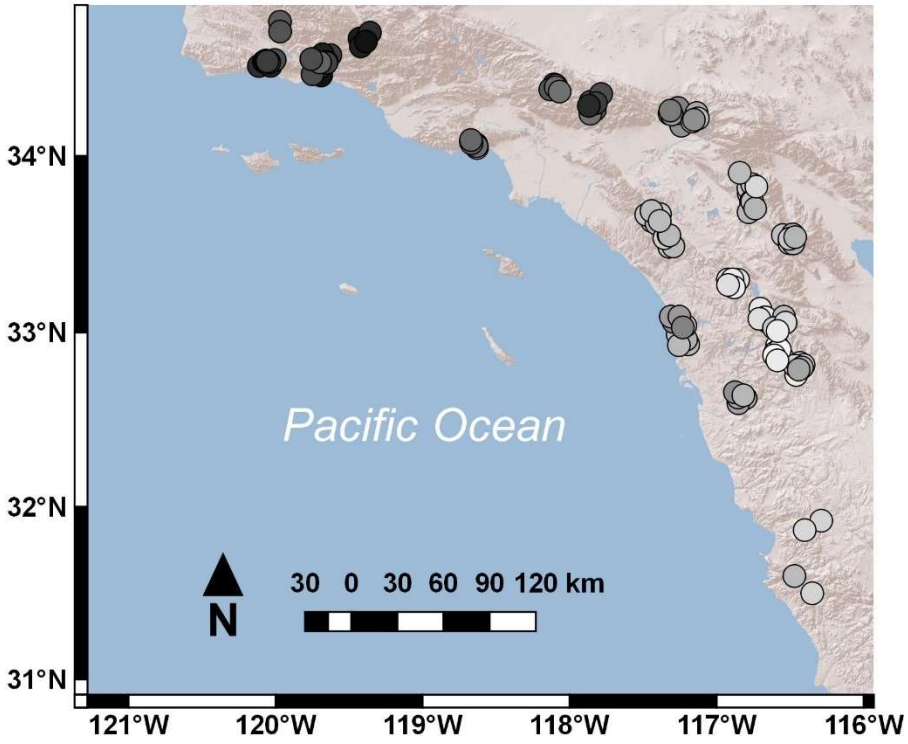


Fig. 3.10. Map showing continuous geographic genetic structure within *A. glandulosa*. Points are shaded in proportion to their value for the first dimension of the MDS on the 4N data set. A random "jitter" value was applied as in Fig. 3.3.

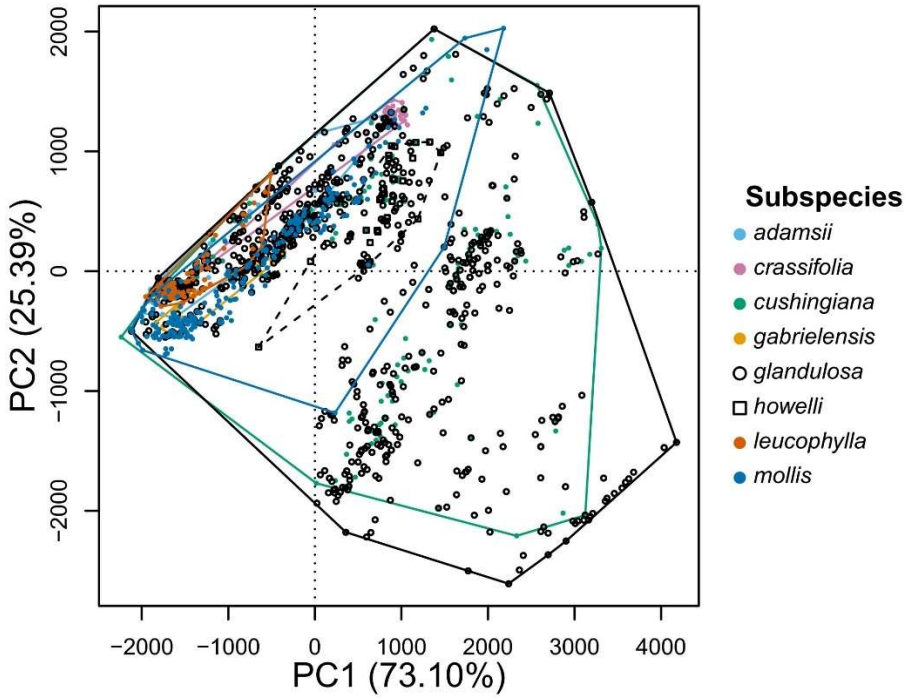


Fig. 3.11. PCA using environmental data for herbarium records for Eastwood manzanita. Points represent individual collection records (n = 1648). Subspecies are distinguished by color and shape. Polygons represent minimum areas that enclose all samples of a given subspecies.

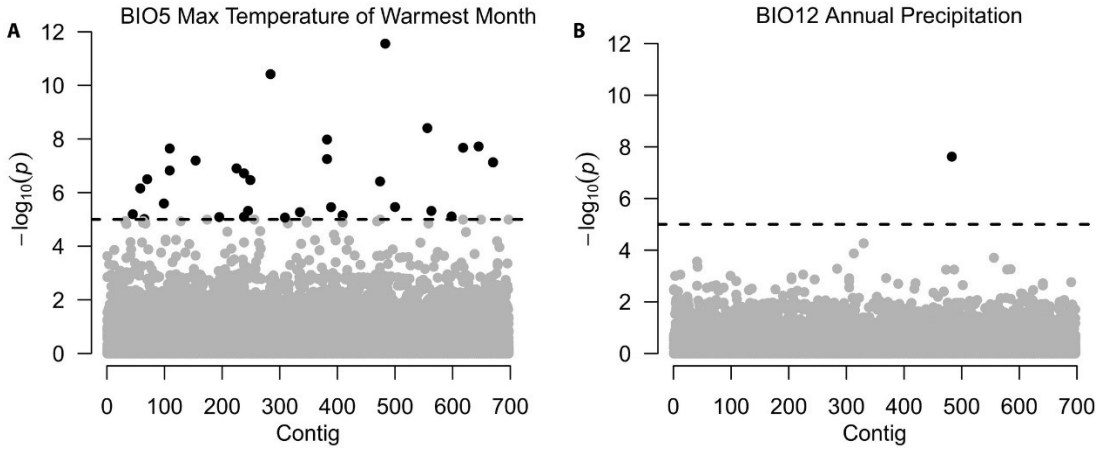


Fig. 3.12. SNPs associated with climatic variables. Points represent SNPs. The dashed line represents the threshold for statistical significance at $P = 1* 10^{-5}$ for the association of SNP and the environmental variable. Solid points show significant association.

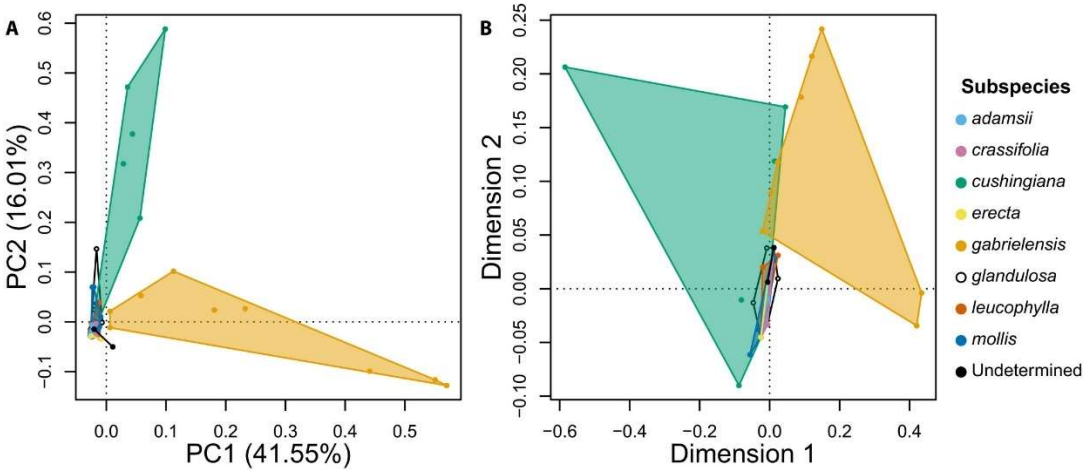


Fig. 13. PCA (a) and MDS (b) using the environment-associated SNP data set. Graphics and colors as in Fig. 3.4a.

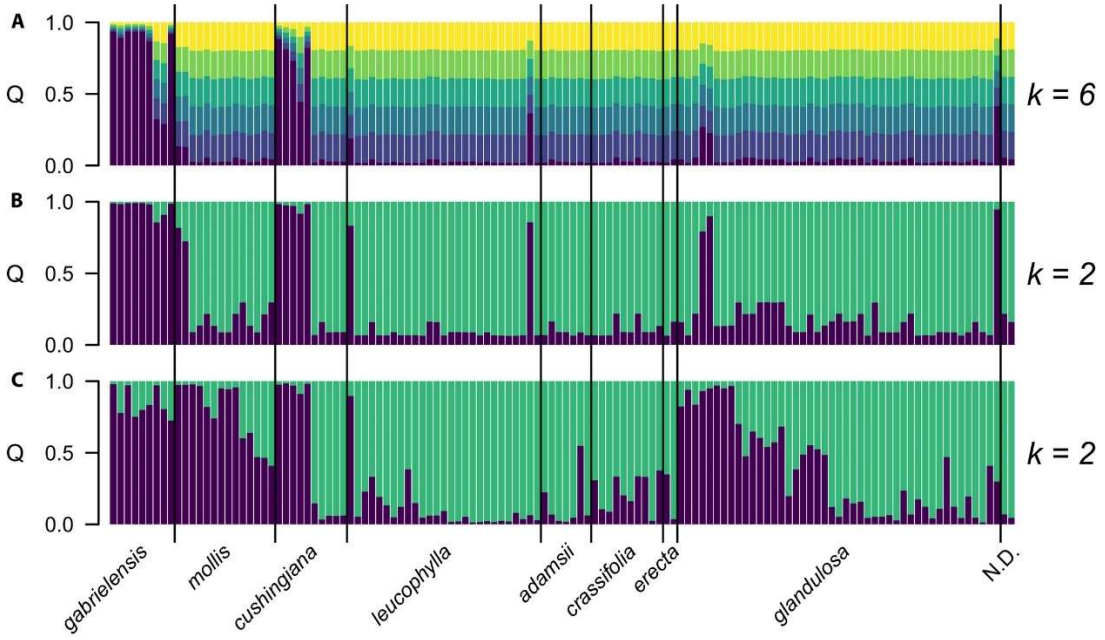


Fig. 3.14. STRUCTURE results for k = 6 (a) and k = 2 (b) for the environment-associated SNP data set in comparison to the STRUCTURE result for k = 2 for the 2N-biallelic data set (c). Graphics and colors as in Fig. 3.5.

Table 3.1. Genes containing the environment-associated SNPs and their predicted functions based on BLAST search results.

Contig	Position on contig	Predicted gene function based on BLAST search results
NB501358:219:HTTTNBGX2:1:11212:18234:14331	227	No significant similarity
NB501358:219:HTTTNBGX2:1:21201:26199:3681	120	traB domain-containing protein (3+ species)
NB501358:219:HTTTNBGX2:2:23110:22431:8467	168	No significant similarity
NB501358:219:HTTTNBGX2:2:23206:7862:12792	152	No significant similarity
NB501358:219:HTTTNBGX2:3:13604:21020:8085	65	PAMP-induced secreted peptide 2-like (Glycine soja)
NB501358:219:HTTTNBGX2:3:13607:8351:2917	213	Probable LRR receptor-like serine/threonine-protein kinase (Camilla sinensis)
NB501358:219:HTTTNBGX2:3:23406:20364:17383	76	Protein PRD1, protein coding (Camilla sinensis)
NB501358:219:HTTTNBGX2:3:23609:23061:8414	97	No significant similarity
NB501358:219:HTTTNBGX2:4:11605:17349:5762	171	No significant similarity
NB501358:219:HTTTNBGX2:4:12504:14054:16959	113	Two-on-two hemoglobin-3, protein encoding (Havea Brasiliensis)
NB501358:222:HLV53BGX3:1:12212:2156:9704	49	Cullin-1-like, protein coding (Camellia sinensis)
NB501358:222:HLV53BGX3:1:21307:7354:13740	20	Enhancer of mRNA-decapping protein 4-like, protein encoding
NB501358:222:HLV53BGX3:1:21307:7354:13740	80	Same contig as above
NB501358:222:HLV53BGX3:2:23109:22067:9536	141	7-deoxyloganetic acid glucosyltransferase-like, protein coding

		(Camilla sinensis)
NB501358:222:HLV53BGX3:4:11501:24418:3341	60	No significant similarity
NB501358:222:HLV53BGX3:4:13411:13473:3010	149	Potassium transporter 7-like (Prunus avium)
NB501358:222:HLV53BGX3:4:23610:4999:18598	218	RPC1 gene intron (Rhododendron)
NB501358:225:HLWVCBGX3:1:13101:19142:9099	32	Uncharacterized protein (Camilla sinensis)
NB501358:225:HLWVCBGX3:1:13101:19142:9099	228	Same contig as above
NB501358:225:HLWVCBGX3:1:13308:24706:16849	33	ncRNA
NB501358:225:HLWVCBGX3:1:21302:19179:16801	49	No significant similarity
NB501358:225:HLWVCBGX3:2:22109:23683:5460	39	Inorganic phosphate transporter (6+ species)
NB501358:225:HLWVCBGX3:3:12412:4674:10256	92	Elongator protein complex, protein coding (Camilla sinensis)
NB501358:225:HLWVCBGX3:3:13508:11032:7043	185	Cell-division cycle 5-like protein (Camellia sinensis)
NB501358:225:HLWVCBGX3:3:13605:2074:10926	93	No significant similarity
NB501358:225:HLWVCBGX3:4:11407:6231:10019	18	Phosphatidylinositol/phosphatidylcholine transfer protein
NB501358:225:HLWVCBGX3:4:11407:6231:10019	148	Same contig as above
NB501358:225:HLWVCBGX3:4:12406:6648:5034	110	Uncharacterized protein YwbO (Camilla sinensis)
NB501358:225:HLWVCBGX3:4:12507:22687:6405	48	No significant similarity
NB501358:225:HLWVCBGX3:4:23503:6747:7761	19	Endoglucanase 11, protein coding (Camellia sinensis)
NB501358:228:HNY5YBGX3:1:11202:1281:2373	58	No significant similarity

NB501358:228:HNY5YBGX3:1:12107:7060:12413	124	Uncharacterized
NB501358:228:HNY5YBGX3:1:12107:7060:12413	209	Same contig as above
NB501358:228:HNY5YBGX3:1:13107:17380:4906	140	IQ domain-containing protein
NB501358:228:HNY5YBGX3:1:23108:6598:11794	49	Putative ABC transporter C family member 15 (4+ species)
NB501358:228:HNY5YBGX3:2:13206:13969:2789	271	Peroxidase 55-like (Camilla sinensis)
NB501358:228:HNY5YBGX3:3:12603:9816:20043	209	Probable serine/threonine-protein kinase PBL1 (Brasicca Napus)
NB501358:228:HNY5YBGX3:3:21407:24289:4603	52	No significant similarity
NB501358:228:HNY5YBGX3:3:21412:20402:1561	57	IST1 homolog proetin encoding
NB501358:228:HNY5YBGX3:3:23401:5304:17780	102	No significant similarity
NB501358:228:HNY5YBGX3:4:23607:8416:4204	206	Protein root primordium defective, protein coding (Vitis vinifera, Capsicum annuum)
NB501358:230:HNJNHBGX3:1:11310:9664:5086	203	No significant similarity
NB501358:230:HNJNHBGX3:1:21101:3107:8118	10	No significant similarity
NB501358:230:HNJNHBGX3:1:22210:8110:19109	251	BAG family molecular chaperone regulator 6
NB501358:230:HNJNHBGX3:2:11201:13168:19171	115	No significant similarity
NB501358:230:HNJNHBGX3:2:21307:21454:18093	6	No significant similarity
NB501358:230:HNJNHBGX3:3:13507:16723:16423	117	AP-1 complex subunit gamma-2-like (Olea europaea var. sylvestris)
NB501358:230:HNJNHBGX3:4:11406:23567:18830	9	No significant similarity
NB501358:230:HNJNHBGX3:4:	43	Same contig as above

11406:23567:18830		
NB501358:230:HNJNHBGX3:4: 11406:23567:18830	77	Same contig as above

LITERATURE CITED:

Abraham, Gad, Yixuan Qiu, and Michael Inouye. 2017. "FlashPCA2: Principal Component Analysis of Biobank-Scale Genotype Datasets." *Bioinformatics* 33 (17): 2776–78.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs." *Nucleic Acids Research* 25 (17): 3389–3402.

Andrews, Kimberly R., Jeffrey M. Good, Michael R. Miller, Gordon Luikart, and Paul A. Hohenlohe. 2016. "Harnessing the Power of RADseq for Ecological and Evolutionary Genomics." *Nature Reviews. Genetics* 17 (2): 81–92.

Baldwin, Bruce G., Douglas Goldman, David J. Keil, Robert Patterson, Thomas J. Rosatti, and Dieter Wilken, eds. 2012. *The Jepson Manual: Vascular Plants of California*. Second Edition, Thoroughly Revised and Expanded edition. University of California Press.

Besnier, Francois, and Kevin A. Glover. 2013. "ParallelStructure: A R Package to Distribute Parallel Runs of the Population Genetics Program STRUCTURE on Multi-Core Computers." *PloS One* 8 (7): e70651.

Bowcock, A. M., A. Ruiz-Linares, J. Tomfohrde, E. Minch, J. R. Kidd, and L. L. Cavalli-Sforza. 1994. "High Resolution of Human Evolutionary Trees with Polymorphic Microsatellites." *Nature* 368 (6470): 455–57.

Bradburd, Gideon S., Graham M. Coop, and Peter L. Ralph. 2018. "Inferring Continuous and Discrete Population Genetic Structure Across Space." *Genetics* 210 (1): 33–52.

Brelsford, A., C. Dufresnes, and N. Perrin. 2016. "High-Density Sex-Specific Linkage Maps of a European Tree Frog (Hyla Arborea) Identify the Sex Chromosome without Information on Offspring Sex." *Heredity* 116 (2): 177–81.

Bryant, David, and Vincent Moulton. 2002. "NeighborNet: An Agglomerative Method for the Construction of Planar Phylogenetic Networks." *Lecture Notes in Computer Science*. https://doi.org/10.1007/3-540-45784-4_28.

———. 2004. "Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks." *Molecular Biology and Evolution* 21 (2): 255–65.

Burge, Dylan O., Diane M. Erwin, Melissa B. Islam, Jürgen Kellermann, Steven W. Kembel, Dieter H. Wilken, and Paul S. Manos. 2011. "Diversification of Ceanothus (Rhamnaceae) in the California Floristic Province." *International Journal of Plant Sciences* 172 (9): 1137–64.

Burge, Dylan O., V. Thomas Parker, Margaret Mulligan, and César García Valderamma. 2018. "Conservation Genetics of the Endangered Del Mar Manzanita (*Arctostaphylos glandulosa* Subsp. *crassifolia*) Based On Rad Sequencing Data." *Madroño* 65 (3): 117–30.

Calsbeek, Ryan, John N. Thompson, and James E. Richardson. 2003. “Patterns of Molecular Evolution and Diversification in a Biodiversity Hotspot: The California Floristic Province.” *Molecular Ecology* 12 (4): 1021–29.

Carstens, Bryan C., Tara A. Pelletier, Noah M. Reid, and Jordan D. Satler. 2013. “How to Fail at Species Delimitation.” *Molecular Ecology* 22 (17): 4369–83.

Catchen, Julian, Paul A. Hohenlohe, Susan Bassham, Angel Amores, and William A. Cresko. 2013. “Stacks: An Analysis Tool Set for Population Genomics.” *Molecular Ecology* 22 (11): 3124–40.

Cavender-Bares, Jeannine. 2019. “Diversification, Adaptation, and Community Assembly of the American Oaks (Quercus), a Model Clade for Integrating Ecology and Evolution.” *The New Phytologist* 221 (2): 669–92.

Chatzimanolis, S., and M. S. Caterino. 2007. “Toward a Better Understanding of the ‘Transverse Range Break’: Lineage Diversification in Southern California.” *Evolution; International Journal of Organic Evolution*. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1558-5646.2007.00186.x.

Clarke, K. R. 1993. “Non-Parametric Multivariate Analyses of Changes in Community Structure.” *Austral Ecology*. https://doi.org/10.1111/j.1442-9993.1993.tb00438.x.

Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, et al. 2011. “The Variant Call Format and VCFtools.” *Bioinformatics* 27 (15): 2156–58.

Dent, A. Earl, and Bridgett M. vonHoldt. 2013. “STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method.” *Conservation Genetics Resources* 4 (2): 359–61.

Dray, Stéphane, Anne-Béatrice Dufour, and Others. 2007. “The ade4 Package: Implementing the Duality Diagram for Ecologists.” *Journal of Statistical Software* 22 (4): 1–20.

Dufresne, France, Marc Stift, Roland Vergilino, and Barbara K. Mable. 2014. “Recent Progress and Challenges in Population Genetics of Polyploid Organisms: An Overview of Current State-of-the-Art Molecular and Statistical Tools.” *Molecular Ecology* 23 (1): 40–69.

Evanno, G., S. Regnaut, and J. Goudet. 2005. “Detecting the Number of Clusters of Individuals Using the Software STRUCTURE: A Simulation Study.” *Molecular Ecology* 14 (8): 2611–20.

Falush, Daniel, Matthew Stephens, and Jonathan K. Pritchard. 2003. “Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies.” *Genetics* 164 (4): 1567–87.

Fletcher, Robert J., Jr, Andre Revell, Brian E. Reichert, Wiley M. Kitchens, Jeremy D. Dixon, and James D. Austin. 2013. “Network Modularity Reveals Critical Scales for Connectivity in Ecology and Evolution.” *Nature Communications* 4: 2572.

Forgy, E. W. 1965. “Cluster Analysis of Multivariate Data: Efficiency vs Interpretability of Classifications.” *Biometrics* 21: 768–69.

Forister, M. L., J. A. Fordyce, and A. M. Shapiro. 2004. “Geological Barriers and Restricted Gene Flow in the Holarctic Skipper Hesperia Comma (Hesperiidae).” *Molecular Ecology* 13 (11): 3489–99.

Franklin, Janet. 1998. “Predicting the Distribution of Shrub Species in Southern California from Climate and Terrain-Derived Variables.” *Journal of Vegetation Science* 9 (5): 733–48.

Frichot, Eric, Sean D. Schoville, Guillaume Bouchard, and Olivier François. 2013. “Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models.” *Molecular Biology and Evolution* 30 (7): 1687–99.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2001. *The Elements of Statistical Learning*. Vol. 1. Springer series in statistics New York.

Fulton, Robert E., and Lynn F. Carpenter. 1979. “Pollination, Reproduction, and Fire in California *Arctostaphylos*.” *Oecologia* 38: 147–57.

Garrison, Erik, and Gabor Marth. 2012. “Haplotype-Based Variant Detection from Short-Read Sequencing.” *arXiv [q-bio.GN]*. arXiv. http://arxiv.org/abs/1207.3907.

Gower, J. C. 1966. “Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis.” *Biometrika* 53 (3): 325–28.

Grant, V. 1972. *Plant Speciation*. Vol. 2. New York: Columbia University Press.

Grooten, M., and R. E. A. (eds) Almond. 2018. “Living Planet Report - 2018: Aiming Higher.” World Wildlife Fund.

Hagler, James R., Shannon Mueller, Larry R. Teuber, Scott A. Machtley, and Allen Van Deynze. 2011. “Foraging Range of Honey Bees, Apis Mellifera, in Alfalfa Seed Production Fields.” *Journal of Insect Science* 11: 144.

Haig, Susan M., Erik A. Beever, Steven M. Chambers, Hope M. Draheim, Bruce D. Dugger, Susie Dunham, Elise Elliott-Smith, et al. 2006. “Taxonomic Considerations in Listing Subspecies under the US Endangered Species Act.” *Conservation Biology: The Journal of the Society for Conservation Biology* 20 (6): 1584–94.

Harrison, Nicola, and Catherine Anne Kidner. 2011. “Next-Generation Sequencing and Systematics: What Can a Billion Base Pairs of DNA Sequence Data Do for You?” *TAXON*. https://doi.org/10.1002/tax.606002.

Harrison, Richard G., and Erica L. Larson. 2014. “Hybridization, Introgression, and the Nature of Species Boundaries.” *The Journal of Heredity* 105 Suppl 1: 795–809.

Hennig, Christian, and Bernhard Hausdorf. 2019. “Prabclus: Functions for Clustering and Testing of Presence-Absence, Abundance and Multilocus Genetic Data.” https://CRAN.R-project.org/package=prabclus.

Huson, Daniel H., and David Bryant. 2006. “Application of Phylogenetic Networks in Evolutionary Studies.” *Molecular Biology and Evolution* 23 (2): 254–67.

Jepson, Willis L. 1916. “Regeneration in Manzanita.” *Madroño* 1 (1): 3–12.

Jombart, T., D. Pontier, and A-B Dufour. 2009. “Genetic Markers in the Pconluslayground of Multivariate Analysis.” *Heredity* 102 (4): 330–41.

Jörger, Katharina M., and Michael Schrödl. 2013. “How to Describe a Cryptic Species? Practical Challenges of Molecular Taxonomy.” *Frontiers in Zoology* 10 (1): 59.

Kassambara, Alboukadel, and Fabian Mundt. 2016. “Factoextra: Extract and Visualize the Results of Multivariate Data Analyses.” *R Package Version* 1 (3): 2016.

Kauffmann, M., T. Parker, and M. Vasey. 2015. *Field Guide to Manzanitas: California, North America and Mexico*. Backcountry Press.

Keeley, Jon E., and Robert L. Hays. 1976. “Differential Seed Predation on Two Species of *Arctostaphylos* (Ericaceae).” *Oecologia* 24 (1): 71–81.

Keeley, Jon E., Michael C. Vasey, and V. Thomas Parker. 2007. “Subspecific Variation in the Widespread Burl-Forming *Arctostaphylos glandulosa*.” *Madroño* 54 (1): 42–62.

Keller, Reuben P., Juergen Geist, Jonathan M. Jeschke, and Ingolf Kühn. 2011. “Invasive Species in Europe: Ecology, Status, and Policy.” *Environmental Sciences Europe* 23 (1): 23.

Lachmuth, Susanne, Walter Durka, and Frank M. Schurr. 2010. “The Making of a Rapid Plant Invader: Genetic Diversity and Differentiation in the Native and Invaded Range of Senecio Inaequidens.” *Molecular Ecology* 19 (18): 3952–67.

Li, Heng. 2013. “Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM.” *PREPRINT* 00. http://github.com/lh3/bwa.

Li, Weizhong, and Adam Godzik. 2006. “Cd-Hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences.” *Bioinformatics* 22 (13): 1658–59.

Mace, Georgina M. 2004. “The Role of Taxonomy in Species Conservation.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 359 (1444): 711–19.

MacQueen, J. 1967. *Some Methods for Classification and Analysis of Multivariate Observations*. Edited by L. M. Le Cam and J. Neyman. Vol. 1. Berkeley: University of California Press.

Martien, Karen K., Matthew S. Leslie, Barbara L. Taylor, Phillip A. Morin, Frederick I. Archer, Brittany L. Hancock-Hanser, Patricia E. Rosel, Nicole L. Vollmer, Amélia Viricel, and Frank Cipriano. 2017. "Analytical Approaches to Subspecies Delimitation with Genetic Data." *Marine Mammal Science* 33 (S1): 27–55.

Moore, Christopher M., and Stephen B. Vander Wall. 2015. "Scatter-Hoarding Rodents Disperse Seeds to Safe Sites in a Fire-Prone Ecosystem." *Plant Ecology* 216 (8): 1137–53.

Nei, M., and S. Kumar. 2000. *Kumar, S.: Molecular Evolution and Phylogenetics*.

Osborne, Juliet L., Andrew P. Martin, Norman L. Carreck, Jennifer L. Swain, Mairi E. Knight, Dave Goulson, Roddy J. Hale, and Roy A. Sanderson. 2008. "Bumblebee Flight Distances in Relation to the Forage Landscape." *The Journal of Animal Ecology* 77 (2): 406–15.

Parchman, Thomas L., Zachariah Gompert, Joann Mudge, Faye D. Schilkey, Craig W. Benkman, and C. Alex Buerkle. 2012. "Genome-Wide Association Genetics of an Adaptive Trait in Lodgepole Pine." *Molecular Ecology* 21 (12): 2991–3005.

Parker, V. Thomas. 2015. "Dispersal Mutualism Incorporated into Large-Scale, Infrequent Disturbances." *PloS One* 10 (7): e0132625.

Patten, Michael A. 2015. "Subspecies and the Philosophy of Science." *The Auk* 132 (2): 481–85.

Patten, Michael A., Philip Unitt, and F. Sheldon. 2002. "Diagnosability Versus Mean Differences of Sage Sparrows." *The Auk* 119 (1): 26–35.

Peterson, Brant K., Jesse N. Weber, Emily H. Kay, Heidi S. Fisher, and Hopi E. Hoekstra. 2012. "Double Digest RADseq: An Inexpensive Method for de Novo SNP Discovery and Genotyping in Model and Non-Model Species." *PloS One* 7 (5): e37135.

Pimm, S. L., C. N. Jenkins, R. Abell, T. M. Brooks, J. L. Gittleman, L. N. Joppa, P. H. Raven, C. M. Roberts, and J. O. Sexton. 2014. "The Biodiversity of Species and Their Rates of Extinction, Distribution, and Protection." *Science* 344 (6187): 1246752.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. "Inference of Population Structure Using Multilocus Genotype Data." *Genetics* 155 (2): 945–59.

Purcell, Shaun, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, Julian Maller, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *American Journal of Human Genetics* 81 (3): 559–75.

Ramsey, Justin, and Douglas W. Schemske. 1998. "Pathways, Mechanisms, and Rates of Polyploid Formation in Flowering Plants." *Annual Review of Ecology and Systematics* 29 (1): 467–501.

Razkin, Oihana, Benjamín J. Gómez-Moliner, Katerina Vardinoyannis, Alberto Martínez-Ortí, and María J. Madeira. 2017. "Species Delimitation for Cryptic Species Complexes: Case Study of Pyramidula (Gastropoda, Pulmonata)." *Zoologica Scripta* 46 (1): 55–72.

R Core Team. 2018. "R: A Language and Environment for Statistical Computing." Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Regan, Helen M., Mark Colyvan, and Mark A. Burgman. 2002. "A Taxonomy and Treatment of Uncertainty for Ecology and Conservation Biology." *Ecological Applications: A Publication of the Ecological Society of America* 12 (2): 618–28.

Renwick, Anna R., Catherine J. Robinson, Stephen T. Garnett, Ian Leiper, Hugh P. Possingham, and Josie Carwardine. 2017. "Mapping Indigenous Land Management for Threatened Species Conservation: An Australian Case-Study." *PloS One* 12 (3): e0173876.

Rodzen, Jeff A., Thomas R. Famula, and Bernie May. 2004. "Estimation of Parentage and Relatedness in the Polyploid White Sturgeon (Acipenser Transmontanus) Using a Dominant Marker Approach for Duplicated Microsatellite Loci." *Aquaculture* 232 (1): 165–82.

Rousseeuw, Peter J. 1987. "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis." *Journal of Computational and Applied Mathematics*. https://doi.org/10.1016/0377-0427(87)90125-7.

Schierenbeck, Kristina A. 2017. "Population-Level Genetic Variation and Climate Change in a Biodiversity Hotspot." *Annals of Botany* 119 (2): 215–28.

Schliep, Klaus, Alastair Alastair Potts, David A. Morrison, and Guido W. Grimm. 2016. "Intertwining Phylogenetic Trees and Networks." e2054v1. PeerJ Preprints. https://doi.org/10.7287/peerj.preprints.2054v1.

Sgariglia, Erik A., and Kevin J. Burns. 2003. "Phylogeography of the California Thrasher (Toxostoma Redivivum) Based on Nested-Clade Analysis of Mitochondrial-DNA Variation." *The Auk* 120 (2): 346–61.

Sokal, Robert R. 1979. "Testing Statistical Significance of Geographic Variation Patterns." *Systematic Zoology* 28 (2): 227–32.

Sork, Victoria L., Paul F. Gugger, Jin-Ming Chen, and Silke Werth. 2016. "Evolutionary Lessons from California Plant Phylogeography." *Proceedings of the National Academy of Sciences of the United States of America* 113 (29): 8064–71.

Stobie, Connor Seamus, Carel J. Oosthuizen, Michael J. Cunningham, and Paulette Bloomer. 2018. "Exploring the Phylogeography of a Hexaploid Freshwater Fish by RAD Sequencing." *Ecology and Evolution* 8 (4): 2326–42.

Thomas, J. A., M. G. Telfer, D. B. Roy, C. D. Preston, J. J. D. Greenwood, J. Asher, R. Fox, R. T. Clarke, and J. H. Lawton. 2004. "Comparative Losses of British Butterflies, Birds, and Plants and the Global Extinction Crisis." *Science* 303 (5665): 1879–81.

Tibshirani, Robert, Guenther Walther, and Trevor Hastie. 2001. "Estimating the Number of Clusters in a Data Set Viq the Gap Statistic." *Journal of the Royal Statistical Society Bulletin* 63 (2): 411–23.

Walsh, Jennifer, Irby J. Lovette, Virginia Winder, Chris S. Elphick, Brian J. Olsen, Gregory Shriver, and Adrienne I. Kovach. 2017. "Subspecies Delineation amid Phenotypic, Geographic and Genetic Discordance in a Songbird." *Molecular Ecology* 26 (5): 1242–55.

Wells, Philip V. 1968. "New Taxa, Combinations, and Chromosome Numbers in *Arctostaphylos* (Ericaceae)." *Madroño* 19 (6): 193–210.

Wells, Phillip V. 2000. *Manzanitas of California*. Dept. of Ecology & Evolutionary Biology University of Kansas.

Wells, P. V. 1992. "Four New Species of *Arctostaphylos* from Southern California and Baja California." *Four Seasons* 9 (2): 44–53.

———. 2000. *The Manzanitas of California: Also of Mexico and the World*. P.V. Wells.

Wieslander, A. E., and Beryl O. Schreiber. 1939. "Notes on the Genus *Arctostaphylos*." *Madroño* 5 (1): 38–47.

Zhang, Jiajie, Kassian Kobert, Tomáš Flouri, and Alexandros Stamatakis. 2014. "PEAR: A Fast and Accurate Illumina Paired-End reAd mergeR." *Bioinformatics* 30 (5): 614–20.

Zink, Robert M., Dale F. Lott, and Daniel W. Anderson. 1987. "Genetic Variation, Population Structure, and Evolution of California Quail." *The Condor* 89 (2): 395–405.

Zurbuchen, Antonia, Lisa Landert, Jeannine Klaiber, Andreas Müller, Silke Hein, and Silvia Dorn. 2010. "Maximum Foraging Ranges in Solitary Bees: Only Few Individuals Have the Capability to Cover Long Foraging Distances." *Biological Conservation* 143 (3): 669–76.

APPENDICES:

Appendix 3.1: Eastwood manzanita (*Arctostaphylos glandulosa*) collections included in genetic analyses.

Determination,County,Country,Voucher deposited,Collection number,

A. glandulosa* Eastw. subsp. *adamsii* (Munz) Munz**, San Diego, USA, UCR, Y. Huang, 317-04, Y. Huang, 317-05, Y. Huang, 317-31, Y. Huang, 317-35, Y. Huang, 317-36, Y. Huang, 418-01, Y. Huang, 418-02 ***A. glandulosa* Eastw. subsp. *crassifoila* (Jeps.) P.V. Wells**, San Diego, USA, UCR, Y. Huang, 618-02 ***A. glandulosa* Eastw. subsp. *crassifolia* (Jeps.) P.V. Wells**, San Diego, USA, DAV, Dylan O. Burge 1,729 ***A. glandulosa* Eastw. subsp. *crassifolia* (Jeps.) P.V. Wells**, San Diego, USA, SD, Dylan O. Burge 2,071 ***A. glandulosa* Eastw. subsp. *crassifolia* (Jeps.) P.V. Wells**, San Diego, USA, UCR, A. C. Sanders 42,680, A. C. Sanders 42,682, A. C. Sanders 42,683, A. C. Sanders 42,684, Y. Huang, 618-01, Y. Huang, 618-03, Y. Huang, 618-04, Y. Huang, 618-05 ***A. glandulosa* Eastw. subsp. *cushingiana* (Eastw.) J.E. Keeley, M.C. Vasey & V.T. Parker**, San Deigo, USA, UCR, Y. Huang, 818-04, A. C. Sanders 42,687, A. C. Sanders 42,689, Y. Huang, 818-01, Y. Huang, 818-02, Y. Huang, 818-03, Y. Huang, 818-05 ***A. glandulosa* Eastw. subsp. *cushingiana* (Eastw.) J.E. Keeley, M.C. Vasey & V.T. Parker**, Santa Barbara, USA, UCR, Y. Huang, 319-16 ***A. glandulosa* Eastw. subsp. *cushingiana* (Eastw.) J.E. Keeley, M.C. Vasey & V.T. Parker**, Ventura, USA, UCR, Y. Huang, 718-35, Y. Huang, 718-36, Y. Huang, 718-37, Y. Huang, 718-38, Y. Huang, 718-39 ***A. glandulosa* Eastw. subsp. *erecta* J.E. Keeley, M.C. Vasey & V.T. Parker**, N.A. / Baja CA, MEX, SD, "Dylan O. Burge 2,017, Dylan O. Burge 2,019 ***A. glandulosa* Eastw. subsp. *gabrielensis* (P.V. Wells) J.E. Keeley, M.C. Vasey & V.T. Parker**, Los Angeles, USA, UCR, Y. Huang, 417-01, Y. Huang, 417-02, Y. Huang, 417-03, Y. Huang, 417-04, Y. Huang, 417-05, Y. Huang, 618-12, Y. Huang, 618-13, Y. Huang, 618-14, Y. Huang, 618-15, Y. Huang, 618-16 ***A. glandulosa* Eastw. subsp. *glandulosa, Los Angeles, USA, UCR, A. C. Sanders 42,880, A. C. Sanders 42,880, Y. Huang, 218-01, Y. Huang, 218-03, Y. Huang, 218-04, Y. Huang, 218-05, Y. Huang, 218-06 ***A. glandulosa* Eastw. subsp. *glandulosa***, Riverside, USA, UCR, A. C. Sanders 42,668, A. C. Sanders 42,669, A. C. Sanders 42,670, Y. Huang, 317-09, Y. Huang, 317-10, Y. Huang, 317-11, Y. Huang, 317-12, Y. Huang, 317-13, Y. Huang, 317-14, Y. Huang, 317-15, Y. Huang, 317-16, Y. Huang, 317-17, Y. Huang, 317-18, Y. Huang, 317-19, Y. Huang, 317-20, Y. Huang, 317-21, Y. Huang, 317-22, Y. Huang, 317-23, Y. Huang, 317-24, Y. Huang, 317-25" ***A. glandulosa* Eastw. subsp. *glandulosa***, San Bernardino, USA, UCR, A. C. Sanders 42,813, A. C. Sanders 42,872, A. C. Sanders 42,874, A. C. Sanders 42,875, A. C. Sanders 42,876, A. C. Sanders 42,877, A. C. Sanders 42,878 ***A. glandulosa* Eastw. subsp. *glandulosa***, San Diego, USA, UCR, A. C. Sanders 42,693, A. C. Sanders 42,695, Y. Huang, 818-07, Y. Huang, 818-11, Y. Huang, 818-14 ***A. glandulosa* Eastw. subsp. *glandulosa***, Santa Barbara, USA, UCR, Y. Huang, 319-13, Y. Huang, 718-06, Y. Huang, 718-07, Y. Huang, 718-08, Y. Huang, 718-09, Y. Huang, 718-31, Y. Huang, 718-32, Y. Huang, 718-33, Y. Huang, 718-34 ***A. glandulosa* Eastw. subsp. *leucophylla***

J.E. Keeley, M.C. Vasey & V.T. Parker, Orange, USA, UCR, A. C. Sanders 42,701, A. C. Sanders 42,697, A. C. Sanders 42,698, A. C. Sanders 42,699, A. C. Sanders 42,700, A. C. Sanders 42,702, A. C. Sanders 42,703 ***A. glandulosa* Eastw. subsp. *leucophylla* J.E. Keeley, M.C. Vasey & V.T. Parker**, Riverside, USA, UCR, A. C. Sanders 42,673, A. C. Sanders 42,676 ***A. glandulosa* Eastw. subsp. *leucophylla* J.E. Keeley**, M.C. Vasey & V.T. Parker, San Diego, USA, UCR, Y. Huang, 317-01, Y. Huang, 317-02, Y. Huang, 317-03, Y. Huang, 317-06, Y. Huang, 317-07, Y. Huang, 317-08, Y. Huang, 317-26, Y. Huang, 317-27, Y. Huang, 317-28, Y. Huang, 317-29, Y. Huang, 317-30, Y. Huang, 317-32, Y. Huang, 317-33, Y. Huang, 317-34, Y. Huang, 317-37, Y. Huang, 317-38, Y. Huang, 317-39 ***A. glandulosa* Eastw. subsp. *leucophylla* J.E. Keeley, M.C. Vasey & V.T. Parker**, Santa Barbara, USA, UCR, Y. Huang, 319-18 ***A. glandulosa* Eastw. subsp. *mollis* (J.E. Adams) P.V. Wells**, Los Angeles, USA, UCR, Y. Huang, 618-06, Y. Huang, 618-07, Y. Huang, 618-08, Y. Huang, 618-09 ***A. glandulosa* Eastw. subsp. *mollis* (J.E. Adams) P.V. Wells**, San Bernardino, USA, UCR, A. C. Sanders 42,870, A. C. Sanders 42,871, A. C. Sanders 42,879 ***A. glandulosa* Eastw. subsp. *mollis* (J.E. Adams) P.V. Wells**, Santa Barbara, USA, UCR ,Y. Huang, 318-01, Y. Huang, 318-02, Y. Huang, 318-03, Y. Huang, 318-05, Y. Huang, 318-06, Y. Huang, 319-08, Y. Huang, 319-09, Y. Huang, 319-11, Y. Huang, 319-12, Y. Huang, 319-7 ***A. glandulosa* Eastw. subsp. undetermined**, N.A./Baja CA, MEX, SD, Dylan O. Burge 2,013, Dylan O. Burge 2,015

Appendix 3.2: Collection localities for the five diploid *Arctostaphylos* species collected in southern California. Asterisk represents the coordinates that are not recorded in the field and are approximated after collection.

Species, County, State, Voucher deposited, Collection number

***A. glauca* Lindl.**, San Luis Obispo, CA, DAV, Dylan O. Burge 2,024 ***A. glauca* Lindl.**, San Bernardino, CA, UCR, Yi Huang, 218-01, Yi Huang, 218-02 ***A. glauca* Lindl.**, Santa Barbara, CA, UCR, Yi Huang, 318-04, Yi Huang, 318-09 ***A. glauc*a Lindl.**, Ventura, CA, UCR, Yi Huang, 318-19 ***A. pringlei* Parry**, Riverside, CA, No specimen available, Dylan O. Burge JmsRsr-35, Dylan O. Burge JmsRsr-37 ***A. pungens* Kunth**, San Diego, CA, UCR, A. C. Sanders 43,396, A. C. Sanders 43,401, Yi Huang, 218-06 ***A. purissima* P.V. Wells**, Santa Barbara, CA, UCR, Yi Huang, 718-10, Yi Huang, 718-12, Yi Huang, 718-14 ***A. rainbowesis* J.E. Keeley & Massihi**, San Diego, CA, UCR, A. C. Sanders 43,392, A. C. Sanders 43,394, A. C. Sanders 43,395

CONCLUSION:

In my study of the biogeographic histories of three widespread North American manzanita species (chapter 1), I found best supported hypotheses for two species showing likely areas of origin within (*Arctostaphylos patula*), and outside of (*A. pungens*), the California Floristic Province (CFP). Each of these lends support for one of the two conflicting hypotheses regarding the origins of CFP diversity. For the third species (*A. pringlei*), a single likely area of origin could not be identified, and the most likely biogeographic history indicated is that the species came to be distributed in two parts (one in the CFP, and one outside) due to a vicariance event splitting a previously contiguous, or near-contiguous, range. The study thus did not provide support for any one hypothesis regarding the history of endemic diversity in the CFP, but rather shows that the historical biogeographical context of *Arctostaphylos* evolution may be complex, and areas outside of the CFP may have been important in the diversification of the genus.

In my study of morphological distinguishability and distinctiveness of manzanita taxa, I asked (1) whether currently recognized taxa are distinguishable on a morphological basis alone, without information on geographic origin, and (2) whether the structure of morphological variation corresponds to the current taxonomy. I answered the first question in the affirmative, finding that even while using a dataset of only vegetative traits, taxa were distinguishable to a high degree, with subspecies being less distinguishable, on average, than species taxa. The lesser degree of distinguishability among subspecies, however, is not unexpected, given those taxa are understood to be more genetically, and therefore probably more morphologically, similar to each other than species would be. While I did find high distinguishability overall, I also found that distinguishability of taxa stems from taxonomically diagnostic traits, i.e. specific trait values and conditions linked to specific taxa. This reflects how taxa are described/identified, and does not

correspond to lack of morphological overlap among taxa. I answered the second question in the negative, finding unstable results in attempts to cluster samples based on morphology, indicating that there may be a lack of clear morphological boundaries among taxa, at least when using a vegetative dataset.

In my third chapter, I asked whether subspecies of *A. glandulosa* are genetically distinct, and found the answer is mostly that they are not. One subspecies, *A. glandulosa* subsp. *gabrielensis*, did show genetic differentiation, which may be a result of admixture with another species. I found that instead of forming clusters corresponding to subspecies identity, samples of *A. glandulosa* showed geographic genetic structure, forming a gradient of variation across the sampled range. This suggests that at least in this species, subspecies may only correspond to some set of genotypes responsible for the traits used in description and identification, meaning the subspecies would be of doubtful evolutionary significance.

The results of these studies, taken together, show the promise of revisiting historically challenging groups like *Arctostaphylos* with modern methods, but simultaneously show that modern data or analyses do not necessarily solve historical problems that are anchored in philosophy of biology. For example, questions like how distinct a population or group of populations needs to be, genetically or morphologically, to be considered a species or a subspecies, still need to be considered. Nonetheless, the continued study of difficult groups like *Arctostaphylos* advances our understanding of real-world, natural variation, in all of its complexity, and thus puts us in a better position to consider how we should describe natural variation.

Addendum 1. Summary of Contributions to California Conservation Genomics Project and Future Plans for Continued Research.

During my dissertation research, I did extensive work on an additional project, which had not produced data by the time of the completion of my dissertation, due to long-term delays at an outside sequencing facility. This addendum is a note describing the work I did on that project, and the work I plan to do when data are received.

This project is part of a much larger, state-funded initiative called the California Conservation Genomics Project (CCGP) (Shaffer et al. 2022). The ultimate goal of the CCGP is to produce a map of genetic diversity across the state, calculated from a sample of plant, animal and fungal life. This map would serve as a basis for the discovery of areas of the state with low or high genetic diversity across diverse and distantly related lineages, and give conservation practitioners and state policy-makers new evidence to take actions aimed at conserving and preserving California's diversity. Specifically, the goal is to calculate measures of genetic diversity using whole genome (re)sequencing (WGS) data, following the assembly of a chromosome-scale genome assembly for each species, or genus.

In order to achieve this goal, landscape genomic studies were undertaken for species across the tree of life. We sampled a widespread manzanita species, bigberry manzanita, (*Arctostaphylos glauca* Lindl.) across its range (Fig. 4.1), and six narrow endemic species, all of which occur within or near the range *A. glauca*. These species are *A. pajaroensis* (J.E.Adams ex McMinn) J.E.Adams, *A. montereyensis* Hoover, *A. luciana* P.V.Wells, *A. pechoensis* (Dudley ex Abrams) Dudley ex Munz*, A. rainbowensis* J.E.Keeley & A.Massihi*,* and *A. otayensis* Wiesl. & B. Schrieb.

We collected 117 samples of *A. glauca* and approximately 15 samples of each narrow endemic species (Fig. 4.1). I procured permits for collections, and planned and conducted the majority of collections for this project, in association with Yi Huang, Tito Abbo, and Angela Buehlman. I extracted DNA from the majority of these samples and sent the DNA and associated data on collection locations to the CCGP core facilities in January of 2023, but due to back-logs in the CCGP sequencing core, the WGS data have not yet been received. However, our reference genome assembly was completed in 2022 (Huang et al. 2022), and was the first of any of the CCGP genome assemblies to be published.

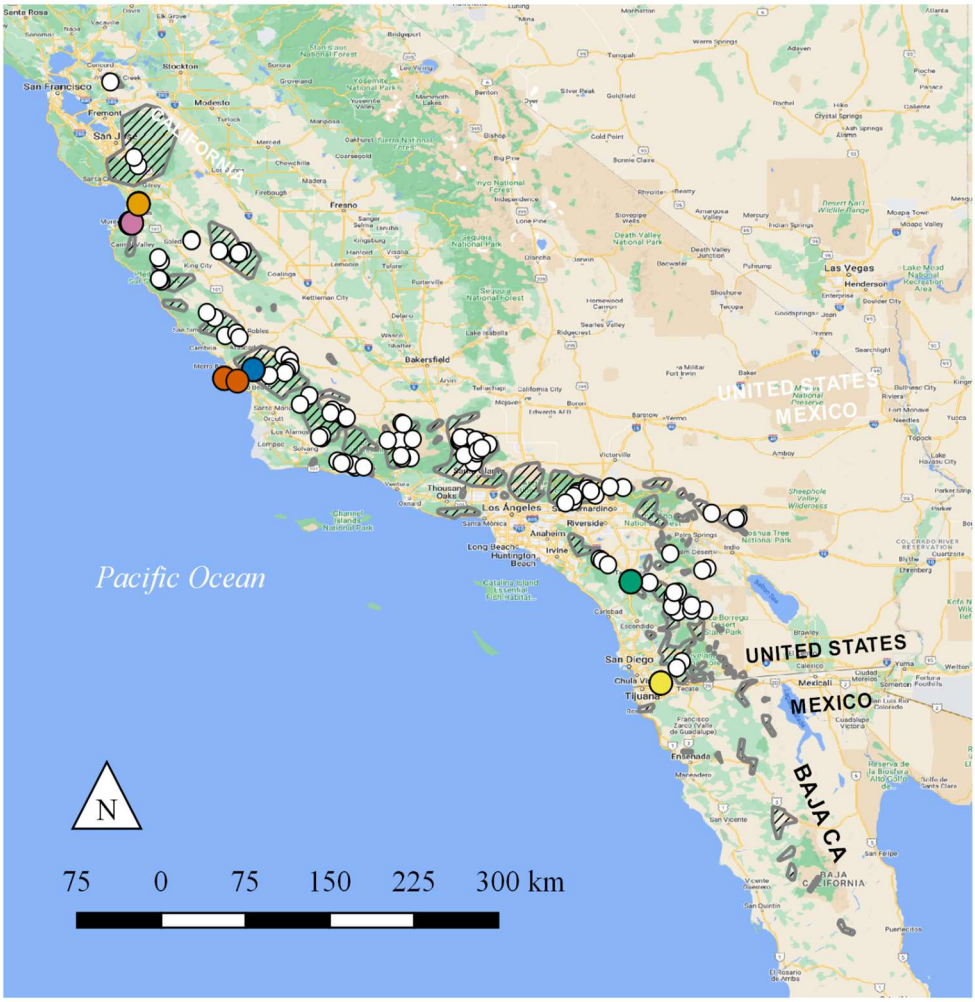
When these data are received, I will conduct landscape genomic analyses on *A. glauca*, to map several measures of genetic diversity (e.g. nucleotide diversity, heterozygosity, and allelic richness) in a rasterized, sliding-window across its range in California. These calculations can now be performed using an R package developed by the CCGP computational core, specifically to enable these kinds of map-based calculations in a regular way across individual CCGP projects (Bishop, Chambers, and Wang 2023). I will calculate these same statistics for each sample of the six narrow endemics to test whether genetic diversity is different in narrow endemic species compared to widespread species (e.g. whether measures of genetic diversity are lower in narrow endemics than for nearby *A. glauca*). Additionally, I plan to do comparisons of the genetic diversity of the narrow endemic species to that of *A. glauca* while correcting for range size.

In addition to the analyses based on estimating population genetic parameters, which are core goals of the CCGP initiative, the *A. glauca* dataset will also allow for more in-depth exploration of landscape patterns of genetic structure, e.g. in relation to geographic boundaries and gaps in the range of *A. glauca.* Insights gleaned from this highly spatially dispersed and rich dataset can help us understand how the landscape of California (topography, climate, etc) impacts

gene flow across the state. This will be of particular interest to anyone studying similar chaparral shrubs of California and the California Floristic Province. I will also have the opportunity to look for genetic loci correlated with environmental variation (e.g. using a genome-wide association study, GWAS). This dataset will be well suited for analysis like a GWAS, as the size of the dataset is large, both in terms of genome coverage and the number of individuals, and the amount of climate variability across the species range is also large.

Finally, looking to my results in chapter one of this dissertation, in which I detected interspecific gene flow across a wide geographic area, it is likely that a species like *A. glauca*, which has a large range and comes into close proximity with numerous other manzanita species, will have some level of interspecific gene flow, as well. Some species are already hypothesized to have a history of admixture with *A. glauca* (Gankin 1967; Keeley 1976; Parker and Vasey 2004), and these data may give us empirical insights into the plausibility of that hypothesized gene flow. While I will not have complementary WGS data for those species suspected to be involved in admixture with *A. glauca*, I will have double-digest restriction site-associated (ddRADseq) data available for those species suspected to have been involved in admixture with *A. glauca*, and that will allow me to create a dataset of common loci by subsampling the data from WGS dataset to correspond to ddRADseq loci. I will also have WGS data for the six narrow endemic species included in the CCGP dataset, and I plan to look for signs of admixture between *A. glauca* and those species as well; the ranges of those species either overlap the range of *A. glauca*, or are close to it, providing potential for interspecific gene flow.

FIGURES:



A. glauca range

A. glauca collections

Narrow endemic species

A. pajaroensis collections

A. pechoensis collections

A. montereyensis collections

A. rainbowensis collections

A. luciana collections

A. otayensis collections

Fig. 4.1. Map showing the range of *A. glauca* (polygons with hatched lines), along with the locations sampled for the CCGP study. The polygons showing the range of *A. glauca* were hand drawn to approximate the range, as evinced by georeferenced herbarium collections from the Consortium of California Herbaria (www.cch2.org) (Consortium of California Herbaria n.d.) and the *Field Guide to Manzanitas (Kauffmann, Parker, and Vasey 2021)*.

LITERATURE CITED:

Bishop, Anusha P., E. Anne Chambers, and Ian J. Wang. 2023. "Generating Continuous Maps of Genetic Diversity Using Moving Windows." *Methods in Ecology and Evolution / British Ecological Society* 14 (5): 1175–81.

Consortium of California Herbaria. n.d. "Complete CCH2 Records Search for *Arctostaphylos*." CCH2 Data Portal. Accessed September 7, 2023. www.cch2.org.

Gankin, R. 1967. "A New Species of *Arctostaphylos* from Santa Barbara County, California." *The Four Seasons*.

Huang, Yi, Merly Escalona, Glen Morrison, Mohan P. A. Marimuthu, Oanh Nguyen, Erin Toffelmier, H. Bradley Shaffer, and Amy Litt. 2022. "Reference Genome Assembly of the Big Berry Manzanita (*Arctostaphylos Glauca*)." *The Journal of Heredity* 113 (2): 188–96.

Kauffmann, Michael Edward, Tom Parker, and Michael Vasey. 2021. *Field Guide to Manzanitas: California, North America, and Mexico, Second Edition*. Backcountry Press.

Keeley, Jon E. 1976. "Morphological Evidence of Hybridization Between *Arctostaphylos Glauca* and *A. Pungens* (Ericaceae)." *Madroño* 23 (8): 427–34.

Parker, V. T., and M. C. Vasey. 2004. "*Arctostaphylos Gabilanensis* (Ericaceae), a Newly Described Auriculate-Leaved Manzanita from the Gabilan Mountains, California." *Madroño* 51 (3): 322–25.

Shaffer, H. Bradley, Erin Toffelmier, Russ B. Corbett-Detig, Merly Escalona, Bjorn Erickson, Peggy Fiedler, Mark Gold, et al. 2022. "Landscape Genomics to Enable Conservation Actions: The California Conservation Genomics Project." *The Journal of Heredity* 113 (6): 577–88.

Addendum 2. Collection and Analysis of Morphological Data in a Classroom Setting.

As part of the methods development for chapter 2 (*Morphological Analysis of Manzanita Species and Subspecies*), I explored a classroom-based approach to data collection. This was motivated by an interest in bringing students into the process of an authentic research project, but was also in finding a method of data collection that would be higher throughput than could be achieved by myself and few others. By involving students in data collection in the classroom environment, I hoped to give them a low-stakes, and accessible, research experience that could spark further interest in research. I began this as a collaborative project with Dr. Diana Jolles of Plymouth State University, a small, public, primarily-undergraduate university in Plymouth, New Hampshire. In spring of 2022, I participated remotely in Diana's Botany course (BL 2070), giving a guest lecture and participating in two meetings of the lab class.

In the guest lecture I presented a brief overview of California plant ecology, an overview of manzanitas, and specific information on the study to which they would be contributing. During the lab activity I provided general instruction on the use of imageJ (Schneider et al., 2012), and then provided each student with a set of links to scanned images of herbarium specimens, publicly posted on the database of the California Consortium of Herbaria (Consortium of California Herbaria, 2023) (see Fig. 5.1 for example). I led the students in measurements of their first specimen, explaining each variable before moving on to the next, and after training them on the process, released them to gather data for several more specimens on their own, encouraging them to ask for help as needed. Each specimen was measured by two students, as a cross-check. Dr. Jolles and an undergrad teaching assistant, Tina Andreski, facilitated the measurement process in-person. Several traits that could not be easily assessed or measured from an image were not included in this activity (e.g. hair length and density). At the end of this activity I

curated the resulting data, checking for numbers reported in the wrong units, and standardizing the encoding of categorical variables. I then designed and led a lab class activity, where in the students installed R, and performed analyses such as Principal Components Analysis (Minchin, 1987) to test the morphological distinction among several taxa included in the dataset, and making basic plots of results.

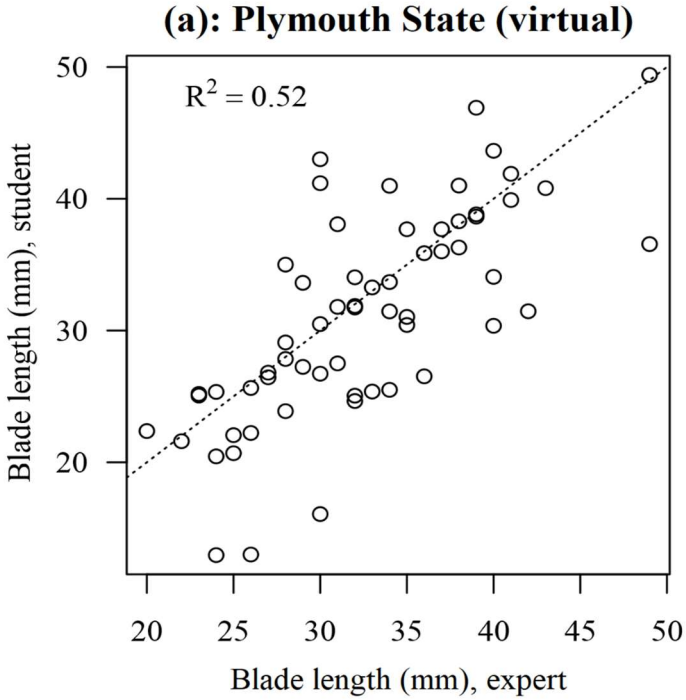
I performed additional analysis of the student data from Plymouth State after the classroom activities, comparing the data the students produced to measurements I made on the same specimens. I found there was substantial error in this data, as compared to the data I collected myself. Additionally, while duplicated student measurements were significantly correlated, it was a weak correlation, due to substantial noise (see example in 2). For some variables, it was clear that students misunderstood measurements (e.g. a leaf midrib length reported as greater than the blade length on the same leaf. Overall, analysis showed that the data could not be relied upon for research purposes (see one example in Fig. 5.3).

In the Spring of 2023, we repeated the activity, this time in Plant Diversity and Evolution (BPSC 133) at UC Riverside. Taking lessons from the previous trial at Plymouth State, I provided more reference materials to guide the students in measuring variables correctly, and I was also able to be present in person to instruct them in the process, and help them with uncertainties as they recorded the actual data. This yielded data more consistent with expert measurements (see example in Fig. 5.2b). In redesigning the activity for implementation in BPSC 133, I made several changes aimed at improving the reliability of the data to be produced, for example providing more diagrams illustrating how to measure variables, as well renaming certain variables, based on feedback from Plymouth State activity. Some or all of those changes improved the quality of the data, although, I cannot say which changes were responsible.

For certain kinds of studies, methods like this may work well to accumulate large amounts of data quickly, however, it requires extensive preparation, and a sizable amount of instruction and oversight to produce reliable data, undercutting the efficiency gains of such classroom-based data collection methods. This trade-off is less of a problem if a large number of observations for a small number of variables are to be recorded, and is a larger problem if many variables are to be recorded, as each variable requires instruction and oversight.

FIGURES:

Fig. 5.1. Example scan image of a herbarium specimen, publicly available on the Consortium of California Herbaria database (cch2.org).


(a): Plymouth State (virtual)
$R^2 = 0.52$
Blade length (mm), student
Blade length (mm), expert
50
40
30
20
20
25
30
35
40
45
50

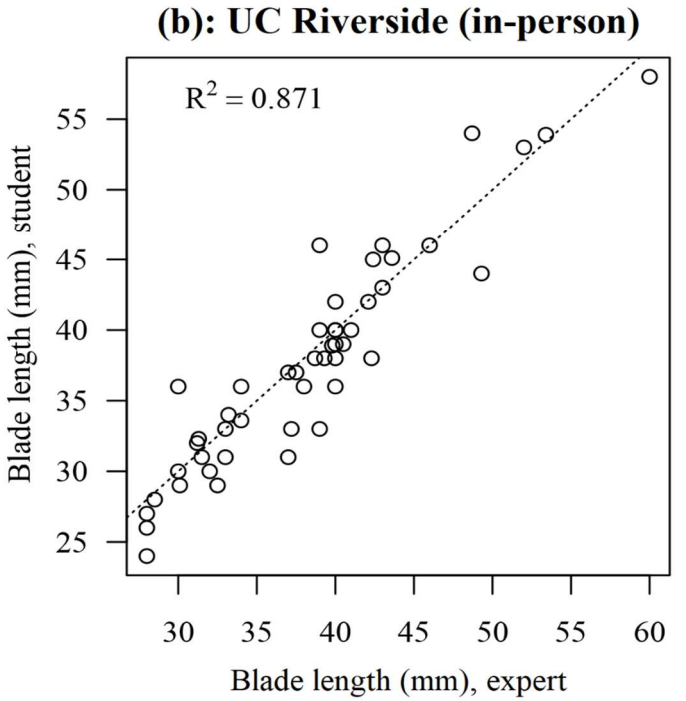
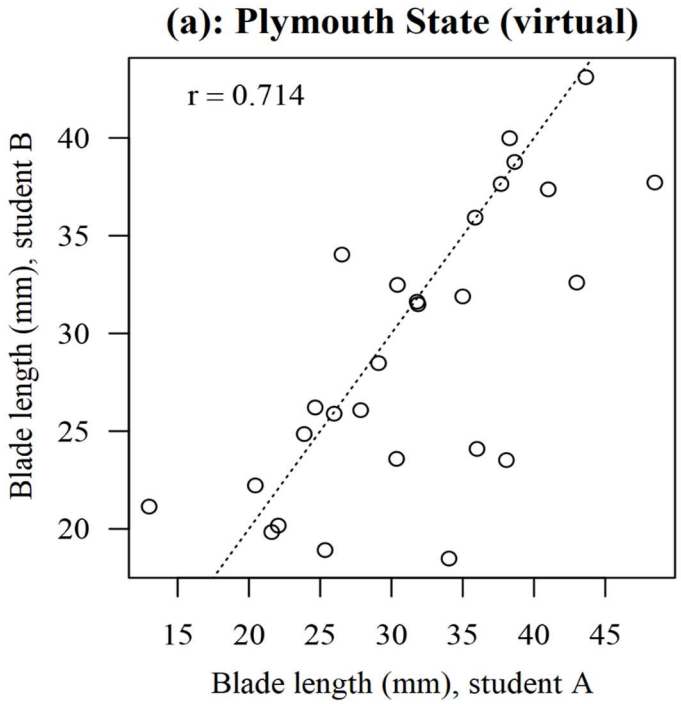

(b): UC Riverside (in-person)
$R^2 = 0.871$
Blade length (mm), student
Blade length (mm), expert
55
50
45
40
35
30
25
30
35
40
45
50
55
60

Fig. 5.2. Scatterplots comparing expert measurements of the length of the leaf blade for the largest leaf per specimen (on x-axis), with averaged student measurement for that same specimen (on y-axis). The plot at the top (a) shows data collected in the virtual instruction format with students in BL 2070 at Plymouth State University, and the plot at bottom (b) shows data collected during the activity with in-person instruction, in BPSC 133 at UC Riverside. R^2 values are calculated from a linear regression predicting each set of student measurements from corresponding expert measurements.


(a): Plymouth State (virtual)
r = 0.714
Blade length (mm), student B
Blade length (mm), student A
20
25
30
35
40
15
20
25
30
35
40
45

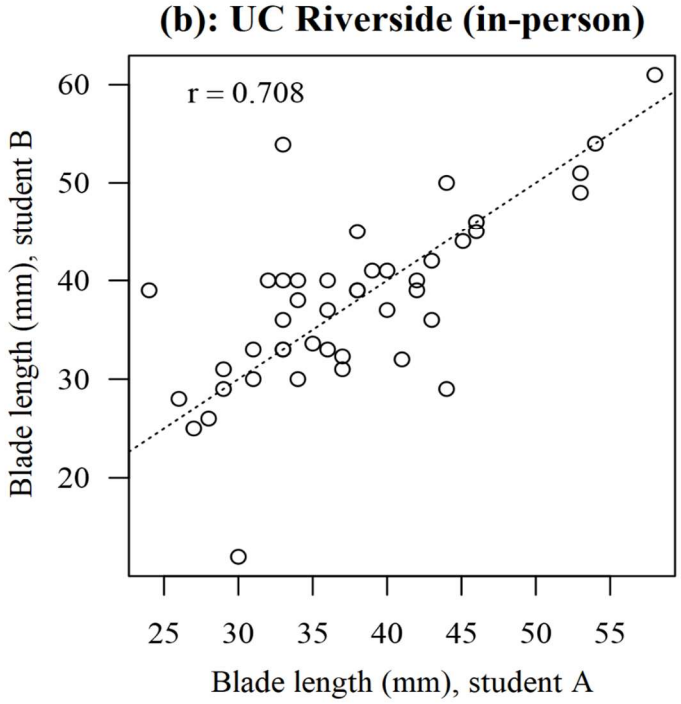

(b): UC Riverside (in-person)
r = 0.708
Blade length (mm), student B
Blade length (mm), student A
20
30
40
50
60
25
30
35
40
45
50
55

Fig. 5.3. Scatterplots comparing duplicated student measurements of the length of the leaf blade of the largest leaf per specimen. The plot at the top (a) shows data collected in the virtual instruction format with students in BL 2070 at Plymouth State University, and the plot at bottom (b) shows data collected during the activity with in-person instruction, in BPSC 133 at UC Riverside. Pearson correlation coefficients (Pearson, 1901) are plotted within each plot.

LITERATURE CITED:

Consortium of California Herbaria. 2023. Complete CCH2 records search for *Arctostaphylos*. *CCH2 data portal*. Website www.cch2.org [accessed 7 September 2023].

Minchin, P. R. 1987. An Evaluation of the Relative Robustness of Techniques for Ecological Ordination. *Vegetatio* 69: 89–107.

Pearson, K. 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2: 559–572.

Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* 9: 671–675.