2023

A Population Level Examination of the Incredible Fruit Diversity in Astragalus Lentiginosus

Quinn Thomas

LOYOLA UNIVERSITY CHICAGO

A POPULATION LEVEL EXAMINATION OF THE INCREDIBLE FRUIT DIVERSITY IN *ASTRAGALUS LENTIGINOSUS*

A THESIS SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

MASTER OF SCIENCE

PROGRAM IN BIOINFORMATICS

BY

QUINN THOMAS

CHICAGO, IL

MAY 2023

ACKNOWLEDGMENTS

I would like to thank everyone who made this research possible, beginning with my mentor Dr. Michael Grillo, without whom this project would not have taken place. I would also like to thank all the members of my committee, Dr. Thomas Sanger, Dr. Yoel Stuart, and Dr. Heather Wheeler for providing me with much needed guidance throughout research. Additionally, I would like to thank Dr. Sara Lipshutz for providing me with further research opportunities to pursue while completing my own research and for the innumerable connections she has provided me with to propel my research forward. I also want to thank the Chicago Botanic Garden for an outstanding collaboration in developing a library prep protocol development. Finally, I would like to thank the team of undergraduates that worked together to help complete the more tedious tasks of this research, namely Olivia Knizka, Sarah Bebla, Nicolle Leon-Arujo, Vivianna Gonzalez and Virginia Addison.

I would also like to thank Loyola University Chicago for providing the funds with which to complete my research and writing. A Research Assistant Fellowship during the 2021-2022 school year allowed me to make remarkable progress on my research.

Finally, I would like to thank my family and friends who have continued to support and encourage me throughout the highs and lows of the last few years. This research would not have been possible without them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Background

Adaptation by natural selection is considered a driving force for generating biodiversity. A long-held goal in evolutionary biology is to identify the drivers of selection in the environment and the phenotypic traits that are targets of selection and confer adaptation (Leinonen et al., 2013). Often, studies of adaptation focus on a limited number of individuals from populations that are already significantly diverged to identify the mechanisms of local adaptation (Belinda & Sgrò, 2017). Despite plausible connections between trait variation and potential agents of selection in the environment, many investigators assume differentiated traits are adaptations without demonstrating a role for selection on the trait or impacts on fitness (Leinonen, 2007). However, phenotypic variation among populations does not solely arise due to divergent natural selection, but results from a combination of evolutionary forces including neutral processes such as isolation by distance and genetic drift, which vary across geographic scales (Borsa et al., 1997; Hutchison & Templeton, 1999). In order for populations to adapt to local environmental conditions there must be mechanisms that limit the homogenizing effects of gene flow among populations. Thus, a comprehensive understanding of adaptation requires a spatially explicit population level approach that encompasses local selective forces and the extent of gene flow, in order to identify key adaptive traits.

Detecting Traits Under Selection

One way to assess the relative contribution of natural selection and neutral processes on shaping phenotypic variation, within and among natural populations, is to compare quantitative trait (Q_{ST} or P_{ST}) and neutral genetic (F_{ST}) variation (Leinonen et al., 2013). Q_{ST} examines the quantitative phenotypic variation between populations, compared to the total variation within and between populations. Q_{ST} can be calculated by the following equation, where σ_B^2 is the variance between populations and σ_W^2 is the variance within populations.

$$Q_{ST} = \frac{\sigma_B^2}{\sigma_B^2 + 2\sigma_W^2}$$

If Q_{ST} is equal to F_{ST}, then the phenotypic variation is what we would expect from neutral processes such as drift. If Q_{ST} is greater than F_{ST}, this indicates that phenotypic variation is greater than patterns of genetic variation among populations, suggesting that directional selection is driving greater divergence than we would expect from drift alone. Q_{ST} values lower than F_{ST} values provide evidence that traits are under stabilizing selection, even in the face of drift (Brommer, 2011; Leinonen et al., 2013). Q_{ST} estimates typically require genetic breeding designs under controlled conditions to limit environmental and maternal variance and estimate additive genetic variance (c) and narrow sense heritability (h^2). This approach is not feasible for many study systems and neglects to reflect phenotypic divergence of natural populations.

P_{ST} is a field-based proxy for Q_{ST} in which phenotypic variation is measured directly from field specimens. While not as robust as Q_{ST} measurements, P_{ST} measures divergence based on phenotypic variance and can provide initial insights to the role of local adaptation on phenotypic variation (Kauffer et al., 2012; Leinonen et al., 2006; Raetmaekers et al., 2007;

Saether et al., 2007). To estimate P_{ST}, heritability values are obtained from relevant systems in the literature and are also interpreted across a range of heritabilities and scalar values.

$$P_{ST} = \frac{c\sigma_B^2}{c\sigma_B^2 + 2h^2\sigma_W^2}$$

Commonly used ratios for unknown *c* and h^2 are 1 (i.e., 100% of variance among populations is due to additive genetic variance) and 0.5 (i.e., 50% of phenotypic variation is genetically based and due to additive genetic variance) respectively (Brommer, 2011; Leinonen et al., 2013). A lower ratio corresponds with a more conservative comparison to F_{ST}. However, because heritability values are not directly measured, it is worth examining the change in P_{ST} across all combinations of a range of heritabilities and scalar values (from 0 to 2). By looking at how P_{ST} changes, we can inspect sensitivity of the mean P_{ST} estimates, relative to mean F_{ST}. How well P_{ST} approximates Q_{ST} depends on how well the parameters *c* and h^2 are estimated, such that if the values of *c* and h^2 are known, then the phenotypic divergence quantified by P_{ST} would equal Q_{ST}.

Population Genomics Using RADSeq

Investigation of the genetic composition of populations has long been an area of interest in biology (Haldane, 1932; Fisher, 1950; Wright, 1984). Recent advances in DNA sequencing technologies make it feasible and affordable to obtain genomic information from large numbers of individuals in non-model systems, setting the stage for applications and advances in population genomics.

Restriction site Associated DNA Sequencing (RADSeq) is one such method that can generate neutral loci in organisms with no reference genome (Davey and Blaxter, 2010; Andrews et al., 2016). In the RADSeq method, genomic DNA is digested using restrictions enzymes.

Adapters with unique barcodes are ligated onto fragmented DNA and multiplexed allowing multiple individuals to be sequenced in a single sequencing run. A common approach to process sequencing reads is through *Stacks*, a bioinformatics pipeline, which was built to track a set of loci in a population, and the alleles present at those loci (Catchen et al., 2013). The general pipeline is to first demultiplex reads and align to one another or a reference genome to generate loci for all individuals. Loci are then added to a single catalog and individuals are mapped to the catalog to determine which haplotype alleles are present at every locus in each individual. SNPs can then be filtered according to presence in proportion of individuals, presence in given populations, SNPs per locus, and more. Methods like RADSeq allow investigators to evaluate gene flow in closely related populations and estimate the extent to which a population has diverged. However, RADSeq prepared data often generates many reads that are shared by only a few individuals, preventing a comprehensive population genetic analysis for all individuals.

Alternatively, double digest RAD sequencing (ddRAD), subsamples the genome by using two enzymes to digest genomic DNA. Using two enzymes permits for a larger percentage of shared reads across individuals thus improving the ability to reconstruct loci in downstream analyses (Peterson et al., 2012). The size selection interval can therefore be varied to impact the number of reads, read coverage, and ultimately the number of resulting SNPs. This flexibility makes ddRADseq a customizable approach for a wide range of purposes.

SNPs generated from ddRADseq can be used to examine genetic differentiation among populations. Fixation indices (or F-statistics) have long been used to describe genetic variance through comparison of variance in allele frequencies between populations (Wright 1984, Chesser 1991). Wright's F_{ST} assumes that the allele frequency per population is known. However, unless all individuals from all populations have been sampled, allele frequencies cannot be known with

certainty. There are several estimators for F_{ST} including Nei's Distance, Weir and Cockerham's, and Hudson's F_{ST} (Nei, 1973; Weir & Cockerham, 1984; Hudson et al., 1992). Here, I evaluate Nei's genetic distance for population pairwise estimates and Weir & Cockerham's F_{ST}. Values for fixation indices will provide insight to *A. lentiginosus* population structure and the scale of gene flow occurring between *A. lentiginosus* populations. Characterization of population structure will allow me to investigate correlations between genetic structure and morphological varieties. I will also compare genetic and geographic distances to observe correlations between population structure and location and signatures for isolation by distance.

Astragalus: The Most Species Rich Plant Genus

Astragalus is the most species-rich plant genus on Earth, encompassing nearly 3,000 species. For context, most plant families contain far fewer than 3,000 species. *Astragalus* is particularly diverse across temperate regions of southwest Asia (ca. 2,000 spp.), the Sino-Himalayas (ca. 400 spp.), western North America (ca. 400 spp.; Linnaeus, 1799), and the Andes (ca.100 spp.; Wojciechowski, 2005; Podlech & Zarre, 2013). Many *Astragalus* species are found in restricted habitats or those requiring edaphic specializations (Wojciechowski et al., 1999). Not only is *Astragalus* strikingly diverse but it has important cultural, agronomic and ecological importance. One species, *A. membranaceous* is used in traditional Chinese medicine for immune system support. Other species of *Astragalus,* termed "Locoweeds", produce or accumulate toxic compounds that are lethal to grazing livestock, costing millions of dollars in economic losses in the United States alone (Cook et al., 2009).

Originating in Eurasia, *Astragalus* has radiated globally since diverging from its sister taxa *Oxytropis* approximately 16 million years ago (Ma) (Azani et al., 2013; Wojciechowski, 2005; Su et al., 2021). Furthermore, phylogenetic analyses suggest that the majority of

Astragalus is monophyletic (Wojciechowski et al., 1999). *Astragalus* has been delineated into 136 old-world (Podlech & Zarre, 2013) and 93 new-world (Barneby, 1964) taxonomic sections. It is estimated that the new world species migrated from Asia as recently as 2.63 million years ago (Su et al. 2021).

Reproductive isolating barriers restrict the homogenizing effect of gene flow, making it an essential process leading to speciation. There are several mechanisms that may serve as reproductive isolating barriers in plants including habitat specialization, ecogeographic isolation, plant-pollinator interactions, mating system differentiation, and temporal isolation. For example, in *Mimulus*, another diverse western North America plant genera, plant pollinators shifts are a strong diversifying mechanism among closely related sister taxa (e.g., species or ecotypes) (Schemske & Bradshaw, 1999; Peter & Johnson, 2014; Byers & Bradshaw, 2021). This does not seem to be the case in *Astragalus*. *Astragalus* flowers, like many Papilionoid legumes, are predominately pollinated by mid-size bees (Cronk, 2006). In fact, of the approximately 400 species of *Astragalus* in western NA, only one is pollinated by hummingbirds (*A. coccineus*). Similarly, shifts in plant mating system from outcrossing to selfing can generate reproductive isolation and promote diversification. For example, two subspecies of *Clarkia*, one primarily outcrossing and the other primarily self-fertilized, occur in sympatry, however, hybrids are rarely observed. While some species of *Astragalus* have been documented to be partially self-compatible (Karron, 1989), reproduction depends primarily on cross-pollination. Thus, unlike other diverse plant groups (e.g., *Mimulus, Aquilegia,* etc.), mating systems and pollinator-shifts do not stand out as obvious drivers of diversification in *Astragalus*.

Astragalus is renowned for its striking diversity in fruit morphology as noted by systematist Rupert Barnaby (1964), *"No other genus of flowering plants exhibits greater variety*

in the form of the fruit". The diversity in *Astragalus* is highly noticeable in the different shapes and characteristics of the fruits of different species and varieties, and fruit morphology is often a diagnostic character for delineating taxa (Barneby, 1964). *Astragalus* fruits are pods that are often inflated, one or two chambered, sometimes mottled or possess trichomes, dry with ripening age, and may fall from the plant when dehiscing seeds. The adaptive significance of *Astragalus* fruit morphological distinction is not immediately clear. Unlike plant-pollinator interactions there is not an evident direct link between fruit morphological variation and reproductive isolation. Thus, it remains unknown whether fruit morphological variation is a driver of diversification in *Astragalus* or if it is shaped by underlying neutral genetic processes.

Beyond the remarkable species diversity throughout the *Astragalus* genus, there is also considerable variation observed within species. One species in particular, *A. lentiginosus*, is the most taxonomically diverse species in the North American flora with over 40 recognized varieties, many of which are distinguished based upon fruit morphological differentiation (Isely, 1998; Barneby, 1964; Knaus, 2010) albeit somewhat arbitrarily, as there are differences in classification depending on the systematist (Barneby, 1964, 1989; Isely, 1998; Welsh, 2007, Knaus, 2010). A few differentiating fruit traits among varieties include pod inflation, pod width, pod thickness, leaf morphology and flower color (Knaus, 2010).

Varieties of *A. lentiginosus* are morphologically distinct when surveyed at geographically distant locations but may become indistinguishable the closer populations are to one another (Barneby, 1964, p. 922). Varieties of *A. lentiginosus* vary considerably in the size of their geographic ranges. For example, var. piscinensis occupies approximately six square kilometers whereas var. fremontii and var. salinus have a range encompassing 25,000 square kilometers and over 50,000 square kilometers respectively (Harrison et al., 2019). This variation in range size

among varieties is also observed among species of *Astragalus*. One hypothesis is that taxa with broad range sizes represent generalists whereas narrowly distributed taxa are likely habitat specialists. While varieties of *A. lentiginosus* appear to be morphologically distinct, they still belong to a single species and thus are likely in the earliest stages of divergence. *A. lentiginsous* is therefore a strong model for studying the drivers of rapid diversification. Furthermore, trait variation within *A. lentiginosus* is also observed throughout the genus (e.g., variation in range size, fruit morphology, etc.), making *A. lentiginosus* an insightful system for understanding the microevolutionary mechanisms of broader evolutionary trends across *Astragalus*.

Although varieties are defined by morphological characteristics, little is known about the actual genetic population structure of *A. lentiginosus*. Preliminary research suggests genetic structure exists between varieties but the extent of the distinction across all varieties isn't known (Knaus et al., 2005; Harrison et al., 2019). These studies are often limited in scope by the number of varieties investigated, the geographic area under study, or the number of loci used to determine genetic structure. A study by Knaus et al. (2005) examined genetic structure between 3 different varieties of *A. lentiginosus* sampled over six sites using AFLPs. Two of these varieties occupied ranges that overlapped (var. lentiginosus and var. salinus) and one variety occupied a range that was geographically distinct from the other two (var. variablis). A principal coordinates analysis based on Dice's genetic distances revealed patterns of distinction between varieties but also suggested patterns of distinction based on sampling location. While patterns of genetic structure were found, only 184 loci were generated across just three varieties and six sites (Knaus et al., 2005). Another study characterized genetic structure of a single, very narrowly distributed *A. lentiginosus* variety (var. piscinensis) at different sampling locations around an 8 square km area of the Nevada Fish Slough (Harrison et al., 2019). This study adapted a

RADSeq/GBS protocol to generate 7274 SNPs for 117 individuals. Not only was structure found at the site level, but there was evidence that structure correlated with abiotic factors. While these studies provide promising insights, it remains unclear whether taxonomic varieties of *A. lentiginosus* represent discrete genetic units and the underlying population genetic structure in this system.

Adaptive Significance of Fruit Morphology

Despite the profound diversity in fruit morphology throughout the genus and within species, the potential adaptive significance of *Astragalus* fruit morphological variation remains unknown. In many plant systems, fruits serve as a means for animal dispersal (Pijl, 1982; Mazer & Wheelright, 1993), however to my knowledge there is no evidence for animal dispersal of fruits or seeds in *Astragalus*. Many *Astragalus* species in western North America have dry inflated fruits, which are likely dispersed by wind in the open environments that they inhabit (Houghton et al. 2020). Abiotic dispersal agents such as wind are expected to function uniformly across populations and are unlikely to select for distinct fruit morphologies. A more likely adaptive role for fruit morphological variation is as a physical defense against insect herbivores.

Pre-dispersal seed herbivory by insects has been documented to have profound impacts on fitness in *Astragalus* (Green & Palmbald, 1975; Martin & Menke, 2012). In *Astragalus*, specialist seed beetles (*Acanthoscelides*) have been documented to have drastic effects on *Astragalus* fitness with seed mortality rates of 74% (Green & Palmbald, 1975) and even as much as 93% (G. Morse pers. comm.). Herbivory and seed predation are not unique to *Astragalus* but are common agents of selection among plants (Althoff, 2014, Agrawal et al., 2018). Ehrlich and Raven (1964) famously proposed that antagonistic coevolution between plants and herbivorous insects is responsible for generating much of the plant diversity on Earth. Common fruit

adaptations known to make feeding more difficult for insect herbivores include waxes, trichomes, husk hardness, sclerophylly, latex deposition and lactiferous structures (Janzen 1971; Fürstenberg- Hägg et al. 2013; Dalin and Björkman 2003; War *et al*. 2018). In some *Astragalus* species such as *A. cibarius*, Green and Palmbald (1975) hypothesize that pubescent pods may prevent beetles from depositing eggs within fruits. It is important to note that insect population size and levels of herbivory often vary across the landscape, establishing variable selection across space and time (Thompson 1994 & 2005; Whitney & Stanton, 2004; Althoff et al., 2014).

Heritability of Fruit Traits

Because of unknown values for c and h^2, P_{ST} values can often be misinterpreted. Therefore, when utilizing P_{ST} estimates there must be careful considerations when making assumptions about for c and h^2 Brommer,2011). The null assumption is that $c = h^2$, however, since one cannot be sure that a system follows the null assumption, P_{ST} should be evaluated across a range of values. One such study that implements this practice in natural populations is by Seymour et al. (2019) where various phenotypic traits in stickleback are observed across a variety of scalar and narrow sense heritability values. For several of their traits, interpretation of the P_{ST} - F_{ST} comparisons depended on which values of c and h^2 were used. Only one trait, defense, showed higher values of P_{ST} compared to F_{ST} for all values of c and h^2. Another study by Mota et al. (2020), looked at fruit shape characteristics in 25 subpopulations of *Dipteryx alata (*Fabaceae). P_{ST} values for these traits were generally low, ranging from 0.22-0.24 but were not compared to F_{ST} values.

Phenotypic heritability for fruit traits in legumes ranges with some species estimated to have low-moderate phenotypic heritability ranging from 0.28-0.55 (chickpeas: Karami, 2011; Petrova, 2021) and other species have moderate-high phenotypic heritability ranging from 0.68-

0.97 (pea, common bean and faba bean: Gómez & Ligarreto, 2012; Delfini et al., 2021; Bora et al., 1998) (Kuzbakova et al., 2022). Pod heritability in the common bean has been shown to have relatively high heritabilities ranging from 0.31-0.91 (García-Fernández et al., 2021). Pod length has also been documented to have high narrow sense heritability in yardlong beans (Kusmiyati et al., 2021). Fruit shape in other families like pecans and tomatoes have shown moderate to high heritabilities with tomatoes having heritability estimates as high as 0.97 (Bhattarai et al., 2022; Zörb et al., 2020). Even though estimates for fruit shapes tend to be moderate to high, it is important to observe P_{ST} across a range of heritabilities.

Analyzing Fruit Shape with Geometric Morphometrics

While fruits have been identified as a delineating characteristic for species and varieties of *Astragalus*, little has been done to quantify fruit phenotypic variation. One study by Knaus (2010) evaluated floral, fruit and vegetative characteristics of fourteen *A. lentiginosus* varieties to determine how well plant characteristics described taxonomic varieties. With relation to fruit shape, pod beak length was identified as the most influential shape trait when delineating varieties. This study evaluated fruit shape with four linear measurements, however these measurements fail to preserve the overall pod shape configurations of the fruit.

Landmark-based geometric morphometrics is now the standard approach to quantify shape and shape variation in a system (Mitteroecker & Gunz, 2009). Homologous landmarks are used for conserved points in a model that can be found in all individuals. However, traditional homologous landmarks are insufficient in capturing curved shapes that lack distinct homologous points. Semilandmarks are used to capture curves and surfaces across individuals that share homologous geometric patterns. Semilandmarking allows shape to be captured in more detail and thus provides insight to shape variation at a finer scale. For example, semilandmarks have

been used to document shape differences between plant varieties in leaf morphologies (Nery et al., 2019; Márquez et al., 2022). Here, I use a Procrustes based approach in which coordinates are aligned based off the centroid and semilandmarks slide are adjusted in relation to their surrounding landmarks (Adams et al., 2022).

Scope

Astragalus lentiginosus is an extremely diverse species and may serve as an important evolutionary model for other plant systems. This study will serve to characterize the population genetic structure of *A. lentiginosus* across the Southwest United States and determine the amount of gene flow occurring between populations of *A. lentiginosus*. Understanding the scale of gene flow is necessary to provide insight to the potential for local adaptation to occur in *A. lentiginosus*. If considerable population structure and reduced gene flow is identified across the landscape, this could allow populations to readily adapt to local selective pressures, like herbivory, and also for neutral genetic differentiation among populations. Profound levels of genetic differentiation between certain varieties could be indicative of distinct species within *A. lentiginosus*, which may be likely given the high degree of fruit morphological variation. These findings will also be of applied significance as numerous morphological varieties are given special conservation status.

Additionally, patterns of morphological variation will be determined across *A. lentiginosus* varieties. Determining the amount of variation across populations of *A. lentiginosus* not only provides insight for how discrete morphological varieties actually are, but also allows us to correlate how much genetic structure varies in comparison to morphological structure. This analysis will provide insight into the variables that may influence fruit morphological variation. Mechanisms such as local adaptation and genetic drift may depend on variables examined here

such as scale of gene flow and sampling location. Comparing P_{ST} to F_{ST} will provide the first evidence as to whether fruit morphological variation in *A. lentiginosus* is adaptive or is driven by neutral evolutionary processes.

This study aims to address the following questions. 1) What is the scale of gene flow and resulting population genetic structure of *Astragalus lentiginosus*? 2) Do morphological varieties represent population genetic structure of *Astragalus lentiginosus*? 3) Do varieties with narrowly distributed range sizes have reduced genetic diversity compared to broadly distributed ranges? Are there correlations between genetic and geographic distances? 4) Does fruit shape variation explain taxonomic varieties? 5) Is there a signature for selection based on comparison of fruit shape variation and genetic variation?

CHAPTER TWO

METHODS

A. lentiginosus Fruit Collections and Seed Germination

In the summer of 2019, *A. lentiginosus* was collected by the Grillo lab from 86 sites in Arizona, California, Nevada, and Utah (Figure 1). This region encompasses approximately 1/3 of the *A. lentiginosus* range in the United States. Field collections consisted of harvesting numerous ripe fruits from up to 20 plants at each site, to be used for morphometric analysis and DNA extraction from seedlings. Most sampled populations of *A. lentiginosus* were identified through online georeferenced herbarium records (e.g., Jepson's eflora; Intermountain Biota etc.). When ripe fruits were not available leaf tissue was collected for DNA extraction. From the 86 sites, 853 *A. lentiginosus* specimens were collected from 18 taxonomic varieties. Fruits from the Summer 2019 collections were imaged for morphometric analysis. Pictures for up to ten fruits per plant were taken from both the top and side view. In total there were 2,505 fruit images that were landmarked for each view. These images represented 288 plants, 35 sites and 14 varieties.

Seeds were harvested from fruit collections and grown in the Loyola greenhouse until they were large enough to extract DNA. Seeds were first scarified by nicking the seed coat with a razor blade. Seeds were then surface sterilized in 30% bleach for 10 minutes and were rinsed three times in sterile DI water. On the final rinse, seeds were imbibed in water for approximately 30 minutes. After imbibing, seeds were placed on a wet kimwipe in a petri dish and incubated at

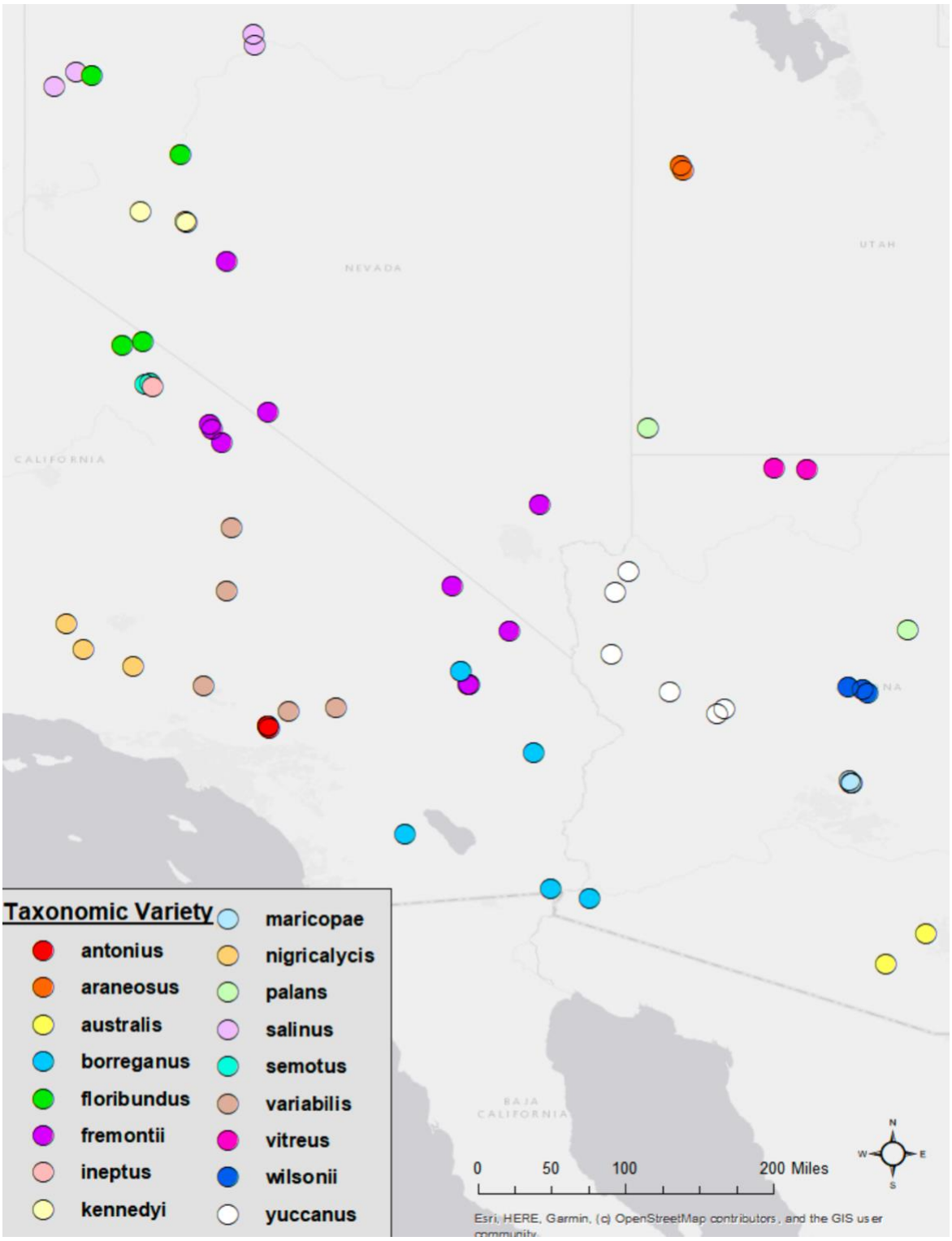

Figure 1. Field Collections of *A.lentiginosus* sites. Data for this paper do not include sites for var. ineptus or var. semotus.

30°C. Once seeds germinated, they were planted in small pots with up to 8 seedlings from each plant collected in the field.

Double Digest RADSeq Library Preparation and Sequence Processing

DNA was extracted from young *A. lentiginosus* leaves using Quick-DNA Plant/Seed Miniprep Kit (Zymo Research Group, California, USA) and eluted in 55uL of MilliQ water. Leaves were collected from a single plant so as not to introduce false heterozygosity. DNA concentration was measured with a Qubit 2.0 Fluorometer using a Qubit HS DNA Assay Kit (Life Technologies, Thermo Fisher Scientific) and kept in -20°C until library preparation.

Here, I utilize double digested RADSeq (ddRAD), an adaptation of RADSeq (Peterson et al. ,2012), to generate SNPs and evaluate the population genetic structure of *A. lentiginosus*. Prior to preparing a complete library for sequencing, an in-silico digestion simulation was completed to estimate the number of fragments to be expected after digestion and size selection. Because *Astragalus* does not have a reference genome, the R package *SimRAD* was used to run a simulated digest of *Glycine max,* a well sequenced legume with a similar estimated genome size to *Astragalus* at ~1Gb (Lepais & Weir, 2014). Simulated digestions were run on a single chromosome of each species and those results were then multiplied to represent the entire genome (Table 1). A simulated digest informs decisions for size selection and allows to estimate sequencing coverage based on the number of individuals (Table 1). Coverage was estimated by taking the expected number of reads from the sequencer and dividing by the number of individuals times the number of expected fragments. The expected coverage for 192 individuals on a NovaSeq instrument with a size selection between 300-700bp is: 1,300,000,000 reads / (192 individuals * 64,000 fragments) = 105X.

Table 1. Summary of Simulated Digest and Size Selection of *Glycine max*. "A" represents a cutsite by the first enzyme and "B" represents a cutsite by the second enzyme. ddRAD only utilizes fragments with one cutsite from both enzymes.

	Values for one chromosome	**Values for entire genome (x20)**
Number of restriction sites for the first enzyme (EcorI)	16530	330600
Number of restriction sites for the second enzyme (MseI)	31986	639720
Number of type AB and BA fragments	16362	327240
Number of type AA fragments	8322	166440
Number of type BB fragments	23729	474580
Fragments between 300 and 700 bp	**3211**	**64220**
Fragments between 400 and 600 bp	1483	29660

A modified version of the original ddRAD-Seq method (Peterson et al., 2012) was used for reduced representation genomic library construction (Diaz-Martin et al., 2023). Adapters are oligonucleotides designed to ligate to the sticky ends of the restriction cut sites of genomic DNA and contain sequencing primer binding sites that initiate sequencing. Additionally, adapters serve to identify individuals with a unique 5bp barcode so that all individuals with unique adapters may be multiplexed and run on a single sequencing lane. Adapters were designed following Peterson et al. (2012) with 48 adapters with a unique barcode designed to ligate to the EcoRI cut site and 1 common adapter designed to ligate to the MspI cut site. Adapters are ordered single-stranded and were annealed by combining equal amount of forward and reverse primer in a 96-well plate and heating them to 97.5°C and cooling slowly until 21°C was reached. PCR primers were also ordered according to Peterson et al. (2012) and are used for PCR amplification and also contain complementary sequences for Illumina sequencing. PCR primers can also be used in

combination with adapters for further multiplexing by incorporating a standard Illumina multiplexing read index.

All samples were diluted to a normalized DNA concentration of approximately 30ng/uL. A total of 192 samples in four libraries were digested using enzymes EcoRI-HF (New England Biolabs, Ipswich, MA. USA) and MspI (New England Biolabs, Ipswich, MA. USA) for 12-18 hours at 37.5°C in a thermocycler to ensure complete digestion. Adapters containing custom barcodes were ligated onto sticky ends of digested product with T4 DNA ligase (New England Biolabs, Ipswich, MA. USA) in a thermocycler at room temperature for 3 hours with a 10-minute heat kill step to denature ligase activity. Following ligation, samples were pooled by library and size selected using AmpureXP beads (Beckman Coulter, Indianapolis, IN). A double-sided size selection was performed following Diaz-Martin et al. (2023) by first adding AmpureXP beads to bind to very large fragments and leave behind fragments less than 700bp in supernatant. The supernatant containing fragments 700bp and under is then transferred and beads are added to attach fragments larger than 300bp. The supernatant is discarded, and DNA is eluted from the beads in 20uL of MilliQ water. Libraries were PCR amplified with 14 cycles and cleaned with AmpureXP beads. After measuring library concentrations, each library was normalized at 40ng/uL and combined in equal volumes. In total, four reduced representation genomic libraries containing 175 individuals were sequenced on an Illumina NovaSeq 6000 at Michigan State University at the MSU Genomics Technology Support Facility for 150bp paired end sequencing.

Assembling Loci and Collecting SNPs

Sequenced libraries were first evaluated with fastQC reports. Upon initial inspection, cut sites in the forward reads and reverse reads failed to be resolved, however the remainder of the read was high quality. This is a common issue with NovaSeq instruments as there is no

nucleotide diversity at the restriction cut site which leads to low quality reads at the cut site position. As a result of these findings, reads were demultiplexed and filtered considering paired end reads and forward reads separately. *Stacks* is a program that takes raw sequencing reads from individuals and aligns reads into "stacks" to generate loci (Catchen et al., 2013). This program was designed for systems that have no reference genome. *Ustacks* aligns identical reads in individuals to build a "stack" and then merges stacks together to create a locus and detect SNPs at that locus. c*stacks* adds the loci that were built in *ustacks* into a catalog and merges shared loci across individuals. s*stacks* maps the loci from individuals in *ustacks* to a catalog locus. *tsv2bam* transposes data to orient it by locus rather than by individual and *gstacks* identifies SNPs for the meta-population at each locus and then genotype each individual at each identified SNP and phase the SNPs into a set of haplotypes.

Libraries were demultiplexed by individual and reads with uncalled bases and low-quality scores were removed using *process_radtags* from *Stacks*. This command takes a map of individuals and their associated barcodes and demultiplexes based on identification of those sequences. Additionally, it checks for restriction enzyme cut sites in the forward and reverse reads and removes reads where the restriction site cannot be found. *A. lentiginosus* does not have a reference genome, therefore the denovo pipeline was used for loci assembly. A subset of 66 samples was taken to test the parameters for M/n = 1-9 prior to running the full dataset. One individual was chosen per sample site so that the catalog would have full representation of all individuals. During parameter testing, individuals with more than 60% missing data were removed from the final dataset so as not to limit the number of SNPs for final consideration. Only four specimens were removed from downstream analyses for this reason.

After loci were built and SNPs were mapped for each individual, loci were filtered using *populations.* Loci that were not present in at least 80% of all individuals were removed following Paris et al. (2017). Additionally, SNPs were removed that had a minor allele frequency (MAF) < 0.05 and a maximum observed heterozygosity of 0.7. Only one SNP per locus was exported for downstream analyses to filter out SNPs that are linked.

Comparative Analysis of Genetic Variation across *A. lentiginosus* Varieties

SNPs generated through the stacks pipeline were used to assess population genetic structure among *A. lentiginosus* specimens. Here, SNPs were used as input for STRUCTURE, and *adegenet* and were run to evaluate admixture proportions for individuals assigned to different genetic populations (K). STRUCTURE uses a naïve Bayesian clustering approach to sort individuals into K different populations based off SNPs at each locus (Pritchard et al., 2000). The more SNPs that are available, the easier it is to differentiate individuals as well as evaluate the amount of admixture across populations. To estimate the ideal number of K populations I used the Evanno's delta K method for which three independent runs for each K = 1-9 was run with a burn-in period of 10,000 followed by 50,000 MCMC steps (Evanno, 2005) using the program Structure Harvester (Earl & vonHoldt, 2012). The Evanno method uses the second rate of change of the log likelihood of K (Ln(K)) to determine the most likely number of genetic populations (Evanno, 2005). Once the Evanno method indicated the optimal level of K, STRUCTURE was rerun with a burn-in period of 50,000 followed by 100,000 MCMC steps.

Fixation indices (or F-statistics) have long been used to describe genetic variance through comparison of variance in allele frequencies between populations (Wright, 1984; Chesser, 1991). I use several estimators of Wright's F-statistics to evaluate the population structure of *A.*

lentiginosus including Nei's D, Weir and Cockerham's F_{ST}, and Hudson's F_{ST} (Nei, 1973; Weir & Cockerham, 1984; Hudson et al., 1992).

adegenet is an R package used to evaluate population statistics and is used primarily as verification for our clustering-based approaches detailed above (Jombart, 2008; Jombart & Ahmed, 2011). Genetic variation was modeled and visualized with a principal component analysis (PCA). Additionally, this package was used to evaluate isolation by distance between sites through regression of genetic distance to geographical distance. For this, I converted all distances between sampling site coordinates to kilometers to generate a matrix of geographic distances. For the matrix of genetic distances, pairwise genetic distances were calculated for each sampling site using Nei's D and normalized those genetic distances with the equation: F_{ST} /(1-F_{ST}). A mantel test was then utilized to compare these two distance matrices with 999 permutations. All samples at a single sampling site were assigned the same GPS coordinates and therefore isolation by distance was evaluated primarily at the site level.

Landmarked based Analysis of *A. lentiginosus* Fruits

Landmark-based geometric morphometrics is a common approach to quantify shape and shape variation in a system (Adams et al., 2013). Homologous landmarks are used for conserved points in a model such as joints. However, traditional homologous landmarks are insufficient in capturing curved shapes. Semilandmarks are used to capture curves and surfaces across individuals that share homologous geometric patterns. The number of semilandmarks used to capture a shape depends on the complexity of that shape and the scale of shape variation across the system. Rather than using equidistance as a method of semilandmarking, I utilize sliding semilandmark which orients points according to the mean Procrustes shape (Gunz &

Mitteroecker, 2013). Semilandmarking allows shape to be captured in more detail and thus provides insight to shape variation at a finer scale.

A. lentiginosus fruits are curved pods with two chambers. Due to its curved nature, there are few points on the fruit that are conserved across all individuals. The landmarking protocol used here was developed to consider the shape of both chambers with equal weight. There are four homologous landmarks from the top view occur along the division between the two chambers with one at the base of the fruit and the other three representing the beak opening at the top. An equal amount of landmarks were assigned for each half of the fruit. In total, thirty landmarks were used to represent the fruit from the top view with twenty-six landmarks being sliding semilandmarks and four being homologous. From the side view, the tip of the beak is the only homologous point on the shape and thus there is only one homologous landmark. Again, thirty total landmarks were used with twenty-nine being sliding semilandmarks and one homologous.

There are several programs designed to produce landmarks for multiple types of imaging analysis. In this study, ImageJ is used to mark and generate landmarks from images captured. While many programs have the capability to produce automatic landmarks with machine learning techniques, due to the nature of the quality of images, landmarks were placed individually by myself and several undergraduate students at Loyola University Chicago. Work produced by undergraduates was checked to ensure consistent landmarking. Additionally, images for a given site were only landmarked by one student so that variation at the site level was minimized. Prior to landmarking, each image the scale was set at 500 pixels per 1 cm to account for size variation. Upon landmarking completion, landmarks were exported as XY coordinates in CSV files.

Characterizing Fruit Shape Variation Between *A. lentiginosus* Groups

Landmark data was read into R and analyzed using the geometric morphometric analysis package *geomorph* (Adams et al., 2022). g*eomorph* was designed to provide all standard tools used in geometric morphometric analyses in a single package. One such tool is a generalized Procrustes analysis (GPA: Gower 1975; Rohlf & Slice 1990) which superimposes the individuals to a common coordinate system by holding constant variation in their position, size and orientation and slides semilandmarks with reference to their neighboring points. Transformed shape variables from the procrustes alignment can be used downstream in multivariate statistical analyses. In geomorph, a procrustes transformation also calculates centroid sizes of each shape variable which can be used for downstream analyses in multiple linear regression and allometry. Size variation is a common indicator of evolutionary diversification and often informs variation in morphological traits (Klingenberg, 2016). Allometry is the relationship of how size and a morphological trait covary. In the field of geometric morphometrics allometry can be evaluated by multivariate regression of shape on centroid size (Klingenberg, 2016). This indicates how well we may predict morphological traits based on the size characteristics. Allometry was tested in this system by investigating the relationship between fruit size and shape using *plot.allometry* (Adams et al., 2022). Multivariate regression is also utilized to examine how well collection sites and variety explain fruit shape variation. Not only are predictors of shape variation examined, but I also explore the morphological disparity between the amount of variation across population identifiers (i.e., sites and varieties).

Landmark data was assessed using a hierarchical analysis in which variation was evaluated at the plant, site, and variety level. To limit the amount of variation occurring within a single plant and also account for differences in the number of fruits representing a single plant,

the average procrustes transformed fruit shape per plant was used for downstream analyses. In each scenario, landmarks were aligned and transformed with a generalized procrustes analysis, and a principal component analysis was then run to evaluate how individuals varied with respect to one another. To determine significance of variation, an ANOVA was performed to evaluate which groups varied significantly. Pairwise tests were done to determine which groups differed from one another. Initial observation of fruit images indicated that a primary source of variation between fruits was in the beak region. For that reason, landmarks 15-17 for each fruit were used to do a standalone measurement of beak opening width and beak opening depth for each fruit. Variation of these measurements between varieties was evaluated using a Tukey test.

Comparative Analysis of Morphological and Genetic Variation in *A. lentiginosus*

Comparison of differentiation in quantitative traits and neutral genetic markers is a common way to discern the relative importance of selective and neutral processes (Leinonen et al., 2013). P_{ST} is an estimation of quantitative trait variation (Q_{ST}) for which we can evaluate phenotypic differentiation in naturally occurring populations (Pujol et al., 2008). Here I utilize R package *Pstat* to calculate P_{ST} (Da Silva & Da Silva, 2018). It is important to note that the two variables that primarily affect P_{ST} values are the ratio of c and h^2. One of the primary advantages of using *Pstat* is that the package accounts for population size in its P_{ST} calculations and allows the user to specify the range of values to test c/h^2. P_{ST} values were calculated using c/h^2 ratio ranging from 0-2 however, it is worth noting that most studies focus on c/h^2 values ranging from 0.5-1 (Brommer, 2011; Lopez et al., 2020). Therefore, this analysis will primarily focus on comparison in this range. Robustness of p-values was checked through bootstrapping and using a 95% confidence interval for all values as suggested by Brommer (2011). P_{ST} was calculated for the first three principal components of the PCA of landmarked coordinates, the beak opening

measurements and the beak depth measurements. Comparison of P_{ST} and F_{ST} was evaluated for sites, varieties and K populations.

CHAPTER THREE

RESULTS

Processing NovaSeq Reads

Sequencing ddRAD libraries of 192 *A. lentiginosus* individuals on one lane of NovaSeq resulted in a total of ~805 million paired end reads. FastQC reports indicate a failure to capture restriction cut sites for both EcorI and MspI (Figure 2) due to the low diversity at these sites. However, read quality for all other bases, including those at the end, was high.

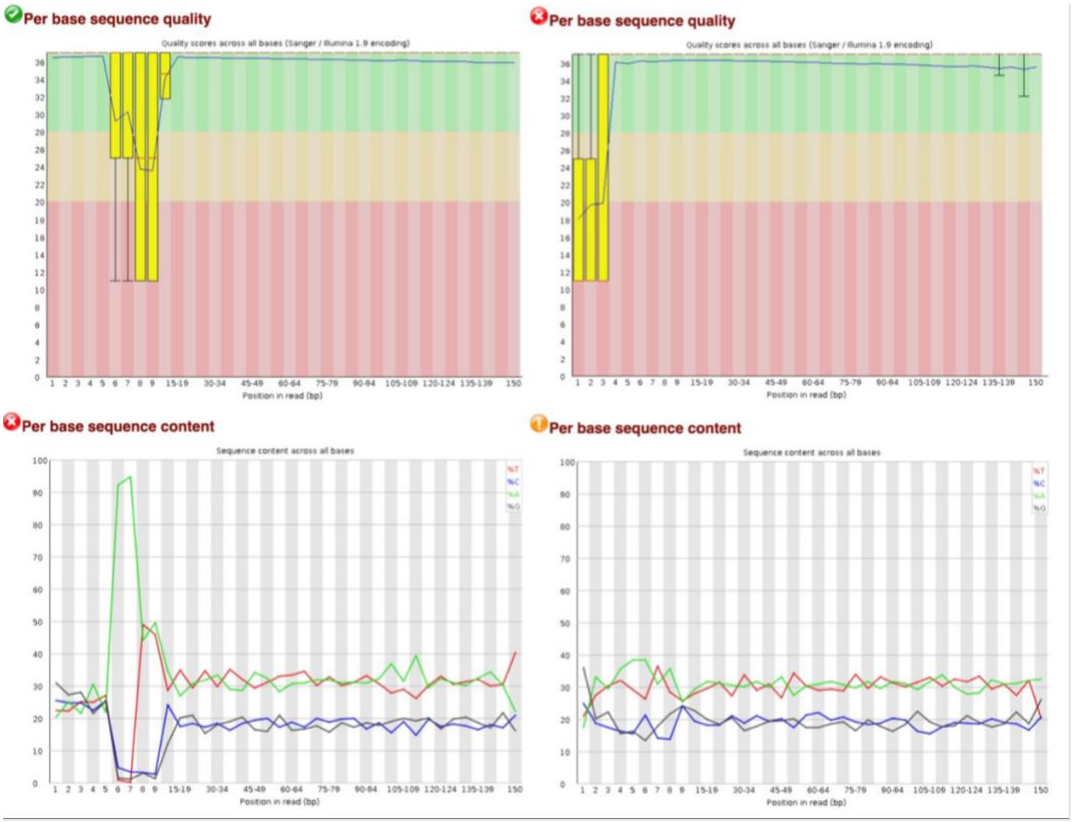

Figure 2. FastQC. Per base sequence quality and per base sequence content for forward and reverse reads in Library 1.

Recognition of cut sites are required for demultiplexing in STACKS. Given the inability to recognize cut sites, I troubleshooted demultiplexing using various pipeline settings involving single or paired end reads (Table 2). For each demultiplexing run, there was a relatively low number of barcodes not found and low-quality reads which is in agreement with the FastQC report. For all runs that did not check for the restriction cut site, over 99% of reads were retained. There are two reasons for why restriction cut sites are unable to be found; the first as previously mentioned, is that read quality at the cut site is extremely low and unable to be resolved and the second is that there was either incomplete digestion and/or non-specific ligation. Retained reads were used to evaluate SNPs produced by the default pipeline (Table 3).

Table 2. Summary of Output from process_radtags. Bolded pipelines were evaluated in downstream analyses.

Pipeline	Total Sequences	Barcode not found	Low Quality	Cut site not found	Retained reads
Paired, -r -c –q, check for both cut sites	805,821,086	4,439,476	154,063	546,926,834	254,300,713
Paired, -r -c -q, only check for EcorI	**805,821,086**	**4,439,476**	**265,563**	**211,801,001**	**589,315,046**
Paired, -r -c -q, --disable-rad-check	805,821,086	4,439,476	539,673	0	800,841,937
Single, -r -c –q, check for both cut sites	402,910,543	2,219,738	152,431	211,801,001	188,737,373
Single, -r –c –q, --disable-rad-check	402,910,543	2,219,738	426,541	0	400,264,264

Table 3. Summary of Output from denovo_map. Bolded pipelines were evaluated in downstream analyses.

Process_radtags	denovo_map	# reads after process_radtags	populations	popmap	# loci/SNPs
Paired, -r -c -q	-M 3 -m 3 -n 3	254,300,713 out of 805,821,086	-r80	1 population	399
Paired, -r -c -q, only check EcorI	**-M 3 -m 3 -n 3**	**589,315,046 out of 805,821,086**	**-r80**	**1 population**	**5965**
Paired, -r -c -q, --disable-rad-check	-M 3 -m 3 -n 3	800,841,937 out of 805,821,086	-r80	1 population	11
Single, -r –c –q, --disable-rad-check	-M 3 -m 3 -n 3	400,264,264 out of 402,910,543	-r80	1 population	8

Evaluation of SNPs produced from the denovo pipeline with different sets of reads coming from the demultiplexing stage indicate that using paired reads and only checking for the cut site on the forward read produces the largest number of SNPs (Table 3). Parameter testing of the subset of 66 individuals indicates that parameters where M/n = 4 maximizes the number of total loci (Figure 2). It is worth noting that there is a relatively high number of heterozygous loci which typically results in reduced SNP counts after filtering. After rerunning stacks at the optimized parameters and filtering based on minor allele frequency and maximum observed heterozygosity, a total of 5047 SNPs were used for downstream analyses.

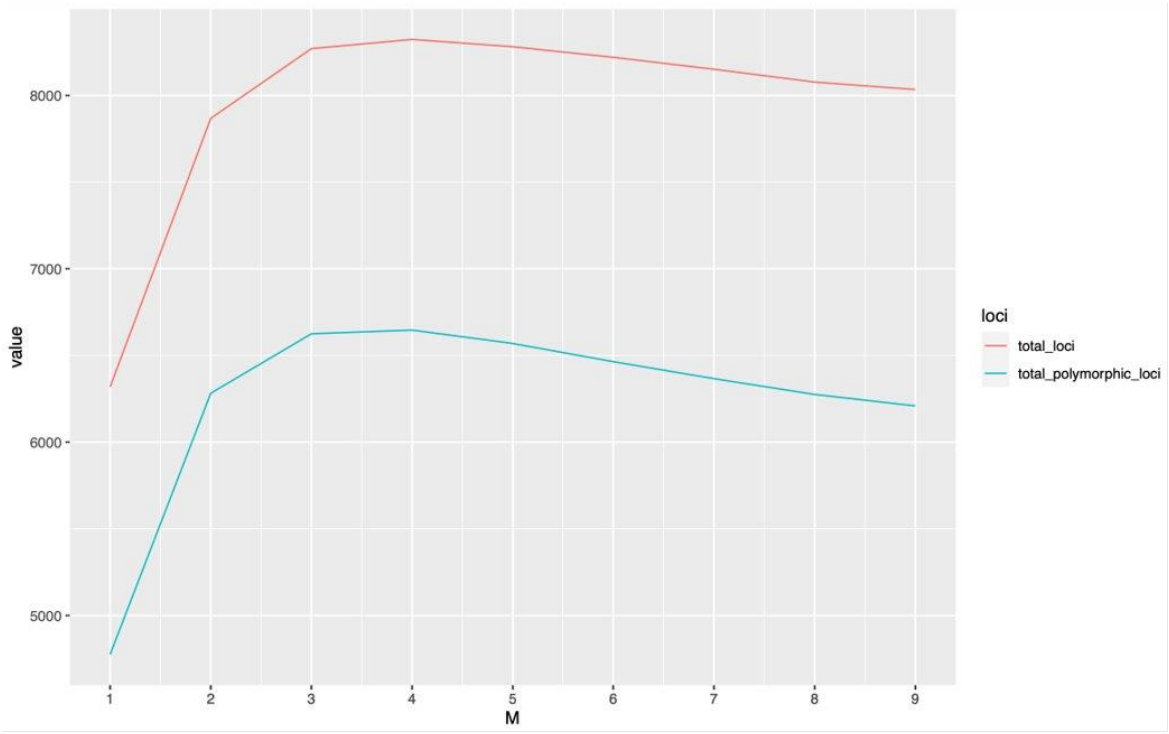

Figure 3. Total Number of Loci and Polymorphic Loci. Evaluated for each STACKS parameter where M = n = 1-9. The total number of loci generated is maximized with the parameter of M=4.

Population Genomic Structure of *A. lentiginosus*

Using the 5047 SNPs from STACKS, population structure was estimated among all *A. lentiginosus* specimens. Initial investigation of population structure using a PCA revealed three major clusters between the first two principal components. The first two PCs explain 24.5% of genetic variation (Figure 3). Along PC1 (15%) there is a distinction between six varieties (positive values: var. nigricalycis, var. yuccanus, var. borreganus, var. variablis, var. antonius & var. australis) and the remaining ten. Along PC2 (9.5%), there is a distinction between the remaining ten varieties with five varieties (positive: var. fremontii, var. salinus, var. kennedyi, var. lentiginosus & var. floribundus) distinguishing themselves from four (negative: var. wilsonii, var. maricopae, var. vitreus & var. araneosus). One variety, var. palans, spans both the positive and negative area of the second principal component. By and large, individuals from the

same variety cluster together, even when collected from different sampling sites. This is case for fourteen out of sixteen of the varieties.

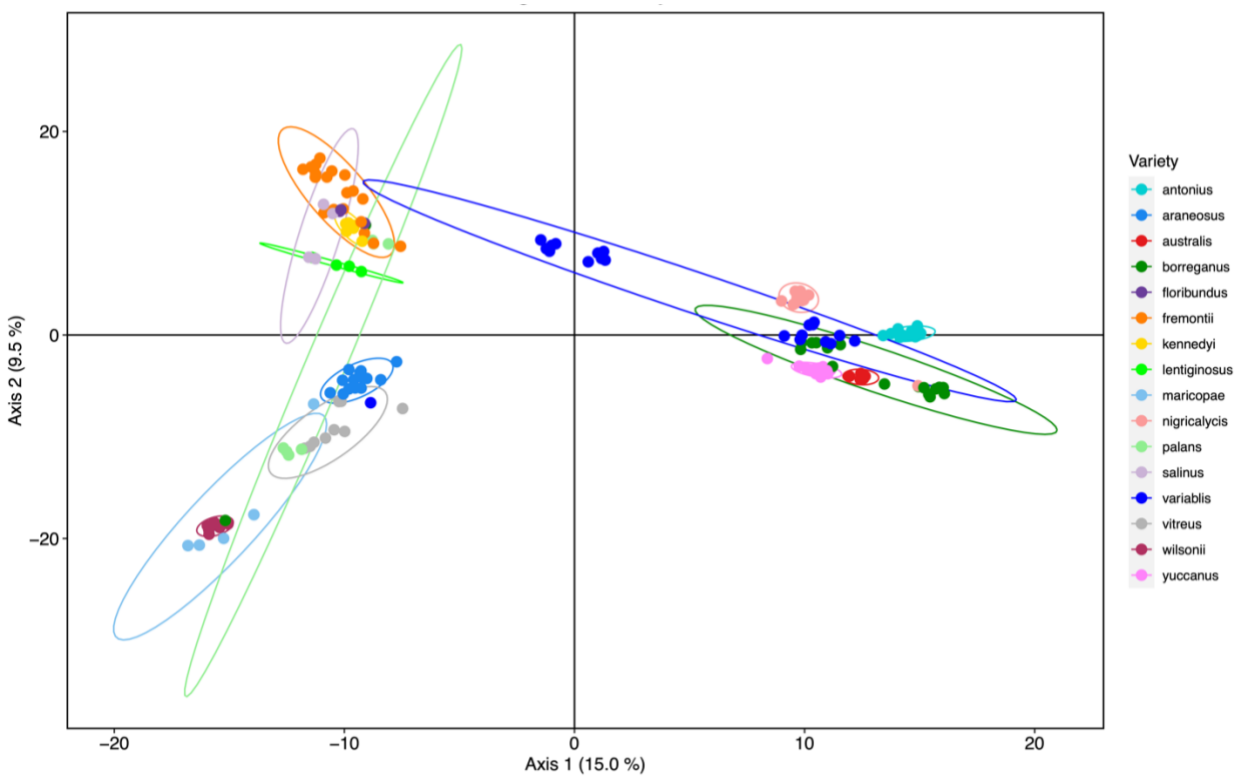

Figure 4. Genetic Principal Component Analysis. PCA of 5047 SNPs for 173 *A. lentiginosus* samples. Points and ellipses are colored by variety. PC1 explains 15% of the variation and PC2 explains 9.5% of the variation.

Genetic structure was further evaluated by bayesian clustering using STRUCTURE (Pritchard et al., 2000). The Evanno method indicates that the most likely K groups for the 5047 SNPs was K = 4 (Figure 4). Interestingly, there are no other values of K that indicated other possible genetic groupings.

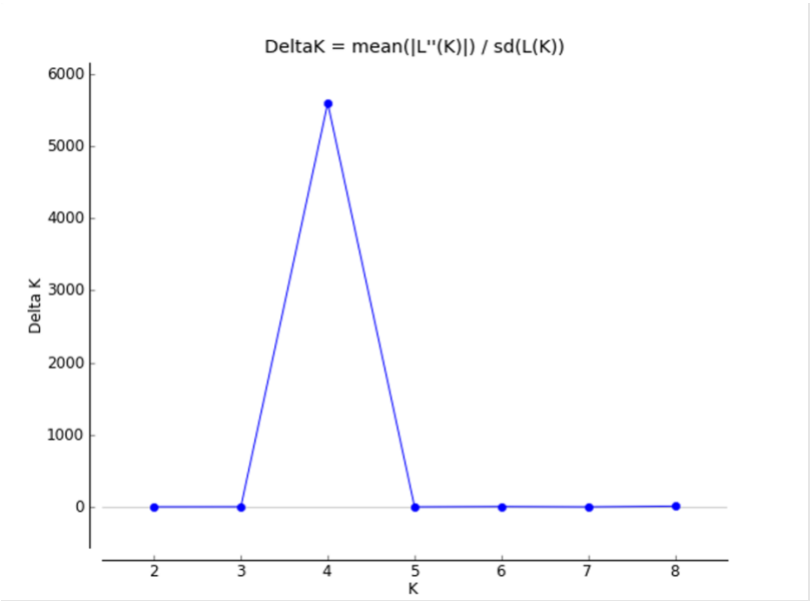

Figure 5. Delta K. Delta K for structure groups where K = 1-9. Created using Structure Harvester (Earl & vonHoldt, 2012). The highest value for Delta K occurs at K = 4, indicating that individuals can be best split into 4 genetic groupings.

STRUCTURE was run for values of K = 2, 4, and 16 (Figure 6). Base levels of variation for K = 2 indicates that of the sixteen varieties, six separate from the other ten varieties. Variation at K=4 indicates that var. borreganus and var. yuccanus as well as var. australis distinguish themselves from other varieties which also holds true for larger values of K. Evaluating STRUCTURE groups for K = 16, the number of taxonomic varieties investigated here, shows that taxonomic varieties are not all genetically distinct from one another. However, for K = 16 there are several varieties that do appear distinct from others including var. antonius, var. kennedyi, and var. nigricalycis. Additionally, there are several varieties that appear to be genetically indistinguishable including var. australis and var. yuccanus, var. wilsonii and var. maricopae, and var. araneosus and var. vitreus.

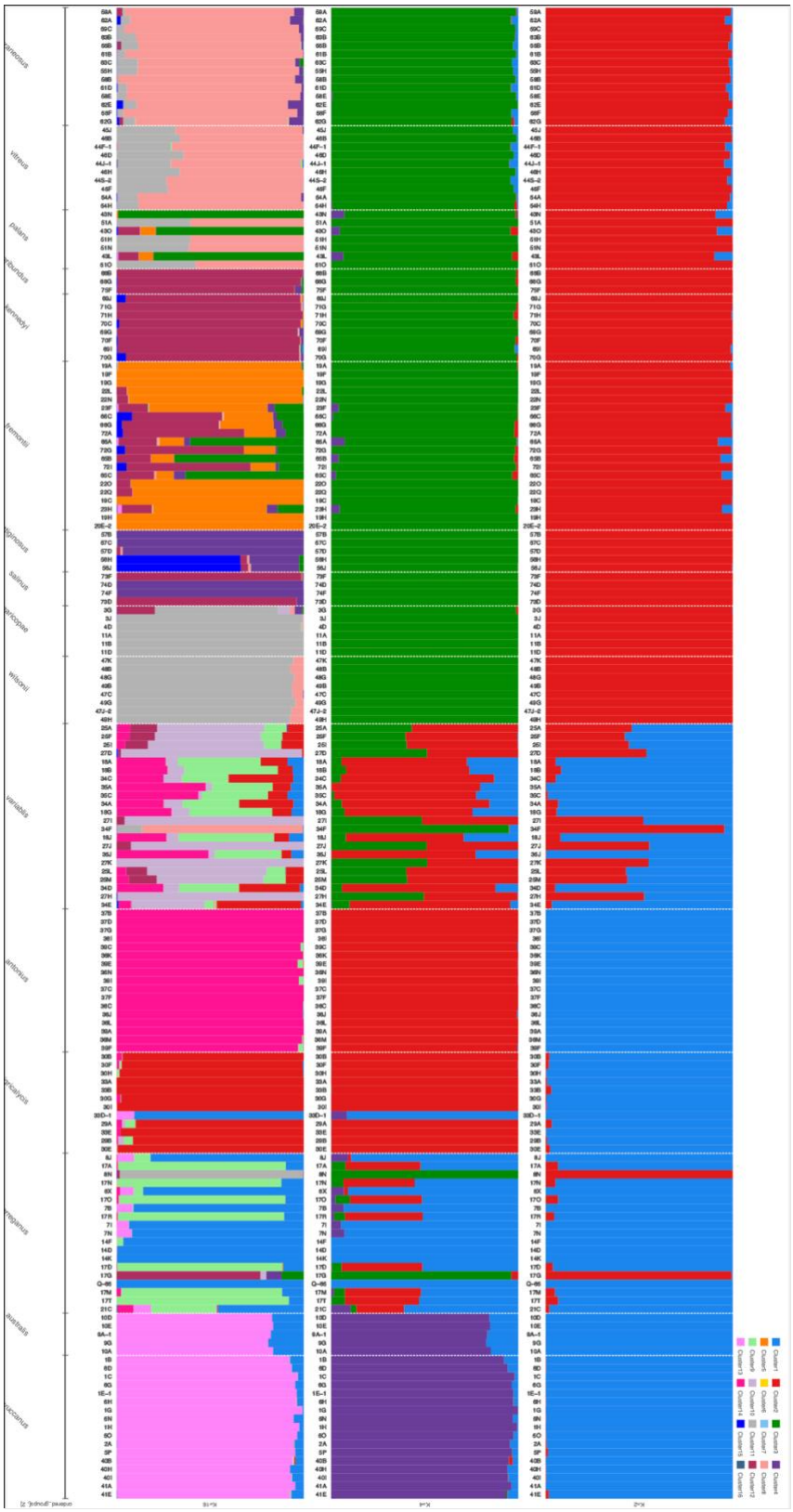

Figure 6. Stacked Structure Plots. Structure plot for K = 2 (top), K = 4 (middle) and K = 16 (bottom). Bars are grouped by variety and labeled by individual. Colors denote assigned K group.

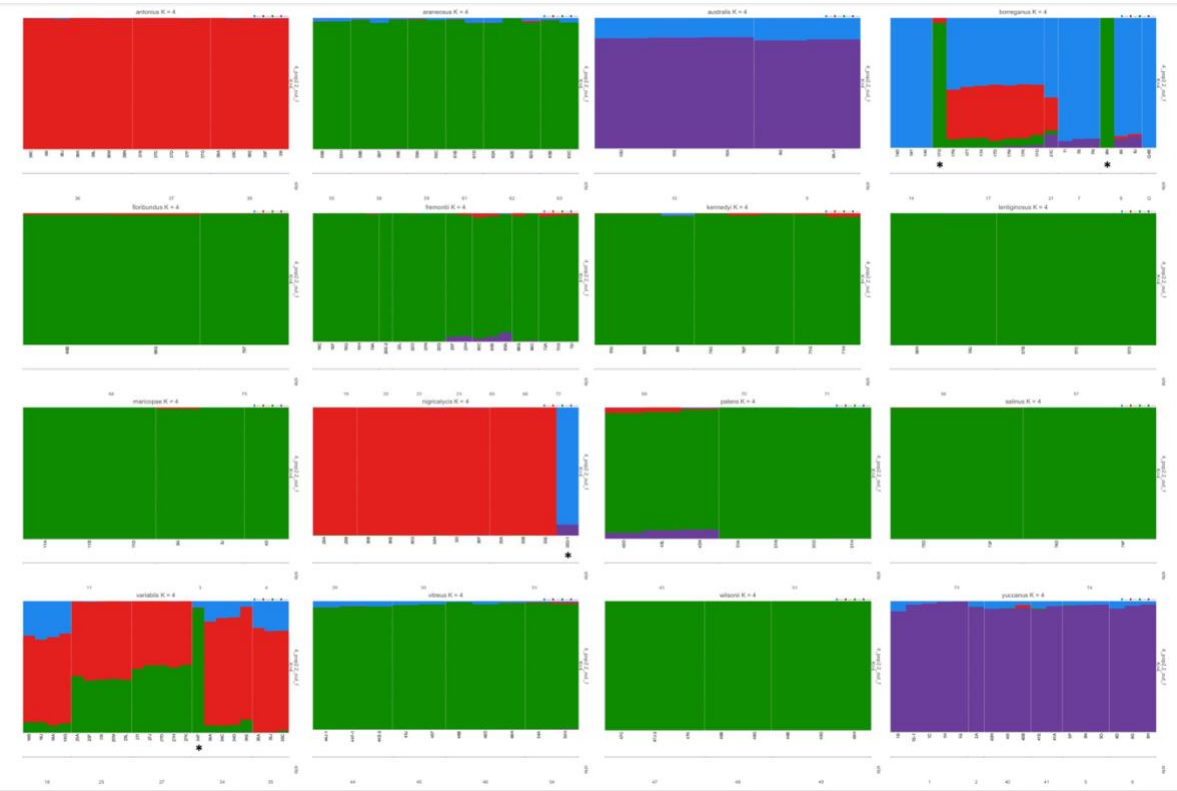

Figure 7. Structure Plots for each Variety at K = 4. Bars are grouped by site and labeled by individual. An asterisk (*) indicates the sample is an outlier.

Varieties explain much of the genetic variation occurring for K = 4 groups, however, I also investigated genetic structure at the site level. Most varieties do not exhibit variation across sites, that is, all sites are in the same genetic cluster, with the exception of var. borreganus and var. variablis. Sites 17 and 8 from var. borreganus show unique levels of admixture compared to other sites within var. borreganus (Figure 7). Sites 25 and 27 also show unique levels of admixture compared to other sites within var. variablis (Figure 7).

Pairwise F_{ST} was calculated at the variety, site and K level according to Weir and Cockerham (1984). Average F_{ST} between sites is 0.43 excluding F_{ST} values for sites with only one plant. The minimum F_{ST} value was –0.15 between sites 8 and 21 and a maximum F_{ST} value of 0.83 between sites 54 and 9 (Table 4). Average F_{ST} between varieties was slightly lower at

0.38 with a minimum pairwise F_{ST} value of 0.11 between var. maricopae and var. wilsonii and a maximum F_{ST} value of 0.69 between var. australis and var. lentiginosus (Table 5). Lastly, average pairwise F_{ST} between K groups was the lowest at a value of 0.29 with the minimum F_{ST} value of 0.23 between K2 and K3 and a maximum F_{ST} value of 0.36 between K1 and K2 (Table 6).

Table 4. Pairwise FST between A. lentiginosus sites according to Weir and Cockerham (1984).

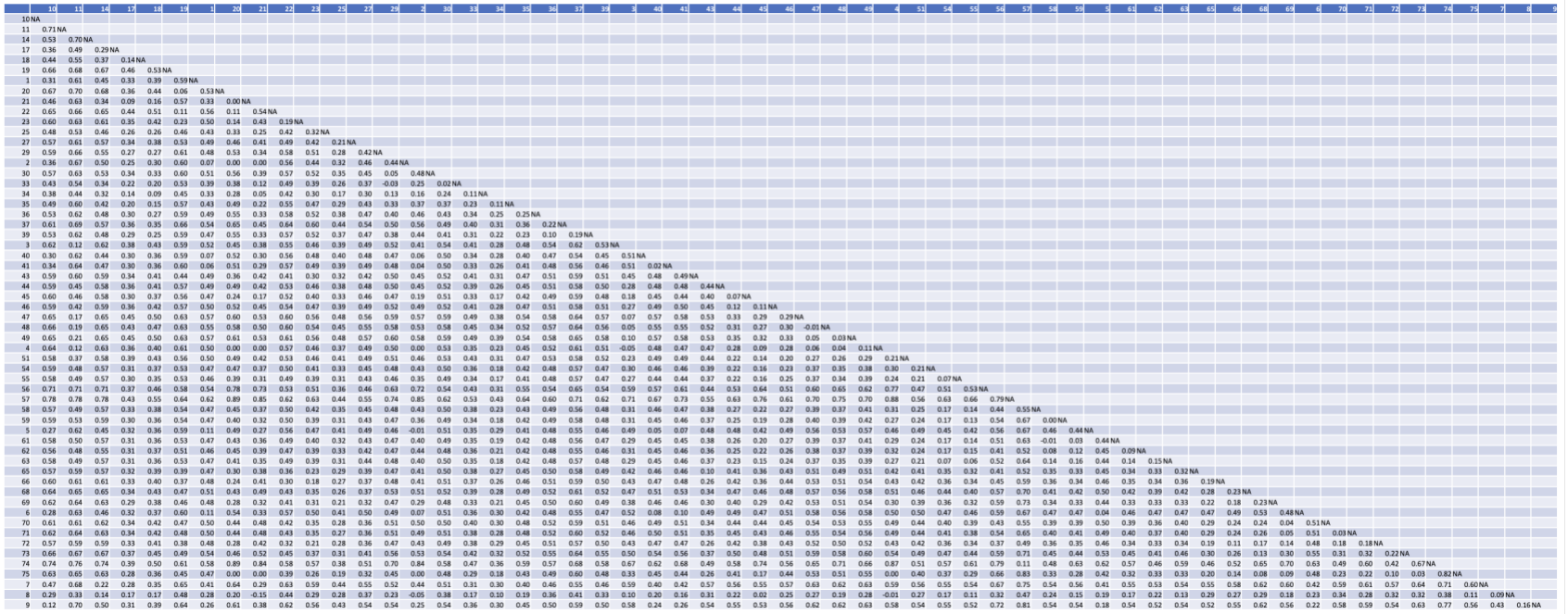

Table 5. Pairwise F_{ST} between *A. lentiginosus* Varieties. F_{ST} is calculated according to Weir and Cockerham (1984).

	antonius	araneousus	australis	borreganus	floribundus	fremontii	kennedyi	lentiginosus	maricopae	nigricalycis	palans	salinus	variablis	vitreus	wilsonii	yuccanus
antonius	NA															
araneousus	0.43798606	NA														
australis	0.48883146	0.4863798	NA													
borreganus	0.27986989	0.30813799	0.26323031	NA												
floribundus	0.49516157	0.33500094	0.60792109	0.35478537	NA											
fremontii	0.47085163	0.31477467	0.49078302	0.36744821	0.19511552	NA										
kennedyi	0.49325796	0.32727953	0.57816836	0.36579555	0.22864267	0.23002309	NA									
lentiginosus	0.53709025	0.33337625	0.69449787	0.39624658	0.53718869	0.31800848	0.43229821	NA								
maricopae	0.5649774	0.34988167	0.64081471	0.42613979	0.57230816	0.48093899	0.54755765	0.6420168	NA							
nigricalycis	0.36104315	0.4221986	0.48980421	0.27926301	0.46124122	0.44369811	0.46146316	0.52513484	0.55156425	NA						
palans	0.41662357	0.12642337	0.4358407	0.27404927	0.22063306	0.22737775	0.2550997	0.29787728	0.28600933	0.38863229	NA					
salinus	0.49283732	0.28455628	0.59100691	0.34890696	0.13309626	0.20702735	0.25101398	0.17946141	0.54744047	0.46414943	0.20319045	NA				
variablis	0.20865325	0.23069713	0.32246056	0.14884143	0.21018176	0.26500472	0.22977113	0.28289214	0.37913974	0.18933351	0.19249126	0.21926474	NA			
vitreus	0.45211887	0.1273139	0.50003539	0.31498696	0.38531725	0.3541304	0.37458794	0.41189722	0.29586244	0.437405	0.11782879	0.34929363	0.25273366	NA		
wilsonii	0.55777746	0.33294445	0.62534165	0.42588435	0.55897744	0.48009913	0.53613905	0.6012413	0.10991408	0.54836639	0.27220387	0.52987919	0.38121589	0.26012418	NA	
yuccanus	0.43803992	0.41411736	0.20641323	0.25553007	0.46625157	0.44029137	0.46227524	0.51399164	0.53906268	0.43262986	0.38064473	0.46561719	0.29456436	0.42772445	0.535321	NA

Table 6. Pairwise F_{ST} between K = 4 Structure Groups. F_{ST} is calculated according to Weir and Cockerham (1984).

	K1	K2	K3	K4
K1	NA			
K2	0.36335449	NA		
K3	0.30774151	0.23072812	NA	
K4	0.26149	0.34667982	0.26822225	NA

Nei's genetic distance was calculated between sites and used to evaluate isolation by distance across all sites of *A. lentiginosus*. Genetic distances between varieties were used to build a neighbor joining tree which highlighted similarities between closely related varieties. A neighbor joining tree indicates that varieties occupying the same geographic area tend to be more genetically similar (Figure 8). Comparison of matrices resulted in a mantel statistic of r = 0.491 (p-value < 0.001) indicating a positive correlation between genetic and geographic distance. Plants found at sites in closer proximity were more likely to be genetically similar (Figure 9). The spatial scale of population genetic structure across the landscape is depicted in Figure 10. *A. lentiginosus* populations from northern collection sites group within a large cohesive genetic

structure whereas *A. lentiginosus* populations from southern collection sites are divided into three structure groups from east to west (Figure 10).

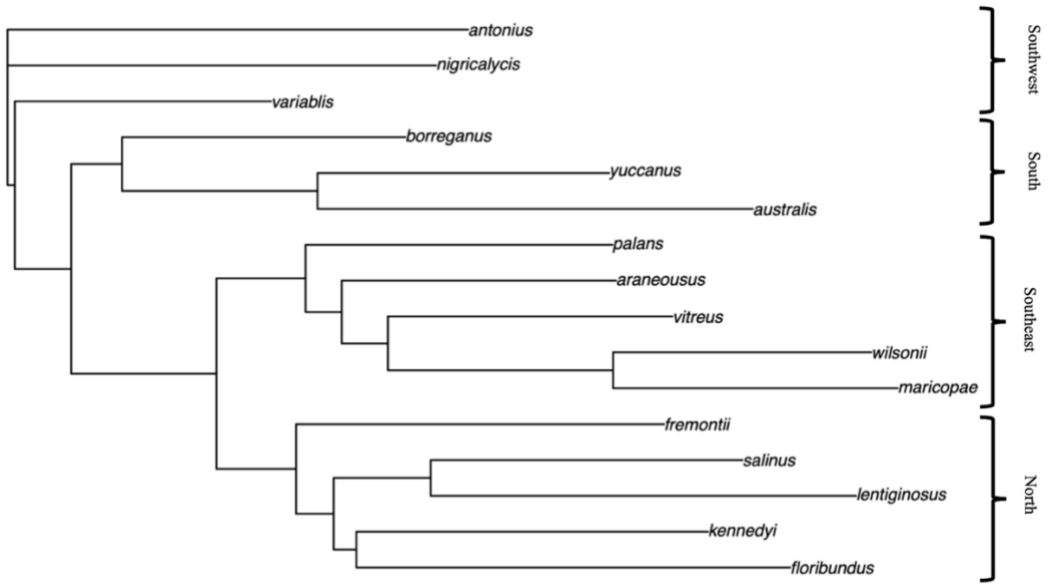

Figure 8. Neighbor Joining Tree between *A. lentiginosus* varieties using Nei's Genetic Distance. Varieties group closest with the varieties found in close proximity.

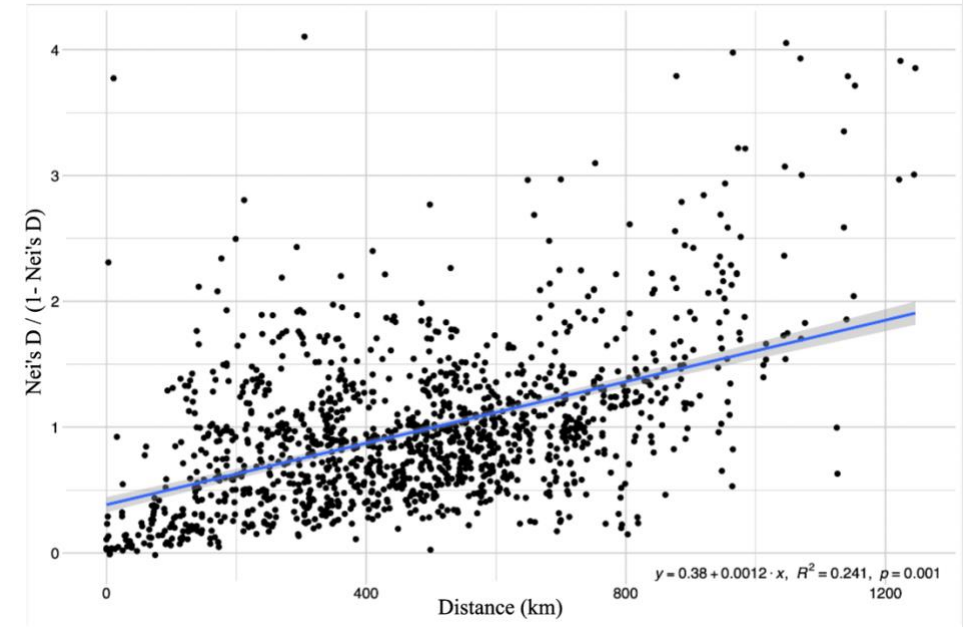

Figure 9. Isolation by Distance Model. Transformed Nei's genetic distance against geographic distance in km.

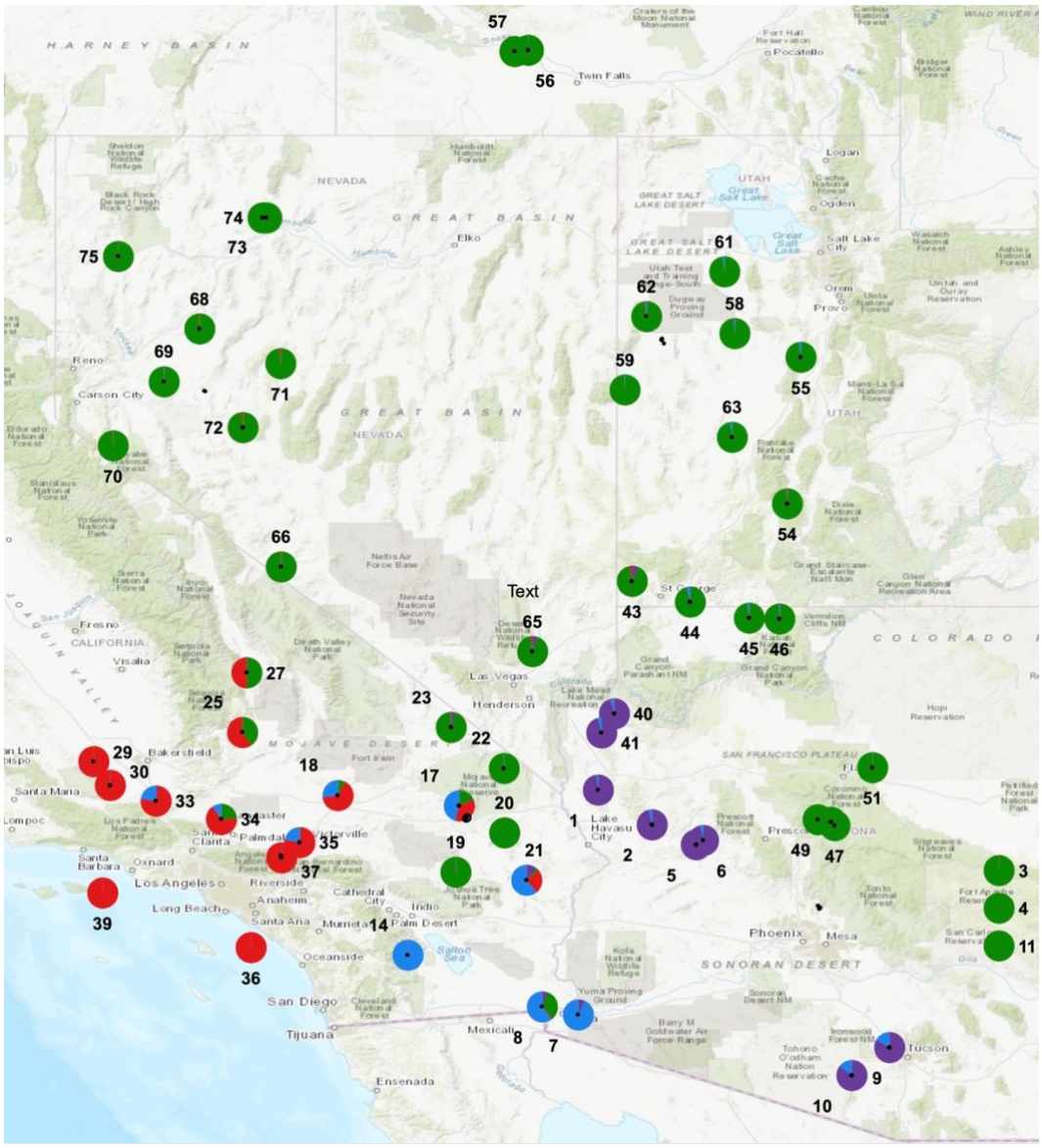

Figure 10. Geographic Map of Population Structure. Population structure of *A. lentiginosus S*ites across southwest United States. Colors correspond to K clusters described in Figure 6.

Lastly, an AMOVA was performed to determine at which level variation is best explained with regards to site and variety. The AMOVA revealed that the majority of genetic variation occurred among varieties (36.12%), and 12.92% at the site level within varieties, 8.5% among samples within site and 42.48% within individuals in the total population (Table 7).

Table 7. A Hierarchical Analysis of Molecular Variance (AMOVA).

	Df	SS	Variance (sigma)	% Variation	Phi-stat	P-value
Between Varieties	15	30443	79.18	36.10	0.36	0.001
Between sites within varieties	43	12697	28.34	12.92	0.20	0.001
Between samples within site	118	15396	18.64	8.50	0.17	0.001
Within samples	177	16492	93.18	42.48	0.58	0.001
Total	353	75029	219.35	100.00	-	-

Fruit Morphometric Architecture of *A. lentiginosus*

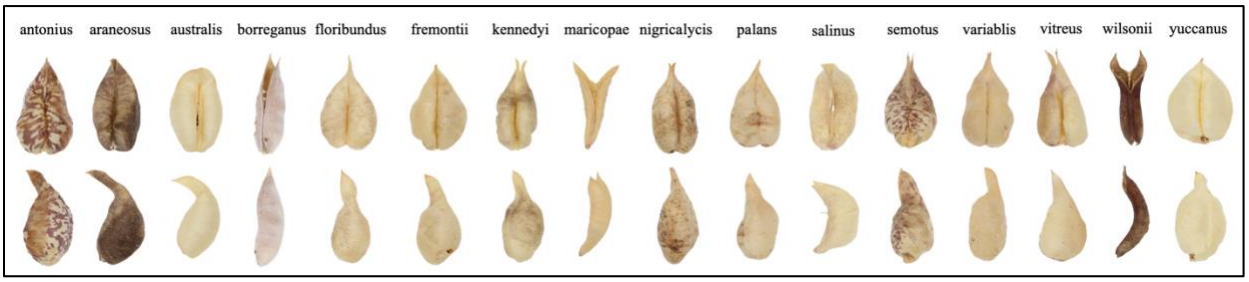

Figure 11. *A. lentiginosus* Fruits. Top and side views of representative fruits for each *A. lentiginosus* variety. Fruits are not scaled to size.

A. lentiginosus fruit shape variation was assessed from top and side view images separately. For the top view, there was a total of 2,505 fruit images from 288 plants and 14 varieties and for the side view, there was a total of 2,365 fruit images from 267 plants and 14 varieties. Examples of fruit morphologies across varieties are depicted in Figure 11. Plots of the procrustes shape variation show four primary regions of variation across plants, two at the top view for beak opening and overall length and two at the side view for beak curvature and pod width (Figure 12).

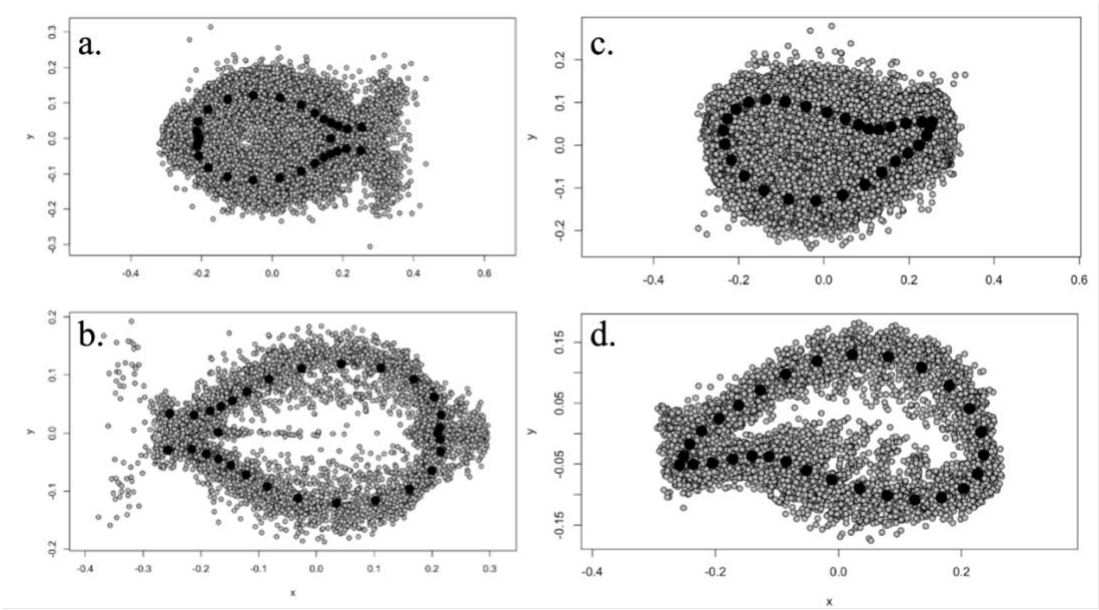

Figure 12. Procrustes Superimposition. (a.) Procrustes superimposition of the top view for all 2,505 fruits and (b.) all 288 plants. (c.) Procrustes superimposition of the top view for all 2,365 fruits and (d.) all 267 plants. Gray dots represent procrustes transformed coordinates of all fruits or plants. Black dots represent the average procrustes transformed shape.

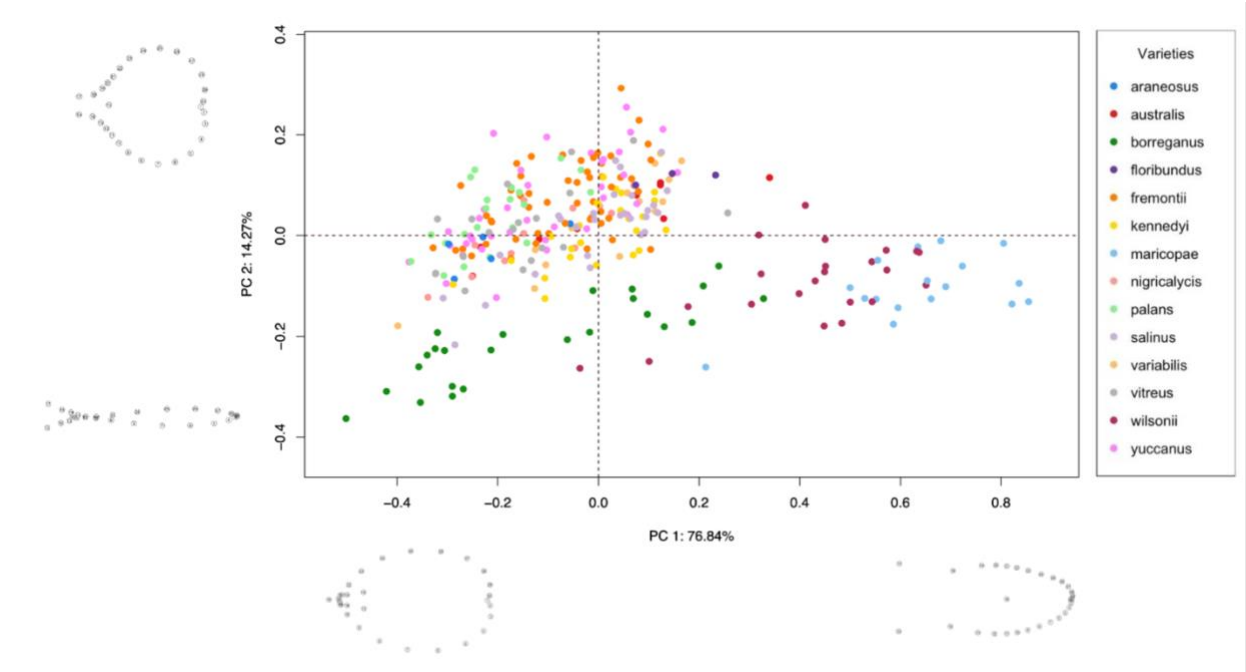

Figure 13. Shape Principal Component Analysis -Top View. PCA of the top view for 288 *A. lentiginosus* plants. Shapes along each axis indicate the maximum and minimum shape for the associated principal component.

A PCA of procrustes transformed coordinates for plants was performed to visualize patterns in shape variation in the top and side view. Over 90% of shape variation of the top view can be explained by the first two principal components. PC1 accounts for 78.84% of shape variation and PC2 accounts for 14.27% of shape variation (Figure 13). Variation along PC1 is attributed to primarily in the pod beak opening whereas variation along PC2 can be attributed to the length and width of the pod. Clustering of varieties according to fruit shape variation on the PCA indicates that only three varieties, var. borreganus, var. maricopae and var. wilsonii, separate from the eleven other varieties. Fruits from these varieties tend to be narrower with deep beak openings (Figure 11).

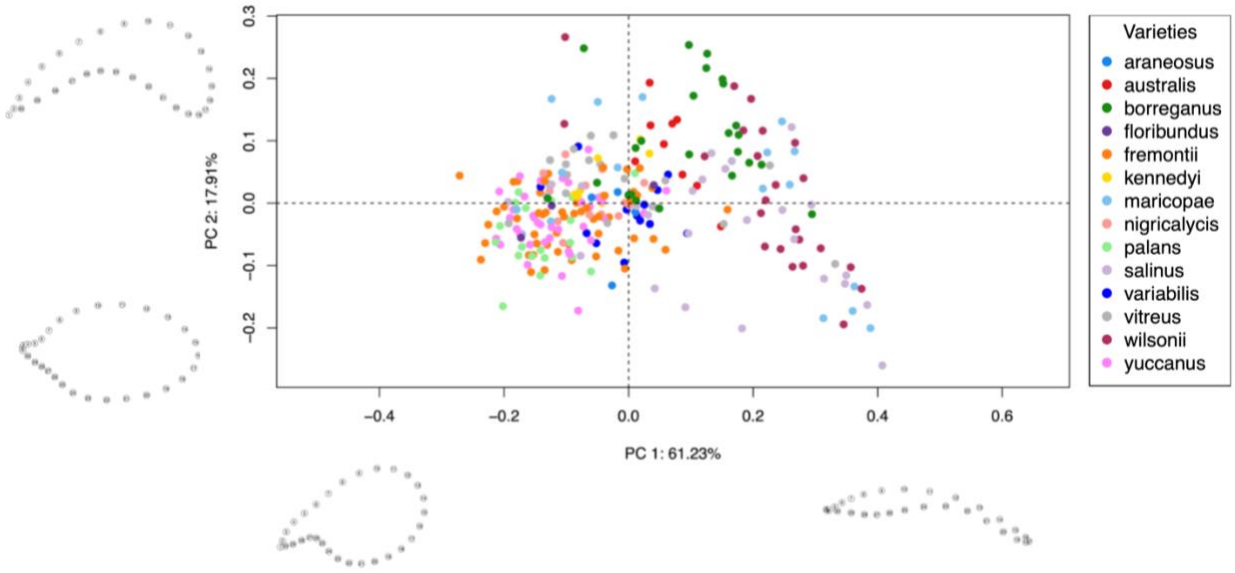

Figure 14. Shape Principal Component Analysis - Side View. PCA of the side view for 267 *A. lentiginosus* plants. Shapes along each axis indicate the maximum and minimum shape for the associated principal component.

Approximately 79% of shape variation in the side view can be explained by the first two principal components. PC1 accounts for 61.23% of shape variation and PC2 accounts for 17.91% of shape variation (Figure 14). Variation along PC1 can be attributed to pod width and length as

well as a little bit of beak curvature whereas variation along PC2 can be seen in beak and pod curvature. Clustering of varieties according to fruit shape variation on the PCA indicates that four varieties, var. borreganus, var. maricopae, var. salinus and var. wilsonii, separate from the 10 other varieties by the first principal component. Fruits from these varieties tend to be narrower with less prominent inflation in the lower half of the fruit (Figure 11). Variety australis also appears to have some clustering in the upper right quadrant of the PCA.

Table 8. A Hierarchical Analysis of Variance for Shape –Top and Side View. An ANOVA examining correlation between variety and sites within variety to fruit shape using randomized residuals in a permutation procedure (RRPP).

Top View	**Df**	**SS**	**MS**	**Rsq**	**F**	**Z**	**Pr(>F)**	
variety	13	16.2348	1.24883	0.63435	39.5928	13.135	0.001	**
variety:site	29	1.6301	0.05621	0.06369	1.7821	2.9984	0.001	**
Residuals	245	7.7278	0.03154	0.30195				
Total	287	25.5927						
Side View								
variety	13	6.1703	0.47464	0.59287	35.5182	11.8292	0.001	**
variety:site	28	1.2439	0.04443	0.11952	3.3245	7.0957	0.001	**
Residuals	224	2.9934	0.01336	0.28761				
Total	265	10.4076						

A nested ANOVA was performed to understand how well varieties and sites explain shape variation (Table 8). The ANOVA for the top view indicates that average fruit shape between at least two varieties differ and that average fruit shape between at least two sites within a variety differ significantly. While both independent variables are significant, the R square value for varieties is much larger than the R square value for sites indicating that variety plays a more prominent role in explaining variation of fruit shape. The ANOVA for the side view has similar results to that of the top view. Results indicate that average fruit shape between at least

two varieties differ and that average fruit shape between at least two sites in one variety differ. The R square value for varieties and sites within varieties explains 59.3% and 12.0% of fruit shape variation (Table 8).

Table 9. Shape Pairwise Differences between Varieties – Top and Side View. Pairwise differences between average fruit shape of *A. lentiginosus* varieties (lower diagonal) and the associated P-value (upper diagonal) for the top view (top) and side view (bottom). Significant distances are bolded.

	araneosus	australis	borreganus	floribundus	fremontii	kennedyi	maricopae	nigricalycis	palans	salinus	variabilis	vitreus	wilsonii	yuccanus
araneosus		0.109	0.112	0.054	0.188	0.094	0.001	0.807	0.759	0.120	0.114	0.696	0.001	0.321
australis	0.245		0.010	0.423	0.373	0.815	0.001	0.135	0.039	0.904	0.911	0.188	0.001	0.214
borreganus	0.214	**0.286**		0.013	0.001	0.001	0.001	0.025	0.002	0.001	0.003	0.010	0.001	0.001
floribundus	0.404	0.180	**0.434**		0.159	0.318	0.002	0.069	0.025	0.283	0.371	0.071	0.037	0.108
fremontii	0.177	0.104	**0.292**	0.238		0.109	0.001	0.164	0.062	0.194	0.270	0.279	0.001	0.631
kennedyi	0.231	0.056	**0.260**	0.190	0.107		0.001	0.052	0.003	0.821	0.981	0.065	0.001	0.047
maricopae	**0.863**	**0.646**	**0.777**	**0.543**	**0.738**	**0.656**		0.001	0.001	0.001	0.001	0.001	0.010	0.001
nigricalycis	0.076	0.184	**0.213**	0.337	0.120	0.172	**0.799**		0.535	0.088	0.095	0.868	0.001	0.352
palans	0.079	**0.231**	**0.275**	**0.375**	0.143	**0.225**	**0.860**	0.079		0.010	0.026	0.456	0.001	0.163
salinus	0.216	0.044	**0.259**	0.200	0.089	0.039	**0.663**	0.153	**0.206**		0.943	0.126	0.001	0.108
variabilis	0.229	0.051	**0.265**	0.190	0.102	0.025	**0.657**	0.170	0.221	0.037		0.158	0.001	0.165
vitreus	0.091	0.163	**0.235**	0.319	0.091	0.155	**0.793**	0.044	0.082	0.137	0.151		0.001	0.517
wilsonii	**0.648**	**0.434**	**0.564**	**0.356**	**0.529**	**0.441**	**0.234**	**0.587**	**0.649**	**0.451**	**0.442**	**0.580**		0.001
yuccanus	0.146	0.136	**0.282**	0.273	0.042	**0.136**	**0.766**	0.090	0.107	0.115	0.130	0.065	**0.558**	

	araneosus	australis	borreganus	floribundus	fremontii	kennedyi	maricopae	nigricalycis	palans	salinus	variabilis	vitreus	wilsonii	yuccanus
araneosus		0.115	0.033	0.953	0.937	0.614	0.489	0.607	0.113	0.022	0.365	0.002	0.876	0.971
australis	0.171		0.431	0.151	0.002	0.270	0.020	0.001	0.254	0.023	0.093	0.001	0.039	0.004
borreganus	**0.204**	0.079		0.064	0.001	0.067	0.001	0.001	0.177	0.026	0.043	0.001	0.002	0.001
floribundus	0.059	0.172	0.199		0.947	0.724	0.344	0.428	0.134	0.039	0.399	0.002	0.729	0.854
fremontii	0.045	**0.181**	**0.206**	0.049		0.273	0.045	0.056	0.001	0.001	0.048	0.001	0.133	0.533
kennedyi	0.093	0.112	0.149	0.088	0.090		0.248	0.043	0.217	0.030	0.471	0.001	0.663	0.261
maricopae	0.097	**0.156**	**0.185**	0.124	**0.097**	0.099		0.006	0.034	0.010	0.475	0.001	0.205	0.065
nigricalycis	0.086	**0.238**	**0.276**	0.113	0.098	**0.157**	**0.153**		0.001	0.001	0.004	0.001	0.047	0.177
palans	0.159	0.092	0.086	0.159	**0.152**	0.104	**0.115**	**0.232**		0.136	0.412	0.001	0.011	0.001
salinus	**0.228**	**0.187**	**0.157**	**0.228**	**0.211**	**0.180**	**0.173**	**0.298**	0.107		0.085	0.015	0.001	0.001
variabilis	0.112	**0.127**	**0.137**	0.120	**0.101**	0.082	0.063	**0.182**	0.065	0.125		0.001	0.140	0.048
vitreus	**0.325**	**0.233**	**0.213**	**0.334**	**0.320**	**0.278**	**0.244**	**0.392**	**0.189**	**0.166**	**0.222**		0.001	0.001
wilsonii	0.054	**0.136**	**0.180**	0.079	0.067	0.060	0.076	**0.110**	**0.131**	**0.209**	0.091	**0.296**		0.349
yuccanus	0.039	**0.178**	**0.210**	0.062	0.036	0.093	0.095	**0.079**	**0.159**	**0.228**	**0.111**	**0.326**	0.054	

Because the ANOVA indicates significant variation at the variety level for both the top and side view, a pairwise comparison of mean fruit shape between varieties was made. Pairwise distances between fruit shape means of the top view indicates that there are three varieties that consistently differ in shape compared to the remaining eleven varieties, var. borreganus, var. maricopae and var. wilsonii (Table 9). These findings agree with the observations from the PCA. Another variety that shows some significant differences of shape mean between other varieties is var. palans. While var. maricopae and var. wilsonii have significantly different shape means, the distance between their shape means is the smallest when comparing each of them to other

varieties. The largest significant difference in shape mean was 0.863 between var. maricopae and var. araneosus.

Pairwise comparisons of fruit shape means of the side view indicate more significant differences between varieties than found in differences of shape from the top view. Four varieties stood out as being distinct from the others (significantly different from >8 varieties), var. australis, var. borreganus, var. nigricalycis, and var. vitreus. The maximum significant difference between fruit shapes was 0.392 between var. vitreus and var. nigricalycis.

Table 10. Shape Variance Pairwise Differences between Varieties – Top and Side View. Pairwise differences between variances of the side view of *A. lentiginosus* varieties (lower diagonal) and the associated P-value (upper diagonal) for the top view (top) and side view (bottom). Significant distances are bolded.

	araneosus	australis	borreganus	floribundus	fremontii	kennedyi	maricopae	nigricalycis	palans	salinus	variabilis	vitreus	wilsonii	yuccanus
araneosus		0.013	0.001	0.854	0.952	0.955	0.770	0.823	0.910	0.039	0.776	0.002	0.073	0.628
australis	**0.050**		0.028	0.038	0.001	0.001	0.001	0.003	0.001	0.164	0.001	0.223	0.085	0.001
borreganus	**0.080**	**0.030**		0.006	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.208	0.001	0.001
floribundus	0.004	**0.046**	**0.076**		0.807	0.795	0.633	0.964	0.894	0.108	0.677	0.011	0.205	0.560
fremontii	0.001	**0.051**	**0.081**	0.005		0.998	0.689	0.690	0.750	0.001	0.684	0.001	0.001	0.394
kennedyi	0.001	**0.051**	**0.081**	0.005	0.000		0.708	0.695	0.793	0.002	0.698	0.001	0.004	0.472
maricopae	0.004	**0.055**	**0.084**	0.009	0.004	0.004		0.487	0.534	0.001	0.992	0.001	0.003	0.816
nigricalycis	0.003	**0.047**	**0.077**	0.001	0.004	0.004	0.008		0.914	0.004	0.503	0.001	0.019	0.302
palans	0.002	**0.049**	**0.078**	0.002	0.002	0.002	0.006	0.001		0.001	0.586	0.001	0.008	0.327
salinus	**0.034**	0.017	**0.046**	0.030	**0.034**	**0.034**	**0.038**	**0.031**	**0.032**		0.003	0.002	0.525	0.001
variabilis	0.005	**0.055**	**0.085**	0.009	0.004	0.004	0.000	0.008	0.006	**0.038**		0.001	0.012	0.854
vitreus	**0.067**	0.017	0.013	**0.063**	**0.067**	**0.067**	**0.071**	**0.064**	**0.065**	**0.033**	**0.071**		0.002	0.001
wilsonii	**0.028**	0.022	**0.052**	0.024	**0.029**	**0.029**	**0.032**	0.025	**0.026**	0.006	**0.032**	**0.039**		0.001
yuccanus	0.007	**0.057**	**0.087**	0.011	0.006	0.006	0.002	0.010	0.008	**0.041**	0.002	**0.074**	**0.035**	

	araneosus	australis	borreganus	floribundus	fremontii	kennedyi	maricopae	nigricalycis	palans	salinus	variabilis	vitreus	wilsonii	yuccanus
araneosus		0.780	0.820	0.601	0.977	0.798	0.066	0.905	0.141	0.556	0.184	0.354	0.085	0.281
australis	0.005		0.503	0.795	0.719	0.991	0.012	0.639	0.022	0.290	0.057	0.134	0.014	0.066
borreganus	0.004	0.009		0.463	0.662	0.549	0.016	0.867	0.054	0.612	0.114	0.331	0.022	0.183
floribundus	0.009	0.004	0.013		0.642	0.795	0.042	0.518	0.071	0.270	0.091	0.179	0.043	0.117
fremontii	0.001	0.004	0.004	0.009		0.777	0.002	0.771	0.005	0.332	0.019	0.083	0.001	0.013
kennedyi	0.005	0.000	0.008	0.005	0.004		0.019	0.648	0.051	0.321	0.065	0.180	0.018	0.096
maricopae	0.032	**0.037**	**0.028**	**0.041**	**0.032**	**0.036**		0.009	0.477	0.095	0.419	0.128	0.807	0.094
nigricalycis	0.002	0.007	0.002	0.011	0.002	0.006	**0.030**		0.033	0.495	0.079	0.246	0.011	0.141
palans	0.025	**0.029**	0.021	0.034	**0.025**	0.029	0.007	**0.023**		0.279	0.849	0.325	0.610	0.340
salinus	0.010	0.015	0.007	0.020	0.011	0.015	0.021	0.009	0.014		0.342	0.714	0.131	0.600
variabilis	0.022	0.027	0.018	0.032	**0.023**	0.027	0.010	0.020	0.002	0.012		0.523	0.515	0.544
vitreus	0.015	0.019	0.011	0.024	0.015	0.019	0.017	0.013	0.010	0.004	0.008		0.166	0.846
wilsonii	0.029	**0.034**	**0.025**	**0.039**	**0.030**	**0.034**	0.003	**0.027**	0.005	0.019	0.007	0.015		0.130
yuccanus	0.016	0.021	0.013	0.026	0.017	0.021	0.015	0.014	0.008	0.006	0.006	0.002	0.013	

Morphological disparity was tested to evaluate differences between the amount of variance between varieties. For the top view, there are five varieties, var. australis, var. borreganus, var. salinus, var. vitreus and var. wilsonii, that show consistent significant difference between variances compared to other varieties (Table 10). The average variances for these

varieties are 0.099, 0.128, 0.082, 0.115, and 0.076 respectively. These variances are higher compared to the remaining 9 varieties that have variances closer to 0.045. For the side view, there are two varieties- var. maricopae and var. wilsonii- that show consistent significant difference between variances compared to other varieties (Table 10). The average variances for these varieties are 0.047 and 0.044 respectively. Var. palans also showed some significant differences in variance and had the next largest variance at 0.039. These variances are higher compared to the remaining eleven varieties that have variances ranging from 0.005 to 0.031.

Table 11. Regression between Size and Shape – Top and Side View.

Top View	Df	SS	MS	Rsq	F	Z	Pr(>F)
log(Csize)	1	0.201	0.201	0.008	2.266	1.244	0.117
Residuals	286	25.392	0.089	0.992			
Side View							
log(Csize)	1	0.075	0.075	0.007	1.889	1.190	0.124
Residuals	265	10.457	0.039	0.993			
Total	266	10.531					

A linear model was also used to test how well shape variation is explained by size. Tests were insignificant for both the top view and the side view suggesting there is no significant correlation between fruit size and shape (Table 11). Not only is the model insignificant, but Rsq values indicate that centroid size poorly explains fruit shape. These results suggest that size does not play a role in fruit shape variation.

Because it was noted that variation in beak opening, and beak depth was a defining characteristic of variation in PC1, I evaluated differences in beak opening and beak depth without influence of overall shape using Tukey tests. Both var. maricopae and var. wilsonii are significantly larger in both beak opening and beak depth than other varieties (Figure 15). While var. borreganus shares the narrow fruit shape similar to var. wilsonii and var. maricopae, it

differentiates from both varieties in its beak characteristics. Other patterns indicate var. araneosus has smaller beak opening and depth than several varieties (Figure 15). Many varieties have enough variation within beak opening and beak depth to be indistinguishable from other varieties in beak characteristics.

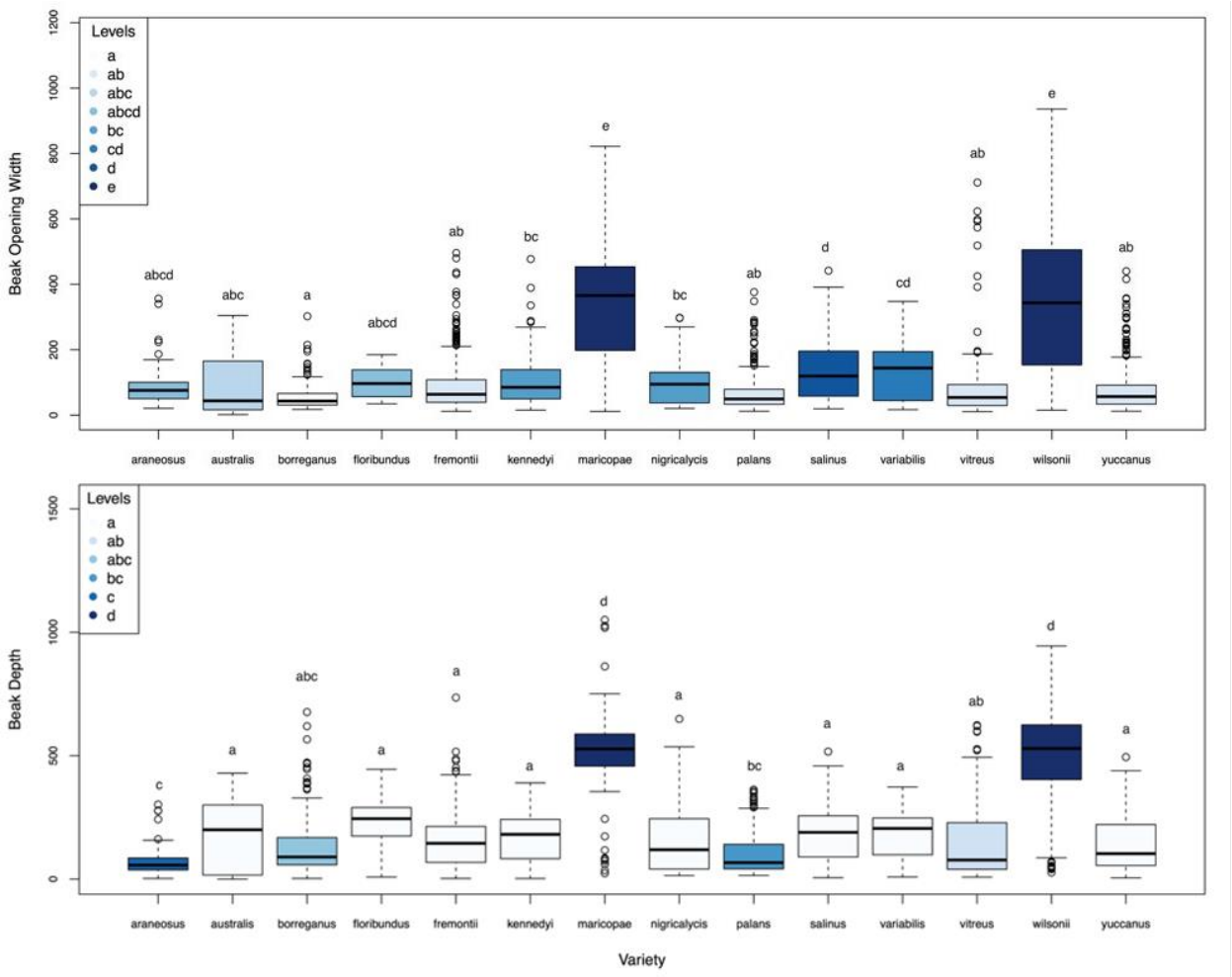

Figure 15. Differences in Beak Opening and Depth. Tukey test to evaluate differences between fruit beak opening width and fruit beak opening depth between varieties of *A. lentiginosus.* Significance groups are denoted by color and letters above bars.

Comparative Analysis between Fruit Shape Variation and Genetic Variation

Mean P_{ST} among sites was calculated for shape variables for each view of the *A. lentiginosus* fruits. P_{ST} was evaluated using a ratio of 1 for additive genetic variance to narrow

sense heritability as recommended by Brommer (2011). Mean P_{ST} values of shape variables are all larger than mean F_{ST} for both the top view and the side (Figure 16).

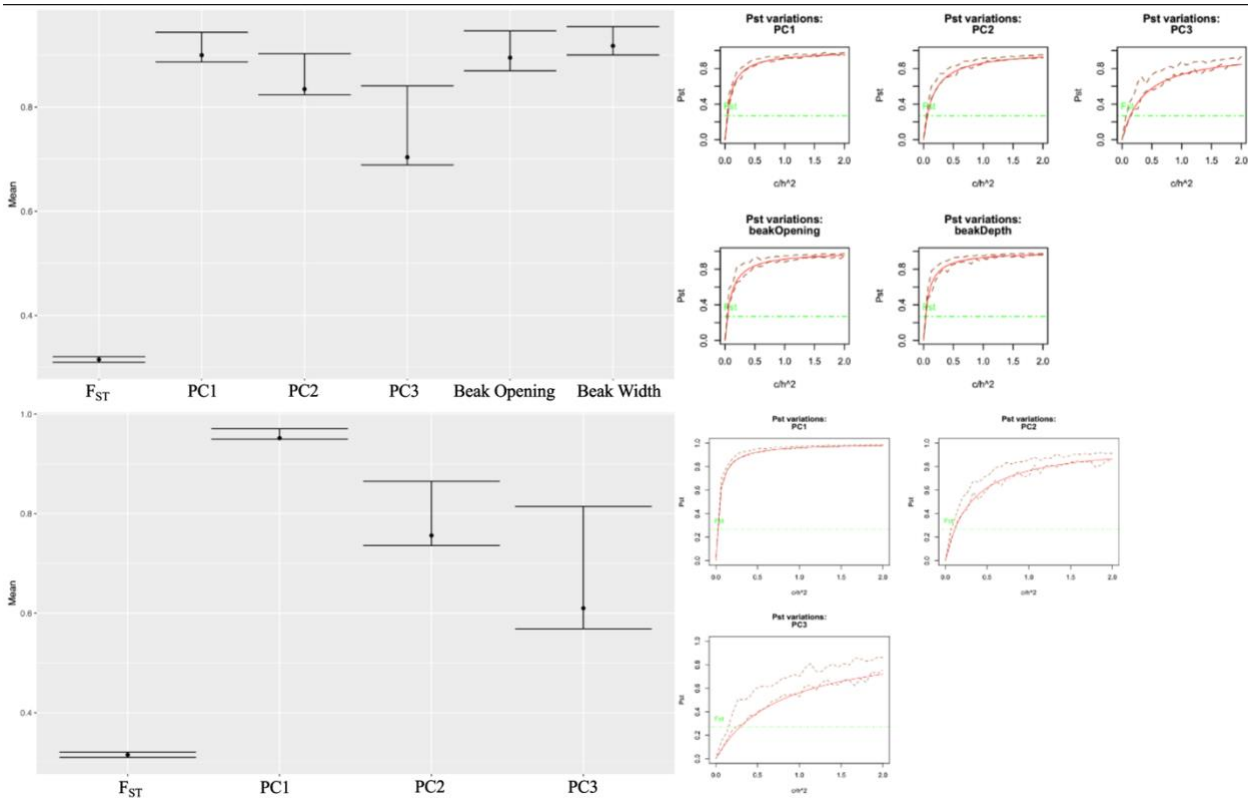

Figure 16. Mean P_{ST} and P_{ST} variation. Mean P_{ST} and F_{ST} for each top view shape variables (top left), side view shape variables (bottom left) among sites of *A. lentiginosus* and P_{ST} variations across c/h^2 ratios of 0-2 for top view (top right) and side view (bottom right).

Robustness of P_{ST} estimates was tested for c/h^2 ratios ranging from 0.0-2.0 and was found to be greater than mean F_{ST} for all values of c/h^2 higher than 0.38 (Figure 16). When removing PC3, all P_{ST} estimates were greater than F_{ST} for all c/h^2 values greater than 0.2.

CHAPTER FOUR

DISCUSSION

A long-held goal in evolutionary biology is to understand the forces that shape phenotypic variation. The interaction of evolutionary forces like natural selection and neutral processes like isolation by distance across a wide heterogeneous landscape provides insight on how phenotypic variation in a system arises (Borsa et al., 1997; Hutchison & Templeton, 1999). As the most taxonomically diverse species in North American flora, *A. lentiginosus* and its corresponding varieties offer a unique opportunity to investigate mechanisms that support differentiation among widespread populations. In this study, population genomic analyses using over 5000 neutral genetic markers has identified four primary genetic clusters and significant isolation by distance. Moreover, much of the genetic variation occurs among varieties and there is evidence that a majority of the taxonomic varieties represent discrete genetic entities. Key morphological characteristics relating to pod inflation and beak opening can distinguish several taxonomic varieties from one another. Comparison of genetic and fruit morphological variation suggests that *A. lentiginosus* fruit morphological variation is likely driven by selection.

Population Genomic Structure of *A. lentiginosus*

The overall genetic structure of *A. lentiginosus* was best described by four genetic groupings according to STRUCTURE. There was considerable agreement between how the PCA and STRUCTURE identified groups, with PC1 explaining divisions for all structure groups. The existence of genetic structure in this system is unsurprising as similar studies have found structure across large and fine scale populations of *A. lentiginosus* (Knaus et al., 2005; Harrison

et al., 2019). AMOVA partitioned most of the genetic variance occurring among varieties, indicating that taxonomic varieties are genetically distinct entities on at least some level. Indeed, for fourteen of the sixteen varieties investigated, individuals within a taxonomic variety clustered with one another for both PCA and STRUCTURE results. Results here agree with findings from Knaus et al. (2005) which found that taxonomic varieties of *A. lentiginosus* have unique genetic structure. Knaus examined three varieties, var. salinus, var. lentiginosus and var. variablis with populations coming from two sites per variety. There, it was found that each variety was genetically distinct from one another, even for var. salinus and var lentiginosus for which distributions overlap. Interestingly, although var. salinus and var. lentiginosus are geographically close to one another, var. salinus was more genetically differentiated from the other two varieties. In contrast with Knaus, in the present study, individuals from var. salinus and var. lentiginosus clustered more closely together with respect to each other than when compared to individuals from var. variablis. Structure for these three varieties was also not found to be unique in that individuals could not be assigned into a taxonomic variety based on the STRUCTURE results.

While individuals within a variety cluster with one another, those clusters sometimes overlap with individuals from another variety, indicating that not all varieties are unique in their genetic structure. This study included sixteen varieties of *A. lentiginosus*, when examining sixteen levels of K in STRUCTURE several, but not all varieties, correspond to structure groups, indicating that while many varieties are distinct entities there is still considerable gene flow among several varieties. At K=16, there are two patterns that emerge across varieties; Varieties can be described as distinct meaning all individuals within the variety are assigned to the same cluster. If all individuals in a variety are assigned to a single cluster and there are no other

varieties assigned to that cluster, the variety may also be considered unique to other varieties. These patterns are supported by results from the neighbor joining tree.

There are two varieties, var. antonius and var. nigricalycis, that are both genetically distinct and unique. These varieties are assigned distinct clusters when run in STRUCTURE and have consistently high F_{ST} values when compared with other varieties. These varieties also appear to be further separated from most other varieties according to the neighbor joining tree. Var. nigralycis is documented to have yellow flowers whereas var. antonius has light pink flowers. Interestingly, these varieties are close in geographic distance to one another, separated by less than 200km. Even so, they are distinct from one another, likely because var. antonius is found at a much higher elevation than var. nigricalycis. This is supported by research from Harrison et al. (2019) which suggests that genetic variation and abiotic variation are linked to one another.

There are many varieties that are genetically distinct from other varieties but not unique in their genetic structure. For example, var. maricopae and var. wilsonii are assigned to the same cluster when K=16 and have low F_{ST} values between each other, however, F_{ST} values between other varieties are high and no other varieties are assigned to their cluster in STRUCTURE. These two varieties are only 100km apart and group together in the neighbor joining tree more closely than with any other variety. fruit shape characteristics for var. maricopae and var. wilsonii are highly similar, with both having narrower pods and wider beak openings than all other varieties. It is recommended to combine these two taxonomic varieties into a single merged variety.

Another pair of varieties that appear to be highly similar to one another are var. australis and var. yuccanus. Not only do the varieties share the same assigned cluster for K=16 but they

are two varieties that separate from the remaining fourteen varieties at K=4. These two varieties are further apart in geographic distance than previously mentioned varieties, however there is evidence for sufficient gene flow despite this distance. Neighbor joining tree also these two varieties to be separate from other distinct varieties. These two varieties share similar morphological characteristics with both having inflated pods. Again, it is recommended that these two varieties likely belong to a single genetic group and should therefore be merged into one taxonomic variety.

Variety araneousus and var. vitreus are another pair of varieties that appear to cluster closely together. The neighbor joining tree indicates that these two varieties are closely related to var. maricopae, var. wilsonii and var. palans. These varieties may be considered similar enough to one another to be categorized as a single taxonomic variety. Additionally, var. floribundus and var. kennedyi are highly similar to one another and may also be considered as a single taxonomic variety.

The remaining six varieties showed inconsistent genetic structure at the variety level for K=16 and tended to have lower F_{ST} values when compared to other varieties. Further investigation of some of these varieties like var. lentiginosus and var. salinus may provide evidence for new taxonomic classifications. Alternatively, other varieties like var. borreganus and var variablis may not be actual genetic entities. Both varieties contain very high levels of admixture unlike other varieties and sites for each variety grouped with different genetic clusters in STRUCTURE and PCA. Both varieties are adjacent to multiple other varieties with no clear geographic barriers between them. These varieties also span larger geographic ranges compared to other varieties indicating that geographic location may play a more important role in underlying genetic structure than prescribed taxonomic variety. It is thus recommended that

classification for these varieties is abandoned. This also demonstrates that isolation by distance may play a prominent role in shaping the genetic structure of *A. lentiginosus*. Similar patterns of population structure according to geographic location were found for the diverse widespread Californian species *Mimulus guttatus* (Twyford et al., 2020). However, clustering programs like STRUCTURE that do not incorporate a spatial component may detect artificial genetic clusters when rather, spatial correlation from IBD better characterizes genetic structure (Guillot et al., 2009). However, when investigated for a variety with a restricted range, Harrison et al. (2019) demonstrated that there is considerable structure at the site level, even in close proximity.

Isolation by distance models indicate geographic distance significantly explains variation among *A. lentiginosus* sites. For example, sampling site 3 and 11, both identified as var. maricopae, are only three kilometers apart and Nei's distance between the two is 0.12 whereas sites 11 (var. maricopae) and 57 (var. lentiginosus) are 1047 kilometers apart and Nei's distance between the two is 0.80. There are some exceptions to this trend, such as sites 30 (var. nigricalycis) and 37 (var. antonius) are only 186 kilometers apart but they have a genetic distance of 0.49 whereas sites 25 (var. variablis) and 66 (var. fremontii) are 197 kilometers apart and have a genetic distance of 0.29. Sites 19 (var. fremontii) and 72 (var. fremontii) are over 500km apart but have a relatively low genetic distance of 0.37. In cases where genetic distance and geographic distances do not correlate, there is reason to believe there may be some barriers to gene flow. For example, site 37 (var. antonius) is found at a much higher elevation than site 30 (var. nigricalycis), even though they are geographically close to one another. It is important to note that genetic distances at the site level may be artificially inflated due to low sample size at all sites therefore minimizing the variation found within sites (Willing et al., 2012).

Characterization of population structure is of particular importance for conservation studies. Of the varieties sampled in this study, var. antonius is the only one listed as endangered. We confirm that this is likely a genetically unique variety of *A. lentignosus* and thus should maintain its endangered conservation status.

Fruit Shape Architecture of *A. lentiginosus*

This study evaluates morphometric architecture of *A. lentiginosus* fruits from both the top and side view. The primary components of shape variation are related to beak opening (top view), pod inflation (top and side view) and pod curvature (side view). Beak opening, and depth were confirmed to be defining morphological characteristics for a few varieties (Figure 13). It is of note that for all fruits that have very wide beak openings, none also have an inflated pod (Figure 11). I predict this is due to differing mechanisms of seed dispersal. It is thought that fruits with inflated pods also have papery thin shells and thus disperse seeds by the fruit blowing in the wind and eventually breaking down to release seeds (Augspurger, 1989). This is the case with other legumes like *Crotalaria* where inflated pods act as balloons and are likely dispersed by wind (Le Roux et al., 2011). In contrast, for fruits with wide beak openings, the "wings" formed by the opening may act as a sail, allowing fruits to be blown further distances by wind (Houghton et al., 2020). This has been investigated as an adaptive dispersal mechanism in other *Astragalus* species (Houghton et al., 2020). One explanation for this, termed "informed dispersal", is that plants adapt to changes in local environmental conditions by modifying dispersal strategy (Clobert et al., 2009). It is currently unclear whether morphological variation in beak opening in *A. lentiginosus* is a result of informed dispersal or the rather the result of other adaptive pressures. Nonetheless, P_{ST}-F_{ST} analysis indicates that beak opening is under selection in *A. lentiginosus* (see below).

For both the top and side view, shape alone cannot distinguish between all fourteen taxonomic varieties. Rather, shape appears to be a good predictor for some varieties such as var. maricopae, var. wilsonii, var. borrreganus and var. vitreus. Other varieties appear to have morphometric variation that overlaps with one another. These results are similar to findings from Knaus (2010) which suggested that morphometric shape PCs do not discreetly describe Barneby's varieties of *A. lentiginosus* but rather describe trends within varieties and between similarly classified varieties. Based on observations during field collections and information from other researchers, it would be worthwhile to investigate additional *A. lentiginosus* fruit characteristics such as pod thickness and pod color patterns as well as evaluate fruit shape from a three-dimensional perspective to learn how characteristics of the top and side view combined may further elucidate patterns of morphological differentiation.

Population Genetic Signature of Selection Shaping Fruit Morphology in *A. lentiginosus*

Comparison of P_{ST} to F_{ST} indicates high potential for local adaptation, supporting the hypothesis that fruit shape may be an adaptive trait in *A. lentiginosus*. P_{ST} values were consistently higher than F_{ST} across all ranges of c/h^2 and for all population stratifications.

Fruit shape has been demonstrated to be an adaptive trait in other species. Common fruit adaptations known to make feeding more difficult for insect herbivores include waxes, trichomes, husk hardness, sclerophylly, latex deposition and lactiferous structures (Janzen, 1971; Fürstenberg-Hägg et al., 2013; Dalin & Björkman, 2003; War et al., 2018). Insect herbivores have been documented to have profound impacts on fitness in *Astragalus* (Green & Palmbald, 1975; Martin and Menke 2012). Insect population size and levels of herbivory often vary across the landscape, establishing variable selection across space and time (Thompson 1994 & 2005; Whitney & Stanton, 2004; Althoff et al., 2014). Correlation between insect populations and

herbivory in *A. lentiginosus* populations to the genetic and morphological variation may further elucidate whether insects shape their structure.

Specific to legumes, pod shattering is a frequent selective trait of agronomic interest (Bandillo et al., 2017; Ogutcen et al., 2018). One study found several SNPs associated with pod shattering in soybeans, indicating that shattering was more strongly selected for in particularly arid regions (Bandillo et al., 2017). This may also be the case for var. maricopae and var. wilsonii for which we noted considerably large beak openings. Environmental variables have been documented to have a correlation with genetic and morphological variation (Harrison et al., 2019; Knaus 2010) although it has not yet been verified for these varieties.

This study investigates the population structure of *A. lentiginosus* and directly compares it to the morphological variation found. Further research on morphological variation in *Astragalus lentiginosus* populations, and how it correlates with environmental differences, can provide a deeper understanding of factors driving *Astragalus* diversification.

REFERENCE LIST

Adams D, Collyer M, Kaliontzopoulou A, Baken E (2022). “Geomorph: Software for geometric morphometric analyses. R package version 4.0.4.” https://cran.r-project.org/package=geomorph.

Adams, D. C., Rohlf, F. J., Slice, D. E. (2013). A field comes of age: geometric morphometrics in the 21st century. Hystrix, the Italian Journal of Mammalogy, 24(1), 7-14. https://doi.org/10.4404/hystrix-24.1-6283

Agrawal, A.A., Hastings, A.P., Fines, D.M., Bogdanowicz, S. and Huber, M. (2018), Insect herbivory and plant adaptation in an early successional community*. Evolution, 72: 1020-1033.

Althoff, David M., Kari A. Segraves, and Marc TJ Johnson. "Testing for coevolutionary diversification: linking pattern with process." Trends in Ecology & Evolution 29.2 (2014):82-89.

Andrews, K., Good, J., Miller, M. et al. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet 17, 81–92 (2016).

Baken E, Collyer M, Kaliontzopoulou A, Adams D (2021). “geomorph v4.0 and gmShiny: enhanced analytics and a new graphical interface for a comprehensive morphometric experience.”

Barneby, Rupert C. "Atlas of North American Astragalus." Mem NY Bot Gard 13 (1964): 1-1188.

Bhattarai, G., Bonhomme, V. & Conner, P. Image-based morphometric analysis reveals moderate to highly heritable nut shape traits in pecan. Euphytica 218, 102 (2022). https://doi.org/10.1007/s10681-022-03049-1

Blanquart F, Gandon S, Nuismer SL (2012). The effects of migration and drift on local adaptation to a heterogeneous environment. J Evol Biol 25: 1351–1363.

Bora, G. C., Gupta, S. N., Tomer, Y. S., and Singh, S. (1998). Genetic variability, correlation and path analysis in faba bean (Vicia faba). Ind. J. Agric. Sci. 68, 212–214.

Borsa, P., Blanquer, A. & Berrebi, P. Genetic structure of the flounders Platichthys flesus and P. stellatus at different geographic scales. Marine Biology 129, 233–246 (1997).

Bradburd, Gideon & Coop, Graham & Ralph, Peter. (2018). Inferring Continuous and Discrete Population Genetic Structure Across Space. Genetics. 210. genetics.301333.2018. 10.1534/genetics.118.301333.

Brommer, J. E. (2011). Whither Pst? The approximation of Qst by Pst in evolutionary and conservation biology. Journal of Evolutionary Biology, 24, 1160–1168. https://doi.org/10.1111/j.1420-9101.2011.02268.x

Byers, Kelsey & Bradshaw, H.. (2021). Rational Design of a Novel Hawkmoth Pollinator Interaction in Mimulus Section Erythranthe. Frontiers in Ecology and Evolution. 9. 10.3389/fevo.2021.658710.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A. and Cresko, W.A. (2013), Stacks: an analysis tool set for population genomics. Mol Ecol, 22: 3124 3140.

Chesser R. K. (1991). Influence of gene flow and breeding tactics on gene diversity within populations. Genetics, 129(2), 573–583.

Clobert, J., Le Galliard, J. F., Cote, J., Meylan, S., & Massot, M. (2009). Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. Ecology letters, 12(3), 197–209. https://doi.org/10.1111/j.1461-0248.2008.01267.x

Collyer ML, Adams DC (2018). "RRPP: An R package for fitting linear models to high-dimensional data using residual randomization. ."

Collyer ML, Adams DC (2021). "RRPP: Linear Model Evaluation with Randomized Residuals in a Permutation Procedure, R package version 1.1.2." https://cran.r-project.org/package=RRPP.

Cronk, Quentin. (2006). Legume flowers bear fruit. Proceedings of the National Academy of Sciences of the United States of America. 103. 4801-2. 10.1073/pnas.0601298103.

Da Silva, S. B., & Da Silva, A. (2018). Pstat: An R package to assess population differentiation in phenotypic traits. R Journal, 10, 447–454. https://doi.org/10.32614/RJ-2018-010

Dalin, P., Björkman, C. Adult beetle grazing induces willow trichome defence against subsequent larval feeding. Oecologia 134, 112–118 (2003). https://doi.org/10.1007/s00442-002-1093-3

Davey JW, Blaxter ML. RADSeq: next-generation population genetics. Brief Funct Genomics. 2010 Dec;9(5-6):416-23. doi: 10.1093/bfgp/elq031.

Delfini, J., Moda-Cirino, V., Dos Santos Neto, J., Ruas, P. M., Sant'Ana, G. C., Gepts, P., & Gonçalves, L. S. A. (2021). Population structure, genetic diversity and genomic selection signatures among a Brazilian common bean germplasm. Scientific reports, 11(1), 2964. https://doi.org/10.1038/s41598-021-82437-4

Diaz-Martin, Z., Fant, J., Havens, K., Cinea, W., Tucker Lima, J. M., & Griffith, M. P. (2023). Current management practices do not adequately safeguard endangered plant species in conservation collections. Biological Conservation, 280, 109955. https://doi.org/10.1016/j.biocon.2023.109955

Earl, Dent A. and vonHoldt, Bridgett M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources vol. 4 (2) pp. 359-361 doi: 10.1007/s12686-011-9548-7

Ehrlich, Paul R., and Peter H. Raven. "Butterflies and plants: a study in coevolution." Evolution 18.4 (1964): 586-608.

Evanno, G., Regnaut, S. and Goudet, J. (2005), Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology, 14: 2611-2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Fisher, R. A. (1950). "Gene Frequencies in a Cline Determined by Selection and Diffusion". Biometrics. 6 (4): 353–361.

Fürstenberg-Hägg, Joel et al. “Plant defense against insect herbivores.” International journal of molecular sciences vol. 14,5 10242-97. 16 May. 2013, doi:10.3390/ijms140510242

García-Fernández, C., Campa, A., Garzón, A. S., Miklas, P., & Ferreira, J. J. (2021). GWAS of pod morphological and color characters in common bean. BMC plant biology, 21(1), 184. https://doi.org/10.1186/s12870-021-02967-x

Gentry, Alwyn H. "Changes in plant community diversity and floristic composition on environmental and geographical gradients." Annals of the Missouri botanical garden (1988): 1-34.

Gilbert and Levine (2017), Ecological drift and the distribution of species diversity, Proceedings of the Royal Society B: Biological Sciences, 284(1855).

Gómez, G., & Ligarreto, G.A. (2012). Analysis of genetic effects of major genes on yield traits of a pea (Pisum sativum L.) cross between the Santa Isabel x WSU 31 varieties.

Gower, J.C. (1975) Generalized Procrustes Analysis. Psychometrika, 40, 33-50. http://dx.doi.org/10.1007/BF02291478

Green, Thomas W., and Ivan G. Palmbald. "Effects of insect seed predators on Astragalus cibarius and Astragalus utahensis (Leguminosae)." Ecology 56.6 (1975): 1435-1440.

Guillot, G., Leblois, R., Coulon, A., & Frantz, A. C. (2009). Statistical methods in spatial genetics. Molecular Ecology, 18, 4734–4756. https://doi.org/10.1111/j.1365-294X.2009.04410.x

Gunz, P., Mitteroecker, P. (2013). Semilandmarks: a method for quantifying curves and surfaces. Hystrix, the Italian Journal of Mammalogy, 24(1), 103-109. https://doi.org/10.4404/hystrix-24.1-6292

Gunz, P., Ramsier, M., Kuhrig, M., Hublin, J.-J. and Spoor, F. (2012), The mammalian bony labyrinth reconsidered, introducing a comprehensive geometric morphometric approach. Journal of Anatomy, 220: 529-543. https://doi.org/10.1111/j.1469-7580.2012.01493.x

Haldane, J. B. S. (1932). The causes of evolution. Macmillan.

Harrison, Joshua G., et al. "Rarity does not limit genetic variation or preclude subpopulation structure in the geographically restricted desert forb Astragalus lentiginosus var. piscinensis." American journal of botany 106.2 (2019): 260-269.

Houghton, S., Stevens, M. T., and Meyer, S. E.. 2020. Pods as sails but not as boats: dispersal ecology of a habitat-restricted desert milkvetch. American Journal of Botany 107(6): 864– 875.

Hudson, R.R., Slatkin, M., & Maddison, W. P. (1992). Estimation of levels of gene flow from DNA sequence data. Genetics, 132(2), 583–589. https://doi.org/10.1093/genetics/132.2.583

Hutchison, D.W. and Templeton, A.R. (1999), Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. Evolution, 53: 1898-1914.

Isely, D. 1998. Native and naturalized Leguminosae (Fabaceae) of the United States. Provo: M. L. Bean Life Science Museum, Brigham Young University.

J. Catchen, A. Amores, P. Hohenlohe, W. Cresko, and J. Postlethwait. Stacks: building and genotyping loci de novo from short-read sequences. G3: Genes, Genomes, Genetics, 1:171-182, 2011.

Janzen, Daniel H. "Seed predation by animals." Annual review of ecology and systematics 2.1 (1971): 465-492.

Jombart T (2008). “adegenet: a R package for the multivariate analysis of genetic markers.” Bioinformatics, 24, 1403-1405. doi:10.1093/bioinformatics/btn129.

Jombart T, Ahmed I (2011). “adegenet 1.3-1: new tools for the analysis of genome-wide SNP data.” Bioinformatics. doi:10.1093/bioinformatics/btr521.

Jordon-Thaden, I. E., Beck, J. B., Rushworth, C. A., Windham, M. D., Diaz, N., Cantley, J. T., Martine, C. T., & Rothfels, C. J. (2020). A basic ddRADseq two-enzyme protocol performs well with herbarium and silica-dried tissues across four genera. Applications in plant sciences, 8(4), e11344. https://doi.org/10.1002/aps3.11344

Kaeuffer, R., Peichel, C. L., Bolnick, D. I., & Hendry, A. P. (2012). Parallel and nonparallel aspects of ecological, phenotypic, and genetic divergence across replicate population pairs of lake and stream stickleback. Evolution; international journal of organic evolution, 66(2), 402–418. https://doi.org/10.1111/j.1558-5646.2011.01440.x

Karami, E. (2011). Genetic analysis of some agronomic characters in chickpea (Cicer aretinum L.).

Karron, J.D. (1989), Breeding Systems And Levels Of Inbreeding Depression In Geographically Restricted And Widespread Species Of Astragalus (Fabaceae). American Journal of Botany, 76: 331-340. https://doi.org/10.1002/j.1537-2197.1989.tb11319.x

Kessler, Michael. "Patterns of diversity and range size of selected plant groups along an elevational transect in the Bolivian Andes." Biodiversity & Conservation 10.11 (2001): 1897-1921.

Klingenberg CP, Duttke S, Whelan S, Kim M (2012) Developmental plasticity, morphological variation and evolvability: a multilevel analysis of morphometric integration in the shape of compound leaves. J Evol Biol 25:115–129

Klingenberg, C.P. Size, shape, and form: concepts of allometry in geometric morphometrics. Dev Genes Evol 226, 113–137 (2016). https://doi.org/10.1007/s00427-016-0539-2

Knaus, Brian J. "Morphometric architecture of the most taxon-rich species in the US flora: Astragalus lentiginosus (Fabaceae)." American journal of botany 97.11 (2010): 1816-1826.

Knaus, Brian J., Rich C. Cronn, and Aaron Liston. "Genetic characterization of three varieties of Astragalus lentiginosus (Fabaceae)." Brittonia 57.4 (2005): 334-344.

Kusmiyati, Florentina & Anwar, Syaiful & Jamiati, Urip & Herwibawa, Bagus. (2021). Segregation ratio and heritability of agronomic character in yardlong bean F2 population. IOP Conference Series: Earth and Environmental Science. 803. 012027. 10.1088/1755-1315/803/1/012027.

Kuzbakova M, Khassanova G, Oshergina I, Ten E, Jatayev S, Yerzhebayeva R, Bulatova K, Khalbayeva S, Schramm C, Anderson P, Sweetman C, Jenkins CLD, Soole KL, Shavrukov Y. Height to first pod: A review of genetic and breeding approaches to improve combine harvesting in legume crops. Front Plant Sci. 2022 Sep 16;13:948099. doi: 10.3389/fpls.2022.948099. PMID: 36186054; PMCID: PMC9523450.

Leinonen, T., Cano, J. M., Makinen, H., & Merilä, J. (2006). Contrasting patterns of body shape and neutral genetic divergence in marine and lake populations of threespine stickle-backs. Journal of Evolutionary Biology, 19, 1803–1812. https://doi. org/10.1111/j.1420-9101.2006.01182.x

Leinonen, T., McCairns, R. J. S., O'Hara, R. B., & Merilä, J. (2013). QST– FST comparisons: Evolutionary and ecological insights from genomic heterogeneity. Nature Reviews Genetics, 14, 179. https://doi. org/10.1038/nrg3395

Lepais O, Weir JT. SimRAD: an R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. Mol Ecol Resour. 2014 Nov;14(6):1314-21. doi: 10.1111/1755-0998.12273. Epub 2014 May 26. PMID: 24806844.

Lepais O, Weir JT. SimRAD: an R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. Mol Ecol Resour. 2014 Nov;14(6):1314-21. doi: 10.1111/1755-0998.12273. Epub 2014 May 26. PMID: 24806844.

Lopez, Javier & Scheepens, J.F. & Lara-Romero, Carlos & Ruiz-Checa, Raquel & Tabarés, Pablo & Iriondo, Jose. (2019). Past selection shaped phenological differentiation among populations at contrasting elevations in a Mediterranean alpine plant. Environmental and Experimental Botany. 170. 103894. 10.1016/j.envexpbot.2019.103894.

MacLeod, Norman. "PalaeoMath: Part 27 - Going Round the Bend II: Extended Eigenshape Analysis." PalaeoMath: Part 27 - Going Round the Bend II: Extended Eigenshape Analysis | The Palaeontological Association, Palaeontology Newsletter, Dec. 2012, www.palass.org/publications/newsletter/palaeomath-101/palaeomath-part-27-going-round-bend-ii-extended-eigenshape-analysis.

Márquez F, Lozada M, Idaszkin YL, González-José R, Bigatti G. Cannabis Varieties Can Be Distinguished by Achene Shape Using Geometric Morphometrics. Cannabis Cannabinoid Res. 2022 Aug;7(4):409-414. doi: 10.1089/can.2020.0172. Epub 2021 May 31. PMID: 34061660; PMCID: PMC9418367.

Martin, E.F., Meinke, R.J. Variation in the demographics of a rare central Oregon endemic, Astragalus peckii Piper (Fabaceae), with fluctuating levels of herbivory. Popul Ecol 54, 381–390 (2012).

Mazer, S.J., & Wheelwright, N.T. (1993). Fruit size and shape: Allometry at different taxonomic levels in bird-dispersed plants. Evolutionary Ecology, 7, 556-575.

Mitteroecker P, Gunz P, Windhager S, Schaefer K (2013) A brief review of shape, form, and allometry in geometric morphometrics, with applications to human facial morphology. Hystrix 24:59–66

Morente Lopez, Javier & Scheepens, J.F. & Lara-Romero, Carlos & Ruiz-Checa, Raquel & Tabarés, Pablo & Iriondo, Jose. (2019). Past selection shaped phenological differentiation among populations at contrasting elevations in a Mediterranean alpine plant. Environmental and Experimental Botany. 170. 103894. 10.1016/j.envexpbot.2019.103894.

Mota, Elias & Novaes, Carolina & Silva, Luciana & Chaves, Lázaro. (2020). Structure of the phenotypic variability of fruit and seeds of Dipteryx alata vogel (Fabaceae). Revista Brasileira de Fruticultura. 42. 10.1590/0100-29452020003.

Nei M. (1973). Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences of the United States of America, 70(12), 3321–3323. https://doi.org/10.1073/pnas.70.12.3321

Nery, Eduardo Koerich & Fiaschi, Pedro. (2019). Geometric Morphometrics Dismiss the Polymorphic Hydrocotyle quinqueloba (Araliaceae) from the Neotropics. Systematic Botany. 44. 451-469. 10.1600/036364419X15561132273558.

Paris, J.R., Stevens, J.R. and Catchen, J.M. (2017), Lost in parameter space: a road map for stacks. Methods Ecol Evol, 8: 1360-1373. https://doi.org/10.1111/2041-210X.12775

Peter CI, Johnson SD. A pollinator shift explains floral divergence in an orchid species complex in South Africa. Ann Bot. 2014 Jan;113(2):277-88. doi: 10.1093/aob/mct216. Epub 2013 Oct 9. PMID: 24107684; PMCID: PMC3890387.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (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. https://doi.org/10.1371/journal.pone.0037135

Petrova, S. (2021). Chickpea plant model for the climatic conditions of the Sadovo region according to the yield components. Bulg. J. Agric. Sci., 27 (3), 531–535

Pijl, L. van der. (1972). Principles of dispersal in higher plants. Berlin ; New York : Springer-Verlag

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000 Jun;155(2):945-59. doi: 10.1093/genetics/155.2.945. PMID: 10835412; PMCID: PMC1461096.

Pujol, B., Wilson, A. J., Ross, R. I., & Pannell, J. R. (2008). Are Q(ST)-F(ST) comparisons for natural populations meaningful? Molecular Ecology, 17(22), 4782–4785. https://doi.org/10.1111/j.1365-294X.2008.03958.x

Raeymaekers, J. A. M., Van Houdt, J. K. J., Larmuseau, M. H. D., Geldof, S., & Volckaert, F. A. M. (2007). Divergent selection as revealed by PST and QTL-based FST in three-spined stickleback (Gasterosteus aculea- tus) populations along a coastal-inland gradient. Molecular Ecology, 16, 891–905. https://doi.org/10.1111/j.1365-294X.2006.03190.x

Rohlf, F. & Slice, Dennis. (1990). Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks. Systematic Zoology. 39. 10.2307/2992207.

Rohlf, F.J. (1999) Shape Statistics: Procrustes Superimpositions and Tangent Spaces. Journal of Classification, 16, 197-223. https://doi.org/10.1007/s003579900054

Sæther, S. A., Fiske, P., Kålås, J. A., Kuresoo, A., Luigujõe, L., Piertney, S. B., ... Höglund, J. (2007). Inferring local adaptation from QST–FST comparisons: Neutral genetic and quantitative trait variation in European populations of great snipe. Journal of Evolutionary Biology, 20, 1563–1576.

Schemske, Douglas & Bradshaw, Harvey. (1999). Schemske DW, Bradshaw Jr HD. Pollinator preference and the evolution of floral traits in monkeyflowers (Mimulus). Proc Natl Acad Sci USA 96: 11910-11915. Proceedings of the National Academy of Sciences of the United States of America. 96. 11910-5. 10.1073/pnas.96.21.11910.

Seymour, M, Räsänen, K, Kristjánsson, BK. Drift versus selection as drivers of phenotypic divergence at small spatial scales: The case of Belgjarskógur threespine stickleback. Ecol Evol. 2019; 9: 8133– 8145. https://doi.org/10.1002/ece3.5381

Skinner, M.M. and Gunz, P. (2010), The presence of accessory cusps in chimpanzee lower molars is consistent with a patterning cascade model of development. Journal of Anatomy, 217: 245-253. https://doi.org/10.1111/j.1469-7580.2010.01265.x

Sun, Y, Bossdorf, O, Grados, RD, Liao, Z, Müller-Schärer, H. Rapid genomic and phenotypic change in response to climate warming in a widespread plant invader. Glob Change Biol. 2020; 26: 6511– 6522. https://doi.org/10.1111/gcb.15291

Thompson, John N. The coevolutionary process. University of Chicago Press, 1994.

Thompson, John N. The geographic mosaic of coevolution. University of Chicago Press, 2005.

Twyford, A. D., Wong, E. L. Y., & Friedman, J. (2020). Multi-level patterns of genetic structure and isolation by distance in the widespread plant Mimulus guttatus. Heredity, 125(4), 227–239. https://doi.org/10.1038/s41437-020-0335-7

van Heerwaarden, B. and Sgrò, C.M. (2017), The quantitative genetic basis of clinal divergence in phenotypic plasticity. Evolution, 71: 2618-2633.

Vicente M.J., Segura F., Aguado M., Migliaro D., Franco J.A., Martínez-Sánchez J.J., (2011) Genetic diversity of Astragalus nitidiflorus, a critically endangered endemic of SE Spain, and implications for its conservation. Biochemical Systematics and Ecology, Volume 39(3), , Pages 175-182, ISSN 0305-1978, https://doi.org/10.1016/j.bse.2011.03.005.

Vieira, Márcia & Mayo, Simon & Andrade, Ivanilza. (2014). Geometric morphometrics of leaves of Anacardium microcarpum Ducke and A. occidentale L. (Anacardiaceae) from the coastal region of Piauí, Brazil. Brazilian Journal of Botany. 37. 315-327. 10.1007/s40415-014-0072-3.

War, Abdul R & Taggar, Gaurav & Hussain, Barkat & Taggar, Monica Sachdeva & Nair, Ramakrishnan & Sharma, H.C.. (2018). Plant defence against herbivory and insect adaptations. AoB PLANTS. 10. 10.1093/aobpla/ply037.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. Evolution, 38(6), 1358–1370. https://doi.org/10.2307/2408641

Welsh, S. 2007. North American species of Astragalus Linnaeus (Leguminosae): a taxonomic revision. Brigham Young University Press, Provo, UT, USA.

Whitney, K., & Stanton, M. (2004). Insect Seed Predators as Novel Agents of Selection on Fruit Color. Ecology, 85(8), 2153-2160.

Willing, E. M., Dreyer, C., & van Oosterhout, C. (2012). Estimates of genetic differentiation measured by F(ST) do not necessarily require large sample sizes when using many SNP markers. PloS one, 7(8), e42649. https://doi.org/10.1371/journal.pone.0042649

Wojciechowski, Martin F. "Astragalus (Fabaceae): a molecular phylogenetic perspective." Brittonia 57.4 (2005): 382-396.

Wright, Sewall (1984). Evolution and the Genetics of Populations: Genetics and Biometric Foundations New Edition. University of Chicago Press.

Wrozyna C, Neubauer TA, Meyer J, Piller WE (2016) Shape Variation in Neotropical Cytheridella (Ostracoda) Using Semilandmarks-Based Geometric

Morphometrics: A Methodological Approach and Possible Biogeographical Implications. PLOS ONE 11(12): e0168438. https://doi.org/10.1371/journal.pone.0168438

VITA

Quinn Thomas was born and raised in Minneapolis, Minnesota. Before attending Loyola University Chicago, she attended Drake University, where she earned a Bachelor of Science in Biology, in 2019. Following graduation, she spent a year volunteering in Malawi working to support vulnerable persons reach economic freedom.

While at Loyola, Thomas received the Research Assistant Fellowship in her second year of schooling. She went on to work part-time in the Lipshutz Lab as a Lab Technician while finishing her master's degree in her third year.

Currently, Thomas continues to work for the Lipshutz Lab while looking to begin her career in biotech.