Population genetics of the widespread *Boechera laevigata* (Brassicaceae) and comparisons with its rare congener *B. perstellata*

By

Annie Lindsey

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science in Biology

Austin Peay State University

August 2021

Thesis Committee:

Dr. Carol Baskauf, Committee Chair

Dr. Mollie Cashner

Dr. Dwayne Estes

Population genetics of the widespread *Boechera laevigata* (Brassicaceae) and comparisons with its rare congener *B. perstellata*

Annie Lindsey

August 2021

Approved:

Carol J. Baskauf

Dr. Carol Baskauf, Committee Chair

Dr. Mollie Cashner, Committee Member

L. Dwayne Estes

Dr. Dwayne Estes, Committee Member

Chad Brooks

Dr. Chad Brooks, Associate Provost and Dean, College of Graduate Studies

Statement of Permission to Use

In presenting this thesis in partial fulfillment of the requirements for the Master of Science in Biology at Austin Peay State University, I agree that the library shall make it available to borrowers under the rules of the library. Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgement of the source is made. Permissions for extensive quotation or reproduction of the thesis may be granted by my major professor, or in his/her absence, by the Head of the Interlibrary Services when, in the opinion of either, the proposed use of the material is for scholarly purposes. Any copying or use of the material in this thesis for financial gain shall not be allowed without my written permission.

Annie Lindsey

09/14/2021
Date

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1655521. Additional funding for this research was provided by the APSU Graduate Student Research Support Grant and the APSU Department of Biology.

There are a number of people I would like to thank for their help in the completion of this thesis. First, I would like to thank my advisor, Dr. Carol Baskauf, for her unwavering support and guidance through this endeavor. Without her advice, instruction, and boundless dedication, I would not have been able to complete this project. I would also like to thank my committee members, Dr. Mollie Cashner and Dr. Dwayne Estes, for their insights and assistance in completing this thesis.

Without the many volunteers who aided in sample collections, this research would not have been possible. I would like to extend my gratitude to Bonnie and Joe Isaac who were kind enough to collect samples from Pennsylvania, Zach Irick who graciously collected two Tennessee populations for this research, Steve Baskauf, Lydia Deason and Dylan Fitch who accompanied me in sample collections, Evan Raskin who also collected two populations from North Carolina, Kevin England who collected from the Alabama populations, Jason Whittle who collected the Ohio populations, Elizabeth Bach and Jenn Simmons who collected the IL-2 population, Erin Faulkner who collected samples from the IL-1 population, Theo Witsell and Molly Robinson who collected samples from the Arkansas population, and Anna and Lark Malcom who collected from the Maryland population. Furthermore, I would like to thank Lydia Deason, Dylan Fitch, Cameron Kelly, and Stephen Thrasher for their assistance in the lab. With their help, I was able to quickly process the more than 300 samples that were collected for this research.

Lastly, I would like to express my appreciation to Dr. Michael Windham and Patrick Alexander for their expert insights into the *Boechera* genus. Without such insights, my analyses may not have been reflective of the actual biology of *Boechera laevigata*. The information they so graciously contributed, especially Dr. Michael Windham who provided me with unpublished findings, helped to guide the direction of my analyses and give me a stronger foundation upon which to support my findings. Words cannot fully express my gratitude to all the individuals who assisted me through the course of this research.

ABSTRACT

A population genetics study of *Boechera laevigata*, the most widespread *Boechera* species in the eastern United States, was conducted using 15 populations from 11 states throughout its range. Sixteen polymorphic microsatellite loci were resolved for the species. Three populations exhibited nearly fixed heterozygosity at multiple loci, which was attributed to apomictic reproduction based on this study's data and other *Boechera* research. Because of this likelihood of clonal reproduction, only unique multilocus genotypes (MLG) were used for most analyses of the apomictic populations. From 300 total samples, 153 unique MLGs were identified, with no MLGs shared across populations. Percent polymorphic loci at the population level ranged from 6.3-93.8%, with sexual populations averaging 40% and apomictic populations averaging 71%. Observed heterozygosity was lower than expected heterozygosity for sexual populations (H_o = 0.052, He = 0.169), while the opposite was true for apomictic populations (H_o = 0.679, H_e = 0.441), resulting in a high F_{IS} of 0.709 for sexual populations (probably due to inbreeding) and an extremely negative F_{IS} of -0.927 for apomictic populations, as is commonly found with clonal reproduction. While 69% of the genetic variability in the sexual populations was due to differences among the populations, the majority of genetic variation in the apomictic populations was found within individuals. Widespread species often harbor greater genetic diversity than rare species, so this trend was tested by comparing *B. laevigata* with the federally endangered *B. perstellata*, which was previously found to have extremely low levels of genetic diversity. Such congeneric comparisons of genetic diversity provide phylogenetic context and a more nuanced understanding of evolutionary forces that have shaped a species. Considering the 11 loci assayed in both species, sexual *B. laevigata* populations averaged higher allelic richness, observed heterozygosity, and expected heterozygosity than its rare congener, although

differences did not reach statistical significance. The rare species had significantly greater population differentiation than *B. laevigata*, possibly reflective of the very disjunct distribution of *B. perstellata*. This study provides context to the low diversity reported for the endangered *B. perstellata* as well as adding to the growing literature on the model genus *Boechera*.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

Introduction

Understanding how genetic diversity is distributed in a natural population can provide considerable insight to evolutionary processes and how these have impacted the genetic diversity of a species. Genetic diversity is essential for allowing adaptation in a changing environment, and a lack of genetic variation can severely limit a species' evolutionary potential (Frankel 1970, Lande 1988, 1993, Reed and Frankham 2003, Frankham 2005, Hughes et al. 2008). The absence of genetic diversity can even be detrimental on a shorter time scale as a lack of genetic variation may mean that the homogeneous population has no natural defenses against certain pathogens, potentially resulting in significant rates of mortality in a population within a single generation, a problem which can occur when monocultures of crops are grown (Huenneke 1991). Larger populations are able to maintain a greater number of alleles in their gene pool than smaller populations, which may be able to harbor only a fraction of the same alleles. This reduced genetic diversity can lead to inbreeding, reduced reproductive fitness, and by extension, an increased risk of extinction (Charlesworth and Charlesworth 1987, Grueber et al. 2008). By studying population genetic principles, one can make predictions of how a species may change over time, given different evolutionary pressures. Additionally, studying and comparing the population genetics of congeneric species allows the shared evolutionary histories of related species to be accounted for and can provide a more accurate view of the evolutionary pressures affecting a species. In particular, the comparison of rare and widespread congeners can provide context to the frequently reported low genetic diversity levels observed in rare species (Hamrick and Godt 1989, Karron 1991, 1997, Gitzendanner and Soltis 2000, Cole 2003). These

comparisons can provide a greater understanding of the factors that contribute to rarity and the factors that aid in a widespread species' ability to inhabit an extensive geographic range.

Boechera perstellata is a federally endangered species with an extremely narrow and disjunct distribution in the southeastern United States (USFWS 1997, USFWS 2018). A recent population genetics study of this species found it to have extremely low genetic variability at the population level (Baskauf et al. 2014). While such results may not be surprising for a rare species, the question remains if such a result is the consequence of rarity or if there is phylogenetic context that can elucidate these findings. Studies of two western *Boechera* congeners found that the rare species did generally have lower levels of genetic diversity as compared to the widespread species. However, the rare western species had significantly higher levels of observed heterozygosity compared to the widespread species (Song et al. 2006, Song and Mitchell-Olds 2007), thus results were not entirely consistent with the typical expectation of reduced genetic variation in rare species. Overall, though, genetic diversity appears to be low for both the western rare and widespread *Boechera* species, and even more so for the rare eastern *B. perstellata*, potentially suggesting that the genus itself has limited genetic variability.

Evidence indicates that the eastern and western *Boechera* species form separate lineages (Kiefer et al. 2009, Alexander et al. 2013), and while much research has been done on western *Boechera* species, little has been done for the eastern species. Thus, to better understand and contextualize the low levels of genetic diversity found in *B. perstellata,* a population genetics study of *Boechera laevigata*, the most widespread *Boechera* in the eastern United States (Weakely 2015) was conducted. The large geographic range of *B. laevigata* starkly contrasts with that of its rare congener *B. perstellata.* The goal of this study was to investigate the population genetics of *B. laevigata* and compare its genetic diversity levels and population

genetic structure to that of the rare *B. perstellata*. It was expected that *B. laevigata* would exhibit greater levels of genetic diversity compared to that of *B. perstellata*, as is typically the case for rare and widespread species pairs. A lesser degree of population differentiation was also expected for the widespread species due to the larger connectivity of *B. laevigata* populations compared to the disjunct nature of *B. perstellata* populations. Such a study will provide phylogenetic context to the reduced levels of genetic diversity seen in *B. perstellata.* In addition, this study will contribute to the growing information available about the *Boechera* genus, which is becoming a "model system" for non-agricultural plants due to its large geographic range in undisturbed natural habitats, relatively simple genome, ability to easily produce hybrids, and occurrence of diploid apomicts (Schranz et al. 2005, Rushworth et al. 2011).

Boechera laevigata

Description

Boechera laevigata (Muhl. Ex Willd.) Al-Shehbaz (Brassicaceae), commonly known as smooth rockcress, is a monocarpic, facultative biennial herb that inhabits patchily distributed rock outcrops in deciduous forests of the eastern United States and Canada (Al-Shehbaz 1988). Plants occur as immature vegetative rosettes before developing a single, racemose inflorescence during their mature life stage (Figure 1). Cauline leaves are glabrous and oblanceolate with a distinct auriculate clasping base with margins presenting in multiple ways, ranging from slightly dentate to ruffled (Figure 2) (Al-Shehbaz 1988). Flowers are tetramerous, small, and white (Figure 3), and ovaries develop into siliques upon maturity (Figure 4). Figure 5 details the distribution of *B. laevigata* in the eastern United States.

Figure 1. Mature *Boechera laevigata* individual emerging from a rock outcrop (photo courtesy of Kevin England).

Figure 2. Cauline leaves of *Boechera laevigata* (photo courtesy of Kevin England).

Figure 3. Racemose inflorescence of *Boechera laevigata* (photo courtesy of Kevin England).

Figure 4. Immature silique of *Boechera laevigata* (photo courtesy of Erin Faulkner).

Habitat

Boechera laevigata thrives on calcareous rock outcrops within deciduous forests, although it can be found growing in grounded substrate (Al-Shehbaz 1988). *Boechera laevigata*'s presence in stable hardwood forests is fairly uncommon among facultative biennial weeds such as *Daucus carota* and *Dipsacus sylvestris* (Baskin and Baskin 1979), which tend to occupy primarily open, disturbed habitats (Bloom et al. 1990). The patchy nature of the rock outcrops *B. laevigata* occupy leads to discontinuous populations which can range in size and density. Nevertheless, *B. laevigata* is described as common throughout much of its range, although it is considered rare in some states at the edge of its range (Kartesz 2015).

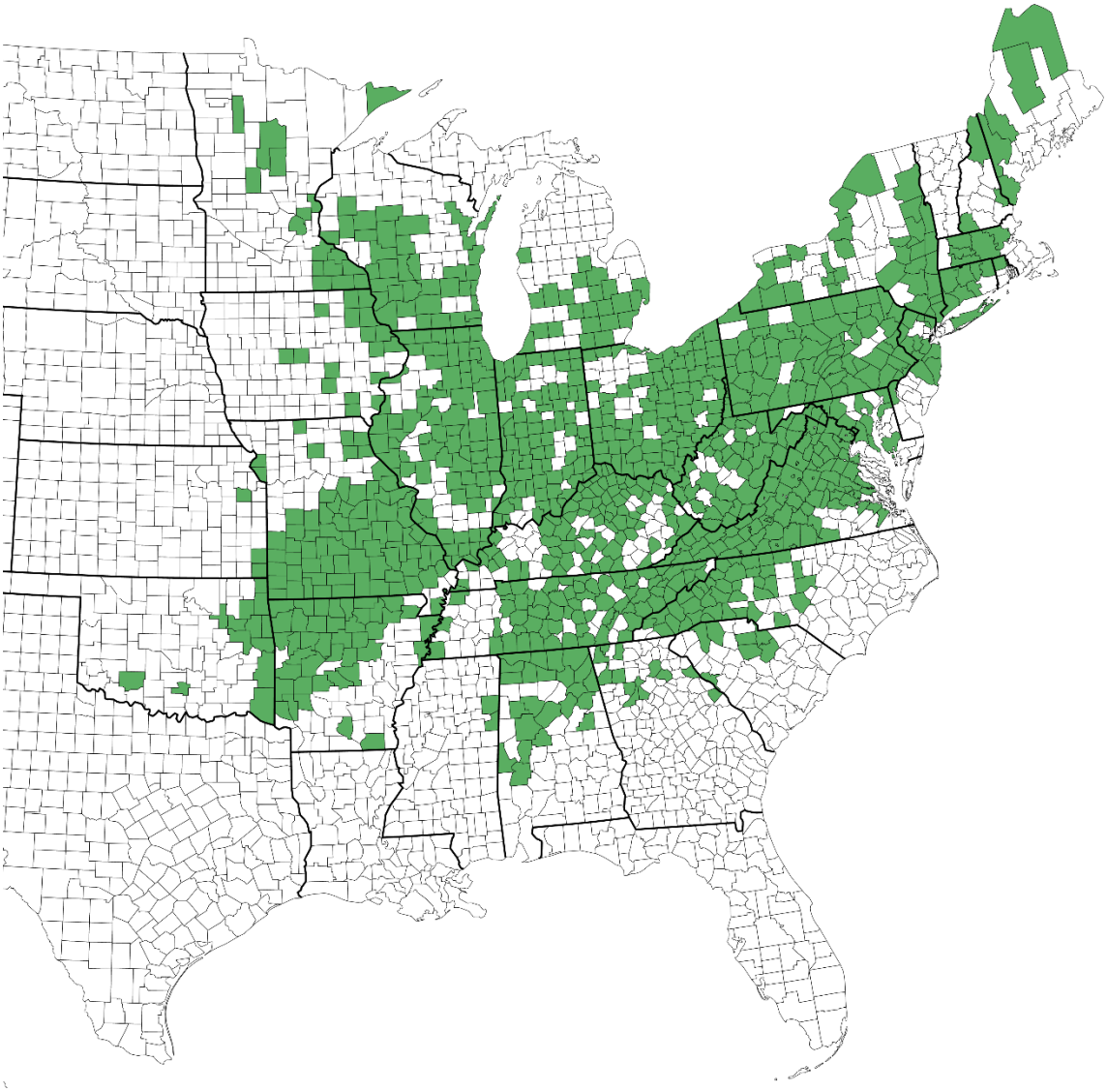


Figure 5. Distribution of *Boechera laevigata* in the eastern United States. Shaded counties indicate documented occurrences of *B. laevigata* as reported by BONAP (Kartesz 2015). The map was created using mapchart.net.

Life History

Boechera laevigata individuals begin as vegetative rosettes and must undergo vernalization before bolting can occur the following spring (Bloom et al. 2002a). Production of reproductive structures is largely size dependent, which can lead to rosettes persisting for multiple years before reaching sexual maturity (Bloom et al. 2002a, 2003a). This strategy decreases mortality risk and increases fitness as *B. laevigata* occupies stressful habitats that are drought prone and nutrient poor (Bloom et al. 2003b); thus, postponing sexual maturity increases the likelihood of successful reproductive opportunities. After bolting, individuals produce flowers from March to June, depending on the geographic area (Bloom et al. 2002a). Mature seeds can develop as early as June, and following seed formation, the parent plant dies. (Bloom et al. 1990).

Seed dispersal begins in July and can continue for up to 19 months (Bloom et al. 1990). Seeds disperse over short distances and tend to stay close to the location of the parent plant (Bloom et al. 2002b). This short-range dispersal has been hypothesized to be an adaptive strategy that allows offspring to inhabit the same suitable environments as their parents, given the irregular nature of the habitats in which the species thrives (Bloom et al. 2002b). Interestingly, Bloom et al. (2002b) noted that seeds tend to disperse south of the parent plant, which may be caused by northerly winds during the winter. Once dispersed, seeds remain dormant until they undergo cold stratification during the winter before germination can occur the following spring (Bloom et al. 1990). Precipitation, leaf litter, and cold temperatures are factors that can severely affect seed germination and juvenile establishment (Bloom et al. 2001). Seeds that do not germinate are added to a persistent seedbank. The presence of such a seedbank is hypothesized to be an adaptation for maintaining a stable population of individuals in the event that stochastic

occurrences, such as a drought, prevent an entire generation from reproducing (Bloom et al. 1990). Furthermore, this strategy allows the establishment of a new generation without the need for immigrants (Bloom et al. 1990).

Breeding Ecology

Extensive work regarding the breeding ecology of *B. laevigata* is limited; however, it is known that *B. laevigata* reproduces sexually via both outcrossing and selfing (Bloom 1988). Additionally, *B. laevigata* has been reported to reproduce asexually via facultative gametophytic apomixis, specifically Hieracium-type apospory (Carman et al. 2019), particularly in populations north of the North American glacial boundary (M. Windham, pers. comm.). Sexual and asexual reproduction have been extensively documented in the genus *Boechera*, which is notorious for its prevalence of diploid and polyploid apomicts that are often the result of hybridization events (Dobeš et al. 2005, Schranz et al. 2005, Beck et al. 2012, Alexander et al. 2015, Windham et al. 2015).

Boechera perstellata

Boechera perstellata is a federally endangered species that inhabits parts of the Bluegrass region in north-central Kentucky and the Central Basin region in north-central Tennessee (Figure 6). *Boechera perstellata* tends to occur on limestone rock outcrops in mesophytic to sub-xeric forests, where it is patchily distributed in small groups or as lone individuals (USFWS 2004, 2018). Currently, there are 47 extant populations in Kentucky and Tennessee, with most populations occurring in Kentucky (USFWS 2018). Extensive research regarding the life history and breeding ecology of *B. perstellata* is limited; however, it is assumed to be pollinated by small flies and bees (USFWS 2004). *Boechera perstellata* has a G2 status and is threatened by habitat loss and competition with native and invasive species. The combination of its very small

population sizes, specialized habitat, and weak competitive ability put it at risk of extinction (USFWS 2004, 2018).

Figure 6. *Boechera perstellata* distribution. Shaded counties indicate documented occurrences of *B. perstellata* as reported in the 5-Year Review by the USFWS (2018). Red dots indicate populations that were sampled in the Baskauf et al. (2014) population genetics study of *B. perstellata*.

Taxonomical Challenges

Boechera is a new world genus consisting of more than 80 primarily inbreeding sexual diploid species and hundreds of hybrids (Li et al. 2017). The western United States harbors most of *Boechera'*s diversity where extensive evidence of hybrids has been documented (Koch et al. 2003, Dobeš et al. 2005, 2006, Schranz et al. 2005, Beck et al. 2012, Alexander et al. 2015, Windham et al. 2015, Schilling et al. 2018). The western and eastern species of *Boechera* have been proposed to belong to distinct clades, with the eastern species being more closely related to the East Asian genus *Borodinia* (Alexander et al. 2013). These findings prompted some taxonomists to support the movement of the eastern *Boechera* species to the genus *Borodinia*.

However, the eastern species *Boechera laevigata* has been documented to hybridize with *B. stricta* in the west (M. Windham, pers. comm.). This inter-clade hybridization in addition to the presence of eastern allotetraploid hybrids has been cited as a reason to preserve the genus *Boechera* for the eastern North American species (Carman et al. 2019). The capability to hybridize and, moreover, for the hybrids to backcross with both parental species demonstrates a complex hybrid network among the North American *Boechera* that more parsimoniously explains the relationship among these species.

Eastern *Boechera* species

There are seven *Boechera* species found in the eastern United States. In addition to *B. laevigata* and *B. perstellata*, *B. missouriensis, B. serotina, B. dentata, B. canadensis,* and *B. burkii* also inhabit the eastern United States. *Boechera burkii*, which is found in five states throughout the Appalachian Mountains, was previously considered a variety of *B. laevigata* (*Arabis laevigata* var. *burkii*); however, it is morphologically distinct from *B. laevigata* as it has linear, non-auriculate, entire leaves compared to the lanceolate, auriculate, and dentate leaves

characteristic of *B. laevigata* (Windham and Al-Shehbaz 2007). Furthermore, *B. serotina* was previously considered a synonym of *Arabis laevigata* var. *burkii* but is distinct in not only its habitat, which is restricted to shale barrens in Virginia and West Virginia, but also morphology, phenology, and geographic range compared to that of *B. burkii* (Windham and Al-Shehbaz 2007). *Boechera missouriensis*, which was also previously treated as a variety of *B. laevigata* (Al-Shehbaz and Windham 2010), is the sister taxon to *B. laevigata*, with both species being sister taxa to *B. serotina* (Kiefer et al. 2009, Alexander et al. 2013). *Boechera perstellata* is the sister taxon to *B. dentata* (Kiefer et al. 2009). Although *B. laevigata* and *B. perstellata* are not sister taxa, they still have a close relationship that can help contextualize the low genetic diversity levels found in *B. perstellata*.

Apomixis in the *Boechera* genus

Asexual reproduction guarantees that successful parental genes are passed to the next generation without interference from allelic recombination, ensuring that offspring are genetically identical to their parents. This comes with the cost of reduced adaptive potential in changing environments. However, asexual reproduction also comes with the advantages of assured reproduction which can be an effective mechanism to facilitate geographic range expansion (Van Dijk 2007).

Apomixis is a form of asexual reproduction unique to angiosperms in which seeds develop without fertilization (Hojsgaard and Hörandl 2019). Apomixis may occur during the sporophytic or gametophytic stage of a plant's life cycle. The former, referred to as sporophytic apomixis or adventitious embryony, involves embryogenesis directly from the maternal sporophyte's ovular tissue, completely bypassing the gametophytic stage (Nogler 1984, Beck et al. 2012). Gametophytic apomixis involves embryo development from an unreduced female

gametophyte and is further divided into aposporous or diplosporous apomixis based on which cell gives rise to the embryo sac (Nogler 1984, Savidan 2001). Aposporous apomicts develop embryo sacs from somatic cells in the ovule while diplosporous apomicts produce embryo sacs from sporocytes that produce unreduced megaspores through a single division rather than meiosis (Nogler 1984, Savidan 2001, Beck et al 2012). Gametophytic apomixis is the more well-documented type of apomixis between the two and eludes typical sexual pathways through three mechanisms. First, a functional female gametophyte is produced without meiosis—a process known as apomeiosis. This is followed by parthenogenesis, or the development of an unfertilized egg cell into an embryo, and concludes with the formation of an endosperm either autonomously or through fertilization with a sperm cell (pseudogamy) (Koltunow and Grossniklaus 2003, Hörandl and Hojsgaard 2012, Hojsgaard and Hörandl 2019).

Gametophytic apomixis is particularly common among polyploid plants which can arise through hybridization between species (allopolyploidy) or through division errors resulting in extra sets of chromosomes (autopolyploidy) (Windham et al. 2015). The co-occurrence of apomixis and polyploidy has often confounded the relationship between the two and made it difficult for researchers to identify what components contribute to the activation of apomixis. As a result, *Boechera* which has a small genome and is the only genus known to have diploid apomicts, has become a model genus for researching apomixis (Rushworth et al. 2011, Brukhin et al. 2019). Apomixis in *Boechera* is facultative and typically diplosporous and pseudogamous (Dobeš et al. 2005, 2006); however, apospory and diplospory have been found to occur in various *Boechera* species, even occurring simultaneously within a single hybrid (Carman et al. 2019). Although there is a close association between hybridization and apomixis among *Boechera* species, the two are not exclusively associated, insinuating that hybridization is not the

only factor that induces apomixis (Lovell et al. 2013, Mau et al. 2015). The relationship between hybridization, polyploidy, and apomixis are still being heavily researched as the reproductive and taxonomical complexities in the North American *Boechera* complex are investigated.

Factors affecting genetic variation

Genetic variation within a species is imperative for allowing adaptation in a dynamic environment (Frankel 1970, Beardmore 1983). Genetically depauperate species often face increased risks of extinction due to the limited variation upon which natural selection can act to improve fitness (Lande 1988, Reed et al. 2003, Frankham 2005). Various evolutionary forces such as mutations, gene flow, genetic drift, natural selection, and nonrandom mating have a large influence on a species' genetic diversity levels (Hartl and Clark 1989).

Mutations are the original source of genetic variation and produce new alleles which may yield beneficial adaptations. The benefits of such mutations are subject to selective pressures which differentially favor certain phenotypes over others. Thus, individuals harboring favored phenotypes have a greater probability of reproducing and passing their genes to their offspring. However, mutations are rarely beneficial and are more often neutral or detrimental, depending on the location of the mutation. The constraints on mutations subsequently limit the extent of evolution in a population or species (Hartl and Clark 1989).

Random genetic drift is an evolutionary force that has a strong impact on small populations due to their reduced gene pool and small effective population size which may result in reduced levels of genetic variation (Barrett and Kohn 1991, Karron 1991). Small populations are especially susceptible to stochastic events such as genetic bottlenecks (an extreme form of genetic drift), and this can further lead to changes in allele frequencies or even allele fixation (Whitlock 2000, England et al. 2003, Bouzat 2010). The effects of bottlenecks can be countered

by gene flow, which is the exchange of alleles among populations. Gene flow leads to reduced population differentiation and increased within-population variation. Because plants are nonmotile, gene flow is facilitated via pollination and seed dispersal, both of which are significantly correlated with genetic diversity levels (Hamrick and Godt 1989, 1996). Lack of gene flow may increase the occurrence of inbreeding in a population, resulting in a genome-wide reduction of heterozygosity. Subsequent genetic erosion caused by genetic drift can increase the prevalence of harmful recessive alleles, thus negatively impacting the evolutionary potential of a population (Frankham 2005, Pfeifer and Jetschke 2006). For example, rare endemic species may have very narrow habitat specificities that severely limit the areas the species can potentially occupy (Kruckeberg and Rabinowitz 1985). If these specific habitats are uncommon or disjunct, then this pattern will be reflected in the distribution of a species that is dependent on these environmental conditions, potentially creating a barrier to gene flow. The combination of few populations and few individuals in each population may lead to difficulties in finding mates, leading to increased levels of inbreeding (Kruckeberg and Rabinowitz 1985, Karron 1987) and a loss of heterozygosity in individuals, perhaps even causing the species to experience inbreeding depression which could further jeopardize reproductive success and shrink population size.

While gene flow may be seen as a solution to avoiding serious cases of inbreeding, it may not always be beneficial, especially in instances where a population is highly adapted to a specific environment. Gene flow between differently adapted populations may lead to the introduction of detrimental allele combinations, resulting in maladapted individuals with reduced fitness (Price and Waser 1979, Edmands 2007). However, such instances are uncommon; thus, in most cases gene flow is beneficial as it impedes genetic erosion (Frankham et al. 2011).

Demographic histories are another important factor influencing the distribution and extent of genetic variability within and among plant populations. Lifespans have been shown to be correlated with genetic diversity, as perennials tend to exhibit greater genetic diversity levels than annuals (Hamrick and Godt 1989). Breeding systems also influence patterns of genetic diversity. Breeding systems in plants are diverse and many species may be able to reproduce via multiple methods. Generally, breeding systems can be classified as outcrossing or inbreeding (Charlesworth 2003, 2006). Outcrossing implies reproduction between unrelated individuals while inbreeding involves mating between relatives, with the most extreme form of inbreeding being selfing. Outcrossing plants typically exhibit higher levels of genetic variation than inbred plants due to the increased possibility of genetic recombination resulting in novel allelic combinations (Hamrick and Godt 1989, 1996).

Clonally reproducing plants can produce whole populations of genetically identical ramets due to the lack of genetic recombination. However, the lack of recombination does not necessarily prevent high levels of genetic variation in these populations. In some cases, asexual reproduction in plants may be the result of hybridization. These events can result in hybrid offspring that carry nonhomologous chromosomes, making crossing over impossible and preventing gene recombination. In such cases, the union of evolutionarily distinct lineages can result in high levels of genome-wide heterozygosity in a hybrid. Furthermore, these high levels of heterozygosity within an individual can be maintained in a population through asexual reproduction, which may be the only feasible method of reproduction in some hybrid lineages. Other hypotheses that address why clonally reproducing organisms can maintain high heterozygosity levels include the Meselson effect which states that in diploid organisms, the absence of sex promotes divergence between alleles at a locus and that homologous

chromosomes will independently accumulate mutations over many generations due to lack of recombination (Welch and Meselson 2000, Balloux et al. 2003). However, conclusive evidence for the Meselson effect has not been found in many species (Schaefer et al. 2006, Maderspacher 2008).

While asexually reproducing species are able to maintain high levels of heterozygosity across multiple loci due to the lack of genetic recombination that maintains genetic diversity within individuals, they do so at the cost of decreased genotypic diversity within a population (Balloux et al. 2003, Halkett et al. 2005). Thus, clonal organisms may display higher levels of allelic diversity at individual loci but fewer multilocus genotypes (Balloux et al. 2003). While the lack of variation among individuals in asexually reproducing species could limit a population's adaptive potential (Niklas and Cobb 2017), asexual reproduction has also been proposed as a strategy to facilitate range expansion, especially in areas of high environmental stress and instability where it may be difficult to find mates or pollinators or where the growing season may be limited (Hörandl 2006, Meloni et al. 2013, Rushworth et al. 2018). Compared to their sexual counterparts, organisms that reproduce asexually have also been suggested to be better colonizers in newly available environments, such as previously glaciated areas, and this pattern has been documented across multiple animal and plant taxa (Verduign et al. 2003, Hörandl 2006, Cosendai et al. 2013, Tilquin and Kokko 2016, Burke and Bonduriansky 2018). Moreover, sexual and asexual populations of the same species may have different geographic distributions, a phenomenon referred to as geographic parthenogenesis (Vandel 1928). The concept of geographic parthenogenesis encompasses patterns of differential distribution such as the occurrence of asexual forms at higher latitudes and altitudes than sexual forms, and/or asexual forms having a wider distribution than sexual forms (Hörandl 2006, Hörandl et al. 2008, Tilquin

and Kokko 2016). However, these patterns are generalizations and may not hold for all organisms. There are many factors that have been proposed to contribute to the pattern of geographic parthenogenesis; however, a single explanation cannot be put forward, and further research that focuses on uncovering the cause of these patterns is warranted (Tilquin and Kokko 2016).

Patterns of genetic variation in comparisons of rare and widespread species are also commonly observed. Often, rare species are found on endangered or threatened species lists due to their alarmingly small numbers, but there are many ways that a species can be classified as rare. Rabinowitz (1981) described seven types of rarity that are defined by aspects of geographic range, habitat specificity, and local population size. The most conventional type of rarity would be attributed to a species having a small geographic range, narrow habitat specificity, and small local population size (Rabinowitz 1981). In contrast, widespread species have comparatively large geographic ranges, although this may also be accompanied by small local populations.

Numerous studies have examined the genetic differences between geographically expansive and restricted species and have found that rare species often have lower levels of genetic diversity than their widespread counterparts (Karron 1987, Hamrick and Godt 1996, Gitzendanner and Soltis 2000, Cole 2003). Because rare species often have small populations and restricted geographic ranges, they are more likely to be negatively impacted by random genetic drift, inbreeding, and demographic stochasticity which, respectively, can lead to a loss of alleles, decreased heterozygosity in individuals, and a further reduction in population size which increases the risk of extinction (Ellstrand and Elam 1993, Gitzendanner and Soltis 2000, Lande 1993). The random changes in allele frequencies via genetic drift can cause small populations to lose alleles, and this reduction in alleles leads to a decrease in genetic polymorphism and

heterozygosity levels. In contrast, widespread species often have larger population sizes and inhabit greater geographic ranges, allowing for a larger gene pool and a decreased chance of events such as inbreeding, loss of alleles through random genetic drift, and the harmful effects of demographic stochasticity.

These diversity patterns are not true for all rare and widespread species, as different life history traits can significantly affect the genetic diversity levels in a species. In a review of allozyme literature, Hamrick and Godt (1989) found that not only geographic range, but also life form, breeding system, and seed dispersal mechanisms had a significant influence on a species' genetic diversity. Annuals, long-lived perennials, plants with mixed-mating breeding systems, wind-pollinated species, and species with animal-dispersed seeds were reported to have the highest levels of allozyme variation at the species level compared to other lifeforms, breeding systems, and modes of seed dispersal (Hamrick and Godt 1989). Taxonomic status is also a significant determinant of genetic diversity with certain families such as Orchidaceae, Poaceae, and predominantly woody families having high genetic diversity (Hamrick and Godt 1996). However, like taxonomic status, life history traits tend to be evolutionarily constrained and often are resistant to change within a lineage (Duminil et al. 2007). This means that species level comparisons lack statistical independence if life history and phylogeny are not accounted for.

Historically, comparisons of rare and widespread species were made with little consideration of phylogeny, which can confound results due to distinct evolutionary pathways and ecological features that influence genetic variability. Studies involving congeneric comparisons have been recommended, to help minimize the confounding factors of phylogeny (Karron 1987, Gitzendanner and Soltis 2000) and such congeneric studies have become the norm in comparative studies of rare and widespread species. In the numerous congeneric studies that

have been published since then, there has been a fairly consistent pattern of rare species having lower levels of polymorphic loci, alleles per locus, and heterozygosity as compared to widespread or common species (Cole 2003). Even when examined at similar spatial scales, widespread species tend to exhibit greater levels of genetic diversity (Gonzales and Hamrick 2005, Gibson et al. 2008).

However, the classification of species as “rare” or “widespread” can be a generalization that offers little insight to the biology of a species and overemphasizes the role of geographic range in regard to genetic diversity (Kunin and Gaston 1993, Gitzendanner and Soltis 2000). For instance, some studies have found rare species to have comparable, if not greater levels of genetic diversity than their widespread congeners. Takahashi et al. (2011) found three rare *Tricyrtis* species in Japan to harbor comparable levels of genetic diversity as a widespread *Tricyrtis*. The low genetic diversity of the widespread species was attributed to the fact that the widespread species selfs and is the result of a relatively recent speciation event as compared to the rare species, all of which outcross and diverged from a common ancestor in the distant past. Thus, these differences in breeding systems had a significant impact on the genetic diversity levels in these species. Another congeneric study found the rare *Helianthus verticillatus* to have significantly higher levels of gene diversity and expected heterozygosity within populations than its widespread congener *H. angustifolius* despite these species having similar life histories (Ellis et al. 2006). Similar findings were reported by Gitzendanner and Solits (2000) in a review of isozyme literature among rare and widespread plant congeners where they noted that rare species do occasionally display comparable or even greater levels of genetic diversity than widespread species. These departures from the typically reported trends indicate it is best to not focus on a single explanatory aspect of genetic diversity, but instead to contextualize a species in terms of

its biogeography, evolutionary history, and ecological influences which can have differential impacts on genetic makeup (Karron 1987, Gitzendanner and Soltis 2000).

Measuring genetic variation

Genetic diversity within a species is assessed through a number of diversity measurements such as the percentage of polymorphic loci, number of alleles per locus, and observed and expected heterozygosity levels. Wright's *F*-statistics evaluate how diversity is distributed within and among populations of a species, giving insights to the partitioning of genetic variability. In addition, various measures of genetic identity and genetic distance display pairwise similarities of genetic data between populations. Clustering methods can uncover population structuring and display similarities in genetic patterns among populations. Such methods can be utilized to hypothesize migratory patterns of a species and determine if there are geographic patterns in genetic makeup.

Percentage of polymorphic loci

A polymorphic locus is defined as a locus where more than one allele is observed; contrastingly, a monomorphic locus is a locus which has only one type of allele (Hartl 2000). However, the occurrence of rare alleles may falsely inflate the reported number of polymorphic loci. Thus, polymorphic loci are sometimes delimitated by an arbitrary frequency cutoff of 0.95 or 0.99 (Hartl and Clark 1989). By providing this cutoff, a greater focused can be placed on loci with allele frequencies greater than what would be expected from random mutations alone (Hartl and Clark 1989).

The percent of polymorphic loci, a ratio of the number of polymorphic loci divided by the total number of loci analyzed, is a diversity measure frequently used in population genetic studies. The percent of polymorphic loci can be calculated at the species and population level.

To determine population level polymorphism, the percent of polymorphic loci is calculated for each population and then averaged across all populations to provide the average polymorphism for a population. At the species level, the percent of polymorphic loci is calculated by determining which loci display any polymorphism and dividing by the total number of loci tested.

Observed and expected heterozygosity

Observed heterozygosity (H_o) is the proportion of heterozygotes observed at a locus. Contrastingly, expected heterozygosity (H_e) is the proportion of heterozygotes expected if a population were in Hardy-Weinberg equilibrium. For a population to be in Hardy-Weinberg equilibrium, there must be no mutations, no gene flow between populations, and no natural selection affecting the gene of interest. Furthermore, the population must be large enough to avoid random genetic drift, and mating must be random. In addition to the conventionally cited assumptions, the Hardy-Weinberg principle only strictly applies to diploid organisms with nonoverlapping generations (Hartl 2000). If all conditions are met, a population's allele and genotype frequencies would remain constant from generation to generation. The constancy of allele frequencies implies that in the absence of evolutionary forces, Mendelian inheritance alone maintains allele frequencies and genetic variation (Hartl and Clark 1989). Although the conditions for Hardy-Weinberg equilibrium are rarely every met in full, the principle is fairly robust to violations of its assumptions and serves as a valuable null model for which measurements can be compared (Hartl and Clark 1989).

Expected genotype frequencies are obtained by multiplying the sample size of a population by the Hardy-Weinberg frequencies. The observed and expected heterozygosity levels are compared via a goodness-of-fit test to determine if the observed genotype frequencies

of a population are comparable to what is expected under Hardy-Weinberg equilibrium conditions. If the data depart from the Hardy-Weinberg proportions, it can be inferred that one or more evolutionary forces are acting on the locus of interest. Often, heterozygote deficits across multiple loci imply inbreeding, while a genome-wide heterozygote excess may be due to allopolyploidy conditions or hybridization, possibly maintained by clonal reproduction (Hartl and Clark 1989; Balloux et al. 2003). A heterozygote excess at a few loci may suggest the possibility of a heterozygote advantage, but this would only be true if the loci of interest codes for a gene, or if the loci are neutral and linked with non-neutral loci. However, other sources of potential error must be considered as well when assessing deviations from Hardy-Weinberg expectations. Artifactual errors such as missing data, null alleles, and genotyping errors may contribute to the apparent departures (Chen et al. 2017). In addition, the erroneous grouping of distinct subpopulations may falsely raise the homozygosity level of a sample and incorrectly suggest the occurrence of inbreeding when the actual cause of such a situation would be population subdivision. The homozygote excess caused by such population structuring is referred to as the Wahlund effect (Wahlund 1928, Hartl and Clark 1989).

Wright's F-statistics

Sewall Wright proposed a set of *F*-statistics, collectively referred to as inbreeding coefficients, which provide information regarding the distribution of genetic diversity within and among populations (Wright 1950; Holsinger and Weir 2009). These parameters measure the degree of departure from the Hardy-Weinberg expectations of random mating (Hartl and Clark 1989). Wright defined three main *F*-statistics which describe varying degrees of population structure: F_{IT}, F_{IS}, and F_{ST} (Wright 1950, Hartl and Clark 1989). F_{IT} is a correlation of gametes within an individual relative to gametes drawn from random in a whole population (Hartl and

Clark 1989; Holsinger and Weir 2009). F_{IS} is associated with the levels of inbreeding within a population (Hartl and Clark 1989). F_{IS} values can range from -1 to 1 with negative values indicating a heterozygote excess and positive values indicating a heterozygote deficiency. A value closer to 1 may be indicative of substantial inbreeding. Negative values may indicate a heterozygote advantage at a specific locus (or close linkage to such a locus). However, as previously mentioned, a high degree of heterozygosity at multiple loci may suggest the possibility of polyploidy or clonal reproduction, (Balloux et al. 2003). An F_{IS} value of 0 suggests that the number of heterozygotes in a population is comparable to what is expected under Hardy-Weinberg equilibrium conditions. In outcrossing populations, F_{IS} values are expected to be near zero (Hartl and Clark 1989).

F_{ST}, known as the fixation index, is a measure of genetic differentiation among populations and is related to the variance in allele frequencies among populations (Holsinger and Weir 2009). F_{ST} values can range from 0 to 1. Lower values indicate a high degree of panmixia and thus, most of the genetic variation in a population can be attributed to individuals within a population being different from each other. In contrast, values close to 1 indicate a low degree of gene flow. In such cases, the majority a species' genetic variation can be attributed to populations being different from each other (among-population variation). High population differentiation may be the result of population isolation preventing gene flow or selective pressures contributing to population differentiation.

Despite the usefulness of Wright's F_{ST}, it was developed under the assumption of purely biallelic loci; and thus, is severely restrained in its breadth of applicability. Various analogues of F_{ST} have been developed including Weir and Cockerham's (1984) ANOVA based estimate and Nei's (1973) G_{ST}, which is an extension of Wright's F_{ST} based on gene diversity within and

among populations. However, these analogues are also constrained by the underlying assumption of biallelic loci. Therefore, in instances where expected heterozygosity levels are high, F_{ST} becomes very small and may falsely suggest weak population structure (Meirmans and Hedrick 2011). Additionally, loci under strong selective pressure or with high mutation rates may have drastically different F_{ST} values from other loci (Meirmans and Hedrick 2011, Whitlock 2011). To avoid such problems, Meirmans (2006) proposed the use of F'_{ST} which utilizes analysis of molecular variance (AMOVA) to standardize F_{ST} measurements by the maximum value of expected heterozygosity at a locus, given the observed within-population diversity (Meirmans and Hedrick 2011). This allows an unbiased comparison across loci and species. However, it has also been argued that F'_{ST} is not a suitable replacement for F_{ST} (Whitlock 2011). In the face of such contradictions, it has been proposed that both F_{ST} and F'_{ST} values be reported in studies for comparative purposes.

Nei's genetic identity and distance

Nei's genetic identity (Nei 1978) is a measure of genetic similarity between population pairs. In contrast, Nei's genetic distance (Nei 1978) is a pairwise measure of genetic divergence between populations. Genetic identity measures range from 0 to 1, with values closer to 1 suggesting a higher degree of similarity between populations while measures close to 0 indicate greater divergence. Genetic distance ranges from 0 to infinity with more closely related populations having a small genetic distance. Populations sharing multiple alleles would exhibit a high genetic identity and a small genetic distance.

Analysis of clonal species

Despite clonality as a life-history strategy being widespread among all forms of life, standardized methods for analyzing the population genetics of clonal species have not been

developed. When assessing clonal plant species, a distinction must be made between genets and ramets. A genet consists of individuals that all have the same genetic composition. Each of these genetically identical individuals are referred to as ramets. If a ramet is mistakenly analyzed as a genet, this may lead to bias through pseudo-replication and violations of assumptions of certain population genetic statistics (Alberto et al. 2005, Bentley and Mauricio 2016, Mandel et al. 2019). To avoid this, the genetic variability of clonal organisms is often based on the number of unique multilocus genotypes. The number of unique multilocus genotypes can be used as a reflection of the true number of genets, and then population genetic analyses can be run on a unique multilocus genotype data set which eliminates any bias that may arise from the inclusion of clones. Although the inclusion of clones would offer a more accurate view of a species' population structure, the exclusion of repeated multilocus genotypes is necessary to avoid misleading statistics (Halkett et al. 2005).

Methods of assaying genetic diversity

The genetic diversity of a species can be assayed in a number of ways. Common methods for analyzing genetic diversity have focused on protein or DNA analysis as these methods allow for direct measures of genetic variation without conflict from environmental influences (Mondini et al. 2009). Protein electrophoresis utilizes allozymes or isozymes to determine genetic differentiation. Although the terms "allozyme" and "isozyme" are often used interchangeably, they technically differ in that allozymes are enzymes with different forms due to being encoded by different alleles at the same locus whereas isozymes are coded by different loci. Isozyme analysis is based on the separation of proteins via electrophoresis resulting in distinct banding patterns. These markers are codominant, allowing for the unambiguous detection of heterozygote individuals and thus proving useful to population genetic studies.

However, isozymes are constrained in the number of different forms they can take and therefore can only provide a limited view of a species' genetic diversity (Mondini et al. 2009).

Molecular methods, including restriction fragment length polymorphism (RFLP), amplified fragment length polymorphisms (AFLPs), and microsatellite markers (also referred to as simple sequence repeats or SSRs), have grown in popularity as they are relatively inexpensive and exhibit higher levels of polymorphism in comparison to isozymes (Liu and Cordes 2004). Microsatellites in particular are popular markers as they are codominant, selectively neutral, and have higher mutation rates than expressed portions of the genome, thus allowing for extensive allelic diversity (Mondini et al. 2009). While these markers often have high rates of polymorphism, there are limitations to the conclusions that can be drawn when using microsatellites to assess genetic variability since neutral markers may not always correlate with adaptive genes, and by extension, the evolutionary potential of a species (Kirk and Freeland 2011). Nevertheless, significant advancements in the field of population genetics have been made thanks to the adaptation of these molecular markers.

CHAPTER II

Materials and Methods

Population Sampling

To provide a representative estimation of genetic diversity levels, populations were sampled throughout *Boechera laevigata*'s range in the eastern United States. Table 1 provides information about the sampled populations. Two populations (the TN-1 and KY populations) were collected by members of our lab. Volunteers contacted through personal communications aided in the collection of the other populations used in this study. Overall, 14 populations were sampled from 11 states, resulting in a total sample size of 304 individual plants. The Ohio population consisted of two groups of plants less than two kilometers apart. After analyzing the data, these separate groups were determined to be two distinct subpopulations. Thus, these subpopulations were split and analyzed separately, resulting in 15 populations. Additionally, population NC-2 was mowed just prior to collection, thus reducing the number of samples that could be collected and resulting in a low sample size. There were also many dead plants at the KY location at the time of collection, which resulted in low sample sizes. Three populations were collected from Tennessee, with distances between these populations ranging from approximately 24 to 40 km. In Alabama, two populations were collected within 6 km of each other. The distances between populations in these states is comparable to the geographic distances of *B. perstellata* populations sampled by Baskauf et al. (2014) in their study. Having comparable distances between the *B. laevigata* populations in relation to the previously studied *B. perstellata* populations allows for comparisons of genetic diversity levels at similar spatial scales between the two species.

Table 1. Geographical location and population information for sampled *Boechera laevigata* populations. "Estimated" population size is a rough estimate by the individual collector of the number of plants present in the population at the time of leaf collection.

State	County	Population abbreviation	GPS coordinates	Estimated population size	Sample size	Collection date
Alabama	Colbert	AL-1	N 34.780817 W 87.666238	45-40	30	March 2020
		AL-2	N 34.749747 W 87.725302	35-40	30	March 2020
Arkansas	Carroll	AR	N 36.43516, W 93.77003	>100	27	April 2019
Illinois	Cook	IL-1	N 42.000568 W 87.464595	50-60	30	May 2019
	Lee	IL-2	N 41.893785 W 89.369955	100-200	25	June 2019
Kentucky	Warren	KY	N 37.14455 W 86.38361	20-40	10	May 2019
Maryland	Washington	MD	N 39.362723 W 77.740191	20-25	14	May 2020
North Carolina	Henderson/ Rutherford	NC-1	N 34.4425 W82.27617	50-150	25	April 2019
	Haywood	NC-2	N 35.4387 W 83.0492	30-50	13	May 2019
Ohio	Summitt	OH-1	N 41.135092 W 81.534681	>30	16	May 2019
		OH-2	N 41.123698 W 81.516280	25-30	6	May 2019
Pennsylvania	Clarion	PA	N 41.348516 W 79.220889	>400	24	July 2018
Tennessee	Cheatham	TN-1	N 36.28553 W 87.07982	>100	34	April 2019
	Montgomery	TN-2	N 36.425360 W 87.289103	>100	10	March 2019
	Cheatham	TN-3	N 36.14726 W 87.00302	>100	10	March 2019

Leaves were collected from mature plants during the springs of 2018, 2019, and 2020, depending on the population. Sample sizes per population ranged from 10-34 individuals. For further information regarding population collection information, see Table 1. As they were collected, leaves for each plant were placed in separate paper envelopes and dried in silica gel.

Voucher specimens were collected for all populations. Vouchers for the Alabama, Arkansas, Kentucky, Pennsylvania, Tennessee, NC-1, and Illinois populations were deposited in the Austin Peay State University herbarium (ASPC). Voucher specimens for the OH-1 and OH-2 populations were deposited in the Kent State University herbarium (KE) and the voucher specimen for the NC-2 population was deposited in the Great Smoky Mountains National Park herbarium (GSMNP). Vouchers for the Maryland population were digitally recorded and can be viewed on iNaturalist (https://www.inaturalist.org/observations/44991266).

DNA Extraction

DNA was extracted from the dried leaf tissues of 305 plants following the E.Z.N.A SP Plant DNA Kit (Omega Bio-tek, INC., Norcross, GA) protocols with slight modifications. Using two tungsten-carbide beads, leaf tissues were ground in microfuge tubes in a Retsch (Newton, PA) MM301 bead mill for 4-5 minutes until leaf tissues were mostly homogenized. Procedures for the binding and washing steps were followed according to the manufacturer's protocol. Modifications to the elution step involved reusing the initial 100 µL of Elution Buffer added to each HiBind® DNA Mini Column for the second elution step to yield greater DNA concentrations. Concentrations were measured with a NanoDrop-1000 Spectrophotometer (Thermo Scientific, Wilmington, DE) then stored at -20°C.

PCR

Using four individuals from 11 populations (excluding AL-1, AL-2, and MD), the 19 primer pairs resolved by Baskauf et al. (2014) for *B. perstellata* were initially tested for transferability to *B. laevigata*. To increase the number of loci analyzed, six additional primer pairs published by Schranz et al. (2007) were selected for testing, using the same individuals

from the first survey stage. In total, 25 primer pairs were surveyed, with only loci polymorphic for *B. laevigata* utilized in this study.

Polymerase chain reaction (PCR) products were fluorescently labeled following the procedures outlined by Shuelke (2000) and modified by Baskauf et al. (2014) who added an 18 bp M-13 tag (5'-CAC GAC GTT GTA AAA CGA-3') to the forward primer and the fluorophore. For locus FO3 and five loci (G03, G06, G08, G09, and H06) not utilized by Baskauf et al. (2014), a 7-base pair short tag (GTTTCTT) was added to the 5' end of the reverse primer to help reduce stutter (S. Bogdanowicz, personal communication). Each PCR reaction was run in a 15 µL solution consisting of 1.5 µL tricine buffer (300 mM tricine, 500 mM KCl, 20mM $MgCl_2$), 0.06 µL dNTPs (25 mM each), 0.04 µM M-13 tagged forward primer (0.6 pmoles), 0.2 µM reverse primer (0.6 pmoles) or 0.2 µM short-tagged reverse primer (0.6 pmoles), 0.2 µM M-13 tagged fluorophore, 0.03 units of *Taq* DNA polymerase (GoTaq Flexi DNA Polymerase, Promega, Madison, WI), and 1 µL DNA template (8.8-386.1 ng/µL). PCR conditions for all primers, excluding ICE4, F03, and H06, were as those described in Baskauf et al. (2014). For primers ICE4 and F03, the PCR conditions were as follows: 95.0°C for 3 minutes; 35 cycles of 94°C for 30 seconds, 50°C for 30 seconds, 72°C for 45 seconds; followed by a 72°C final extension for 20 minutes. For primer H06, the PCR conditions were as follows: 95.0°C for 30 minutes; 35 cycles of 94°C for 30 seconds, 55°C for 30 seconds, 72°C for 45 seconds; followed by a 72°C final extension for 20 minutes.

PCR products were run on a 1% agarose gel for 60 to 75 minutes to verify amplification. PCR products were then placed in a 1:10 dilution with molecular biology grade USP sterile purified water (Mediatech, INC., Manassass, VA). One microliter of the diluted microsatellite solution was added to a 10 µL solution containing GeneScan LIZ-500 size standard (Applied

Biosystems, Foster City, CA) and Hi-Di formamide (Applied Biosystems, Foster City, CA) in a 0.25 : 9.25 ratio. The PCR products and size standard were then heated to 95°C for five minutes and chilled at 4°C for 2 minutes before being shipped overnight on ice to the University of Tennessee Health Science Center (Memphis, TN) for autosequencing on an ABI 3130XL DNA Analyzer (Applied Biosystems, Foster City, CA). Fragments were viewed using GeneMarker v1.97 (SoftGenetics, LLC, State College, PA, 2010) and alleles were manually identified. Allele calls were verified by two people.

Population Analyses

Apomictic reproduction results in offspring identical to the parent. If multiple clones from the same genetic individual are sampled, this can skew population genetic analyses of DNA markers and distort diversity levels in clonal populations. In these cases, a researcher often utilizes a data set which excludes possible "clones" by including each unique multilocus genotype ("unique MLG") only once. Various lines of evidence from this study combined with ongoing studies of *B. laevigata* (M. Windham pers. comm.) and other *Boechera* species by other researchers (Schranz et al. 2005, Song and Mitchell-Olds 2007, Rushworth et al. 2011, Beck et al. 2012, Lovell et al. 2014, Alexander et al. 2015, Li et al. 2017, Rushworth et al. 2018), suggest that while most of the sampled populations are likely reproducing sexually, a few of the populations are probably reproducing apomictically. This would necessitate using a unique MLG only data set for the apomictic populations but a "full" data set (all samples included) for the sexual populations. However, cytological analyses of seeds are necessary to definitively confirm apomictic reproduction, and since the current study did not include such cytological analyses, some tables in Appendix A provide various genetic diversity measures re-calculated using a

"full" data set for the putative apomictic populations and using a "unique MLG" data set for the putative sexual populations.

The fact that the apomictic OH-1 population had only a single unique MLG was problematic for a few analyses when using the "unique-MLG-only" data set (AMOVA, pairwise and overall F_{ST} and standardized F_{ST}, and genetic identity,). In those cases, two OH-1 individuals were actually included in the analysis of the "unique MLG only" data set.

Analysis of clonal and genetic diversity

The R package *Poppr* v4.0.3 (Kamvar et al. 2014) was used to generate a unique MLG dataset. Individuals are assigned as "clones" if the genetic distance between them falls below the established clonal threshold, which was chosen to be 0. Two populations each had one individual with missing data at a single locus, and a third population had two individuals with missing data at that locus. Individuals with missing data were excluded from all analyses.

In *Poppr*, a genotype accumulation curve was generated with 10,000 permutations for each population to determine the minimum number of loci needed to discriminate between unique individuals for that population. Once it was determined that there was enough power to differentiate individuals, the probability of obtaining a particular MLG by chance (P_{gen}) and the probability of encountering the same MLG more than once by chance (P_{sex}) (Parks and Werth 1993) were calculated with MLGsim 2.0 (Stenberg et al. 2003, Ivens et al. 2012 see http://www.rug.nl/fmns-research/theobio/downloads) using the F_{IS} model for calculating P_{gen}. This model accounts for deviations from Hardy-Weinberg equilibrium and gives a more conservative estimate of P_{sex} (Arnaud-Haond et al. 2007). Small P_{gen} values indicate a low probability of obtaining the same MLG by chance. Further, a significant P_{sex} value suggests that

an MLG observed more than once in a population is likely due to clonal reproduction rather than to sexual reproduction (Stenberg et al. 2003).

Clonal diversity measures for the unique MLG only data set were calculated using *Poppr*. These clonal diversity measures included the number of unique multilocus genotypes (*MLG*), the expected number of multilocus genotypes (*eMLG*) at the smallest sample size greater than or equal to 10 based on rarefaction, the Shannon-Wiener index of MLG diversity (also called the Shannon-Wiener or the Shannon-Weaver index) which uses a natural logarithm, Simpson's index, and evenness.

For both the full data set and the unique MLG only data set, GenAlEx v6.503 (Peakall and Smouse 2006, 2012) was used to calculate genetic diversity statistics such as percentage of polymorphic loci (*P*), observed (H_o) and Nei's (1978) unbiased expected heterozygosity (H_e, abbreviated as uHe in GenAlEx). It would be appropriate to include the full data set for any sexually reproducing population (referred to as simply "sexual" from here on), which we suspect includes most of our sampled populations. However, the OH-1 population and both the Illinois populations are likely reproducing apomictically (referred to as "apomictic" populations from here on), in which case the diversity statistics for these populations would be best calculated over only the unique MLG. Apomixis cannot be completely dismissed as a mode of reproduction in some of the other presumably sexual populations without cytological analysis, so the diversity analyses were carried out on a full data set and unique MLG only data set for all populations as was done by Bentley and Mauricio (2016) and Mandel et al. (2019).

GENEPOP v4.7.5 (Rousset 2020) was used to conduct chi-square goodness-of-fit tests to determine significant deviations from Hardy-Weinberg expectations using 10,000 Markov Monte Carlo permutation on a full data set only. Sequential Bonferroni corrections were performed to

account for multiple comparisons. GENEPOP was also used to calculate linkage disequilibrium (LD) between loci pairs, checking for non-random associations between alleles at the loci according to the composite linkage disequilibrium test of Weir (1996)., using both the full data set and the unique MLG only data set. GENEPOP uses 10,000 Markov Chain Monte Carlo permutations to identify and test for significance of the non-random associations. Because the inclusion of ramets can skew heterozygosity estimates (Douhovnikoff and Leventhal 2016), deviations from Hardy-Weinberg expectations were calculated on both the full and unique MLG only data sets in all populations. Additionally, clonal reproduction can lead to a genome-wide non-random association of loci; thus, Halkett et al. (2005) suggest that LD estimates be performed with and without repeated multilocus genotypes. Accordingly, LD was tested on both the full and MLG only data sets. Sequential Bonferroni corrections were performed for linkage disequilibrium tests to account for multiple comparisons.

In addition to pairwise tests of linkage disequilibrium, de Meeus and Balloux (2004) recommend multiple loci estimates for linkage disequilibrium in species that display asexual reproduction. Therefore, the standardized index of association ($\bar{r}_d$ or *rBarD*) (Agapow and Burt 2001) was calculated using the R package *Poppr* (Kamvar et al. 2014) on multilocus genotypes. This index has been used to identify clonal reproduction in populations and tests how likely individuals that are the same at one locus are the same at other loci while also accounting for the number of loci tested (Agapow and Burt 2001).

Additionally, statistical differences in genetic diversity measures between the two reproductive modes, utilizing a full data set for the sexual populations and a unique MLG only data set for the apomictic populations (using two individuals from the OH-1 population), were examined using FSTAT ver. 2.9.4 (Goudet 2003) for several parameters including: the number

of alleles per locus (allelic richness) observed heterozygosity, and expected heterozygosity ("gene diversity"), F_{ST}, and F_{IS}. FSTAT's estimate of allelic richness utilizes a rarefaction method that calculates the expected number of alleles per locus at the smallest number of individuals sampled for a locus, which bypasses the dependence of the number of alleles per locus on sample size. FSTAT calculates all the estimates listed above, weighted by sample size; then for the statistical testing it employs a permutation scheme (1000 permutations were used) that randomly allocates individual samples to different groups (populations) while keeping the number of samples per group constant. For an unbiased estimate of gene diversity (expected heterozygosity), FSTAT utilizes Nei's (1987) equation 7.39 rather than Nei's (1978) equation.

Analysis of population genetic structure

Population differentiation was analyzed via analysis of molecular variance (AMOVA) in GenAlEx v6.503 (Peakall and Smouse, 2006, 2012), using the codominant allelic distance matrix input option. This option was also used to estimate Wright's *F*-statistics (Weir and Cockerham 1984) to determine genetic variability within and among populations. Estimates of F_{IS} were carried out using FSTAT ver. 2.9.4 (Goudet 2003). Jackknifed means for the sexual and apomictic populations were tested for differences from zero using a one-sample *t*-test. Because inclusion of the presumed apomictic populations would skew the overall *F*-statistic estimates of the other 12 populations due to differing reproductive methods, these populations were analyzed separately, as in Cosendai et al. (2013). In addition, pairwise F_{ST} comparisons and pairwise standardized F_{ST} (F'_{ST}) (Meirmans 2006) comparisons among all populations, including the apomictic ones, were estimated from the AMOVA using 999 Markov Chain Monte Carlo permutations using GenAlEx v6.503.

Nei's (1978) genetic identities and genetic distances were calculated using a full data set for the presumed sexual populations and a unique MLG only data set (with two OH-1 samples) for the presumed apomictic populations, although analyses for a full data set and a unique MLG only data set for all populations were also performed.

Using the distance matrix with data standardization, a principal coordinate analysis (PCoA) was generated to graphically represent the genetic distance matrix created from the full data set for sexual populations and a unique MLG only data set for apomictic populations.

Parametric approaches for determining population substructure often assume populations are in Hardy-Weinberg equilibrium and that there is no linkage disequilibrium (Alhusain and Hafez 2018). If these assumptions are violated, nonparametric methods can be utilized which are more appropriate for clonal or partially clonal species as well as inbreeding species, which often violate the assumptions of some traditional clustering software such as STRUCTURE (Pritchard et al. 2000), which assumes LD and departures from Hardy-Weinberg proportions are due to population structure rather than other factors that may contribute to these deviations. A discriminant analysis of principal components (DAPC) is a multivariate analysis that combines principal component analysis with discriminant analysis to identify and describe genetic clusters of related individuals (Jombart 2010). This is done through maximizing among group differences and minimizing within-group variation. DAPC was conducted using the R package *adegenet* v2.1.3 (Jombart 2008) for which a full data set was used for the sexual populations and a unique MLG only data set was used for the apomictic ones. To determine the number of genetic clusters (*K*) in the data, the function *fviz_nbclust*, using the silhouette method, from the R package *Factoextra* v1.0.7 (Kassambara and Mundt 2020) was used. The optimal number of *K* is associated with the largest average silhouette width. To determine the number of principal

components (PCs) to retain for the analysis, cross-validation procedures were carried out using the function *xvalDapc*. Cross-validation divides the data into a training set, comprised of 90% of the data, and a validation set, containing the remaining 10% of the data. Members included in the validation set are selected via stratified random sampling which ensures at least one individual from each population is represented in the validation and training sets. Varying numbers of PCs are retained and tested on the training set. For each level of PC retention, sampling procedures were repeated 1000 times. The best number of PCs to retain was associated with the lowest root mean square error.

Comparison of genetic diversity with Boechera perstellata

Genetic diversity levels for the 12 "sexual" *B. laevigata* populations were compared with levels reported for the rare *Boechera perstellata* (Baskauf et al. 2014), which is considered a sexually reproducing species. The apomictic *B. laevigata* populations were excluded from this comparison. Although *B. laevigata* has been assayed for 16 loci and Baskauf et al. (2014) report on *B. perstellata* data from 19 loci, the comparison between the two species is based on only the 11 microsatellite loci that were assayed for both species (loci B07, B11, C02, C03, E09, E11, F03, ICE4, ICE5, R3_02, and R3_35). As had been done for each species' complete set of loci, the percentage of polymorphic loci (P), number of alleles per locus (A), number of private alleles (A_p), and observed (H_o) and Nei's (1978) unbiased expected heterozygosity (H_e) were recalculated over just this set of 11 loci for both species using GenAlEx v6.503 (Peakall and Smouse 2006, 2012). GenAlEx was also used to estimate F_{ST} and F_{IS} values using AMOVA.

In addition, statistical differences in allelic richness, observed heterozygosity, expected heterozygosity ("gene diversity"), F_{ST}, and F_{IS} between the species were examined using FSTAT ver. 2.9.4 (Goudet, 2003). GenAlEx and FSTAT handle missing data differently, so a data set

without missing data at these 11 loci was utilized for the species' comparisons. A total of 19 *B. perstellata* individuals had some missing data. Because *B. perstellata* populations have extremely low levels of within population genetic diversity, most missing genotypes occurred at loci that were monomorphic for that population, thus the missing genotype could be reasonably inferred. Only one *B. perstellata* individual had to be excluded from the data set as it was missing data at locus that was polymorphic in its population.

CHAPTER III

Results

Of the 25 microsatellite loci tested, 16 polymorphic loci were resolved for *B. laevigata*. One locus (F03) often amplified weakly and did not amplify at all for four individuals spread across three different populations, so these four individuals were removed from the data set. However, all other individuals amplified at all loci. The lack of missing data at the other loci suggests that null alleles are unlikely to be an issue for them, especially considering that *B. laevigata* appears to inbreed extensively, as is discussed further in this section. Characterization of these loci is provided in Table 2. Population level allele frequencies for all loci are available in Appendix A Table 1.

Using the full data set, chi-squared goodness-of-fit tests found genotype frequencies deviated significantly from Hardy-Weinberg expectations at all polymorphic loci for all populations except TN-2, TN-3 and OH-2 (Appendix A Table 2). In the TN-2 population, one of seven polymorphic loci deviated from expectations, while in TN-3 four of eight polymorphic loci did not meet expectations. No significant deviations were found in the OH-2 population following sequential Bonferroni corrections.

After sequential Bonferroni corrections to account for multiple tests, only four of the 15 populations (AL-1, MD, PA, TN-1) had significant linkage disequilibrium (LD) for any pairs of loci, as tested by GENEPOP, and the significant pairs of linked loci were different for each population (Appendix A Table 3). AL-1 had the most evidence of LD, with 71 pairs of loci found to have significant LD although only 35 were significant when analyzing the unique MLG only data set. The MD and PA populations each had only one locus pair with significant LD, although for the PA population the pair was not significantly linked when only MLGs were analyzed, probably owing to the few MLGs in this population. For TN-1, five locus pairs

displayed significant LD, with four of these locus pairs significantly linked when analyzing unique MLGs. Only two of the apomictic populations (IL-1 and IL-2) could be tested for pairwise linkage disequilibrium for which there was no linkage disequilibrium when utilizing either the unique MLG or the full data set.

Table 2. *Boechera laevigata* microsatellite characterization including primer sequence, repeat motif, number of alleles, range of pcr product sizes for the alleles, and the fluorophore tag used in the study.

Locus	Primer sequence	Repeat motif	No. of alleles	Size (bp)	Fluorophore
B07	F-CGGGAAGATTCAGCAGGTAA R-TCCTTTCCTCTCTTTATCCATCA	$(TTG)_3$	4	169-179	PET
B11	F-CCAAAGCAGTGACCAAAACA R-GAGCAGCATCAGGAGAAACC	$(CT)_2$	6	169-187	PET
C02	F-CTCGGTCTCCTCCATTACCA R-CGTTGTTTGGTGTCTGCATC	$(GA)_2$	5	158-178	NED
C03	F-CTCGGTCTCCTCCATTACCA R-CGTTGTTTGGTGTCTGCATC	$(GA)_2$	5	161-181	PET
E11	F-CATTTGGCTGTCCATGTTGA R-AGGGGTACAAGTGGTGGTTG	$(CT)_2$	2	186-192	PET
F03*	F-TCCGCAAAACTAAAAGGCTTA R-CCATCTTCACTTCCCGATGA	$(TC)_2$	25	233-295	VIC
G03*	F-CGCCTCCATTTTATCTTCCA R-GTTGGTAACGCCGAATCTGT	$(GA)_{16}$	6	245-265	VIC
G06*	F-TGTGCAGTTAAAGCCATCCA R-GCCCCCAAATCAACCTCTAT	$(CT)_{14}$	28	259-328	NED
G08*	F-CAGGAGCTGAATGAACTTTGG R-TGAGCCAGCAGAGCTTAACA	$(AG)_{20}$	9	250-278	6FAM
G09*	F-CCCCATAGCTTTTTCTTCCA R-CCAGTCGTGATGTGTTTTAGAGA	$(AT)_{13}$	9	235-254	VIC
H06*	F-TGCATTTCACCGTTTCATTT R-TAATTTTTCCCCGCTCATTT	$(AT)_{16}$	8	234-252	6FAM
ICE4	F-CACGAGGAATCTGGCATGGTCG R-AGCGATTGCAAGCGGCTCAAG	$(CT)_2$	2	194-198	6FAM
ICE5	F-CTTGCAACCGCCAACTCAATCG R-CCTGTCTCGCTCCCGCACG	$(GT)_2$	3	203-210	PET
R3_02	F-TTAGTGCTCCAAACCCTTCG R-TTCCAGGCGAGTGAGAAGAT	$(CT)_2$	5	128-142	NED
R3_35	F-TCATCGCCTGCAAGTAACAA R-CCAGAGGATCTTATCGGTGTAA	$(AG)_2$	14	144-178	6FAM

*Indicated loci had a short tag (GTTTCTT) attached to the 5’ end of the reverse primer.

The standardized index of association ($\bar{r}_d$ or *rbarD*) as tested by *Poppr*, which is a multilocus LD test that assesses how likely individuals that are the same at one locus will be the same at other loci accounting for the number of loci tested thus making it independent of sample size, was significant in seven populations, including five of the sexual populations (AL-1, NC-1, PA, TN-1, TN-2), and two of the apomictic populations (IL-1, IL-2) (Table 3). Note that this index could not be calculated for apomictic population OH-1 due to the presence of only a single unique MLG. These results indicate that there is significant linkage disequilibrium among the tested microsatellite loci in these seven populations. Beyond the contrasting results in the two LD analyses for the apomictic populations from IL, there was overlap but not complete agreement in the LD test results for the sexual populations. Three of the sexual populations having significant LD for the pairwise GENEPOP analysis are also significant in this multilocus analysis, but previously mentioned MD is not significant here, and one sexual population that was not significant for the pairwise GENEPOP analysis (TN-2) was significant for the multilocus standardized index of association LD test. Note that significant linkage disequilibrium between loci can result from a lack of genetic recombination, as would occur for clonal reproduction, such as the apomictic *B. laevigata* populations. However, similar $\bar{r}_d$ results could also occur if a sexual population is engaging in extensive inbreeding (Nordborg and Innan 2002, Hudson 2004), as appears to be the case for the sexual *B. laevigata* populations sampled.

Clonal diversity analyses

Genotype accumulation curves generated for each population determined that the 16 loci analyzed in this study had enough power to discriminate between clonal individuals in 9 of the 15 populations sampled (Appendix B Figure 1). P_{gen} values for all populations were extremely small, ranging from 1.42×10^{-30} to 1.443×10^{-8}, thus supporting the discriminatory power of our 16

Table 3. Clonal diversity statistics for *Boechera laevigata* populations using a clonal threshold of 0. IL-1, IL-2, and OH-1 are considered to be apomictic populations, with all other populations considered to be sexually reproducing. Standard error is given in parentheses.

Population	**Sample size (*N*)**	**Multilocus genotype *(MLG)***	**Expected MLG[1] *(eMLG)***	**Shannon-Wiener Index[2] *(H)***	**Simpson's Index *(D)***	**Evenness *(E)***	**Standardized index of association[3] *(rbarD)***
Sexual populations							
AL-1	30	20	8.00	2.71	0.898	0.627	0.438**
AL-2	30	8	4.96	1.69	0.758	0.710	-0.025
AR	27	21	8.91	2.91	0.936	0.830	0.001
KY	10	7	7.00	1.89	0.840	0.938	-0.085
MD	13	13	10.00	2.56	0.923	1.000	-0.019
NC-1	23	16	8.36	2.63	0.915	0.835	0.066*
NC-2	12	2	1.83	0.29	0.153	0.543	NA
OH-2	6	3	3.00	1.01	0.611	0.898	-0.189
PA	24	2	1.42	0.17	0.080	0.459	1.000*
TN-1	34	33	9.92	3.49	0.969	0.983	0.089**
TN-2	10	9	9.00	2.16	0.880	0.952	0.095*
TN-3	10	10	10.00	2.30	0.900	1.000	0.053
Sexual means	**19.1**	**12.0** (0.6)	**6.9** (0.9)	**2.0** (0.3)	**0.739** (0.088)	**0.815** (0.054)	**0.119** (0.104)
Apomictic populations							
IL-1	30	5	2.56	0.67	0.298	0.442	0.400**
IL-2	25	3	1.80	0.33	0.150	0.446	0.167**
OH-1	16	1	1.00	0.00	0.000	N/A	N/A
Apomictic means	**23.7**	**3.3** (0.2)	**1.8** (0.5)	**0.3** (0.2)	**0.149** (0.086)	**0.444** (0.002)	**0.284** (0.117)

[1]Expected number of unique multilocus genotypes at the lowest common sample size that is greater than or equal to 10 (i.e. as if all population sample sizes had been "10" here), based on rarefaction,

[2]Shannon-Wiener Index of MLG diversity

[3]*P*-value < 0.05 is designated by *, *P*-value < 0.001 is designated by *

loci to distinguish clones and providing evidence that there was a low chance of getting the same genetic profile by chance. From the P_{gen} values, P_{sex} was calculated for each unique MLG. Over all 15 populations, there were 29 significant P_{sex} values (Appendix A Table 4). Significant P_{sex} values suggest that individuals with these MLGs were derived by clonal reproduction. MD and TN-3 were the only populations for which clonal reproduction could be ruled out according to the P_{sex} calculations. All three apomictic populations (IL-1, IL-2, and OH-1) and two sexual populations (NC-2 and PA) each had one dominant MLG that was possessed by the majority of individuals in their respective populations (Appendix A Table 4). It should be noted that P_{gen} and P_{sex} use MLGs to determine how likely clonal reproduction is, and while repeated MLGs are often a sign of clonal reproduction, it is not definitive proof of clonal reproduction.

Diversity statistics that are appropriate even for clonal populations are reported in Table 3, including the number of unique multilocus genotypes (MLGs). From a sample of 300 individuals, 153 unique multilocus genotypes were identified across all populations using a clonal threshold of 0. The number of unique multilocus genotypes expected per population (*eMLG*) if standardized for a population sample size of 10 ranged from 1.42 to 10.00 for the presumed sexual populations with 2 – 33 MLGs actually observed, and from 1.00 to 2.56 for the presumed apomictic populations with 1 – 5 actually observed. All TN populations had very few repeated MLGs, with population TN-3 having none repeated and TN-1 and TN-2 having only one repeated. Note that although sample size is small for TN-2 and TN-3, TN-1 has the largest sample size (34) of all the *B. laevigata* populations sampled. The number of unique multilocus genotypes in the sexual populations of AL-1, AR, KY, MD, NC-1, and OH-2 were more than half the sample size of each population. The remaining three sexual and all three of the apomictic populations had very few unique MLGs, considering their respective sample sizes. The fewest

unique MLGs were found in the apomictic OH-1 population, which had a single MLG shared by all 16 plants sampled. Despite being presumably sexual, populations NC-2 and PA each had only two unique MLGs, with just one individual in each population having a different MLG from the others.

The Shannon diversity index, Simpson's index, and the evenness for populations are also reported in Table 3. For the sexual populations, values of the Shannon diversity index (*H*) ranged from 0.17 to 3.49 and values for Simpson's index (*D*) ranged from 0.080 to 0.969, with the lowest diversity found in populations NC-2 and PA for both indices. Population TN-1 exhibited the highest degree of diversity (H = 3.49, D = 0.969), for both indices followed by, AR (H = 2.91, D = 0.936). Evenness among the sexual populations was highest in TN-3 and MD, reflecting the absence of repeated MLGs in these populations, and evenness was also very high for TN-1. As was consistent with the diversity indices, evenness was lowest in populations NC-2 and PA, each of which had one dominant MLG. For the apomictic populations, values for the Shannon index ranged from 0.00 to 0.67 and Simpson's index ranged from 0.00 to 0.298. For both indices, OH-1 exhibited no diversity while population IL-1 had the greatest diversity of the apomictic populations. Evenness was low among the apomictic populations due to the low number of MLGs and the dominance of a single MLG in each population.

Standard Genetic Diversity Estimates

Table 4 lists population level averages for various standard genetic diversity measures using the full data set for sexual populations and a unique MLG only data set for the three apomictic populations. Population level polymorphism ranged from 6.3-93.8%, with presumed sexual populations accounting for both the lowest and the highest % polymorphism and

Table 4. Genetic variability averaged across 16 polymorphic loci for 15 *Boechera laevigata* populations, using a full data set for sexual populations and a unique MLG only data set for apomictic populations. Standard error is given in parentheses. Mean sample size (*N*), mean number of alleles per locus (*A*), number of private alleles (A_p), percentage of polymorphic loci (*P*), mean direct count of heterozygosity (H_o), Nei (1978) unbiased estimate of mean expected heterozygosity (H_e).

Population	***N***	**MLG**	***A***	A_P	***P***	H_o	H_e
Sexual populations (using a full data set)							
AL-1	30	20	2.5 (0.2)	6	93.8	0.040 (0.008)	0.435 (0.045)
AL-2	30	8	1.3 (0.1)	1	31.3	0.015 (0.011)	0.056 (0.033)
AR	27	21	1.6 (0.3)	3	31.3	0.037 (0.019)	0.108 (0.055)
KY	10	7	1.3 (0.2)	3	25.0	0.025 (0.019)	0.089 (0.041)
MD	13	13	2.0 (0.5)	8	37.5	0.048 (0.023)	0.185 (0.073)
NC-1	23	16	1.9 (0.3)	3	43.8	0.033 (0.014)	0.161 (0.052)
NC-2	12	2	1.1 (0.1)	0	6.3	0.000 (0.000)	0.009 (0.009)
OH-2	6	3	1.2 (0.1)	0	18.8	0.063 (0.063)	0.083 (0.046)
PA	24	2	1.1 (0.2)	0	12.5	0.000 (0.000)	0.010 (0.010)
TN-1	34	33	4.1 (0.7)	18	81.3	0.228 (0.035)	0.482 (0.065)
TN-2	10	9	1.7 (0.2)	4	50.0	0.100 (0.041)	0.191 (0.057)
TN-3	10	10	1.9 (0.3)	5	50.0	0.031 (0.018)	0.216 (0.067)
Sexual means	**19.1** (0.7)	**12.0** (0.6)	**1.8** (0.1)	**4.3** (2.0)	**40.1** (7.6)	**0.052** (0.009)	**0.169** (0.018)
Apomictic populations (using a unique MLG only data set)							
IL-1	30	5	2.1 (0.24)	4	81.3	0.788 (0.099)	0.472 (0.060)
IL-2	25	3	1.7 (0.18)	1	56.3	0.500 (0.129)	0.350 (0.081)
OH-1	16	1	1.8 (0.11)	0	75.0	0.750 (0.112)	0.750 (0.075)
Apomictic means	**23.7** (0.9)	**3.3** (0.2)	**1.9** (0.1)	**1.7** (1.2)	**70.8** (7.5)	**0.679** (0.067)	**0.524** (0.055)

averaging about 40%. All three presumed apomictic populations were highly polymorphic, averaging about 71%. Sexual populations also had the highest and lowest population averages for number of alleles per locus (a range of 1.1– 4.1) as compared to apomictic populations (1.7–2.1), although the averages across the two groups of populations were similar. Apomictic populations were all highly heterozygous, with the average observed heterozygosity for apomictic populations (0.679) exceeding the average for sexual populations (0.052) by an order of magnitude. For individual apomictic populations, observed heterozygosity ranged from 0.500 to 0.788, as compared to a range of 0 to 0.228 for sexual populations. Expected heterozygosity values overlapped for the two groups of populations but were generally higher for the apomictic populations. Observed heterozygosity levels were less than expected under Hardy-Weinberg equilibrium for all sexual populations, resulting in a high inbreeding coefficient for the sexual populations (F_{IS} = 0.709, t (15) = 25.321, P < 0.001). In contrast, apomictic populations had fixed or nearly fixed heterozygosity at polymorphic loci, resulting in a very negative F_{IS} value (F_{IS} = -0.790, t(13) = 10.676, P < 0.001) due to observed heterozygosities being much higher than expected heterozygosities).

Although all loci assayed were polymorphic at the species level for *B. laevigata*, not all loci were polymorphic within each population (Table 4). Among sexual populations, AL-1 exhibited the highest level of polymorphism (93.8%), followed by TN-1 (81.3%). Both populations had very few loci that were fixed for only one allele (1 and 3 loci, respectively), and they had the highest average number of alleles per locus (2.5 and 4.1 alleles per locus, respectively). All other sexual populations had polymorphism levels at or below 50% and showed allele fixation at more than half of the loci, with correspondingly low average numbers

of alleles per locus. NC-2 had the lowest percentage of polymorphic loci at 6.3%, being fixed at all loci but one, followed by PA with 12.5% polymorphic loci, being fixed at all but two loci.

Excluding NC-2, OH-2, PA, and OH-1, all other populations had at least one private allele (Table 4, Appendix A Table 5). TN-1 had by far the most private alleles, with 18 private alleles across 11 loci. Most private alleles in populations were found at two loci – F03 and G06 – for which there were 12 private alleles across six populations and 14 private alleles across eight populations, respectively.

Among the sexual populations, both observed and expected heterozygosity were highest in the TN-1 population (H_o = 0.228, H_e = 0.482, Table 4). TN-2 had the second highest level of observed heterozygosity whereas AL-1 had the second highest expected heterozygosity. Most other sexual populations had observed heterozygosity values less than 0.050, and no heterozygotes were observed in the NC-2 or PA populations.

Although each of the apomictic populations exhibited polymorphism at half or more of the analyzed loci, the levels of polymorphism did not exceed that of the AL-1 and TN-1 populations. All apomictic populations displayed fixed heterozygosity as well as allele fixation at multiple loci. IL-1 displayed fixed heterozygosity at eight loci and allele fixation at three loci. IL-2 displayed fixed heterozygosity at six loci and allele fixation at seven loci, and OH-1 exhibited fixed heterozygosity at 12 loci and allele fixation at four loci.

Statistical comparisons of some of these genetic diversity measures using FSTAT found allelic richness (P = 0.043), observed heterozygosity (P = 0.005), and F_{IS} (P = 0.005) to be significantly higher in the apomictic populations relative to the sexual ones (Table 5). Gene diversity (expected heterozygosity) was the only diversity measurement not found to differ significantly (P = 0.220) between the differing reproductive modes.

Table 5. Statistical comparison of genetic diversity statistics averaged across 16 loci for 12 sexual and 3 apomictic *Boechera laevigata* populations, as calculated and tested via FSTAT. Values were calculated using a full data set for the sexual populations and a unique MLG only data set for the apomictic populations.

Statistic	**Sexual populations**	**Apomictic populations**	***P*-value**
Allelic richness (R_S)[1]	1.169	1.524	0.043
Observed heterozygosity (H_O)	0.060	0.694	0.005
Gene diversity (H_S)[2]	0.206	0.387	0.220
Inbreeding coefficient (F_{IS})	0.709	-0.792	0.005
Fixation index (F_{ST})	0.693	0.321	0.062

[1]FSTAT utilizes a rarefaction method that calculates the expected number of alleles per locus at the smallest number of sampled individuals at a locus)
[2]Nei's (1987) gene diversity, which is equivalent to Nei's (1978) expected heterozygosity (H_e)

As mentioned previously, it is appropriate to derive these standard population genetic diversity estimates from the full data set (all samples) for sexual populations but to utilize only unique multilocus genotypes for populations that are reproducing clonally (the apomictic populations), and Table 4 reports the results of these analyses. This study's categorization of "sexual" and "apomictic" populations is based on the genetic patterns observed in the data as well as information from other *Boechera* researchers. However, since cytological analyses of seeds have not been carried out to directly prove sexual reproduction vs. apomixis, Appendix A Table 6 also shows results that were calculated the opposite way (unique MLG only data utilized for sexual populations, full data set for apomictic populations). Thus, when considering only unique MLGs for sexual populations, observed and expected heterozygosity levels were increased relative to the full data set but overall patterns remained unchanged, and although the inbreeding coefficient decreased somewhat (F_{IS} = 0.677, $t(15)$ = 24.179, P < 0.001), the value was still quite high. Using the full data set for the apomictic populations had mixed effects on the population-level heterozygosities, with little or no effect on observed heterozygosity (H_o only increased somewhat for IL-1) but a decrease in expected heterozygosity across all three

populations, resulting in an even more negative inbreeding coefficient (F_{IS} = -0.965, $t(13) = 64.3$, $P < 0.001$).

Population genetic structure

Figure 7 shows that among the sexual populations, 69% of genetic variation could be attributed to populations being different from one another (F_{ST} = 0.693, estimated from AMOVA, $P < 0.001$) using the full data set. Among-individual variation accounted for 22% of the genetic variation while within-individual variation accounted for 9% of the total genetic variation. The three apomictic populations displayed divergent results. Differences between apomictic populations when using a unique MLG only data set accounted for only 21% of the total genetic variation (F_{ST} = 0.321, estimated from AMOVA, $P = 0.010$), while within individual variation accounted for the rest (79%) of the observed variation. There was virtually no among-individual variation in the apomictic populations, in agreement with the low number of unique multilocus genotypes in these populations. Most of the genetic diversity in these populations is harbored in an individual's genome, as reflected in the exceedingly high heterozygosity levels for these three populations. Despite the seemingly large differences in F_{ST} values between sexual and apomictic populations, they were not statistically different, although values did approach significance ($P = 0.062$, Table 5). Meirmans' (2006) standardized F'_{ST} shows the same trends as F_{ST}, with higher population differentiation for sexual than apomictic populations (0.871 vs. 0.521, respectively).

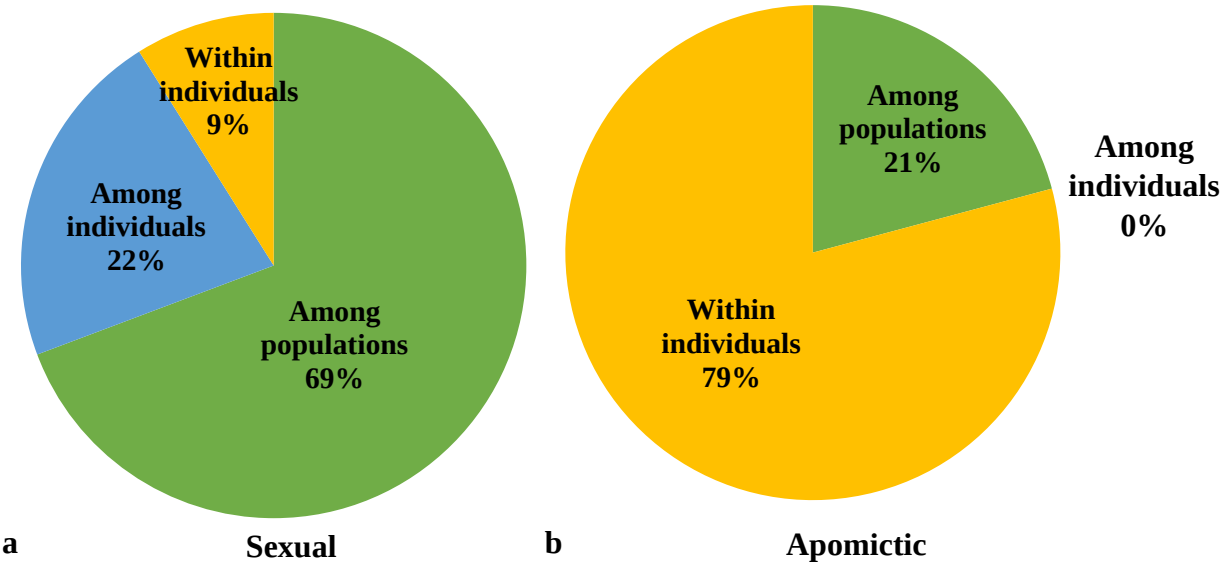


Figure 7. AMOVA results for (a) 12 sexual *Boechera laevigata* populations using a full data set and (b) three apomictic populations using a unique MLG only data set.

Appendix B Figure 2 shows the AMOVA results when using the opposite data set for each group (unique MLG only data set for the sexual populations and the full data set for the apomictic populations), although this analysis would not be appropriate if the group designations of "sexual" and "apomictic" are correct. Although specific values shift (e.g. decreases in F_{ST} and F'_{ST} for sexual populations and increases in both values for apomictic populations), the overall trends are the same when comparing sexual and apomictic populations for these AMOVA results and for Meirman's F'_{ST} (sexual $F'_{ST} = 0.781$, apomictic $F'_{ST} = 0.586$).

Estimates of Nei's (1978) unbiased genetic identity displayed a substantial degree of genetic differentiation among most of the sampled populations, with no relationship to geographic proximity. Table 6 shows genetic identities and genetic distances using the full data set for sexual populations and the unique MLG only data set for apomictic populations. For example, despite being separated by roughly 6 km, populations AL-1 and AL-2 shared a low

proportion of alleles, reflected by their low genetic identity value of 0.355 (genetic distance 1.035). In contrast, AL-1 shared the highest proportion of alleles with populations from distant states: OH-2 (genetic identity 0.732, genetic distance 0.312) and NC-1 (genetic identity 0.714, genetic distance 0.337). Population AL-2, NC-2, and PA had consistently low genetic identities with all other populations. Furthermore, the apomictic IL-1 and OH-1populations displayed a high degree of genetic similarity (genetic identity 0.961, genetic distance 0.040), although the same degree of genetic similarity was not seen between population IL-2 and the other apomictic populations. Genetic identity and distance patterns were consistent when using a full data set for all populations and a unique MLG only data set for all populations. These results are provided in Appendix A Table 7 and Table 8, respectively.

Pairwise F_{ST} estimates (Table 7) and F'_{ST} estimates (Table 8) reflected the same similarity patterns as Nei's genetic identity. For most populations, pairwise F_{ST} and F'_{ST} values were quite high, indicating a significant amount of population differentiation. Patterns are similar when calculating pairwise F_{ST} and F'_{ST} estimates for a full data set and unique MLG only data set for all populations, for which results are provided in Appendix A Table 9 through Table 12.

Many of the same trends uncovered by Nei's (1978) unbiased genetic identity and pairwise F_{ST} values were reflected in the principal coordinate analysis (PCoA) (Figure 8). Together, the first two coordinates explained about 55% of genetic variation. Data points for individuals within populations AL-1 and TN-1 displayed the greatest spread, reflecting the high amount of genetic variation observed within these populations. The lack of genetic diversity within the NC-2 and PA populations is apparent by the highly overlapping symbols that result in these populations appearing to have few individuals. Such limited within-population variation is also seen among the apomictic populations (OH-1, IL-1, IL-2). PCoA results showed some

Table 6. Pairwise Nei's unbiased genetic identity (below the diagonal) and Nei's unbiased genetic distance (above the diagonal) for *Boechera laevigata* populations using a full data set for sexual populations and a unique MLG only data set for apomictic populations. IL-1, IL-2, and OH-1 are considered to be apomictic populations, with all other populations considered to be sexually reproducing.

Population	**AL-1**	**AL-2**	**AR**	**KY**	**MD**	**NC-1**	**NC-2**	**OH-2**	**PA**	**TN-1**	**TN-2**	**TN-3**	**IL-1**	**IL-2**	**OH-1**
AL-1	-	1.035	0.942	0.758	0.460	0.337	0.685	0.312	0.838	0.896	0.852	0.682	0.616	0.613	0.670
AL-2	0.355	-	0.904	0.977	1.516	1.506	1.410	1.592	1.398	0.655	1.057	0.995	0.615	0.967	0.709
AR	0.390	0.405	-	0.469	0.811	0.748	1.610	0.965	1.580	0.422	0.502	0.397	0.761	0.302	0.659
KY	0.469	0.376	0.626	-	0.865	0.736	1.691	1.015	1.691	0.451	0.338	0.259	0.569	0.479	0.587
MD	0.631	0.220	0.444	0.421	-	0.370	0.760	0.215	0.804	1.019	0.894	0.778	0.628	0.322	0.532
NC-1	0.714	0.222	0.473	0.479	0.691	-	0.523	0.298	0.993	0.993	0.717	0.653	0.697	0.461	0.587
NC-2	0.504	0.244	0.200	0.184	0.468	0.593	-	0.471	0.476	1.482	1.671	1.329	0.733	0.915	0.636
OH-2	0.732	0.203	0.381	0.363	0.806	0.742	0.624	-	0.670	1.159	0.994	0.927	0.652	0.470	0.563
PA	0.433	0.247	0.206	0.184	0.448	0.370	0.621	0.512	-	1.490	1.787	1.517	0.425	0.660	0.228
TN-1	0.408	0.519	0.655	0.637	0.361	0.370	0.227	0.314	0.225	-	0.482	0.375	0.554	0.554	0.393
TN-2	0.426	0.348	0.605	0.713	0.409	0.488	0.188	0.370	0.167	0.617	-	0.246	0.752	0.535	0.483
TN-3	0.506	0.370	0.672	0.772	0.459	0.521	0.265	0.396	0.219	0.687	0.782	-	0.661	0.414	0.568
IL-1	0.540	0.541	0.467	0.566	0.534	0.498	0.481	0.521	0.654	0.575	0.471	0.516	-	0.376	0.040
IL-2	0.542	0.380	0.739	0.619	0.725	0.631	0.400	0.625	0.517	0.575	0.586	0.661	0.686	-	0.277
OH-1	0.512	0.492	0.517	0.556	0.588	0.556	0.529	0.569	0.796	0.675	0.617	0.566	0.961	0.758	-

Table 7. Pairwise comparison of F_{ST} values for *Boechera laevigata* populations using a full data set for sexual populations and a unique MLG only data set for apomictic populations. Significance values are given above the diagonal. IL-1, IL-2, and OH-1 are considered to be apomictic populations, with all other populations considered to be sexually reproducing.

Population	**AL-1**	**AL-2**	**AR**	**KY**	**MD**	**NC-1**	**NC-2**	**OH-2**	**PA**	**TN-1**	**TN-2**	**TN-3**	**IL-1**	**IL-2**	**OH-1**
AL-1	-	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.003	0.006
AL-2	0.663	-	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
AR	0.611	0.870	-	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
KY	0.517	0.898	0.761	-	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
MD	0.403	0.877	0.777	0.768	-	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
NC-1	0.388	0.870	0.771	0.758	0.592	-	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
NC-2	0.545	0.943	0.904	0.942	0.821	0.766	-	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.001
OH-2	0.321	0.923	0.840	0.865	0.499	0.593	0.912	-	0.001	0.001	0.001	0.001	0.001	0.001	0.001
PA	0.641	0.953	0.922	0.958	0.873	0.870	0.973	0.950	-	0.001	0.001	0.001	0.001	0.001	0.001
TN-1	0.404	0.562	0.444	0.396	0.510	0.546	0.613	0.524	0.670	-	0.001	0.001	0.001	0.001	0.001
TN-2	0.499	0.864	0.717	0.627	0.710	0.706	0.885	0.774	0.922	0.373	-	0.001	0.001	0.001	0.001
TN-3	0.452	0.849	0.663	0.541	0.673	0.678	0.858	0.741	0.907	0.318	0.440	-	0.001	0.001	0.001
IL-1	0.343	0.771	0.715	0.623	0.557	0.627	0.764	0.608	0.789	0.303	0.572	0.524	-	0.009	0.065
IL-2	0.352	0.867	0.630	0.688	0.486	0.606	0.885	0.665	0.911	0.314	0.577	0.487	0.370	-	0.027
OH-1	0.326	0.848	0.749	0.720	0.575	0.642	0.873	0.697	0.863	0.217	0.551	0.530	0.185	0.388	-

Table 8. Pairwise comparison of F'_{ST} values for *Boechera laevigata* populations using a full data set for sexual populations and a unique MLG only data set for apomictic populations. IL-1, IL-2, and OH-1 are considered to be apomictic populations, with all other populations considered to be sexually reproducing.

Population	**AL-1**	**AL-2**	**AR**	**KY**	**MD**	**NC-1**	**NC-2**	**OH-2**	**PA**	**TN-1**	**TN-2**	**TN-3**	**IL-1**	**IL-2**	**OH-1**
AL-1	-														
AL-2	0.884	-													
AR	0.851	0.947	-												
KY	0.775	0.962	0.850	-											
MD	0.623	0.973	0.901	0.902	-										
NC-1	0.567	0.971	0.892	0.884	0.717	-									
NC-2	0.775	0.986	0.981	0.989	0.916	0.861	-								
OH-2	0.497	0.984	0.939	0.951	0.593	0.695	0.945	-							
PA	0.850	0.988	0.984	0.992	0.944	0.952	0.983	0.975	-						
TN-1	0.755	0.781	0.647	0.623	0.825	0.835	0.915	0.851	0.929	-					
TN-2	0.785	0.953	0.829	0.733	0.881	0.856	0.979	0.916	0.987	0.615	-				
TN-3	0.720	0.945	0.773	0.644	0.849	0.832	0.963	0.896	0.980	0.532	0.558	-			
IL-1	0.634	0.884	0.874	0.799	0.772	0.822	0.895	0.808	0.872	0.591	0.806	0.761	-		
IL-2	0.625	0.952	0.734	0.813	0.631	0.755	0.956	0.798	0.956	0.586	0.756	0.660	0.597	-	
OH-1	0.626	0.933	0.882	0.859	0.765	0.815	0.941	0.846	0.905	0.439	0.742	0.745	0.317	0.593	-

clustering of populations, that are geographically closer, particularly the TN and KY populations, but this pattern was not consistent. Other same-state populations, like those in AL, NC, or IL did not cluster together despite their relatively close geographic proximity. Thus, geographic distance does not appear to be a consistent factor in determining genetic differentiation among populations. While Figure 8 shows results when using the full data set for sexual populations and the unique MLG only data set for the apomictic populations, whether using a full data set (Appendix B Figure 3) or unique MLG only data set for all populations (analysis not shown), population clustering was consistent, with minimal changes to the amount of variation explained by each coordinate.

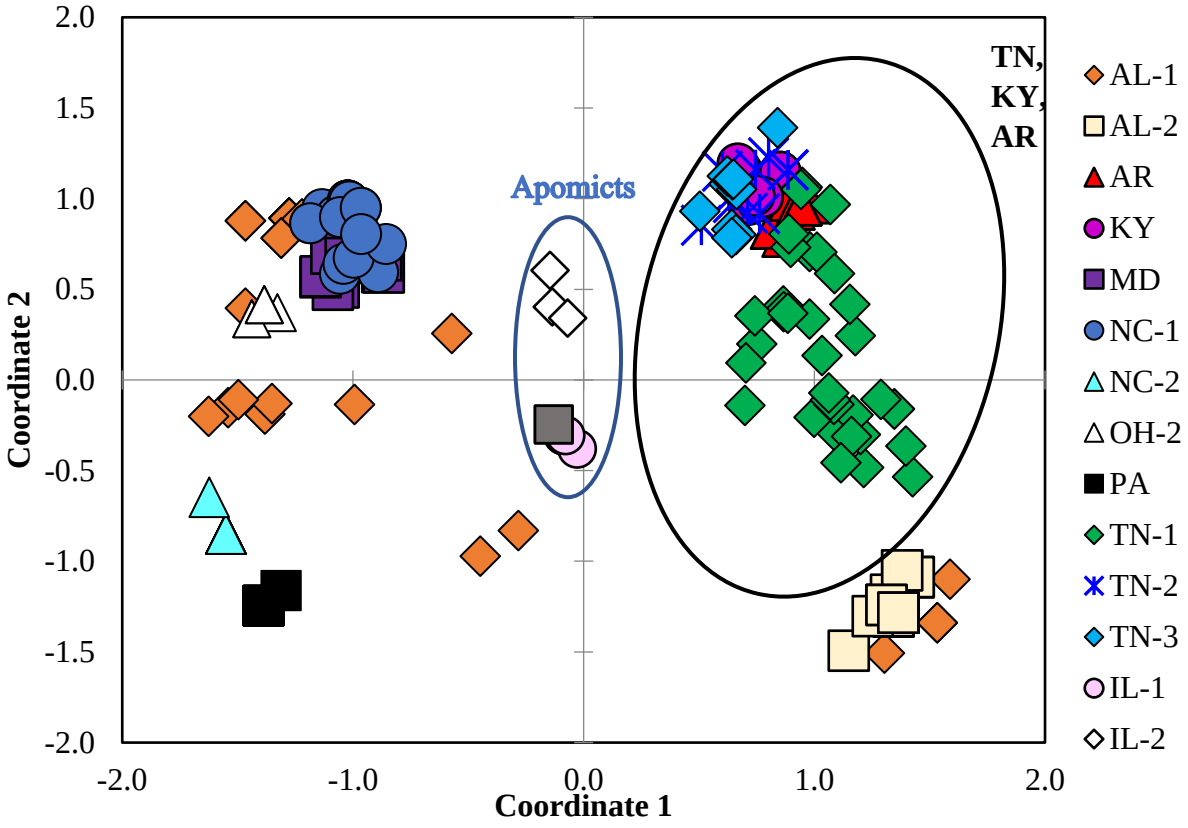


Figure 8. Principal coordinate analysis via distance matrix with data standardization from 15 sampled *Boechera laevigata* populations using a full data set for sexual populations and a unique MLG only data set for apomictic populations. Coordinate 1 explains 33.2% of genetic variation and coordinate 2 explains 22.2% of genetic variation (55.4% total). Encircled groupings do not indicate assigned genetic clusters.

As shown in Appendix B Figure 4, from the discriminant analysis of principal components (DAPC), 13 clusters were chosen from *K*-means clustering as this number of clusters was associated with the greatest average silhouette width. Following cross validation methods, 22 PCs were retained as this was associated with lowest root mean square error, and three discriminant functions were retained as these explained most of the variance in the data.

Clustering results from the DAPC using a full data set for the sexual populations and a unique MLG only data set for the apomictic populations (Figure 9) were similar to those of the PCoA, although with some differences. The TN-1 population was divided into three different genetic clusters, with two individuals joining the genetic cluster formed by all three apomictic populations. Although pairwise genetic identity values were high between TN-2 and TN-3 and these populations did group closely in the PCoA, these were assigned to separate (but spatially close) genetic clusters as TN-3 formed its own cluster and TN-2 clustered with KY. The AL-1 population was divided into three separate clusters, which is consistent with results from the PCoA. Most striking was the separate clustering of AR despite it being closely grouped with the TN and KY populations in the PCoA. Furthermore, the AR population had the highest genetic identity values with AL-1 and AL-2. Thus, AR's genetic similarity to the other sampled populations is inconsistent across three analyses. The PA, NC-1, and NC-2 populations each formed a distinct genetic cluster. Overall, positioning for the MD, NC-1, NC-2, and PA populations were comparable between the two clustering analyses. An analysis for the DAPC using a full data set for all populations was also carried out (Appendix B Figure 6), which showed slightly different results from those presented here. The number of optimal genetic clusters differed (K = 12 when using the full data set for all populations), and AR clustered with the TN and KY populations as was the case for PCoA.

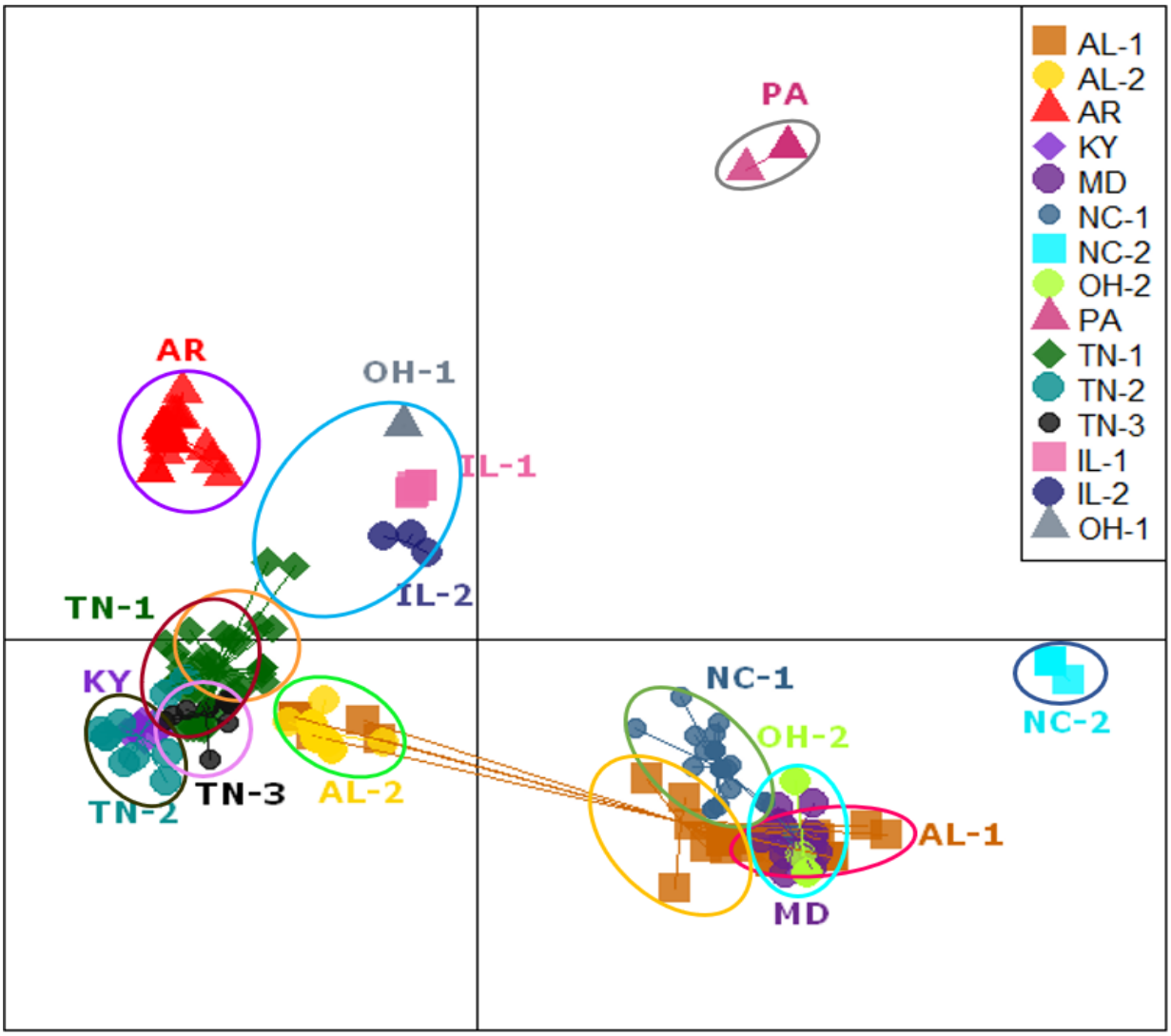


Figure 9. Discriminant analysis of principal components for 15 *Boechera laevigata* populations using a full data set for sexual populations and a unique MLG only data set for apomictic populations. Each encircled group represents a genetic cluster as indicated through *K* means clustering (*K* = 13). A total of 22 principal components and 3 discriminant functions were retained.

Comparing the genetic diversity of *Boechera laevigata* with the rare *B. perstellata*

One goal of this study was to compare the widespread *B. laevigata* with its rare congener *B. perstellata*. Since *B. perstellata* is thought to reproduce sexually, this comparison included only the sexually reproducing *B. laevigata* populations. Baskauf et al. (2014) reported

population-level averages for the rare *B. perstellata* of 21.1% polymorphic loci, 1.3 alleles per locus, observed heterozygosity of 0.002, and expected heterozygosity of 0.059, based on 19 microsatellite loci. In comparison, genetic diversity levels for the sexual *B. laevigata* populations across 16 polymorphic loci averaged 40.1% polymorphic loci, 1.8 alleles per locus, observed heterozygosity of 0.051, and expected heterozygosity of 0.169). However, these genetic diversity estimates for each species are based on an overlapping but not identical set of microsatellites.

To provide a more accurate comparison, the two species were compared using only loci that were assayed in both species. When considering the 11 loci assayed for both species (one of which was monomorphic for *B. perstellata*), genetic diversity levels were once again higher in the sexual *B. laevigata* populations than in *B. perstellata* (Table 9). Examining population-level averages, while the number of alleles per locus was somewhat greater in *B. laevigata* (1.6) compared to *B. perstellata* (1.2), the percentage of polymorphic loci observed in *B. laevigata* populations (31.1%) was almost double that observed in *B. perstellata* (15.6%). The number of private alleles per locus was also greater in *B. laevigata* (2.8 private alleles in *B. laevigata* vs.1.7 private alleles in *B. perstellata*). Furthermore, observed heterozygosity was almost 40 times greater among *B. laevigata* populations ($H_o = 0.040$) than for *B. perstellata* ($H_o = 0.001$), with expected heterozygosity levels for *B. laevigata* ($H_e = 0.128$) exceeding *B. perstellata* ($H_e = 0.021$) by a little less than 5-fold. Despite the large differences between the two species at these 11 analyzed loci, when statistical tests were carried out comparing the species, allelic richness (R_S), observed heterozygosity, and expected heterozygosity (gene diversity) did not differ significantly, although values were approaching significance (Table 10, $P = 0.074$ for R_S, $P = 0.097$ for H_O, $P = 0.079$ for H_S). Because loci that were not assayed for both species were

excluded, fewer loci were used in this comparison than were tested for each species. The limitation in the number of analyzed loci may have contributed to the lack of significance, thus a larger number of loci could increase the power of such an analysis. Furthermore, sexual *B. laevigata* populations were quite variable with some populations having relatively high diversity while others had almost no diversity. The high degree of variance among these populations may have also contributed to the lack of significance for the statistical tests.

The Tennessee and Kentucky populations of the rare species are geographically quite far apart (about 250 km, Baskauf et al. 2014), although distances are much closer for populations within each state. Distances between *B. perstellata* populations in Tennessee ranged from approximately 6 to 32 km, and distances between the Kentucky *B. perstellata* populations ranged from 11 to 32 km. In this study, sampled *B. laevigata* populations were generally farther apart because the species has a much larger geographic range; however, sampling of the widespread species in Tennessee and Alabama was conducted on a scale roughly comparable to the "within-state" sampling of the rare species. Regarding the Tennessee populations of *B. laevigata*, TN-1 and TN-2 were separated by approximately 24 km, TN-1 and TN-3 by approximately 27 km, and TN-2 and TN-3 by approximately 40 km. The two Alabama populations were separated by about 6 km. Examination of values in Table 9 indicates that even when evaluated at a similar spatial scale, the *B. laevigata* populations in Alabama and Tennessee displayed greater levels of polymorphism, number of alleles per locus, and observed and expected heterozygosity than did the *B. perstellata* populations (Table 9).

Table 9. Genetic variability averaged across 11 loci assayed for seven *Boechera perstellata* and 12 "sexual" (only) *B. laevigata* populations. Values for *B. perstellata* are from data collected by Baskauf et al. (2014). Standard error is given in parentheses. Mean sample size (*N*), mean number of alleles per locus (*A*), number of private alleles (A_p), percentage of polymorphic loci (*P*), mean direct count of heterozygosity (H_o), Nei's (1978) unbiased estimate of mean expected heterozygosity (H_e).

Population	***N***	***A***	A_P	***P***	H_o	H_e
Boechera perstellata						
TN-I	33.8 (0.4)	1.3 (0.1)	2	27.27	0.000 (0.000)	0.020 (0.011)
TN-V	31.6 (0.2)	1.3 (0.1)	3	27.27	0.003 (0.003)	0.069 (0.039)
TN-G	32.9 (0.1)	1.0 (0.0)	4	0.00	0.000 (0.000)	0.000 (0.000)
TN-CK	31.9 (0.1)	1.3 (0.1)	2	27.27	0.000 (0.000)	0.027 (0.016)
Mean for TN	**32.6**	**1.2**	**3**	**20.45**	**0.001**	**0.029**
KY-R	30.9 (0.1)	1.1 (0.1)	0	9.09	0.000 (0.000)	0.011 (0.011)
KY-C	26.9 (0.1)	1.1 (0.1)	0	9.09	0.003 (0.003)	0.040 (0.040)
KY-H	14.7 (0.2)	1.1 (0.1)	1	9.09	0.000 (0.000)	0.012 (0.012)
Mean for KY	**24.2**	**1.1**	**0.3**	**9.09**	**0.001**	**0.021**
Mean for *B. perstellata*	**29.0** (0.7)	**1.2** (0.0)	**1.7** (0.6)	**15.58** (4.31)	**0.001** (0.001)	**0.026** (0.009)
Boechera laevigata						
AL-1	30.0 (0.0)	2.3 (0.2)	4	90.91	0.045 (0.010)	0.416 (0.055)
AL-2	30.0 (0.0)	1.1 (0.1)	1	9.09	0.000 (0.000)	0.006 (0.006)
Mean for AL	**30.0**	**1.7**	**2.5**	**50.00**	**0.023**	**0.211**
AR	27.0 (0.0)	1.2 (0.1)	0	18.18	0.010 (0.010)	0.022 (0.016)
KY	10.0 (0.0)	1.3 (0.1)	0	27.27	0.009 (0.009)	0.105 (0.055)
MD	13.0 (0.0)	1.7 (0.5)	6	27.27	0.028 (0.021)	0.155 (0.088)
NC-1	23.0 (0.0)	1.5 (0.4)	2	27.27	0.020 (0.011)	0.113 (0.064)
NC-2	12.0 (0.0)	1.0 (0.0)	0	0.00	0.000 (0.000)	0.000 (0.000)
OH-2	6.0 (0.0)	1.1 (0.1)	1	9.09	0.091 (0.091)	0.050 (0.050)
PA	24.0 (0.0)	1.1 (0.1)	2	9.09	0.000 (0.000)	0.007 (0.007)
TN-1	34.0 (0.0)	3.5 (1.0)	14	72.73	0.195 (0.045)	0.398 (0.081)
TN-2	10.0 (0.0)	1.4 (0.2)	2	36.36	0.055 (0.031)	0.102 (0.051)
TN-3	10.0 (0.0)	1.6 (0.2)	2	45.45	0.027 (0.019)	0.161 (0.071)
Mean for TN	**18.1**	**2.2**	**6**	**51.52**	**0.093**	**0.220**
Mean for *B. laevigata*	**19.1** (0.8)	**1.6** (0.1)	**2.8** (1.5)	**31.06** (7.87)	**0.040** (0.010)	**0.128** (0.019)

Table 10. Statistical comparison of genetic diversity statistics averaged across 11 loci for 7 *Boechera perstellata* and 12 sexual *B. laevigata* populations, as calculated and tested via FSTAT.

Statistic	***B. perstellata***	***B. laevigata***	***P*-value**
Allelic richness (R_S)[1]	1.095	1.424	0.074
Observed heterozygosity (H_O)	0.001	0.046	0.097
Gene diversity (H_S)[2]	0.027	0.159	0.079
Inbreeding coefficient (F_{IS})	0.967	0.711	0.186
Fixation index (F_{ST})	0.956	0.728	0.021

[1]FSTAT utilizes a rarefaction method that calculates the expected number of alleles per locus at the smallest number of sampled individuals at a locus)
[2]Nei's (1987) gene diversity, which is equivalent to Nei's (1978) expected heterozygosity (H_e)

F-statistics based on the 11 in-common loci and estimated from AMOVA indicate that both species exhibited a high degree of genetic population differentiation and potential inbreeding (F_{ST} and F_{IS} significantly different from 0 for both species, $P \leq 0.001$). However, these values were greater in *B. perstellata* ($F_{ST} = 0.924$; $F_{IS} = 0.981$) than in *B. laevigata* ($F_{ST} = 0.728$; $F_{IS} = 0.715$). Although F_{IS} estimates were not significantly different between the two species (Table 10, $P = 0.19$), F_{ST} estimates did differ significantly ($P = 0.02$), indicating that population differentiation was greater among *B. perstellata* populations than *B. laevigata* populations.

CHAPTER IV

Discussion

Clonal diversity

Although there was a total of 300 samples among the 15 populations, only a little more than half of the observed multilocus genotypes were unique. The presumed sexual populations generally had greater genotypic diversity than the apomictic populations, with only two populations (NC-2 and PA) having expected unique MLGs as low as the apomictic populations. The low levels of genotypic diversity seen in the apomictic populations are typical of species that reproduce clonally (Balloux et al. 2003, Halkett et al. 2005). This low level of genotypic diversity among the apomictic populations was also reflected in low diversity values for both the Shannon and Simpson's diversity indices as well as the lowest evenness values. The apomictic IL-1 population had the greatest diversity among the apomictic populations, as indicated by the number of unique MLGs and higher values for both diversity indices. The relatively elevated diversity seen in this population might be attributed to the fact that this area has been part of a restoration project aimed at genetically reconnecting the now fragmented habitat. To overcome the physical barriers now in place due to human development, seeds have been collected along a 19 km stretch bordering the Chicago River and then distributed through the area (E. Faulkner, pers. comm.). Thus, the genetic diversity seen in this population may not be a true reflection of the natural diversity had there been no human interference either by habitat fragmentation or by seed distribution.

Compared to other clonal species, the apomictic *B. laevigata* populations also had lower diversity. Shannon diversity measures for the widespread, invasive kudzu (H = 0.4, Bentley and Mauricio) and the widespread *Fragaria nilgerrensis* (H = 0.8, Lu et al. 2021) were slightly greater t than the average for apomictic *B. laevigata* populations (H = 0.3). Even compared to

endangered clonal plants like *Helianthus verticillatus* (*H* = 2.58, Edwards et al. 2020) or *Betula humilis* (*D* = 0.870, Bona et al. 2019), apomictic *B. laevigata* were less diverse (*D* = 0.086). The lower diversity may be due to a more restricted range among the apomictic populations compared to the other widespread species mentioned here. While the apomictic *B. laevigata* have a larger range compared to the endangered *H. verticillatus*, higher clonal diversity in this endangered species could be a relic of when sexual reproduction was more common in the endangered species, or perhaps sexual reproduction still occurs but very infrequently (Edwards et al. 2020). Although most *Boechera* species are facultative apomicts, sexual reproduction would consist of inbreeding, which would lower genetic diversity. Life form may also have some contribution to the lower clonal diversity compared to *B. humilis*, as this is a shrubby species with a long life span, both characteristics of plants that Hamrick and Godt (1989) found to have high genetic diversity, while *B. laevigata* is an herbaceous, short-lived perennial.

Standard genetic diversity measures

Among sexual *Boechera laevigata* populations, there was very little within-population genetic diversity at the 16 analyzed microsatellite loci. The majority of populations were polymorphic at less than half the loci, and observed heterozygosity levels were low, with heterozygotes completely absent in two populations (NC-2 and PA). In a review of microsatellite studies, Nybom (2004) reported average observed and expected microsatellite heterozygosity levels for a variety of plant life history traits, including widespread angiosperms, short-lived perennials, and selfing species which had expected heterozygosity levels 4 to 6 times higher than the population average for *B. laevigata*. Average observed heterozygosity among sexual *B. laevigata* populations was 10 times lower than that reported for other widespread angiosperms and short-lived perennials, although levels were comparable to that of other selfing plants. Thus,

in comparison to some other plant species with similar life histories, sexual populations of *B. laevigata* have lower levels of genetic diversity.

The low heterozygosity levels seen in the sexual *B. laevigata* populations can likely be attributed to a high rate of self-fertilization. This most extreme form of inbreeding is extremely common among sexually reproducing *Boechera* species (Beck et al. 2012). One consequence of inbreeding is a genome-wide reduction in heterozygosity. A heterozygote deficit was evident in all sampled sexual populations as indicated by the high estimate of F_{IS} (F_{IS} = 0.709).

Many other sexually reproducing diploid *Boechera* species have also been found to have low genetic diversity. Song et al. (2006) reported the widespread *B. stricta* to have even lower observed heterozygosity (Ho = 0.029) and higher levels of inbreeding (F_{IS} = 0.89) than *B. laevigata*. However, the average number of alleles per locus, percent of polymorphic loci, and expected heterozygosity were comparable between *B. stricta* (A = 1.9, H_e = 0.20, P = 41.5%) and the sexual *B. laevigata* populations (A = 1.8, H_e = 0.17, P = 40.1%). Additionally, Alexander et al. (2015) reported some diversity statistics for five sexual diploid *Boechera* species with western ranges. Average observed heterozygosity in *B. laevigata* (0.052) was lower by almost 2- to 5-fold in comparison to *B. fendleri*, *B. gracilipes*, *B. perennans*, and *B. texana* (H_o ranges from 0.098 to 0.269). Only *B. spatifolia* had lower observed heterozygosity estimates (H_o = 0.017). However, all these species had an average number of alleles per locus 2 to 5 times higher (A = 3.3-9.5) than *B. laevigata*. Furthermore, expected heterozygosity estimates were greater by 2 to 3 times in these five *Boechera* species (H_e ranged from 0.300-0.550) compared to *B. laevigata* while F_{IS} estimates were comparable among sexual *B. laevigata* and the other sexual diploid *Boechera* species.

Compared to the sexual populations, the apomictic *B. laevigata* populations had high levels of polymorphism with more than half the analyzed loci being polymorphic and displaying nearly fixed heterozygosity. Among clonally reproducing species, such high levels of heterozygosity are common (Ellstrand and Roose 1987). However, there was low genotypic diversity among individuals in apomictic *B. laevigata* populations due to asexual reproduction. Alexander et al. (2015) reported some diversity statistics for several western diploid apomictic *Boechera* species based on all samples (not a unique MLG only data set), so comparisons with *B. laevigata* use the full data set as well. Observed and expected heterozygosity levels were lower among apomictic *B. laevigata* populations (H_o = 0.686, H_e = 0.356) compared to *B. carrizozoensis*, *B. centrifendleri*, *B. sanluisensis*, and *B. zephyra* (H_o ranged from 0.710 to 0.800, H_e ranged from 0.528 to 0.607). The average number of alleles per locus were almost 2 to 3 times lower among apomictic *B. laevigata* populations (A = 1.9) compared to these other diploid apomictic *Boechera* (A ranged from 3.5 to 6.2). However, estimates of F_{IS} in apomictic *B. laevigata* (F_{IS} = -0.965) are more negative than that reported for the other four apomictic species (F_{IS} ranged from -0.471 to -0.282).

Despite the lack of cytological evidence to confirm that the two *B. laevigata* populations in Illinois and the one in Ohio (OH-1) are apomictic, studies have found high levels of heterozygosity to be associated with apomixis in the genus (Beck et al. 2012). High heterozygosity is often believed to be the product of hybridization events between genetically distinct lineages (Dobes et al. 2006, Beck et al. 2012, Windham et al. 2015), after which the heterozygosity is maintained through apomixis. In addition, unpublished data on seven other *B. laevigata* populations in the northern parts of its range have also been documented to have significant levels of fixed heterozygosity (M. Windham pers. comm.) with one of these

populations confirmed to be apomictic (Carman et al. 2019). The apomictic populations in these regions are believed to have arisen through intraspecific crosses between distant *B. laevigata* lineages (M. Windham, pers. comm.). The location of these populations is presented in Figure 10.

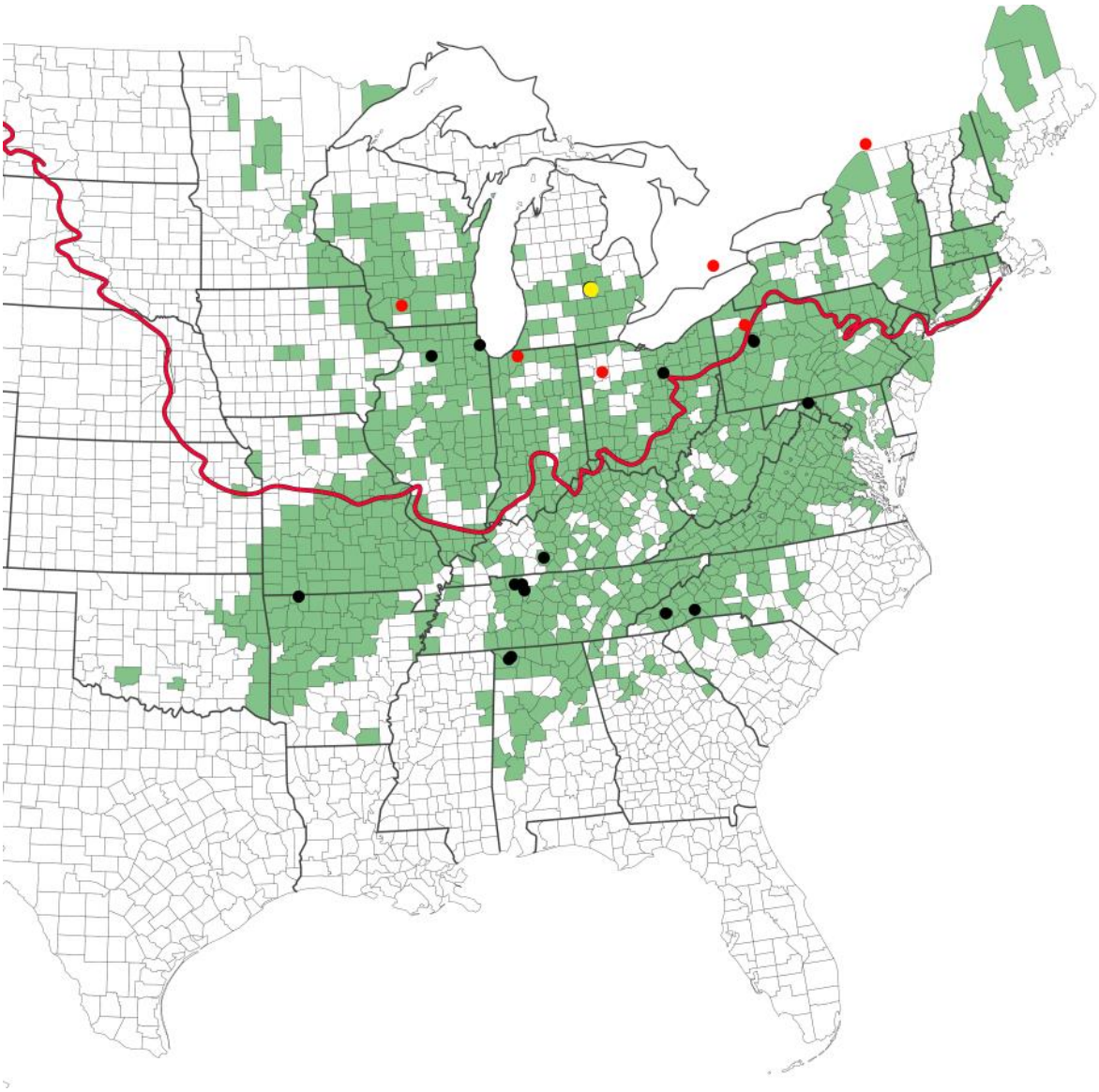


Figure 10. Distribution of *Boechera laevigata* throughout the eastern United States as reported by BONAP (Kartesz 2015). Populations sampled in this study are indicated by black dots. A confirmed apomictic *B. laevigata* population (Carman et al. 2019) is indicated by the yellow dot, and presumed apomictic populations are indicated by red dots (M. Windham, pers. comm.). The extent of the glacial maximum is represented by the red line. The map was created using mapchart.net.

Interestingly, all suspected and confirmed apomictic populations of *B. laevigata* are located within the boundary of the last glacial maximum. Asexual lineages of various species are known to colonize previously glaciated areas (Hörandl 2006, Tilquin and Kokko 2016, Rushworth et al. 2018); so it is often hypothesized that asexual lineages are either better colonizers than sexual lineages due to the benefit of reproductive assurance (since asexual reproduction does not require the presence of a mate or pollinators), their ability to rapidly experience ecological release thus increasing the range of phenotypes and the niche breadth (MacArthur and Wilson 1967, Vrijenhoek and Parker 2009), and/or their ability to outcompete sexual lineages within these previously glaciated areas (Kearney 2005, Hörandl 2006, Tilquin and Kokko 2016). Asexual lineages are often the result of hybridization events, which can lead to a sudden and significant increase in genotypic and phenotypic variation, much more than can happen only through mutation accumulation and recombination in sexual lineages (Anderson and Stebbins 1954).

Reproductive assurance in apomictic lineages is only advantageous if their sexual competitors are self-incompatible (Hörandl 2006). Sexual *B. laevigata* appear to have high levels of selfing, thus eliminating the need for a partner in order to produce offspring. In recently exposed habitats, asexuals may outcompete sexual lineages if they have novel gene combinations that prove more effective in creating new, highly successful phenotypes and well adapted genotypes in these habitats. These phenotypes may be the result of a high degree of plasticity (the “general-purpose genotype” hypothesis) or through the repeated formation of different clonal lineages associated with a variety of phenotypes (the “frozen-niche variation” hypothesis). The general-purpose genotype hypothesis proposes that an asexual lineage can be evolutionarily successful if it has a generalist genotype associated with a highly plastic phenotype that allows it

to occupy a variety of habitats (Baker 1965). The frozen niche-variation hypothesis, states that new apomictic lineages are regularly produced through hybridization events with sexual lineages, and then each apomictic lineages “freezes” the genotypic variation found in each of its sexual parents, producing a range of genotypes and phenotypes (Vrijenhoek 1984, Hörandl 2006). However, not all asexual lineages will prove to be successful as selection will only favor clonal lineages that prove to be highly adapted. The successful gene combinations are then retained through generations due to a lack of recombination, possibly leading to the competitive advantage of asexual lineages over sexual ones, because sexual reproduction can disrupt beneficial allele combinations (Kearney 2005, Tilquin and Kokko 2016). If apomicts have a competitive advantage, it begs the question why asexual populations have not occupied the areas where sexual forms occur. Some have proposed that biotic pressures from pathogens, competitors, and predators in areas inhabited by sexual forms prevent the total dominance of asexual lineages beyond recently exposed areas (Glesener and Tilman 1978, Gibson et al. 2016). There also exists the possibility that sexual lineages are more specially adapted to their habitats than asexuals, giving them an advantage in these already occupied regions and preventing asexual encroachment to these areas (Tilquin and Kokko 2016).

The high within-individual variation seen in the apomictic *B. laevigata* populations may confer a heterozygote advantage in the areas they currently occupy. This may provide a greater degree of phenotypic plasticity that could widen their niche breadth or lead to the occupation of a different niche from their sexual progenitors (Kearney 2005). In their comparison of sexual and asexual *Boechera* species, Rushworth et al. (2018) found asexual lineages to be ecologically differentiated from sexual lineages and have a tendency to occur in areas of greater disturbance and stress. The different environmental pressures experienced by these sexual and asexual

lineages may have led to the ecological differentiation seen. This might be the case for the apomictic and sexual *B. laevigata* populations, although no study has yet investigated whether or not there are ecological differences between the different reproductive forms. Interestingly, the apomictic OH-1 and sexual OH-2 populations are in extremely close proximity to each other (separated by approximately 1.6 km), and their close proximity may provide an excellent study site to exploring the possibility of ecological differentiation. The events that triggered apomixis in these northern *B. laevigata* populations and the reasons for their geographic range are unknown and require further research to understand the evolutionary history behind this reproductive mode shift.

The NC-2 and PA populations had extremely low diversity compared to the other sexual populations, with many of the clonal diversity statistics in Table 5 (number of MLGs, Shannon and Simpson's indices, evenness) being comparable to those of the apomictic populations yet lacking the heterozygosity levels of the apomictic population. The fact that both of these low diversity sexual populations occur in the Appalachian Mountains, raises the question of whether or not stresses or gene flow limitations associated with high elevation have any impact on the genetic diversity of *B. laevigata*. The AR population, which is only about 30 meters lower in elevation than the PA population, did not exhibit the same low diversity of these other two populations, although one might expect less severe conditions at a lower latitude even if the elevation was the similar. Other studies have found reduced genetic diversity at higher altitudes in some plant and animal species (Ohsawa and Ide 2007, Polato et al. 2017). The lack of within population genetic variation in the NC-2 and PA populations may also suggest a historic bottleneck event driven by limited gene flow between the high elevation and the lower elevation

populations that resulted in the almost nonexistent genetic diversity seen today. Further sampling of populations at high elevations may provide more insight.

Population genetic structure

Population structuring was evident for *B. laevigata,* as there were few shared alleles and no shared multilocus genotype across populations. Population differentiation accounted for the greatest share of observed genetic variation for the sexual populations (F_{ST} = 0.687). This high level of population differentiation can likely be attributed to the locally disjunct nature of *B. laevigata*'s rock outcrop habitats, which could make long-distance pollen or seed dispersal difficult. Although no pollination studies have been published for this species, *B. laevigata* has been reported to have gravitational seed dispersal, limiting the ability for long distance gene flow. The level of population differentiation for *B. laevigata* is more than double that reported in Nybom (2004) for other widespread species (F_{ST} = 0.25) and short-lived perennials (F_{ST} = 0.31) but is only slightly higher than population differentiation estimates for selfers (F_{ST} = 0.42). Thus, the high levels of population differentiation among *B. laevigata* populations are not atypical for species with its mode of reproduction. It has been observed (P. Alexander, pers. comm.) that in small populations of selfing *Boechera* lineages, one lineage tends to dominate each population, probably as a result of random genetic drift which would contribute to the high levels of population differentiation.

Among this study's apomictic *B. laevigata* populations, the vast majority of genetic variation was found within individuals, while among population differentiation accounted for less than a quarter of total genetic variation (F_{ST} = 0.21). In a study by Lovell et al. (2014) of *B. spatifolia,* a western species with both sexual and apomictic populations, the apomictic populations were also found to have very weak population structure (F_{ST} = 0.166) compared to

the sexual ones (F_{ST} = 0.360). Because growth chamber and field experiments showed that apomictic *B. spatifolia* populations lacked the adaptive variation to environmental gradients that was evident for sexual populations, Lovell et al. suggested that the low degree of population structure among the apomictic populations may be due to a lack of evolutionary response to directional or purifying selective pressures, which would limit local adaptations. In the case of *B. spatifolia*, the geographic range of the two reproductive types is highly overlapping, so the apomictic populations are exposed to much the same range in environmental variables as the sexual populations. This is in contrast to the situation with *B. laevigata*, for which all the apparently apomictic populations are located at the northern edge of the species' range.

The "general-purpose genotype hypothesis" may explain the low level of population differentiation among apomictic *B. laevigata.* Under this scenario, it could be that a phenotypically plastic clonal lineage of *B. laevigata* has been able to dominate the previously glaciated landscape. However, no studies have been conducted to determine the phenotypic plasticity of these populations.

Since microsatellite are neutral genetic markers, variability at microsatellite loci cannot directly assess the adaptive potential of the apomictic and sexual *B. laevigata* populations and thus no conclusions can be drawn from this study regarding any advantage of apomixis over sexual reproduction in previously glaciated regions, or whether or not apomictic populations have more phenotypic plasticity than sexual populations. It would be interesting for future studies to investigate the relative benefits of the two reproductive modes and whether phenotypic plasticity varies between these forms of *B. laevigata*.

DAPC results found there to be 13 genetic clusters with clustering patterns mostly agreeing with PCoA and genetic identity results. It is surprising, however, that each of the

Tennessee populations were assigned to different genetic clusters, with TN-1 being split between three clusters, especially considering the high degree of genetic similarity among these populations. The AR population formed a cluster separate from the TN and KY populations, contrary to results of the PCoA which displayed significant overlap between these populations.

Comparison with *B. perstellata*

As is consistent with many other studies comparing rare and widespread species (Hamrick and Godt 1989, Gitzendanner and Soltis 2000, Cole 2003), *B. laevigata* had greater levels of genetic diversity than its rare congener *B. perstellata* for the 11 in-common loci. However, it must be noted that differences between the species in allelic richness, observed heterozygosity, and expected heterozygosity did not reach statistical significance and that some sexual *B. laevigata* populations exhibited extremely low levels of genetic variability, comparable to *B. perstellata*. Nevertheless, at least half of the *B. laevigata* populations had higher values for A, H_o, and H_e than even the most genetically diverse *B. perstellata* populations.

Both species had high F_{IS} values, which indicates high levels of inbreeding for both. In addition, both species displayed high levels of population differentiation as indicated by high F_{ST} values, although population structuring was significantly more prevalent among *B. perstellata* populations. The discontinuous habitats of both species and the reliance on gravity for seed dispersal may contribute to the high level of population differentiation observed. The fact that F_{ST} values were the only significant difference between these two congeners is noteworthy considering that other studies have found population differentiation to not be a difference between rare and widespread species (Hamrick and Godt 1989, Gitzendanner and Solits 2000, Cole 2003). However, the significant difference between this study's two *Boechera* species might not be surprising considering that *B. perstellata* populations in Tennessee and Kentucky

are separated by an extremely large distance while *B. laevigata*'s distribution throughout its range is more continuous.

Song and Mitchell-Olds (2007) compared the rare *B. fecunda* with its widespread western congener *B. stricta* and found the widespread species to have greater genetic diversity than the rare species in terms of polymorphic loci, alleles per locus, and expected heterozygosity, with only observed heterozygosity levels being lower for the widespread species (probably due to higher levels of inbreeding). However, differences in observed heterozygosity, expected heterozygosity, and allelic richness were not significant, as was the case when comparing the two *Boechera* species in this study. Thus, both of these comparisons of genetic diversity in rare and widespread *Boechera* congeners fit the expected trends, although in neither case were differences statistically different.

Geographic range has been postulated to be correlated with genetic diversity; however, the heterogeneity of a habitat has also been suggested to influence genetic diversity in a species (Gitzendanner and Solits 2000). For *B. laevigata* and *B. perstellata*, the habitat requirements for these species do not appear to be drastically different, although *B. laevigata* could have a more variable habitat in terms of temperature differences due to its wide latitudinal range

The apomictic *B. laevigata* populations were excluded from the species comparison to because their mode of reproduction differed from the rare species. It should be noted, however, that if the apomictic populations were included, the highly heterozygous nature of these populations would greatly contribute to the overall diversity attributed to the widespread species. Furthermore, the fact that these different modes of reproduction exist for *B. laevigata*, with populations having differing reproductive modes potentially occurring in different parts of the

species range, is a factor that could help increase the geographic range of this widespread species.

Conclusions

Cytological analysis of seeds from the Illinois and Ohio *B. laevigata* populations would be needed to confirm the modes of reproduction and the distribution of apomixis within this widespread species. There does appear to be an interesting correlation between the glacial retreat and apomixis in *B. laevigata*, so further research into the dynamics between historical geography and the distribution of apomictic *B. laevigata* populations could provide insight to the findings reported here. It might also be interesting to do more extensive sampling comparing high and low elevation populations to see whether there are elevational trends in this species' genetic diversity.

With few studies published on the eastern *Boechera* species, this study has added to the growing body of research for the *Boechera* genus and more specifically, the eastern species. Prior to this study, no research regarding the population genetics of *B. laevigata* has been published, although the species has been involved in studies of *Boechera* phylogeny (Kiefer et al. 2009, Kiefer and Koch 2012, Alexander et al. 2013, and Mau et al. 2015) and reproductive mode (Carman et al. 2019). The low levels of genetic diversity within sexual populations and the high heterozygosity levels in apomictic populations of *B. laevigata* are consistent with what generally has been reported for other sexual and apomictic species in this genus. There was a substantial degree of differentiation among the sexual *Boechera laevigata* populations, probably owing to its mode of reproduction and the discontinuous nature of its habitat. In contrast, the apomictic populations were more similar to each other, with the vast majority of their genetic diversity found within individual genotypes. Many of the geographical and genetic patterns found in the

apomictic *B. laevigata* populations are consistent with other studies of asexual lineages with geographical distributions that are distinct from their sexual forms.

As predicted, sexual *B. laevigata* populations exhibited greater genetic diversity overall than did its rare congener *B. perstellata*. Although genetic diversity differences between *B. laevigata* and *B. perstellata* only approached significance when only sexual *B. laevigata* populations are considered, comparisons of this rare and widespread species pair were consistent with the established trends of diversity. This congeneric comparison provides useful context for the low levels of diversity reported for the rare *B. perstellata*, as sexual species in the genus overall seem to lack high genetic variability partly due to the prevalence of inbreeding. This study also adds to the body of literature emphasizing the importance of congeneric comparisons and the genetic diversity patterns of rare and widespread species.

LITERATURE CITED

Agapow, P.M. and A. Burt. 2001. Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes* 1:101-102.

Alhusain, L. and A.M. Hafez. 2018. Nonparametric approaches for population structure analysis. *Human Genomics* 12:25.

Al-Shehbaz, I. A. 1988. The Genera of *Arabideae* (Cruciferae; Brassicaceae) in the Southeastern United States. *Journal of the Arnold Arboretum* 5269:85–166.

Al-Shehbaz, I.A., and M.D. Windham. 2010. *Boechera laevigata*. In: *Flora of North America North of Mexico*. Vol. 7. Flora of North America Editorial Committee (eds), pp. 348. Oxford University Press, New York, NY, USA.

Alberto, F., L. Gouveia, S. Arnaud-Haond, J. L. Pérez-Lloréns, C. M. Duarte, and E. A. Serrão. 2005. Within-population spatial genetic structure, neighbourhood size and clonal subrange in the seagrass *Cymodocea nodosa*. *Molecular Ecology* 14:2669–2681.

Alexander, P. J., M. D. Windham, J. B. Beck, I. A. Al-Shehbaz, L. Allphin, and C. D. Bailey. 2015. Weaving a tangled web: Divergent and reticulate speciation in *Boechera fendleri* sensu lato (Brassicaceae: Boechereae). *Systematic Botany* 40:572–596.

Alexander, P. J., M. D. Windham, J. B. Beck, I. A. Al-Shehbaz, L. Allphin, and C. D. Bailey. 2013. Molecular phylogenetics and taxonomy of the genus *Boechera* and related genera (Brassicaceae: Boechereae). *Systematic Botany* 38:192–209.

Anderson, E., and G.L. Stebbins. 1954. Hybridization as an evolutionary stimulus. *Evolution* 8:378-388.

Arnaud-Haond, S., C.M. Duarte, F. Alberto, and E.A. Serrão. 2007. Standardizing methods to address clonality in population studies. *Molecular Ecology* 16:5115-5139.

Baker, H.G. 1965. Characteristics and modes of origins of weeks. In: *The Genetics of Colonizing Species*, H.G. Baker and G.L. Stebbins (eds.), pp. 147-172. Academic Press, New York, NY, USA.

Balloux, F., L. Lehmann, and T. de Meeûs. 2003. The population genetics of clonal and partially clonal diploids. *Genetics* 164:1635–1644.

Barrett, S.C.H. and J.R. Kohn. 1991. Genetic and evolutionary consequences of small population size in plants: Implications for conservation. In*: Genetics and Conservation of Rare Plants*, D.A. Falk and K.E. Holsinger (eds.), pp. 3-30. Oxford University Press, New York, NY, USA.

Baskauf, C.J., N.C. Jinks, J.R. Mandel, D.E. McCauley. 2014. Population genetics of Braun's Rockcress (*Boechera perstellata*, Brassicaceae), and endangered plant with a disjunct distribution. *Journal of Heredity* 104:124-137.

Baskin, J.M. and C.C. Baskin. 1979. Studies on the autecology and population biology of the monocarpic perennial *Grindelia lanceolata*. *American Midland Naturalist* 102:290-299.

Beardmore JA, 1983. Extinction, survival, and genetic variation. In: *Genetics and Conservation*. C.M. Schonewald-Cox, S.M. Chambers, B. MacBryde, L. Thomas (eds.), pp. 125-151. Benjamin- Cummings, Menlo Park, CA, USA.

Beck, J. B., P. J. Alexander, L. Allphin, I. A. Al-Shehbaz, C. Rushworth, C. D. Bailey, and M. D. Windham. 2012. Does hybridization drive the transition to asexuality in diploid *Boechera*? *Evolution* 66:985–995.

Bentley, K. and R. Mauricio. 2016. High degree of clonal reproduction and lack of large-scale geographic patterning mark the introduced range of the invasive vine, kudzu (*Pueraria montana* var. *lobata*), in North America. *American Journal of Botany* 8:1499-1507.

Bloom, T.C. 1988. The life history of *Arabis laevigata* var. *laevigata* (Muhl.) Poir. (Brassicaceae), M.S. thesis, University of Kentucky, Lexington, KY.

Bloom, T.C., J. M. Baskin, and C. C. Baskin. 1990. Germination ecology of the facultative biennial *Arabis laevigata* variety *laevigata*. *The American Midland Naturalist* 124:214–230.

Bloom, T. C., J. M. Baskin, and C. C. Baskin. 2001. Ecological life history of the facultative woodland biennial *Arabis laevigata* variety *laevigata* (Brassicaceae): Survivorship. *The Journal of the Torrey Botanical Society* 128:93–108.

Bloom, T. C., J. M. Baskin, and C. C. Baskin. 2002a. Ecological life history of the facultative Woodland biennial *Arabis laevigata* variety *laevigata* (Brassicaceae): Reproductive phenology and fecundity. *The Journal of the Torrey Botanical Society* 129:29–38.

Bloom, T. C., J. M. Baskin, and C. C. Baskin. 2002b. Ecological life history of the facultative woodland biennial *Arabis laevigata* variety *laevigata* (Brassicaceae): Seed dispersal. *The Journal of the Torrey Botanical Society* 129:21–28.

Bloom, T. C., J. M. Baskin, and C. C. Baskin. 2003a. Physiological determinants of bolting/ flowering in the facultative woodland biennial, *Arabis laevigata* var. *laevigata* (Brassicaceae). *Castanea* 68:84–87.

Bloom, T. C., J. M. Baskin, and C. C. Baskin. 2003b. Ecological life history of the facultative woodland biennial *Arabis laevigata* variety *laevigata* (Brassicaceae): Effects of leaf litter cover, herbivory, and substrate-type on bolting and fecundity. *The Journal of the Torrey Botanical Society* 130:16–22.

Bona, A., U. Kuelsza, and K.A. Jadwiszczak. 2019. Clonal diversity, gene flow, and seed production in endangered population of *Betula humilis* Schrk. *Tree Genetics and Genomes* 15: 50.

Bouzat, J. L. 2010. Conservation genetics of population bottlenecks: The role of chance, selection, and history. *Conservation Genetics* 11:463–478.

Brukhin, V., J. V. Osadtchiy, A. M. Florez-Rueda, D. Smetanin, E. Bakin, M. S. Nobre, and U. Grossniklaus. 2019. The *Boechera* genus as a resource for apomixis research. *Frontiers in Plant Science* 10:1–19.

Burke, N.W. and R. Bonduriansky. 2018. The geography of sex: sexual conflict, environmental gradients and local loss of sex in facultatively parthenogenetic animals. *Philosophical Transactions of the Royal Society B: Biological Sciences* 373:20170422.

Carman, J. G., M. Mateo de Arias, L. Gao, X. Zhao, B. M. Kowallis, D. A. Sherwood, M. K. Srivastava, K. K. Dwivedi, B. J. Price, L. Watts, and M. D. Windham. 2019. Apospory and diplospory in diploid *Boechera* (Brassicaceae) may facilitate speciation by recombination-driven apomixis-to-sex reversals. *Frontiers in Plant Science* 10:1–14.

Charlesworth, D. 2003. Effects of inbreeding on the genetic diversity of populations. *Philosophical Transactions of the Royal Society B: Biological Sciences* 358:1051–1070.

Charlesworth, D. 2006. Evolution of plant breeding systems. *Current Biology* 16:726–735.

Charlesworth, D., and B. Charlesworth. 1987. Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics* 18:237–268.

Chen, B., J. W. Cole, and C. Grond-Ginsbach. 2017. Departure from Hardy-Weinberg Equilibrium and genotyping error. *Frontiers in Genetics* 8:1–6.

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

Cosendai, A.C., J. Wagner, U. Ladinig, C. Rosche, and E. Hörandl. 2013. Geographical parthenogenesis and population genetic structure in the alpine species *Ranunculus kuepferi* (Ranunculaceae). *Heredity* 110:560-569.

de Meeûs, T. and F. Balloux. 2004. Clonal reproduction and linkage disequilibrium in diploids: a simulation study. *Infection, Genetics, and Evolution* 4:345-351.

Dobeš, C., M. Koch, and T. F. Sharbel. 2006. Embryology, karyology, and modes of reproduction in the North American genus *Boechera* (Brassicaceae): A compilation of seven decades of research. *Annals of the Missouri Botanical Garden* 93:517–534.

Dobeš, C., T. F. Sharbel, and M. Koch. 2005. Towards understanding the dynamics of hybridization and apomixis in the evolution of the genus *Boechera* (Brassicaceae). *Systematics and Biodiversity* 5:321–331.

Douhovnikoff, V. and M. Leventhal. 2016. The use of Hardy-Weinberg Equilibrium in clonal plant systems. *Ecology and Evolution* 6:1173-1180.

Duminil, J., S. Fineschi, A. Hampe, P. Jordano, D. Salvini, G.G. Vendramin, and R.J. Petit. 2007. Can population genetic structure be predicted from life-history traits?. *The American Naturalist*. 169:662-672.

Edmands, S. 2007. Between a rock and a hard place: Evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology* 16:463–475.

Edwards, T.P., R.N. Trigiano, B.H. Ownley, A.S. Windham, C.R. Wyman, P.A. Wadl, and D. Hadziabdic. 2020. Genetic diversity and conservation status of *Helianthus verticillatus*, an endangered sunflower of the southern United States. *Frontiers in Genetics* 11:410.

Ellis, J.R., C.H. Pashley, J.M. Burke, and D.E. McCauley. 2006. High genetic diversity in a rare and endangered sunflower as compared to a common congener. *Molecular Ecology* 15:2345-2355.

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

Ellstrand, N.C. and M.L. Roose. 1987. Patterns of genotypic diversity in clonal plant species. *American Journal of Botany* 74:123-131.

England, P. R., G. H. R. Osler, L. M. Woodworth, M. E. Montgomery, D. A. Briscoe, and R. Frankham. 2003. Effects of intense versus diffuse population bottlenecks on microsatellite genetic diversity and evolutionary potential. *Conservation Genetics* 4:595–604.

Frankel, O.H. 1970. Variation - the essence of life. Sir William Maccleay Memorial Lecture. *Proceedings of the Linnean Society* 95:158-169.

Frankham, R. 2005. Genetics and extinction. *Biological Conservation* 126:131–140.

Frankham, R., J.D. Ballou, M.D.B. Eldridge, R.C. Lacy, K. Ralls, M.R. Dudash, and C.B. Fenster. 2011. Predicting the probability of outbreeding depression. *Conservation Biology* 25:465-475.

Gibson, A.K., J.Y. Xu, and C.M. Lively. 2016. Within-population covariation between sexual reproduction and susceptibility to local parasites. *Evolution* 70:2049-2060.

Gibson, J. P., S. A. Rice, and C. M. Stucke. 2008. Comparison of population genetic diversity between a rare, narrowly distributed species and a common, widespread species of *Alnus* (Betulaceae). *American Journal of Botany* 95:588–596.

Gitzendanner, M. A., and P. S. Soltis. 2000. Patterns of genetic variation in rare and widespread plant congeners. *American Journal of Botany* 87:783–792.

Glesner, R.R., and D. Tilman. 1978. Sexuality and the components of environmental uncertainty: clues from geographic parthenogenesis in terrestrial animals. *American Society of Naturalists* 112:659-673.

Gonzales, E., and J. L. Hamrick. 2005. Distribution of genetic diversity among disjunct populations of the rare forest understory herb, *Trillium reliquum*. *Heredity* 95:306–314.

Goudet, J. 2003. FSTAT, a program to estimate and test gene diversities and fixation indices (v. 2.9.4), http://www.unil.ch/izea/softwares/fstat.html

Grueber, C. E., G. P. Wallis, and I. G. Jamieson. 2008. Heterozygosity-fitness correlations and their relevance to studies on inbreeding depression in threatened species. *Molecular Ecology* 17:3978–3984.

Halkett, F., J. C. Simon, and F. Balloux. 2005. Tackling the population genetics of clonal and partially clonal organisms. *Trends in Ecology and Evolution* 20:194–201.

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

Hamrick, J. L., and M.J.W. Godt. 1996. Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society B: Biological Sciences* 351:1291–1298.

Hartl, D.L. 2000. *A Primer of Population Genetics*. Sinauer Associates Publishing, Sunderland, MA, USA.

Hartl, D.L. and A.G. Clark. 1989. *Principles of Population Genetics*. Sinauer Associates Publishing, Sunderland, MA, USA.

Hojsgaard, D., and E. Hörandl. 2019. The rise of apomixis in natural plant populations. *Frontiers in Plant Science* 10.

Holsinger, K.E. and B.S. Weir. 2009. Genetics in geographically structure populations: defining, estimating and interpreting FST. *National Review of Genetics* 10:639-650.

Hörandl, E. 2006. The complex causality of geographical parthenogenesis. *New Phytologist* 171:525–538.

Hörandl, E., A. C. Cosendai, and E. M. Temsch. 2008. Understanding the geographic distributions of apomictic plants: A case for a pluralistic approach. *Plant Ecology and Diversity* 1:309–320.

Hörandl, E., and D. Hojsgaard. 2012. The evolution of apomixis in angiosperms: A reappraisal. *Plant Biosystems* 143:681–693.

Hudson, R. 2004. Linkage disequilibrium and recombination. In: *Handbook of Statistical Genetics, 2nd Edition*, D.J. Balding, M. Bishop, and C. Cannings (eds.), pp. 662-680. John Wiley & Sons, Ltd, West Sussex, England.

Huenneke, L.F. 1991. Ecological Implications of genetic variation in plant populations. In: *Genetics and Conservation of Rare Plants*, D.A. Falk and K.E. Holsinger (eds.), pp. 31-44. Oxford University Press, USA.

Hughes, A. R., B. D. Inouye, M. T. J. Johnson, N. Underwood, and M. Vellend. 2008. Ecological consequences of genetic diversity. *Ecology Letters* 11:609–623.

Ivens, A.B.F., M. Van de Sander, and J. Bakker. 2012. MLGsim2.0. University of Groningen, Netherlands.

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

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11:94.

Kamvar, Z.N., J.F. Tabima, N.J. Grünwald. 2014. *Poppr*: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 24688859.

Karron, J. D. 1987. A comparison of levels of genetic polymorphism and self-compatibility in geographically restricted and widespread plant congeners. *Evolutionary Ecology* 1:47-58.

Karron, J. D. 1991. Patterns of genetic variation and breeding systems in rare plant species. In: *Genetics and Conservation of Rare Plants*, D.A. Falk and K.E. Holsinger (eds.), pp. 87–98. Oxford University Press, New York, NY. USA.

Karron, J. D. 1997. Genetic consequences of different patterns of distribution and abundance. In: *The Biology of Rarity*. W.E. Kunin, K.J. Gatson (eds.), pp. 174–189. Chapman and Hall, London.

Kartesz, J.T. 2015. The Biota of North America Program (BONAP). *North American Plant Atlas.* (http://bonap.net/napa). Chapel Hill, N.C. [maps generated from Kartesz, J.T. 2015. Floristic Synthesis of North America, Version 1.0. Biota of North America Program (BONAP). (in press)].

Kassambara, A. and F. Mundt. 2020. *Factoextra*: extract and visualize the results of multivariate data analyses. R package version 1.0.7.

Kearney, M. 2005. Hybridization, glaciation and geographical parthenogenesis. *Trends in Ecology and Evolution* 20:495-502.

Kiefer, C., C. Dobeš, and M.A. Koch. 2009. *Boechera* or not? Phylogeny and phylogeography of Eastern North American *Boechera* species (Brassicaceae). *Taxon* 58:1109-1121.

Kiefer, C. and M.A. Koch. 2012. A continental-wide perspective: the genepool of nuclear encoded ribosomal DNA and and single-copy gene sequences in North American *Boechera* (Brassicaceae). *PLoS ONE* 7:e36491.

Kirk, H. and J.R. Freeland. 2011. Applications and implications of neutral *versus* non-neutral markers in molecular ecology. *International Journal of Molecular Sciences* 12:3966-3988.

Koch, M. A., C. Dobes, and T. Mitchell-Olds. 2003. Multiple hybrid formation in natural populations: concerted evolution of the internal transcribed spacer of nuclear ribosomal DNA (ITS) in North American *Arabis divaricarpa* (Brassicaceae). *Molecular Biology and Evolution* 20:338–350.

Koltunow, A. M., and U. Grossniklaus. 2003. Apomixis: a developmental perspective. *Annual Review of Plant Biology* 54:547–574.

Kruckeberg, A. R., and D. Rabinowitz. 1985. Biological aspects of endemism in higher plants. *Annual Review of Ecology and Systematics*. 16:447–479.

Kunin, W. E., and K. J. Gaston. 1993. The biology of rarity: Patterns, causes and consequences. *Trends in Ecology and Evolution* 8:298–301.

Lande, R. 1988. Genetics and demography in biological conservation. *Science* 241:1455–1460.

Lande, R. 1993. Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *The American Naturalist* 142:911–927.

Li, F.W., C.A. Rushworth, J.B. Beck, and M.D. Windham. 2017. *Boechera* microsatellite website: an online portal for species identification and determination of hybrid parentage. Database 2017, baw169. doi:10.1093/database/baw169.

Liu, Z. J., and J. F. Cordes. 2004. DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* 238:1–37.

Lovell, J. T., O. M. Aliyu, M. Mau, M. E. Schranz, M. Koch, C. Kiefer, B. S. Thomas, and M. T. F. Sharbel. 2013. On the origin and evolution of apomixis in *Boechera*. *Plant Reproduction*. 26:309-315.

Lovell, J.T., K. Grogan, T.F. Sharbel, and J.K. McKay. 2014. Mating system and environmental variation drive patterns of adaptation in *Boechera spatifolia*. *Molecular Ecology* 23:4486-4497.

Lu, J. Y. Zhang, X. Dia, K. Yu, X. Dai, P, Qu, M. Crabbe, T. Zhang, and Q. Qiao. 2021. Evaluation of genetic diversity and population structure of *Fragaria nilgerrensis* using EST-SSR markers. *Gene* 796-797: 145791.

MacArthur, R.H. and E.O. Wilson. 1967. *The Theory of Island Biogeography*. Princeton University Press, Princeton, NJ, USA.

Maderspacher, F. 2008. Sex and the drought. *Current Biology* 18:R983-R985.

Mandel, J.R., C.K. Major, R.J. Bayer, and J.E. Moore. 2019. Clonal diversity and spatial genetic structure in the long-lived herb, Prairie trillium. *PLoS ONE* 14:e0224123.

Mau, M., J. T. Lovell, J. M. Corral, C. Kiefer, M. A. Koch, O. M. Aliyu, and T. F. Sharbel. 2015. Hybrid apomicts trapped in the ecological niches of their sexual ancestors. *PNAS* 112:E2357–E2365.

Meirmans, P. G. 2006. Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution* 60:2399–2402.

Meirmans, P. G., and P. W. Hedrick. 2011. Assessing population structure: FST and related measures. *Molecular Ecology Resources* 11:5–18.

Meloni, M., A. Reid, J. Caujapé-Castells, Á. Marrero, J. M. Fernández-Palacios, R. A. Mesa-Coelo, and E. Conti. 2013. Effects of clonality on the genetic variability of rare, insular species: The case of *Ruta microcarpa* from the Canary Islands. *Ecology and Evolution* 3:1569–1579.

Mondini, L., A. Noorani, and M. A. Pagnotta. 2009. Assessing plant genetic diversity by molecular tools. *Diversity* 1:19–35.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences* 70:3321-3323.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583-590.

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

Niklas, K.J., and E.D. Cobb. 2017. The evolutionary ecology (evo-eco) of plant asexual reproduction. *Evolutionary Ecology* 31:317-332.

Nogler, G.A. 1984. Gametophytic apomixis. In: *Embryology of Angiosperms.* B.M. Johri (eds.), pp. 475-518. Springer-Verlag Berlin Heidelberg, Berlin, Germany.

Nordborg, M. and H. Innan. 2002. Molecular population genetics. *Current Opinion in Plant Biology* 5: 69-73.

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

Ohsawa, T. and Y. Ide. 2008. Global patterns of genetic variation in plant species along vertical and horizontal gradients on mountains. *Global Ecology and Biogeography* 17:152-163.

Parks, J.C. and C.R. Werth. 1993. A study of spatial features of clones in a population of bracken fern, *Pteridium aquilinum* (Dennstaedtiaceae). *American Journal of Botany* 80:537-544.

Peakall, R., and P.E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6:288-295.

Peakall, R., and P.E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research- an update. *Bioinformatics* 28:2537-2539.

Pfeifer, M., and G. Jetschke. 2006. Influence of geographical isolation on genetic diversity of *Himantoglossum hircinum* (Orchidaceae). *Folia Geobotanica* 41:3–20.

Polato, N.R., M.M. Gray, B.A. Gill, C.G. Becker, K.L. Casner, A.S. Flecker, B.C. Kondratieff, A.C. Encalada, N.L. Poff, W.C. Funk, and K.R. Zamudio. 2017. Genetic diversity and gene flow decline with elevation in montane mayflies. *Heredity* 119:107-116.

Price, M. V. and Waser N. M. 1979. Pollen dispersal and optimal outcrossing in *Delphinium nelsonii*. *Nature* 277:294–297.

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

Rabinowitz, R. A. 1981. Seven forms of rarity. In: *The Biological Aspects of Rare Plant Conservation*. H. Synge (eds.). pp. 205–217. John Wiley and Sons Ltd, New York, NY, USA.

Reed, D. H., and R. Frankham. 2003. Correlation between fitness and genetic diversity. *Conservation Biology* 17:230–237.

Reed, D. H., E. H. Lowe, D. A. Briscoe, and R. Frankham. 2003. Fitness and adaptation in a novel environment: Effect of inbreeding, prior environment, and lineage. *Evolution* 57:1822–1828.

Rousset, F. 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* 8:103-106.

Rushworth, C. A., B. H. Song, C. R. Lee, and T. Mitchell-Olds. 2011. *Boechera*, a model system for ecological genomics. *Molecular Ecology* 20:4843–4857.

Rushworth, C. A., M. D. Windham, R. A. Keith, and T. Mitchell-Olds. 2018. Ecological differentiation facilitates fine-scale coexistence of sexual and asexual *Boechera*. *American Journal of Botany* 105:2051–2064.

Savidan, Y. 2001. Gametophytic apomixis. In: *Current Trends in the Embryology of Angiosperms*. S.S. Bhojwani and W.Y. Soh (eds.). pp. 419-433. Kluwer Academic Publishers, Norwell, MA, USA.

Schaefer, I., K. Domes, M. Heethoff, K. Schneider, I. Schön, R.A. Norton, S. Scheu, and M. Maraun. 2006. No evidence for the 'Meselson effect' in parthenogenetic oribatid mites (Oribatida, Acari). *Journal of Evolutionary Biology* 19:184-193.

Schilling, P. M., Z. Gompert, F.W. Li, M.D. Windham, and P.G. Wolf. 2018. Admixture, evolution, and variation in reproductive isolation in the *Boechera puberula* clade. *BMC Evolutionary Biology* 18:1-14.

Schranz, E. M., C. Dobeš, M. A. Koch, and T. Mitchell-Olds. 2005. Sexual reproduction, hybridization, apomixis, and polyploidization in the genus *Boechera* (Brassicaceae). *American Journal of Botany* 92:1797–1810.

Schranz, M.E., A.J. Windsor, B.H. Song, A. Lawton-Rauh, and T. Mitchell-Olds. 2007. Comparative genetic mapping in *Boechera stricta*, a close relative of *Arabidopsis*. *Plant Physiology*. 144:586-598.

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18:233-234.

Song, B., M. Clauss, A. Pepper, T. Mitchell-Olds. 2006. Geographic patterns of microsatellite variation in *Boechera stricta*, a close relative of *Arabidopsis*. *Molecular Ecology* 15:357-369.

Song, B. and T. Mitchell-Olds. 2007. High genetic diversity and population differentiation in *Boechera fecunda*, a rare relative of *Arabidopsis*. *Molecular Ecology* 16:4079-4088.

Stenberg, P., M. Lundmark, and A. Saura. 2003. MLGSIM: a program for detecting clones using a simulation approach. *Molecular Ecology Notes* 3:329-331.

Takahashi, Y., H. Takahashi, and M. Maki. 2011. Comparison of genetic variation and differentiation using microsatellite markers among three rare threatened and one widespread toad lily species of *Tricyrtis* section *Flavae* (Convallariaceae) in Japan. *Plant Species Biology* 26:13–23.

Tilquin, A., and H. Kokko. 2016. What does the geography of parthenogenesis teach us about sex? *Philosophical Transactions of the Royal Society B: Biological Sciences* 371.

[USFWS] U.S. Fish and Wildlife Service. 1997. Recovery Plan for *Arabis perstellata* (Braun's Rockcress). Atlanta (GA): USFWS. p. 21.

[USFWS] U.S. Fish and Wildlife Service. 2004. Endangered and threatened wildlife and plants; Designation of critical habitat for *Arabis perstellata* (Braun's Rock-cress). Fed Reg. 69:31459-31496.

[USFWS] U.S. Fish and Wildlife Service. 2018. Braun's Rockcress (*Arabis perstellata*) 5-Year Review: Summary and Evaluation. Frankfort (KY): US Fish and Wildlife Service Southeast Region, Kentucky Ecological Services Field Office. 39 pp.

Vandel, A. 1928. La parthe´nogene`se ge´ographique. Contribution a` l'e´tude biologique et cytologique de la parthe´nogene`se naturelle. *Bulletin Biologique de la France et de la Belgique*. 62, 164–281.

Van Dijk, P. J. 2007. Potential and realized costs of sex in dandelions, *Taraxacum officinale* s.l. In: *Apomixis: Evolution, Mechanisms, and Perspectives*. E. Horandl, U. Grossniklaus, P.J. Van Dijk, T.F. Sharbel (eds.), pp. 137-158. Gantner, Rugell, Austria.

Verduign, M.H., P.J. Van Dijk, and J.M.M. Van Damme. 2003. Distribution, phenology and demography of sympatric sexual and asexual dandelions (*Taraxacum officinale* s.l.): geographic parthenogenesis on a small scale. *Biological Journal of the Linnean Society* 82:205-218.

Vrijenhoiek, R.C. 1984. Ecological differentiation among clones: the frozen niche variation model. In: *Population biology and evolution,* K. Woermann, V. Loeschcke (eds.), Springer, pp. 217-231. Berlin, Germany.

Vrijenhoek, R.C. and E. Davis Parker. 2009. Geographical parthenogenesis: General purpose genotypes and frozen niche variation. In: *Lost sex*. I. Schön, K. Martens, and P. Van Dijk (eds.), Springer Publications, pp. 99-130. Berlin, Germany.

Wahlund, S. 1928. Zusammensetzung von Population und Korrelationserscheinung vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas* 11:65–106.

Weakley, A. 2015. *Flora of the Southern and Mid-Atlantic States.* Chapel Hill, NC: University of North Carolina at Chapel Hill.

Weir, B.S. 1996. *Genetic Data Analysis II*. Sinauer Associates Publishing, Sunderland, MA. USA.

Weir, B.S. and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358-1370.

Welch, D. M., and M. Meselson. 2000. Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science* 288:1211–1215.

Whitlock, M. C. 2000. Fixation of new alleles and the extinction of small populations: Drift load, beneficial alleles, and sexual selection. *Evolution* 54:1855–1861.

Whitlock, M. C. 2011. G′ST and D do not replace FST. *Molecular Ecology* 20:1083–1091.

Windham, M.D., and I.A. Al-Shehbaz. 2007. New and noteworthy species of *Boechera* (Brassicaceae) III: Additional sexual diploids and apomictic hybrids. *Harvard Papers in Botany* 12:235-257.

Windham, M. D., J. B. Beck, F. W. Li, L. Allphin, J. G. Carman, D. A. Sherwood, C. A. Rushworth, E. Sigel, P. J. Alexander, C. D. Bailey, and I. A. Al-Shehbaz. 2015. Searching for diamonds in the apomictic rough: A case study involving *Boechera lignifera* (Brassicaceae). *Systematic Botany* 40:1031–1044.

Wright, S. 1950. The genetical structure of populations. *Nature* 166:247-249.

Appendix A

Appendix A Table 1. Allele frequencies at 16 loci for all *Boechera laevigata* data. Sample sizes for each locus are provided. Allele numbers represent pcr product sizes in base pairs. Loci are listed out of alphabetical order to allow a single locus to be visible on one page. IL-1, IL-2, and OH-1 are considered to be apomictic populations, with all other populations considered to be sexually reproducing.

Locus	N/ Allele size	AL-1	AL-2	AR	KY	MD	NC-1	NC-2	OH-2	PA	TN-1	TN-2	TN-3	IL-1	IL-2	OH-1
B07	**N**	30	30	27	10	13	23	12	6	24	34	10	10	30	25	16
	169	**0.567**	0.000	0.000	0.000	**1.000**	**1.000**	**1.000**	**1.000**	**0.958**	0.000	0.000	0.000	**0.500**	**0.500**	**0.500**
	170	0.200	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.042**	0.000	0.000	0.000	0.000	0.000	0.000
	178	**0.133**	**0.967**	**0.963**	**1.000**	0.000	0.000	0.000	0.000	0.000	**1.000**	**0.900**	**1.000**	**0.500**	**0.500**	**0.500**
	179	**0.100**	**0.033**	**0.037**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.100**	0.000	0.000	0.000	0.000
B11	**N**	30	30	27	10	13	23	12	6	24	34	10	10	30	25	16
	169	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.515**	0.000	0.000	0.500	0.000	**0.500**
	171	**0.217**	**1.000**	**1.000**	0.000	0.000	0.000	0.000	0.000	0.000	**0.368**	0.000	0.000	0.000	0.000	0.000
	173	**0.783**	0.000	0.000	**1.000**	0.000	**0.870**	0.000	0.000	0.000	**0.118**	**1.000**	**1.000**	0.000	0.000	0.000
	175	0.000	0.000	0.000	0.000	**1.000**	**0.130**	**1.000**	**1.000**	0.000	0.000	0.000	0.000	0.000	**0.500**	0.000
	185	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000	**0.017**	0.000	**0.500**
	187	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.483**	**0.500**	0.000
C02	**N**	30	30	27	10	13	23	12	6	24	34	10	10	30	25	16
	158	**0.233**	**1.000**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	162	**0.533**	0.000	**1.000**	**1.000**	**1.000**	**1.000**	0.000	**1.000**	0.000	**0.426**	**1.000**	**1.000**	**0.500**	**1.000**	**0.500**
	164	0.000	0.000	0.000	0.000	0.000	0.000	**1.000**	0.000	**1.000**	**0.015**	0.000	0.000	**0.500**	0.000	**0.500**
	166	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.559**	0.000	0.000	0.000	0.000	0.000
	178	**0.233**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C03	**N**	30	30	27	10	13	23	12	6	24	34	10	10	30	25	16
	161	**0.233**	**1.000**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	165	**0.533**	0.000	**1.000**	**1.000**	**1.000**	**1.000**	0.000	**0.500**	0.000	**0.426**	**1.000**	**1.000**	**0.500**	**1.000**	**0.500**
	167	0.000	0.000	0.000	0.000	0.000	0.000	**1.000**	**0.500**	**1.000**	**0.015**	0.000	0.000	**0.500**	0.000	**0.500**
	169	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.559**	0.000	0.000	0.000	0.000	0.000
	181	**0.233**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Appendix A Table 1 (cont.)

Locus	N/ Allele size	AL-1	AL-2	AR	KY	MD	NC-1	NC-2	OH-2	PA	TN-1	TN-2	TN-3	IL-1	IL-2	OH-1
E09	**N**	30	30	27	10	13	23	12	6	24	34	10	10	30	25	16
	186	**0.267**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	192	**0.733**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
E11	**N**	30	30	27	10	13	23	12	6	24	34	10	10	30	25	16
	199	**0.583**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	201	**0.417**	**1.000**	**1.000**	**0.800**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.691**	**1.000**	**0.900**	**0.517**	**1.000**	**1.000**
	203	0.000	0.000	0.000	**0.200**	0.000	0.000	0.000	0.000	0.000	**0.235**	0.000	0.100	0.000	0.000	0.000
	205	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.000	**0.483**	0.000	0.000
	207	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.059**	0.000	0.000	0.000	0.000	0.000
G03	**N**	30	30	27	10	13	23	12	6	24	34	10	10	30	25	16
	245	**0.233**	**1.000**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	249	**0.550**	0.000	**1.000**	**1.000**	**1.000**	**1.000**	0.000	**1.000**	0.000	**0.456**	**1.000**	**1.000**	**0.500**	**1.000**	**0.500**
	251	0.000	0.000	0.000	0.000	0.000	0.000	**1.000**	0.000	**1.000**	**0.015**	0.000	0.000	**0.500**	0.000	**0.500**
	253	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.529**	0.000	0.000	0.000	0.000	0.000
	263	**0.183**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	265	**0.033**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
G08	**N**	30	30	27	10	13	23	12	6	24	34	10	10	30	25	16
	250	**0.233**	**0.167**	**0.074**	0.000	**0.308**	**0.217**	**0.083**	**0.167**	**0.042**	**0.029**	0.000	**0.100**	0.000	**0.040**	0.000
	252	0.000	0.000	**0.667**	0.000	**0.038**	0.000	0.000	0.000	0.000	**0.412**	0.000	**0.100**	0.000	0.000	0.000
	254	**0.100**	**0.833**	0.000	0.000	0.000	**0.043**	**0.917**	0.000	**0.958**	**0.191**	**0.100**	**0.300**	**0.500**	**0.960**	**0.500**
	255	**0.400**	0.000	0.000	0.000	**0.654**	**0.739**	0.000	**0.833**	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	256	**0.267**	0.000	**0.259**	**1.000**	0.000	0.000	0.000	0.000	0.000	**0.279**	**0.050**	**0.500**	0.000	0.000	0.000
	258	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.250**	0.000	0.000	0.000	0.000
	263	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.600**	0.000	0.000	0.000	0.000
	270	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.088**	0.000	0.000	0.000	0.000	0.000
	278	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.500**	0.000	**0.500**

Appendix A Table 1 (cont.)

Locus	N/ Allele size	AL-1	AL-2	AR	KY	MD	NC-1	NC-2	OH-2	PA	TN-1	TN-2	TN-3	IL-1	IL-2	OH-1
G09	**N**	30	30	27	10	13	23	12	6	24	34	10	10	30	25	16
	235	0.000	0.000	**0.870**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	237	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000	0.000	0.000	**0.500**	0.000	0.000	0.000	0.000	0.000
	239	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.015**	0.000	0.000	0.000	0.000	0.000
	242	**0.183**	0.000	**0.130**	0.000	0.000	0.000	0.000	0.000	0.000	**0.176**	**0.700**	**1.000**	0.000	0.000	0.000
	244	0.000	0.000	0.000	0.000	**0.962**	0.000	0.000	0.000	0.000	**0.015**	0.000	0.000	0.000	**0.480**	0.000
	246	**0.817**	0.000	0.000	0.000	0.038	**0.870**	**1.000**	**1.000**	0.000	0.000	**0.300**	0.000	0.000	0.000	0.000
	248	0.000	**0.983**	0.000	0.000	0.000	**0.087**	0.000	0.000	0.000	**0.132**	0.000	0.000	**0.517**	**0.020**	**0.500**
	250	0.000	**0.017**	0.000	0.000	0.000	**0.043**	0.000	0.000	**1.000**	0.000	0.000	0.000	**0.483**	**0.500**	**0.500**
	254	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.162**	0.000	0.000	0.000	0.000	0.000
H06	**N**	30	30	27	10	13	23	12	6	24	34	10	10	30	25	16
	234	0.000	0.000	0.000	0.000	0.000	**0.739**	**1.000**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	236	0.000	0.000	**1.000**	0.000	0.000	0.000	0.000	0.000	0.000	**0.103**	**0.200**	0.000	0.000	**0.500**	0.000
	238	0.000	**0.017**	0.000	**1.000**	0.000	0.000	0.000	0.000	0.000	0.000	**0.800**	**0.400**	**0.500**	0.000	**0.500**
	240	**0.217**	**0.983**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.029**	0.000	0.000	0.000	0.000	0.000
	242	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.529**	0.000	**0.600**	0.000	0.000	0.000
	244	0.000	0.000	0.000	0.000	0.000	**0.043**	0.000	0.000	0.000	**0.279**	0.000	0.000	0.000	0.000	0.000
	246	**0.783**	0.000	0.000	0.000	**1.000**	**0.217**	0.000	**1.000**	**1.000**	0.000	0.000	0.000	**0.500**	**0.500**	**0.500**
	252	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.059**	0.000	0.000	0.000	0.000	0.000
ICE4	**N**	30	30	27	10	13	23	12	6	24	34	10	10	30	25	16
	194	**0.817**	0.000	0.000	0.000	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.000	0.000	0.000	**0.500**	**0.500**	**0.500**
	198	**0.183**	**1.000**	**1.000**	**1.000**	0.000	0.000	0.000	0.000	0.000	**1.000**	**1.000**	**1.000**	**0.500**	**0.500**	**0.500**
ICE5	**N**	30	30	27	10	13	23	12	6	24	34	10	10	30	25	16
	203	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.700**	**0.100**	0.000	0.000	0.000
	206	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.662**	**0.300**	**0.900**	**1.000**	**1.000**	**1.000**
	210	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.338**	0.000	0.000	0.000	0.000	0.000

Appendix A Table 1 (cont.)

Locus	N/ Allele size	AL-1	AL-2	AR	KY	MD	NC-1	NC-2	OH-2	PA	TN-1	TN-2	TN-3	IL-1	IL-2	OH-1
R3_02	**N**	30	30	27	10	13	23	12	6	24	34	10	10	30	25	16
	128	**0.800**	0.000	0.000	0.000	**0.192**	**1.000**	**1.000**	**1.000**	0.000	0.000	0.000	**0.050**	0.000	0.000	0.000
	130	0.000	0.000	0.000	0.000	**0.808**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	132	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000	**0.500**	**0.500**	**0.500**
	140	0.000	0.000	**1.000**	**1.000**	0.000	0.000	0.000	0.000	0.000	**0.353**	**0.750**	**0.900**	0.000	**0.500**	0.000
	142	**0.200**	**1.000**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.647**	**0.250**	**0.050**	**0.500**	0.000	**0.500**
R3_35	**N**	30	30	27	10	13	23	12	6	24	34	10	10	30	25	16
	144	0.000	0.000	**1.000**	0.000	0.000	0.000	0.000	0.000	0.000	**0.176**	0.000	0.000	0.000	**1.000**	0.000
	146	**0.767**	0.000	0.000	0.000	**0.462**	0.000	**1.000**	**1.000**	**1.000**	0.000	0.000	0.000	0.000	0.000	0.000
	148	0.000	0.000	0.000	0.000	**0.538**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	150	0.000	0.000	0.000	0.000	0.000	**0.739**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	152	0.000	0.000	0.000	0.000	0.000	**0.261**	0.000	0.000	0.000	**0.044**	0.000	0.000	0.000	0.000	0.000
	156	0.000	0.000	0.000	**0.200**	0.000	0.000	0.000	0.000	0.000	**0.515**	**0.950**	0.000	0.000	0.000	**1.000**
	158	**0.233**	**1.000**	0.000	**0.800**	0.000	0.000	0.000	0.000	0.000	**0.088**	0.000	0.000	**1.000**	0.000	0.000
	160	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.050**	0.000	0.000	0.000	0.000
	162	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.029**	0.000	0.000	0.000	0.000	0.000
	164	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.300**	0.000	0.000	0.000
	166	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.029**	0.000	**0.400**	0.000	0.000	0.000
	168	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.300**	0.000	0.000	0.000
	176	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.044**	0.000	0.000	0.000	0.000	0.000
	178	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.074**	0.000	0.000	0.000	0.000	0.000

Appendix A Table 1 (cont.)

Locus	N/ Allele size	AL-1	AL-2	AR	KY	MD	NC-1	NC-2	OH-2	PA	TN-1	TN-2	TN-3	IL-1	IL-2	OH-1
F03	**N**	30	30	27	10	13	23	12	6	24	34	10	10	30	25	16
	233	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000	0.000
	239	0.000	0.000	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000	**0.029**	0.000	**0.600**	0.000	0.000	0.000
	241	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.029**	0.000	0.000	0.000	0.000	0.000
	242	0.000	**1.000**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	244	0.000	0.000	0.000	**0.650**	0.000	**0.043**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.500**	0.000
	245	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.074**	0.000	0.000	0.000	0.000	0.000
	246	0.000	0.000	0.000	0.000	**0.115**	**0.087**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	248	0.000	0.000	**0.907**	0.000	0.000	**0.587**	0.000	0.000	0.000	**0.147**	0.000	0.000	0.000	0.000	0.000
	250	0.000	0.000	**0.093**	0.000	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000	**0.500**	0.000	**0.500**
	252	**0.400**	0.000	0.000	**0.350**	0.000	0.000	0.000	0.000	0.000	**0.015**	0.000	0.000	0.000	0.000	0.000
	254	**0.600**	0.000	0.000	0.000	**0.077**	0.000	0.000	0.000	0.000	**0.544**	0.000	**0.400**	0.000	0.020	0.000
	256	0.000	0.000	0.000	0.000	**0.154**	0.000	0.000	0.000	0.000	**0.059**	0.000	0.000	0.000	0.000	0.000
	258	0.000	0.000	0.000	0.000	0.000	**0.152**	0.000	0.000	0.000	**0.044**	0.000	0.000	0.000	**0.480**	0.000
	259	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.029**	0.000	0.000	0.000	0.000	0.000
	263	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.015**	0.000	0.000	0.000	0.000	0.000
	265	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.017**	0.000	0.000
	269	0.000	0.000	0.000	0.000	0.000	**0.130**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	273	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000	0.000	**0.433**	0.000	0.000
	275	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.033**	0.000	0.000
	276	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.017**	0.000	0.000
	280	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.015**	0.000	0.000	0.000	0.000	**0.500**
	290	0.000	0.000	0.000	0.000	**0.231**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	292	0.000	0.000	0.000	0.000	**0.115**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	294	0.000	0.000	0.000	0.000	**0.077**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	295	0.000	0.000	0.000	0.000	**0.231**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Appendix A Table 1 (cont.)

Locus	N/ Allele size	AL-1	AL-2	AR	KY	MD	NC-1	NC-2	OH-2	PA	TN-1	TN-2	TN-3	IL-1	IL-2	OH-1
G06	**N**	30	30	27	10	13	23	12	6	24	34	10	10	30	25	16
	259	0.000	0.000	0.000	0.000	0.000	**0.022**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	263	0.000	0.000	0.000	0.000	0.000	**0.848**	**1.000**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	264	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.017**	0.000	0.000
	265	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.700**	0.000	0.000	0.000
	266	0.000	0.000	0.000	0.000	0.000	**0.130**	0.000	**0.333**	**1.000**	0.000	0.000	0.000	0.000	**0.500**	**0.500**
	267	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.074**	**0.400**	**0.150**	0.000	0.000	0.000
	268	**0.750**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.483**	0.000	0.000
	269	0.000	**0.617**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.088**	**0.550**	0.000	0.000	0.000	0.000
	270	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.500**	0.000
	271	0.000	0.000	**0.296**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	272	0.000	0.000	0.000	0.000	**0.538**	0.000	0.000	**0.667**	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	276	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.412**	0.000	0.000	**0.500**	0.000	**0.500**
	278	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.015**	0.000	0.000	0.000	0.000	0.000
	280	**0.067**	0.000	**0.241**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	281	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.162**	**0.050**	0.000	0.000	0.000	0.000
	283	0.000	**0.383**	0.000	0.000	**0.077**	0.000	0.000	0.000	0.000	**0.029**	0.000	0.000	0.000	0.000	0.000
	285	**0.183**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.176**	0.000	0.000	0.000	0.000	0.000
	287	0.000	0.000	0.000	0.000	**0.154**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	290	0.000	0.000	0.000	**0.100**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	292	0.000	0.000	0.000	**0.850**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	293	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.044**	0.000	0.000	0.000	0.000	0.000
	294	0.000	0.000	0.000	0.050	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	295	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.100**	0.000	0.000	0.000
	297	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.050**	0.000	0.000	0.000
	301	0.000	0.000	**0.037**	0.000	**0.077**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	303	0.000	0.000	**0.130**	0.000	**0.077**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	311	0.000	0.000	0.000	0.000	**0.077**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	328	0.000	0.000	**0.296**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Appendix A Table 2. Polymorphic loci for each *Boechera laevigata* population that deviate significantly from Hardy-Weinberg expectations after Bonferroni corrections. IL-1, IL-2, and OH-1 are considered to be apomictic populations, with all other populations considered to be sexually reproducing. Tests were ran using a full data set for all populations.

Population	**Number of Polymorphic loci**	**Loci deviating from Hardy-Weinberg expectations** ($P < 0.05$)
AL-1	15	B07, B11, C02, C03, E09, E11, F03*, G03, G06, G08, G09, H06, ICE4, R3_02, R3_35
AL-2	3	B07, G06, G08
AR	5	B07, G06, G08, G09
KY	4	E11, F03, R3_35
MD	5	F03, G06, G08, R3_02, R3_35
NC-1	7	B11, F03, G06, G08, G09, H06, R3_35
NC-2	1	G08
PA	2	B07, G08
TN-1	13	B11, C02, C03, E11, F03, G03, G06, G08, G09, H06, ICE5, R3_02, R3_35
TN-2	7	G06
TN-3	8	F03, G08, H06, R3_35
IL-1	13	B07, B11, C02, C03, E11, F03, G03, G06, G08, G09, H06, ICE4, R3_02
IL-2	9	B08, B11, F03, G06, G08, G09, H06, ICE4, R3_02
OH-1	12	B07, B11, C02, C03, F03, G03, G06, G08, G09, H06, ICE4, R3_02

Appendix A Table 3. *Boechera laevigata* loci pairs with significant linkage disequilibrium (Weir's 1996 test of composite linkage disequilibrium), as tested by GENEPOP. Bolded loci indicate loci pairs that were significantly linked when analyzing both a full and unique MLG only ("MLG") data set. Loci not bolded were only significantly linked for the data set indicated. Bonferroni corrected *P*-values are indicated. The four populations listed are the only ones among the 15 sampled populations showing significant linkage disequilibrium between pairs of loci.

Population	Loci pairs with significant linkage, P < 0.001 for both "full" and "MLG" data sets, unless indicated otherwise
AL-1 (67 pairs with significant linkage disequilibrium)	**C02 and C03**
	B11 and E09
	C02 and E09
	C03 and E09
	C02 and G03
	C03 and G03
	E09 and G03
	B07 and G06
	B11 and G06
	C02 and G06[1]
	C03 and G06[1]
	E09 and G06
	B07 and G09
	B11 and G09
	C02 and G09[1]
	C03 and G09[1]
	G06 and G09
	C02 and H06
	C03 and H06
	E09 and H06
	G03 and H06
	G06 and H06
	G09 and H06[1]
	B07 and ICE4[1]
	B11 and ICE4
	C03 and ICE4[1]
	G06 and ICE4[1]
	G08 and ICE4[1]
	G09 and ICE4
	H06 and ICE4[1]
	B11 and R3_02[1]
	C03 and R3_02[1]
	E09 and R3_02
	G06 and R3_02
	G09 and R3_02[1]
	H06 and R3_02

Appendix A Table 3 (cont.).

Population	Loci pairs with significant linkage, P < 0.001 for both "full" and "MLG" data sets, unless indicated otherwise
AL-1 (67 pairs with significant linkage disequilibrium)	**ICE4 and R3_02**
	C02 and R3_35
	C03 and R3_35
	E09 and R3_35
	G03 and R3_35
	H06 and R3_35
	B11 and C02[1]
	B11 and H06
	B07 and H06[1]
	C02 and R3_02
	R3_02 and R3_35
	B11 and R3_35[1]
	E11 and R3_02[1]
	G03 and R3_02[1]
	B07 and G08[1]
	B11 and C03[1]
	G03 and ICE4[1]
	B11 and G08[1]
	G09 and R3_35[1]
	G08 and G09[1]
	E09 and ICE4[1]
	B07 and R3_02[1]
	ICE4 and R3_35[1]
	B07 and B11[1]
	B07 and E09[1]
	G06 and G08[1]
	C02 and ICE4[1]
	G08 and R3_02[1]
	B07 and R3_35[1]
	B08 and C02[1]
	E09 and F03[2]
MD	**F03 and G06**
PA	B07 and G08[1] (P < 0.05)
TN-1 (6 pairs with significant linkage disequilibrium)	**C02 and C03**
	C02 and G03
	C03 and G03
	F03 and G08[1]
	C02 and R3_35[1]
	G08 and R3_35[2]

[1]Not significant for unique MLG only data set.
[2]Loci pairs were significantly linked when analyzing unique MLGs, but not when analyzing a full data set.

Appendix A Table 4. Significant ($P < 0.05$) P_{sex} values generated by MLGsim 2.0 software using the F_{IS} model to account for Hardy-Weinberg deviations at *Boechera laevigata* loci. P_{sex} values indicate the probability of encountering a particular MLG more than once by chance which are calculated from P_{gen} values, which indicate the probability of obtaining a particular MLG by chance.

Population	No. of individuals with a particular MLG	P_{gen} (F_{IS})	P_{sex} (F_{IS})
AL-1	2	4.47×10^{-10}	0
	8	1.08×10^{-9}	0
	2	7.48×10^{-9}	2.50×10^{-12}
	2	4.6210^{-13}	8.88×10^{-16}
AL-2	3	1.47×10^{-12}	3.89×10^{-15}
	12	1.63×10^{-12}	1.11×10^{-15}
	2	5.28×10^{-14}	0
	3	5.85×10^{-14}	0
	7	9.19×10^{-13}	1.30×10^{-14}
AR	4	7.34×10^{-9}	1.03×10^{-14}
	3	7.34×10^{-9}	1.03×10^{-14}
	2	1.13×10^{-8}	5.75×10^{-12}
KY	2	2.21×10^{-9}	2.14×10^{-13}
	2	4.14×10^{-9}	7.62×10^{-13}
	2	1.39×10^{-10}	4.00×10^{-15}
NC-1	3	1.59×10^{-9}	1.27×10^{-14}
	4	2.19×10^{-9}	6.99×10^{-15}
	2	1.06×10^{-9}	4.16×10^{-14}
	2	7.63×10^{-11}	1.04×10^{-14}
NC-2	11	1.10×10^{-13}	0
OH-2	2	1.40×10^{-11}	4.66×10^{-15}
	3	1.58×10^{-11}	0
PA	23	6.29×10^{-16}	1.10×10^{-14}
TN-1	2	2.33×10^{-12}	0
TN-2	2	3.99×10^{-11}	0
IL-1	25	1.23×10^{-29}	0
	2	1.54×10^{-30}	0
IL-2	23	7.95×10^{-22}	0
OH-1	16	1.79×10^{-28}	0

Appendix A Table 5. Private alleles per locus in all *Boechera laevigata* populations. Allele numbers represent pcr product sizes in base pairs.

Population	Total private alleles per population	# loci with private alleles	Locus	Allele	Freq
AL-1	6	5	C02	178	0.233
			C03	181	0.233
			E09	186	0.267
			E11	199	0.583
			G03	263	0.183
				265	0.033
AL-2	1	1	F03	242	1.000
AR	3	2	G06	271	0.296
				328	0.296
			G09	235	0.870
KY	3	1	G06	290	0.100
				292	0.850
				294	0.050
MD	8	4	F03	290	0.231
				292	0.115
				294	0.077
				295	0.231
			G06	287	0.143
				311	0.071
			R3_02	130	0.821
			R3_35	148	0.500
NC-1	3	3	F03	269	0.130
			G06	259	0.020
			R3_35	150	0.760

Appendix A Table 5 (cont.).

Population	Total private alleles per population	# loci with private alleles	Locus	Allele	Freq
TN-1	18	11	C02	166	0.559
			C03	169	0.559
			E11	207	0.059
			F03	241	0.029
				245	0.074
				259	0.029
				263	0.015
			G03	253	0.529
			G06	278	0.015
				293	0.044
			G08	270	0.088
			G09	239	0.015
				254	0.162
			H06	252	0.059
			ICE5	210	0.338
			R3_35	162	0.029
				176	0.043
				178	0.071
TN-2	4	3	F03	233	1.000
			G08	258	0.250
				263	0.600
			R3_35	160	0.050
TN-3	5	2	G06	265	0.700
				295	0.100
				297	0.050
			R3_35	164	0.300
				168	0.300
IL-1	4	2	F03	265	0.017
				275	0.033
				276	0.017
			G06	264	0.017
IL-2	1	1	G06	270	0.500

Appendix A Table 6. Genetic variability of averaged across 16 polymorphic loci for 15 *Boechera laevigata* populations, using what would be "wrong" data sets for all populations below (a unique MLG only data set for the "sexual populations" and a full data set for the "apomictic populations"), given their "sexual" and "apomictic" designations. However, these data have been provided because these populations' designations as "sexual" and "apomictic" have not been confirmed by cytological analysis of seeds. Sample size (*N*), number of unique multilocus genotypes (MLG), mean direct count of heterozygosity (H_o), Nei (1978) unbiased estimate of mean expected heterozygosity (H_e). Standard error is given in parentheses.

Population	***N***	***MLG***	H_o	H_e
Sexual populations (using a unique MLG only data set)				
AL-1	30	20	0.059 (0.011)	0.485 (0.046)
AL-2	30	8	0.031 (0.018)	0.089 (0.041)
AR	27	21	0.048 (0.025)	0.117 (0.058)
KY	10	7	0.036 (0.028)	0.099 (0.046)
MD	13	13	0.048 (0.023)	0.185 (0.073)
NC-1	23	16	0.047 (0.020)	0.189 (0.059)
NC-2	12	2	0.000 (0.000)	0.042 (0.042)
OH-2	6	3	0.063 (0.063)	0.104 (0.056)
PA	24	2	0.000 (0.000)	0.083 (0.057)
TN-1	34	33	0.235 (0.036)	0.482 (0.065)
TN-2	10	9	0.111 (0.045)	0.204 (0.061)
TN-3	10	10	0.031 (0.018)	0.216 (0.067)
Sexual means	**19.1**	**12.0** (0.6)	**0.059** (0.009)	**0.191** (0.019)
Apomictic populations (using a full data set)				
IL-1	30	N/A	0.808 (0.100)	0.419 (0.052)
IL-2	25	N/A	0.500 (0.129)	0.262 (0.065)
OH-1	16	N/A	0.750 (0.112)	0.387 (0.058)
Apomictic meanss	**23.7**		**0.686** (0.067)	**0.356** (0.035)

Appendix A Table 7. Pairwise Nei's unbiased genetic identity (below the diagonal) and Nei's unbiased genetic distance (above the diagonal) for all *Boechera laevigata* populations using a full data set. IL-1, IL-2, and OH-1 are considered to be apomictic populations, with all other populations considered to be sexually reproducing.

Population	**AL-1**	**AL-2**	**AR**	**KY**	**MD**	**NC-1**	**NC-2**	**OH-2**	**PA**	**TN-1**	**TN-2**	**TN-3**	**IL-1**	**IL-2**	**OH-1**
AL-1	-	1.035	0.942	0.758	0.460	0.337	0.685	0.312	0.838	0.896	0.852	0.682	0.658	0.700	0.772
AL-2	0.355	-	0.904	0.977	1.516	1.506	1.410	1.592	1.398	0.655	1.057	0.995	0.690	1.019	0.811
AR	0.390	0.405	-	0.469	0.811	0.748	1.610	0.965	1.580	0.422	0.502	0.397	0.826	0.368	0.761
KY	0.469	0.376	0.626	-	0.865	0.736	1.691	1.015	1.691	0.451	0.338	0.259	0.628	0.542	0.689
MD	0.631	0.220	0.444	0.421	-	0.370	0.760	0.215	0.804	1.019	0.894	0.778	0.691	0.381	0.634
NC-1	0.714	0.222	0.473	0.479	0.691	-	0.523	0.298	0.993	0.993	0.717	0.653	0.762	0.530	0.689
NC-2	0.504	0.244	0.200	0.184	0.468	0.593	-	0.471	0.476	1.482	1.671	1.329	0.796	0.932	0.738
OH-2	0.732	0.203	0.381	0.363	0.806	0.742	0.624	-	0.670	1.159	0.994	0.927	0.674	0.539	0.665
PA	0.433	0.247	0.206	0.184	0.448	0.370	0.621	0.512	-	1.490	1.787	1.517	0.484	0.684	0.330
TN-1	0.408	0.519	0.655	0.637	0.361	0.370	0.227	0.314	0.225	-	0.482	0.375	0.616	0.625	0.494
TN-2	0.426	0.348	0.605	0.713	0.409	0.488	0.188	0.370	0.167	0.617	-	0.246	0.818	0.594	0.585
TN-3	0.506	0.370	0.672	0.772	0.459	0.521	0.265	0.396	0.219	0.687	0.782	-	0.723	0.477	0.670
IL-1	0.518	0.501	0.438	0.534	0.501	0.467	0.451	0.510	0.616	0.540	0.441	0.485	-	0.479	0.205
IL-2	0.496	0.361	0.692	0.582	0.683	0.589	0.394	0.583	0.505	0.535	0.552	0.620	0.620	-	0.431
OH-1	0.462	0.444	0.467	0.502	0.531	0.502	0.478	0.514	0.719	0.610	0.557	0.512	0.814	0.650	-

Appendix A Table 8. Pairwise Nei's unbiased genetic identity (below the diagonal) and Nei's unbiased genetic distance (above the diagonal) for all *Boechera laevigata* populations using a data set of only unique multilocus genotypes. IL-1, IL-2, and OH-1 are considered to be apomictic populations, with all other populations considered to be sexually reproducing.

Population	**AL-1**	**AL-2**	**AR**	**KY**	**MD**	**NC-1**	**NC-2**	**OH-2**	**PA**	**TN-1**	**TN-2**	**TN-3**	**IL-1**	**IL-2**	**OH-1**
AL-1	-	0.848	0.990	0.837	0.527	0.387	0.661	0.351	0.798	0.919	0.897	0.758	0.643	0.687	0.355
AL-2	0.428	-	0.895	0.968	1.487	1.454	1.452	1.545	1.412	0.662	1.047	0.995	0.610	0.962	0.357
AR	0.371	0.408	-	0.445	0.805	0.736	1.575	0.945	1.514	0.424	0.489	0.391	0.755	0.297	0.306
KY	0.433	0.380	0.641	-	0.850	0.712	1.649	0.987	1.627	0.433	0.313	0.249	0.572	0.467	0.212
MD	0.591	0.226	0.447	0.427	-	0.344	0.725	0.209	0.816	1.013	0.879	0.778	0.628	0.322	0.185
NC-1	0.679	0.234	0.479	0.490	0.709	-	0.496	0.274	0.990	0.969	0.694	0.644	0.670	0.431	0.209
NC-2	0.516	0.234	0.207	0.192	0.484	0.609	-	0.428	0.511	1.497	1.636	1.335	0.755	0.926	0.309
OH-2	0.704	0.213	0.389	0.373	0.811	0.761	0.652	-	0.659	1.142	0.961	0.913	0.640	0.451	0.205
PA	0.450	0.244	0.220	0.196	0.442	0.371	0.600	0.517	-	1.487	1.744	1.509	0.449	0.681	0.000
TN-1	0.399	0.516	0.654	0.649	0.363	0.379	0.224	0.319	0.226	-	0.474	0.375	0.546	0.550	0.043
TN-2	0.408	0.351	0.613	0.731	0.415	0.500	0.195	0.382	0.175	0.622	-	0.242	0.737	0.523	0.124
TN-3	0.468	0.370	0.676	0.780	0.459	0.525	0.263	0.401	0.221	0.687	0.785	-	0.661	0.414	0.222
IL-1	0.526	0.543	0.470	0.564	0.534	0.512	0.470	0.527	0.638	0.579	0.478	0.516	-	0.376	0.000
IL-2	0.503	0.382	0.743	0.627	0.725	0.650	0.396	0.637	0.506	0.577	0.593	0.661	0.686	-	0.000
OH-1	0.701	0.700	0.736	0.809	0.831	0.811	0.734	0.814	1.110	0.958	0.884	0.801	1.359	1.072	-

Appendix A Table 9. Pairwise comparison of F_{ST} values for all *Boechera laevigata* populations using a full data set. Significance values are given above the diagonal. IL-1, IL-2, and OH-1 are considered to be apomictic populations, with all other populations considered to be sexually reproducing.

Population	**AL-1**	**AL-2**	**AR**	**KY**	**MD**	**NC-1**	**NC-2**	**OH-2**	**PA**	**TN-1**	**TN-2**	**TN-3**	**IL-1**	**IL-2**	**OH-1**
AL-1	-	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
AL-2	0.663	-	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
AR	0.611	0.870	-	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
KY	0.517	0.898	0.761	-	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
MD	0.403	0.877	0.777	0.768	-	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
NC-1	0.388	0.870	0.771	0.758	0.592	-	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
NC-2	0.545	0.943	0.904	0.942	0.821	0.766	-	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
OH-2	0.321	0.923	0.840	0.865	0.499	0.593	0.912	-	0.001	0.001	0.001	0.001	0.001	0.001	0.001
PA	0.641	0.953	0.922	0.958	0.873	0.870	0.973	0.950	-	0.001	0.001	0.001	0.001	0.001	0.001
TN-1	0.404	0.562	0.444	0.396	0.510	0.546	0.613	0.524	0.670	-	0.001	0.001	0.001	0.001	0.001
TN-2	0.499	0.864	0.717	0.627	0.710	0.706	0.885	0.774	0.922	0.373	-	0.001	0.001	0.001	0.001
TN-3	0.452	0.849	0.663	0.541	0.673	0.678	0.858	0.741	0.907	0.318	0.440	-	0.001	0.001	0.001
IL-1	0.393	0.627	0.612	0.522	0.509	0.560	0.600	0.517	0.579	0.360	0.527	0.498	-	0.001	0.001
IL-2	0.478	0.785	0.585	0.626	0.516	0.605	0.749	0.612	0.761	0.431	0.596	0.543	0.431	-	0.001
OH-1	0.426	0.726	0.663	0.596	0.540	0.595	0.665	0.568	0.613	0.327	0.516	0.525	0.232	0.449	-

Appendix A Table 10. Pairwise comparison of F_{ST} values for all *Boechera laevigata* populations using an MLG only data set. Significance values are given above the diagonal. IL-1, IL-2, and OH-1 are considered to be apomictic populations, with all other populations considered to be sexually reproducing.

Population	**AL-1**	**AL-2**	**AR**	**KY**	**MD**	**NC-1**	**NC-2**	**OH-2**	**PA**	**TN-1**	**TN-2**	**TN-3**	**IL-1**	**IL-2**	**OH-1**
AL-1	-	0.001	0.001	0.001	0.001	0.001	0.003	0.007	0.002	0.001	0.001	0.001	0.001	0.001	0.006
AL-2	0.506	-	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002
AR	0.594	0.825	-	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
KY	0.490	0.851	0.734	-	0.001	0.001	0.003	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
MD	0.416	0.810	0.762	0.749	-	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
NC-1	0.371	0.802	0.745	0.718	0.546	-	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
NC-2	0.357	0.889	0.861	0.883	0.701	0.627	-	0.004	0.033	0.001	0.001	0.001	0.002	0.003	0.028
OH-2	0.257	0.878	0.819	0.838	0.441	0.501	0.762	-	0.007	0.001	0.001	0.001	0.001	0.001	0.006
PA	0.386	0.877	0.854	0.871	0.709	0.731	0.786	0.782	-	0.001	0.001	0.001	0.010	0.005	0.031
TN-1	0.383	0.453	0.424	0.367	0.510	0.512	0.513	0.487	0.504	-	0.001	0.001	0.001	0.001	0.001
TN-2	0.477	0.780	0.690	0.575	0.700	0.665	0.780	0.727	0.775	0.362	-	0.001	0.001	0.001	0.001
TN-3	0.450	0.758	0.638	0.499	0.673	0.644	0.740	0.698	0.742	0.319	0.425	-	0.001	0.001	0.001
IL-1	0.315	0.615	0.689	0.585	0.557	0.574	0.565	0.541	0.465	0.301	0.554	0.524	-	0.006	0.067
IL-2	0.335	0.766	0.601	0.648	0.486	0.545	0.702	0.584	0.630	0.313	0.556	0.487	0.370	-	0.031
OH-1	0.286	0.729	0.727	0.677	0.575	0.584	0.659	0.616	0.467	0.217	0.527	0.530	0.185	0.388	-

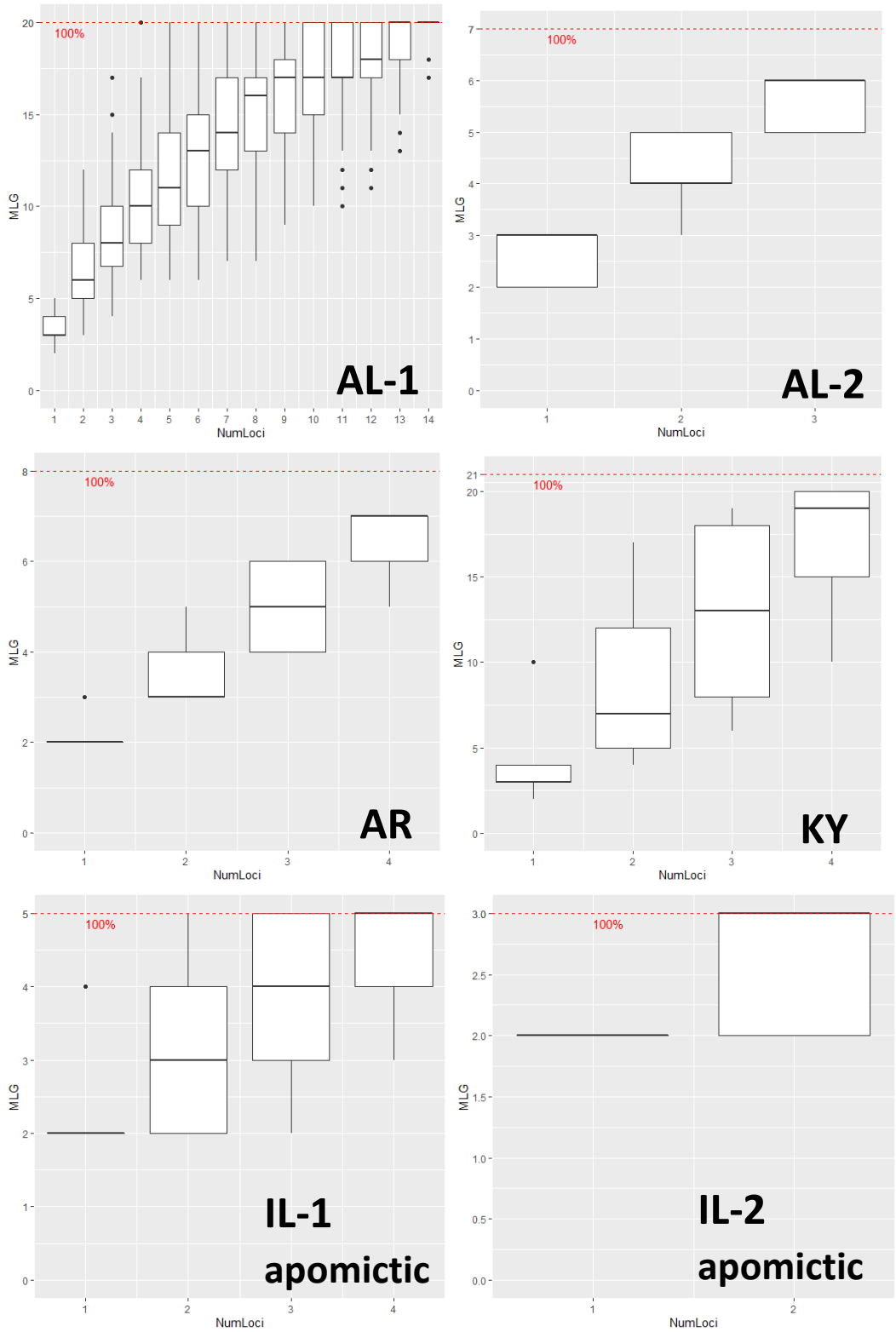
Appendix A Table 11. Pairwise comparison of F'_{ST} values for all *Boechera laevigata* populations using a full data set. IL-1, IL-2, and OH-1 are considered to be apomictic populations, with all other populations considered to be sexually reproducing.

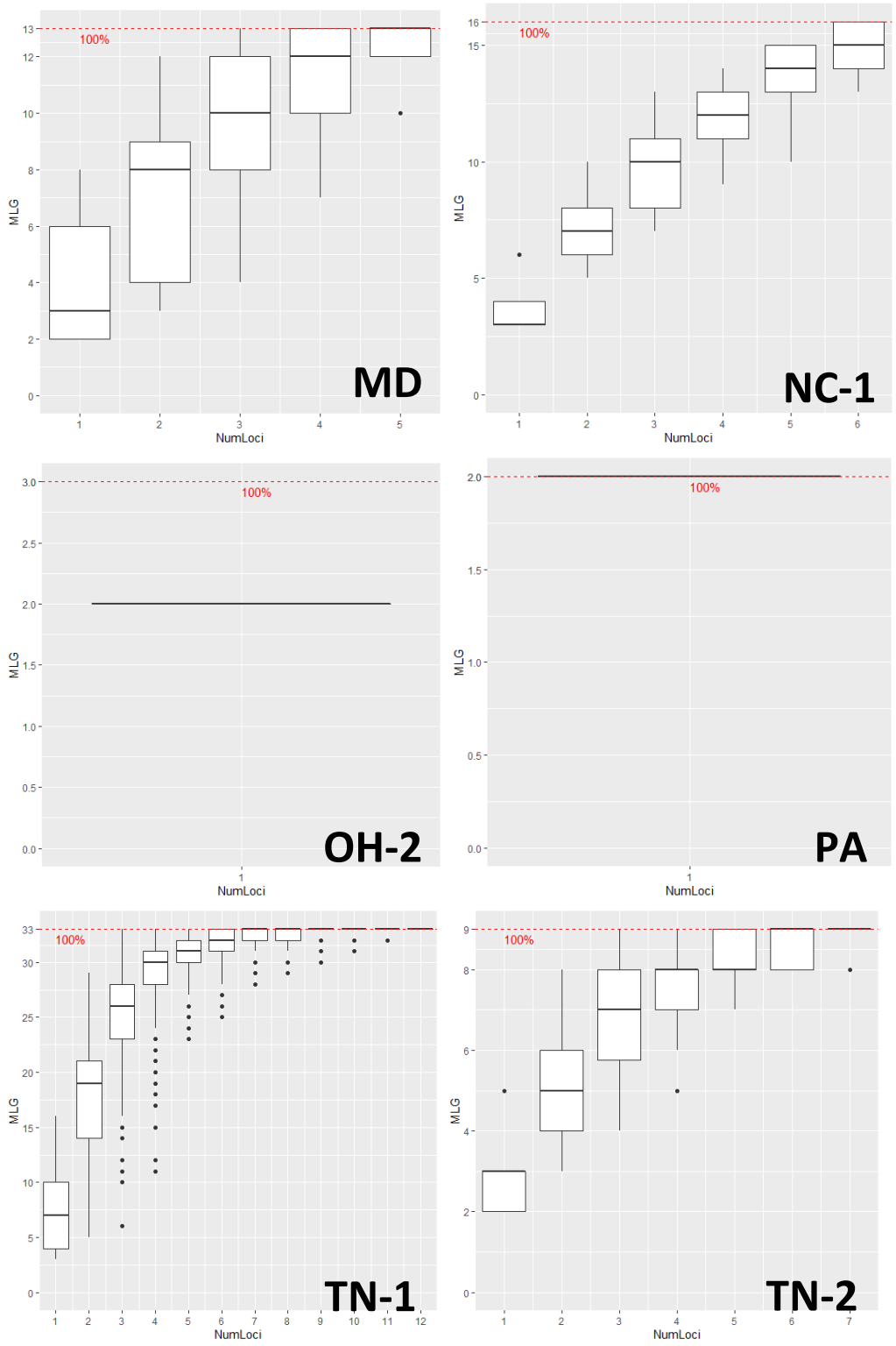
Population	**AL-1**	**AL-2**	**AR**	**KY**	**MD**	**NC-1**	**NC-2**	**OH-2**	**PA**	**TN-1**	**TN-2**	**TN-3**	**IL-1**	**IL-2**	**OH-1**
AL-1	-														
AL-2	0.884	-													
AR	0.851	0.947	-												
KY	0.775	0.962	0.850	-											
MD	0.623	0.973	0.901	0.902	-										
NC-1	0.567	0.971	0.892	0.884	0.717	-									
NC-2	0.775	0.986	0.981	0.989	0.916	0.861	-								
OH-2	0.497	0.984	0.939	0.951	0.593	0.695	0.945	-							
PA	0.850	0.988	0.984	0.992	0.944	0.952	0.983	0.975	-						
TN-1	0.755	0.781	0.647	0.623	0.825	0.835	0.915	0.851	0.929	-					
TN-2	0.785	0.953	0.829	0.733	0.881	0.856	0.979	0.916	0.987	0.615	-				
TN-3	0.720	0.945	0.773	0.644	0.849	0.832	0.963	0.896	0.980	0.532	0.558	-			
IL-1	0.686	0.819	0.835	0.752	0.759	0.799	0.827	0.762	0.751	0.655	0.796	0.761	-		
IL-2	0.742	0.923	0.715	0.784	0.671	0.769	0.903	0.776	0.881	0.700	0.778	0.718	0.653	-	
OH-1	0.733	0.882	0.846	0.803	0.760	0.800	0.846	0.784	0.731	0.589	0.735	0.761	0.384	0.648	-

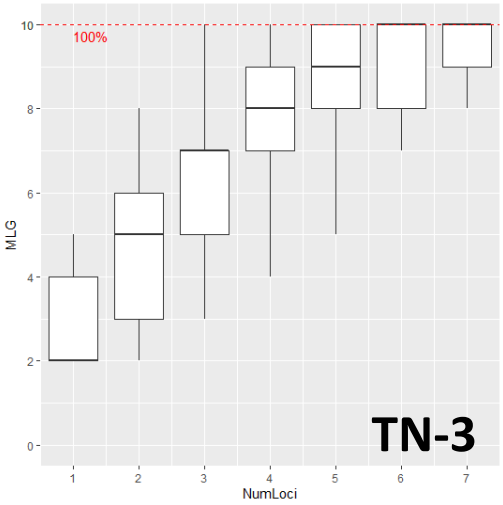
Appendix A Table 12. Pairwise comparison of F'_{ST} values for all *Boechera laevigata* populations using an MLG only data set**.** IL-1, IL-2, and OH-1 are considered to be apomictic populations, with all other populations considered to be sexually reproducing.

Population	**AL-1**	**AL-2**	**AR**	**KY**	**MD**	**NC-1**	**NC-2**	**OH-2**	**PA**	**TN-1**	**TN-2**	**TN-3**	**IL-1**	**IL-2**	**OH-1**
AL-1	-														
AL-2	0.790	-													
AR	0.853	0.928	-												
KY	0.779	0.943	0.829	-											
MD	0.658	0.957	0.893	0.892	-										
NC-1	0.578	0.953	0.877	0.861	0.676	-									
NC-2	0.627	0.974	0.971	0.977	0.849	0.764	-								
OH-2	0.443	0.974	0.929	0.939	0.537	0.612	0.842	-							
PA	0.688	0.970	0.967	0.974	0.866	0.896	0.867	0.885	-						
TN-1	0.751	0.721	0.633	0.593	0.824	0.817	0.882	0.830	0.879	-					
TN-2	0.786	0.922	0.810	0.688	0.875	0.831	0.956	0.894	0.959	0.604	-				
TN-3	0.741	0.910	0.755	0.605	0.849	0.812	0.928	0.875	0.940	0.533	0.543	-			
IL-1	0.626	0.803	0.861	0.779	0.772	0.790	0.808	0.772	0.677	0.586	0.794	0.761	-		
IL-2	0.639	0.913	0.712	0.786	0.631	0.707	0.885	0.743	0.817	0.584	0.740	0.660	0.597	-	
OH-1	0.605	0.879	0.871	0.833	0.765	0.775	0.843	0.802	0.622	0.438	0.724	0.745	0.317	0.593	-

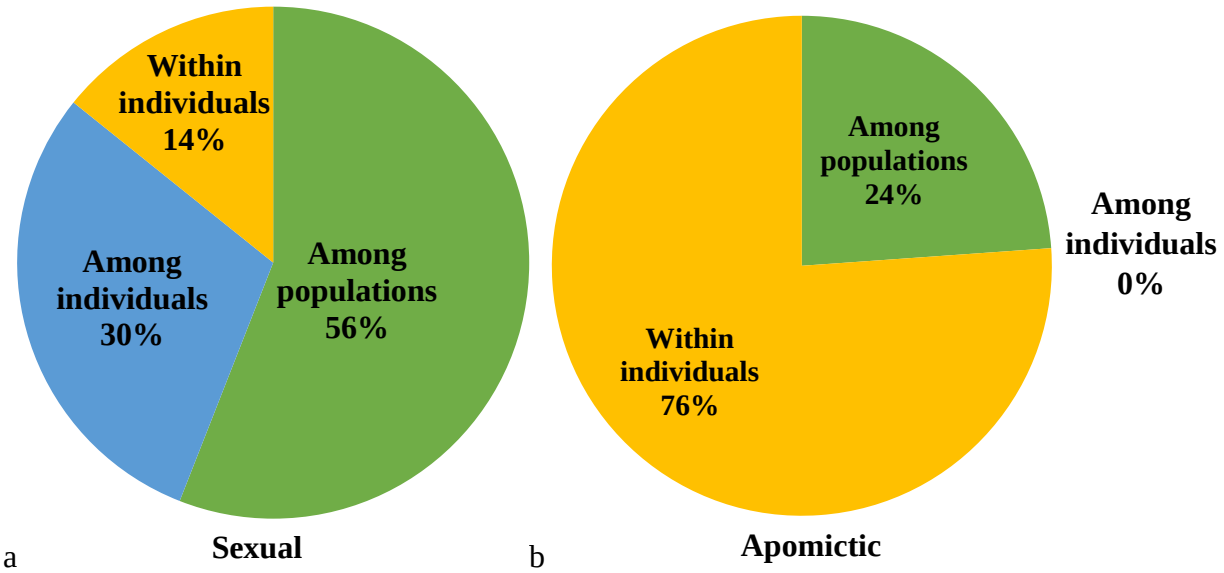
Appendix B



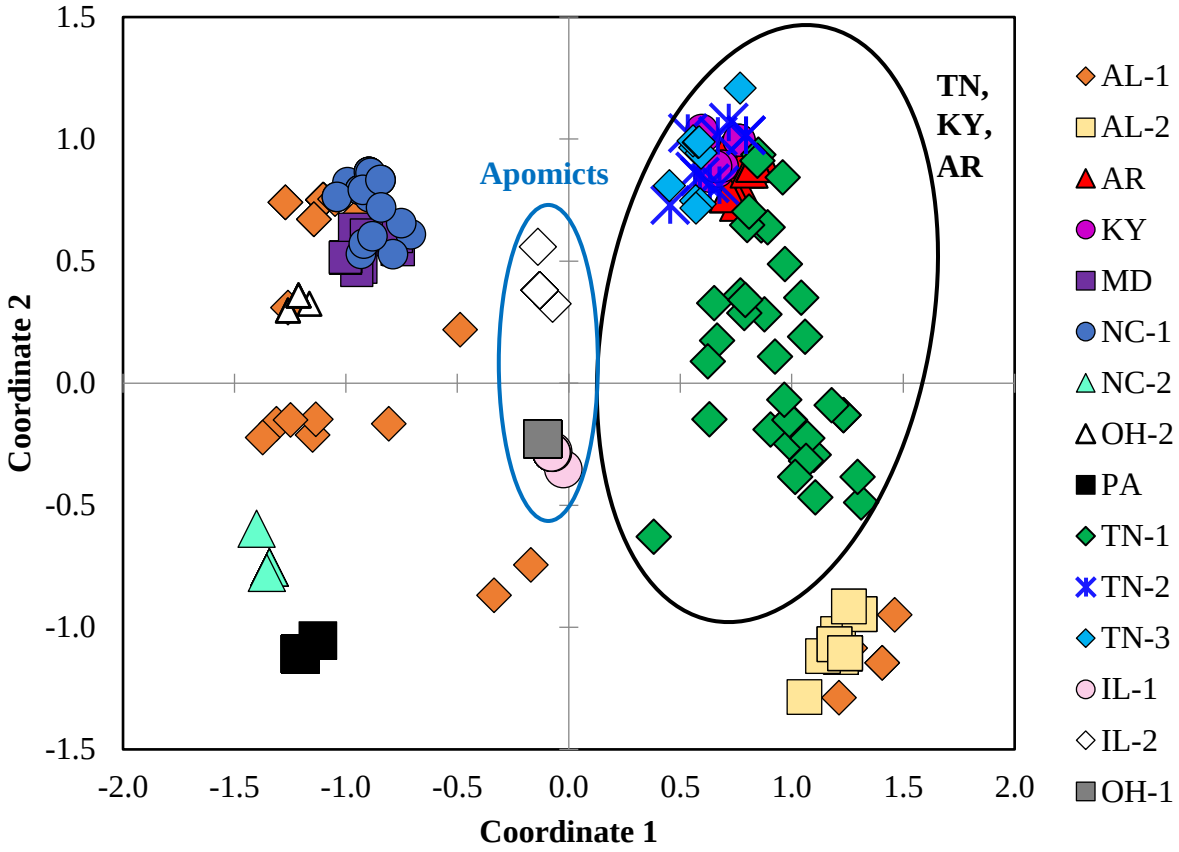

MD
NC-1
OH-2
PA
TN-1
TN-2
100%
MLG
NumLoci



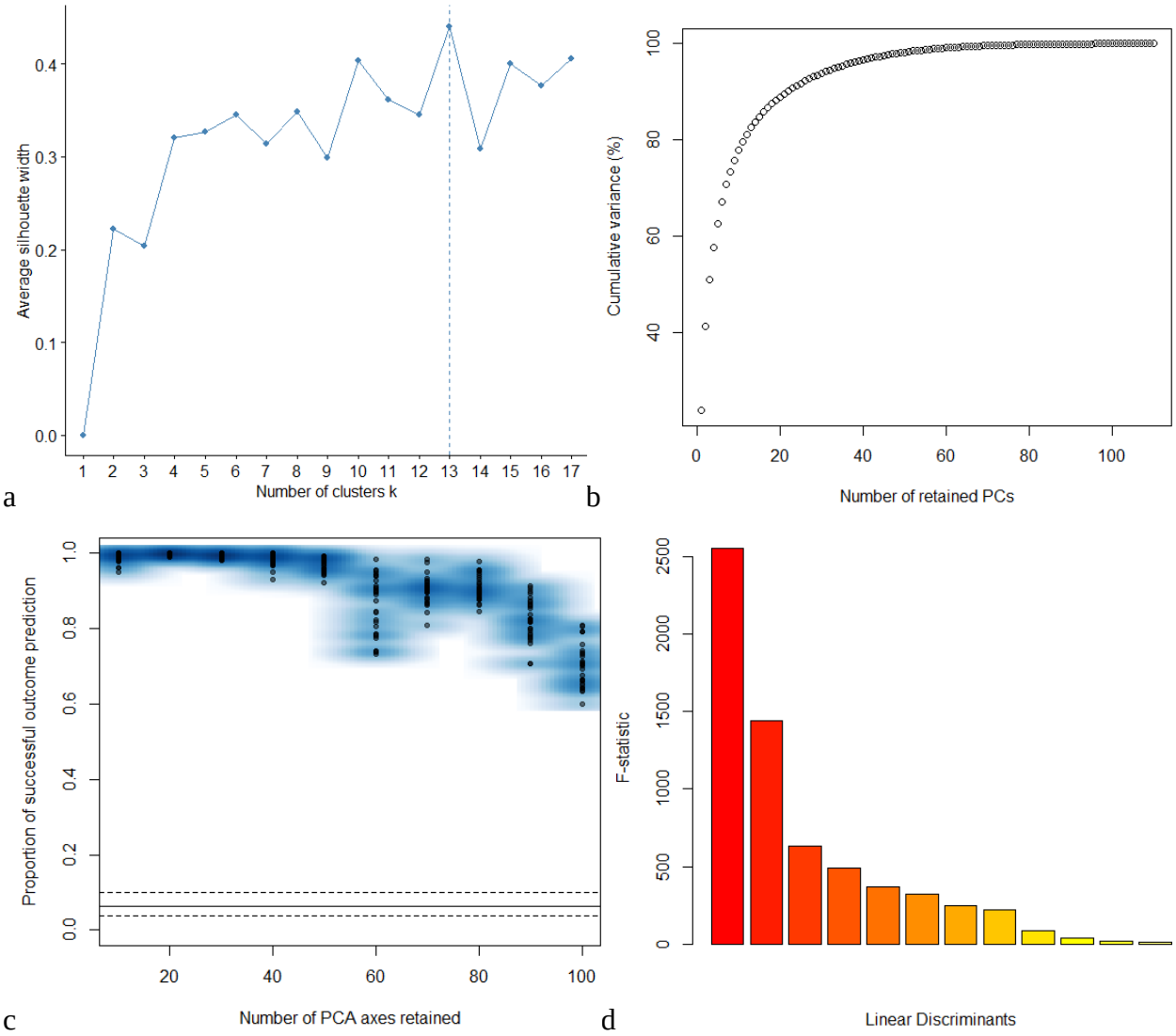
Appendix B Figure 1. Genotype accumulation curves for 13 *B. laevigata* populations. Curves indicate whether the number of polymorphic loci in each population is sufficient to discriminate between clonal individuals in a population by randomly sampling loci without replacement and finding the number of observed multilocus genotypes. The above curves indicate that the loci utilized in this study provided adequate power to discriminate individuals for nine of 15 populations (AL-1, IL-1, IL-2, MD, NC-1, PA, TN-1, TN-2, and TN-3) but power was inadequate to discriminate individuals for four populations (Al-2, AR, KY, OH-2). Curves for OH-1 and NC-2 could not be generated due to the lack of genotypic diversity in these populations.



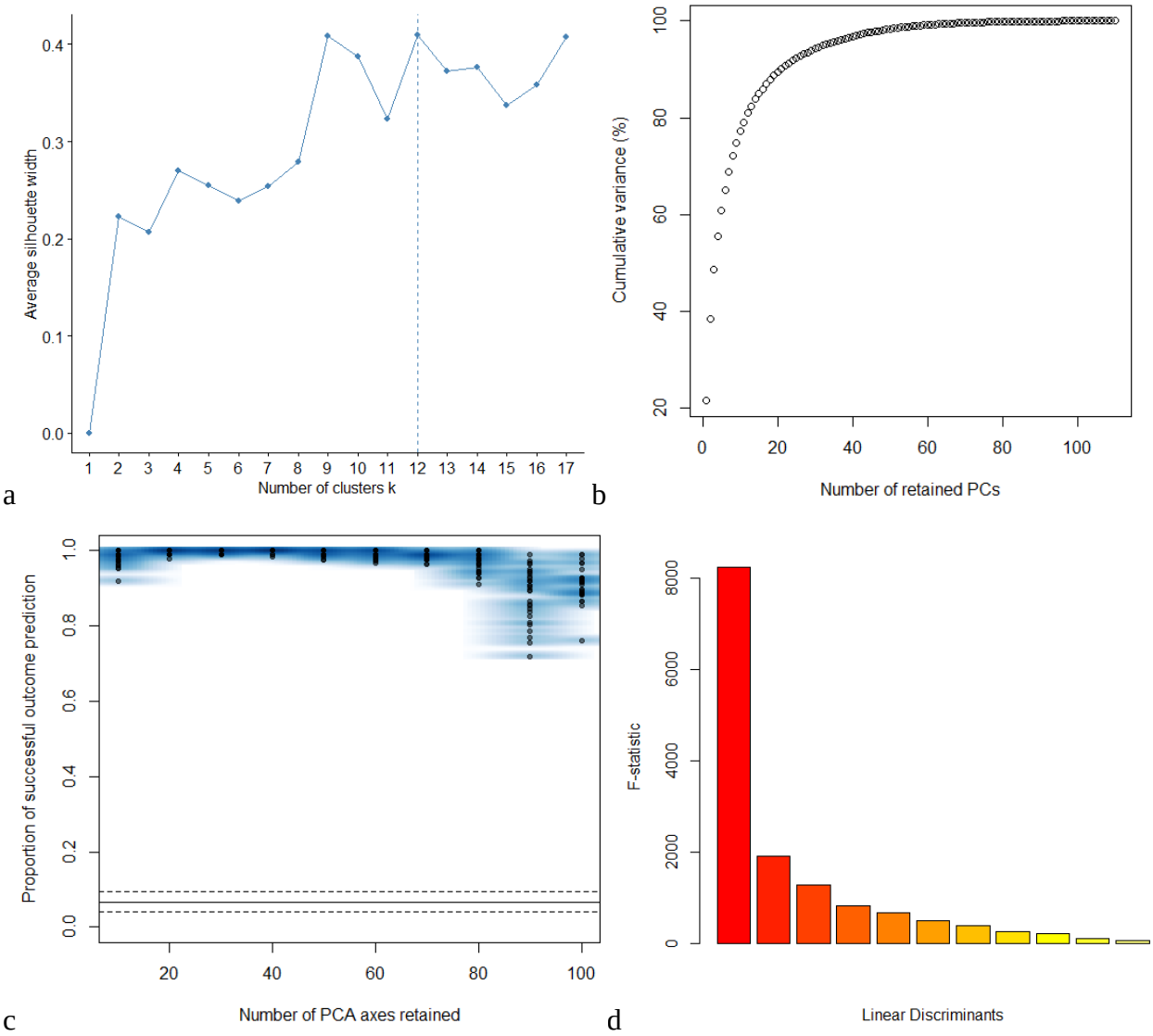
Appendix B Figure 2. AMOVA results for *Boechera laevigata* for the (a) 12 sexual populations using a unique MLGs only data set and (b) three apomictic populations using a full data set -- thus using what would be "wrong" data sets given their "sexual" and "apomictic" designations. However, these data have been provided because these populations' designations as "sexual" and "apomictic" have not been confirmed by cytological analysis of seeds.



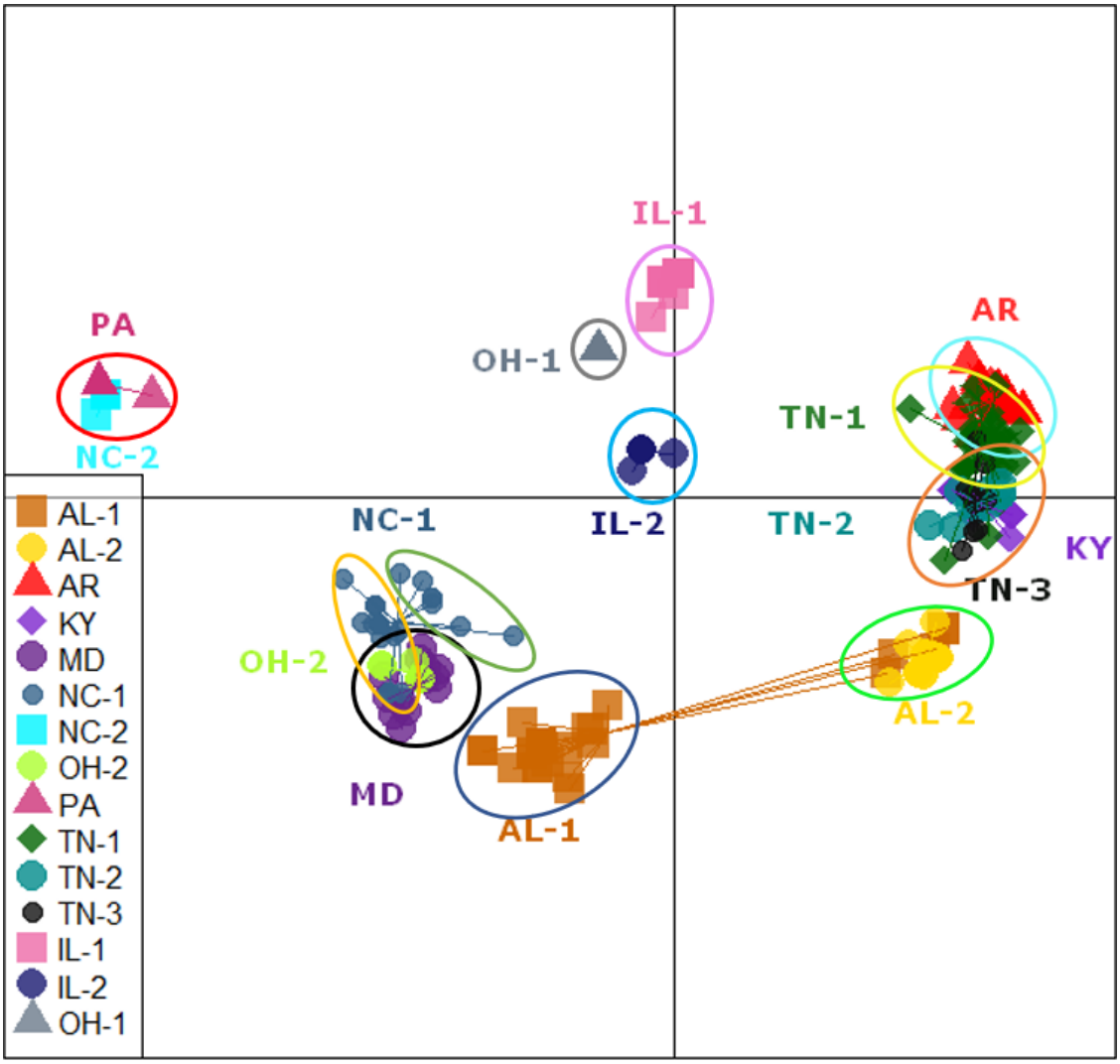
Appendix B Figure 3. Principal coordinate analysis via distance matrix with data standardization from 15 sampled *Boechera laevigata* populations using a full data set. Coordinate 1 explains 30.6% of genetic variation and coordinate 2 explains 21.1% of genetic variation (51.7% total).



Appendix B Figure 4. Genetic data for the following analyses included the full data set for sexual *Boechera laevigata* populations and the unique MLG only data set for apomictic populations. (a) Optimal number of genetic clusters (*K*=13), determined by K-means cluster analysis and indicated by the dotted line using the silhouette method in the R package *Factoextra* (Kassambara and Mundt 2020). Genetic data for the analysis included the (b) Genetic variance explained by principal components analysis (PCA): Number of PCs vs cumulative variance. For the *find.clusters* function, all PCs were retained (120 PCs). (c) Cross-validation results from DAPC for *Boechera laevigata*. The optimal number of PCA axes to retain was 22, as this was the value associated with the lowest root mean square error (root mean square = 0.0075). Root mean square error values are not provided in this graph. (d) Discriminant functions. A total of three discriminant functions were retained for the DAPC for K=13.



Appendix B Figure 5. Genetic data for these analyses included the full data set for all *Boechera laevigata* populations. (a) Optimal number of genetic clusters (*K* = 12), determined by K-means cluster analysis and indicated by the dotted line using the silhouette method in the R package *Factoextra* (Kassambara and Mundt 2020). (b) Genetic variance in explained by PCA: Number of PCs vs cumulative variance. For the *find.clusters* function, all PCs were retained (120 PCs). (c) Cross-validation results from DAPC for *K* = 12. The optimal number of PCA axes to retain was 29, as this was the value associated with the lowest root mean square error (root mean square = 0.0047). Root mean square error values are not provided in this graph. (d) Discriminant functions. A total of three discriminant functions were retained for the DAPC for *K* = 12.



Appendix B Figure 6. Discriminant analysis of principal components for 15 *Boechera laevigata* populations using a full data set. Each encircled group represents a genetic cluster as indicated through *K* means clustering (*K* = 12). A total of 29 principal components and 3 discriminant functions were retained.