Population Genetics and Bumble Bee Conservation: Saving Species by Thinking Small

Cady Greenslit

Follow this and additional works at: https://lux.lawrence.edu/luhp

Part of the Biology Commons, Ecology and Evolutionary Biology Commons, Entomology Commons, and the Genetics Commons

Recommended Citation
https://lux.lawrence.edu/luhp/117

This Honors Project is brought to you for free and open access by Lux. It has been accepted for inclusion in Lawrence University Honors Projects by an authorized administrator of Lux. For more information, please contact colette.brautigam@lawrence.edu.
Population Genetics and Bumble Bee Conservation

Saving Species by Thinking Small

Lawrence University

Cady Greenslit

Research Mentor: Israel del Toro
Abstract:

Pollinators have experienced a severe decline over the last few decades, and this includes bumble bee populations (g. *Bombus*). Bumble bees are important native pollinators, and here I investigate the health of local populations with molecular tools. The field of conservation genetics has created useful methodology for investigating the health and informing management strategies of threatened populations. This work investigates and describes the applications of population genetics, which uses span across the board. These applications are then brought back into the context of bumble bee conservation, and how they fit into the experimental plan I originally designed. I designed an experiment that utilized microsatellite loci to understand population dynamics of bumble bees (*Bombus ssp.*). Microsatellites are DNA segments that are highly variable within a population. Specimens were obtained via net collection throughout the summer of 2017. This thesis goes into the development of my experimental protocol and the future steps that would need to be taken to fully understand the local population dynamics of *Bombus* species. This project will go into the applications of research projects that identify areas in which genetic exchange is taking place, highlighting areas of importance for habitat protection and maintaining threatened populations. The aspects of genetic exchange between populations that essential for species survival over time are also discussed.
Table of Contents:

Abstract .......................................................................................................................... 2
Introduction ...................................................................................................................... 9-31
  Population Genetics and Conservation Case Studies ................................................. 7-17
  Primer on Landscape Genetics .................................................................................. 17-20
  Microsatellites ............................................................................................................. 20-28
  Bumble Bee Colony Structure and the Effects on Genetics ........................................ 28-30
  My Hypothesis ............................................................................................................ 30-31
Material and Methods ................................................................................................... 32-53
  Field Collection and Species Identification ............................................................... 32-35
  DNA Extraction .......................................................................................................... 35-41
  Primer Development ................................................................................................... 41-43
  PCR .............................................................................................................................. 43-50
  Gel Electrophoresis ..................................................................................................... 50-51
  “A Poor Man’s Approach to Genotyping…” (Schuelke, 2000) ................................. 51-53
Results .......................................................................................................................... 54-62
  Species Diversity and Distribution ............................................................................ 54-56
  DNA Extraction .......................................................................................................... 56-57
  Establishing Final Protocol ........................................................................................ 57-59
  The Final Protocol ....................................................................................................... 59-62
Discussion/Conclusion ................................................................................................. 63-69
  Microsatellite Length Analysis .................................................................................. 62
  Statistics for Microsatellite Analysis ......................................................................... 63-64
  Overlaying Results on a Geographic Map ................................................................... 64-65
  Future Directions ........................................................................................................ 66-67
  Broader Implications .................................................................................................. 67-68
References ..................................................................................................................... 69-73

Greenslit 3
Introduction:

Pollinators play an integral role in the food production system in the United States. The monetary value of domestic pollination services provided by native insects is valued at over $3 million (Losey & Vaughan, 2006). Bumble bees (g. *Bombus*) perform a significant amount of the pollination services that humans rely on, from large scale agriculture to single home gardens. The value of bumble bees is also shown directly a study that documented how fruit production increased with native bee diversity. Native bees increase fruit production, whereas honey bee do not, even when the population density is held constant (Mallinger & Gratton, 2015). The authors attributed this pattern to the different methods of pollination associated with bumble and other native bees in contrast to honey bees. Honey bees are more likely to pollinate apple trees with multiple blooming flowers that provide a large supply of food. In contrast, native bees were shown to have no such pollination preference for higher floral count (Mallinger & Gratton, 2015). Due to this lack of preference, trees with a lower floral count still get pollinated which increases the overall crop production of apples. A lack of preference in pollination not only increases fruit production in the case of apple orchards, but it also increases gene flow within the orchard. Therefore, there is not just immediate increase in food production, but also an increase in genetic variability of apple tree populations.
For instance, if honeybees were the only pollinators, then the apple trees that produce many blooming flowers at once would make up the majority of trees able to pass on their genes, and genetic diversity would decrease. A large portion of gene flow in flowering plants is due to pollination (Elistrand, 1992). This situation creates the same type of dilemma that occurs with artificial selection performed by agriculturists trying to increase yield. While increasing the number of trees that produce a lot of flowers may seem beneficial due to their ability to produce more fruit, by decreasing genetic diversity of a population, the number of alleles that could possibly be beneficial in a different environment decreases. As I will expand on later, populations that have experienced a decrease in genetic diversity may be less likely to survive climate change or stochastic events such as drought or disease, due to a decreased ability to adapt. This is one reason of many that the health of native bee populations, such as bumble bees, has a large effect on agricultural systems (Mallinger & Gratton, 2015). The decline of bumble and other native bees could lead to declines in food production in the United States.

Along with increasing food production, native bees are also important to non-agricultural native ecosystems due to their coevolution with plant species that depend on *Bombus* pollination (Borrell, 2005; Whittall & Hodges, 2007). Plants that are pollinated by long tongued bees have developed flowers with long nectar spurs, making food resources only available to species with long enough tongues to access their nectar (Borrell, 2005). This increases pollinator fidelity, which can increase pollination efficiency as their pollen is more likely to be transferred to another flowers of the same species (Borrell, 2005). There are even phenotypic differences within a species that are pollinated by both bumble bees and hummingbirds, but within different parts of their range (Whittall & Hodges, 2007). This pollination specialization and the resulting phenotypic and genetic differences supports the hypothesis that different pollinators can drive a
speciation event (Whittall & Hodges, 2007). As these differences accumulate, the various genetic lines become more reliant on their specific pollinator (Whittall & Hodges, 2007). Therefore, native plants that have evolved a close relationship with their bee pollinators can be heavily affected by native bee decline. This explains why many wild plant species have shown to have decreased reproductive success in areas in which bumble bee populations are declining (Potts et al., 2017).

The value of native bees has also been documented in urban landscapes (Hall et al., 2017). Though cities are sometimes perceived as voids of biodiversity, previous studies have shown that urban areas can have higher bee biodiversity than surrounding rural areas (Hall et al., 2017). Systematic pesticide and herbicide application, and restricted temporal food resources due to monoculture make rural habitat less suitable for native bees (Hall et al., 2017). Other factors contributing to the decline of bees within rural landscapes include mechanization of agriculture, agricultural intensification, and an increase in acreage being farmed in some areas (Herrmann, Westphal, Moritz, & Steffan-Dewenter, 2007). In contrast, urban areas are filled with floral resources, both native and nonnative, that bloom at different times of the year (Hall et al., 2017). Therefore, there is a dynamic relationship between urban agriculture and aesthetic garden that support these pollinators, and the pollinators that allow these urban plants to thrive.

Bumble bee populations in the United States have declined over the past two decades. In general, this decline has been attributed to land use changes, loss of perennial habitats, and habitat fragmentation (Herrmann et al., 2007; Potts et al., 2017). This news is especially concerning because of the documented increase in susceptibility to pathogens in declining bumble bee populations and species (Cameron et al., 2011). In combination with the aforementioned drivers of diversity loss. Bumble bee populations demonstrate a low resilience to
climate change, which reduces total species ranges and can lead to local and regional extinctions (Kerr et al., 2015). Population ranges are shrinking in the southern range extent without gaining any northern range as compared to shifting their entire range northward as climate patterns change(Kerr et al., 2015), leading to an ever-increasing shrinkage of total species range. As the effects of climate change continue to become more severe, we can expect a decrease in bumble bee populations. I designed applied techniques in population and landscape genetics to explore the shifting dynamics in *Bombus* populations.

Population Genetics and Conservation Case Studies

Population genetics integrates ecology and evolutionary biology(Lande, 1988; Wan, Wu, Fujihara, & Fang, 2004) because it is necessary to both understand the ecology of the organism being studied, and the traits, corresponding genes and gene families that contribute to the fitness of individuals within a population (Lande, 1988). By using principles from both, it is possible to determine what level of genetic differentiation between individuals is necessary to maintain a healthy population. As I will go into later, genetic analysis of and between populations is enriched by an understanding of the ecology of a species. Conversely, ecological dynamics, such as mating behaviors and vagility, are much better understood when genetic data is taken into account (Elistrand, 2018; McCracken, Johnson, & Sheldon, 2001).

Population genetics contributes to conservation by focusing on the levels of genetic diversity that confer fitness to populations of organisms. These genetic components are the alleles that express traits that increase the survival of an individual and the families of genes that work together to contribution to one or many functions within an organism. Every version of a trait has both positive and negative effects on fitness, which is why most healthy populations...
have a mix of alleles. Another aspect attributed to healthy populations is a variety of haplotypes. By having multiple combinations of alleles (haplotypes), there are many different combinations of phenotypes or traits within a population. It is important to have many different versions of traits and trait combinations within a population, especially considering the environmental stochasticity that an individual organism can experience. A trait that can be beneficial in one environment, can be quite detrimental in another. Population genetics studies the composition of populations to determine if there is a mix of alleles and haplotypes that are associated with a healthy population and define what a healthy mix of alleles would look like. Different levels of genetic diversity correspond to different levels of adaptability and fitness within an environment (Geist, 2014). As an example, previous studies have shown that bumble bee populations with greater genetic diversity have lower parasite loads and have greater reproductive success (Baer & Schmid-Hempel, 1999). Inbreeding occurs when closely related individuals mate. Inbred colonies are usually smaller than ones with greater genetic diversity (Herrmann et al., 2007), demonstrating the usefulness of experiments that investigate genetic variability in bumble bee populations.

In the conservation world, population genetics have been used to determine the minimum viable population for a species (Chase, Kesseli, & Bawa, 2018). This is a measure of the minimum number of reproducing individuals a population needs to persist for the foreseeable future. Small populations are more prone to use lose genetic information due to random events (e.g. extreme weather or disturbance events). When this loss of genetic diversity is accompanied by loss of habitat, or another detrimental effects, this can create what is called an extinction vortex (Geist, 2014). This refers to a set of conditions that make it almost certain that a species will go extinct (Geist, 2014). Inbreeding is more common in small populations, which can lead to
a decrease in individual fitness, due to an increased expression of recessive traits (Lande, 1988). While not all recessive traits are deleterious, many have the possibility to be so. Dominant traits can also be detrimental to the survival of an individual, but detrimental dominant alleles are usually lost due to natural selection. Deleterious recessive genes remain in the populations because heterozygotes who express the dominant phenotype are able to survive.

Populations that have experienced a bottleneck usually have a lower amount of heterogeneity within the population (Maruyama & Fuerst, 1985). This loss of alleles can be detrimental within a population because it can eliminate the its ability to adapt. The ability to adapt to new circumstances is becoming ever the more relevant with continued human modification of habitats and the worsening effects of climate change (Kerr et al., 2015). When rare alleles are lost that may confer traits that would allow a population to survive the stochastic changes of today’s biosphere, this could be the difference between the survival or the decline of a species. This is especially pertinent because it has been shown that lack of gene flow can become a negative feedback loop of population decline (Ellis & Goulson, 2006). Small populations lose more unique alleles, and then population size decreases due to this loss of alleles. This has been seen in bumblebee colonies, as colony size decreases with decreased genetic diversity (Herrmann et al., 2007).

The field of population genetics has many applications. To fully understand what I could be able to do with the data I planned to collect, I found research papers that detailed these applications. I then use these experiments as case studies, and reflect on how the concepts from these works can be applied to examination of bumble bee genetics.
While utilizing population genetics for conservation is useful, there are other practical (i.e. economic) applications. This usually focuses on species that have a direct value to humans. Wild populations of domesticated plant species, such as grapevines, have been monitored and preserved as a genetic repository for their agriculturally useful relatives (Grassi et al., 2006). This study also had conservation applications, such as identifying important refugia or habitat and mapping out genetic variability (Grassi et al., 2006), demonstrating that studies can fulfill more than one of the applications shown in Figure 2. Other population genetics studies focus on traits that lead to high agricultural output in domestic plant species, in order to create a more productive genetic line (Pandey et al., 2017). Individuals involved in agriculture or any propagation of species for human consumption can use population genetics to maximize efficiency of the effort put into tending this species, and to maintain the health of these species. While bumble bees do not directly produce an item of value, there are domesticated
populations (Whiteley, 2017) that should also be monitored in the same way. Population genetics could also possibly be used to breed better pollinators.

One of the most typical conservation applications for population genetics occurs in zoos and other captive breeding facilities. These are both examples of places that are usually working with a small number of breeding individuals. Genetics is used to monitor these captive populations via measuring both inbreeding and outbreeding depressions. Inbreeding depressions are the negative fitness effects experienced by offspring of individuals that are too closely related, such as the expression of negative recessive traits. An outbreeding depression is when the offspring of two distantly related individuals are less fit due to how genetically different their parents are. Population genetics can be used to create guidelines in maintaining captive populations, such as the minimum number of founder individuals, and a recommended total population size (Witzenberger & Hochkirch, 2011). Many species are in danger of going extinct in the wild, so that when their captive populations are experiencing constraints, such as small initial populations, they are heavily monitored to maintain the health and viability of these populations. Since captive breeding facilities are controlled environments, the same data can be used to test population health can also be used to understand more about population genetics and how to maintain the health of natural populations, such as bumble bees.

One study that utilizes conservation genetics to understand species with both managed and wild populations focuses on Anadromous Pacific salmon (*Oncorhynchus spp.*) (Waples & Teel, 2018). These two types of populations create two genetic pools that experience different selective pressures. Captive individuals are selectively bred to increase profits for the individuals raising them, while wild individuals are selected for survival in natural settings. The decline of wild populations as well as an increase in hatchery populations was hypothesized to be a cause of
accelerated genetic change in the species (Waples & Teel, 2018). Rapid genetic change can negatively affect the species due to a rapid loss in gene complexes, or functionally or evolutionarily related genes, that have evolved over the lifespan of the species and are necessary for adapting to the environment. The genetic changes seen over time were shown to be significantly higher in hatcheries than those observed during the same time period in wild populations (Waples & Teel, 2018). This is concerning because hatchery eggs or adults could get released into the wild and reproduce. This would weaken the genetic background in wild populations and the species as a whole (Waples & Teel, 2018). This is one example of a study that can be used to inform management and hatchery practices. Though it may be a question of when rather than if hatchery salmon get out and interbreed with wild individuals (Waples & Teel, 2018), it is still important to be aware of the possible effect it may have on the species.

This same framework can be applied to other farmed species that have wild populations as well, such as bumble bees. One study investigated the concern that the genes from commercial bumble bees would migrate into wild populations (both B. impatiens) (Whiteley, 2017). Like the previous study, this worry was fueled by the fact that genes that are artificially selected for when animals are domesticated may not always confer fitness in a natural setting. The bee study found that there was no significant gene flow between wild populations and those used for commercial purposes (Whiteley, 2017). This demonstrates that populations can coexist next to each other without significant gene flow occurring. This is good news because it shows that hybridization between wild and domestic bees, and the subsequent loss of important gene complexes, is not another area of high concern for this threatened genus. If there was significant gene flow, between farmed and natural populations, the Pacific Salmon study could be modified for bumble bees, and managers could emphasize the importance of genetic diversity in domestic populations.
Population genetics can also be used to understand the evolutionary relationship between species, as with the case in a study that focuses on the mottled duck (*Anas Fulvigula*), endemic to the Gulf Coast (McCracken et al., 2001). Studies like this can identify new species who were formerly thought of as only one, identifying new units that should be analyzed individually for conservation purposes. It can also inform the conditions in which species evolved and became genetically isolated. The evolutionary history of these species of ducks was postulated by comparing a mitochondrial sequence (McCracken et al., 2001). It showed that species in this geographic area descend from a common ancestor. This study also demonstrated that there is repeated hybridization occurring between modern species. That lead to the prediction that there is an extant physical barrier to genetic exchange between these species (McCracken et al., 2001). This is because if a species can currently interbreed, and has a shared evolutionary history, their speciation is not due to a genetic or physiological barrier. By understanding the conditions that influence speciation, modern factors that create genetic isolation can be understood. This is a valuable alternative method to the one I based my original experimental design on.

Population genetics is also used to analyze the role of a species’ mobility in gene flow. *Podocnemis expansa* is a species of giant Amazonian river tortoise that used to be common throughout their native region (Pearse et al., 2006). The metapopulation dynamics of these species were only understood when genetic analysis was performed (Pearse et al., 2006). These tortoises have high vagility, and this could have feasibly lead to high genetic exchange between neighboring populations (Pearse et al., 2006). Genetic analysis revealed that each riverbank represented an isolated breeding population (Pearse et al., 2006). This study also showed that this genetic isolation was historic and not due to modern habitat fragmentation (Pearse et al., 2006). This is important information for those in charge of designing and implementing conservation
and management measures for this species as now every breeding population must be thought of independently. In general, it is important to think about the ecological factors that influence gene flow. This includes the mobility of the organism (Elistrand, 2018). Specifically, bumble bees are able to fly 20km a day while foraging and queens have been known to travel 80km a day during the springtime (Osborne et al., 2018). This information informed my original hypothesis.

Population genetics has also been used to improve the study of amphibians. Without genetic analysis, migration studies would have little options besides cutting several toes off an individual, making them recognizable later (Jehle & Arntzen, 2002). This process can be stressful to the individual and may have effects on later fitness (Jehle & Arntzen, 2002). Genetics can also be used to identify illegally smuggled individuals, once genetic signatures of habitat locations have been established (Jehle & Arntzen, 2002). This is an important method for helping combat the illegal pet trade (Jehle & Arntzen, 2002). Amphibians are experiencing many challenges that are contributing to their decline (Jehle & Arntzen, 2002), and population genetics has helped researchers get a better hold of that situation. Innovations like these have also been achieved in the study of nesting density, foraging range (Knight et al., 2005), and specimen collection (Holehouse, Hammond, & Bourke, 2003) of bumble bees.

When keystone species are the subject of genetic analysis, population genetics can become a part of the effort to restore ecosystems (Geist, 2014). Freshwater pearl mussels (*Margaritifera margaritifera*) are integral parts of freshwater systems with their capacity to process particles, making nutrients bioavailable for other organisms, and mixing sediment within the water column (Geist, 2014). An understanding of genetic diversity plays two roles within the conservation of fresh water mussels. In an effort to farm this species that is both important ecologically and that can provide an economic resource, methods have been developed to culture
this species (Geist, 2014). It is necessary for these culturing programs to maintain good genetic diversity in order to maintain the health of these populations, but to also make an effort to mimic the diversity seen in the species in natural populations (Geist, 2014). Maintaining genetic diversity is equally important in maintaining healthy domestic populations of bumble bees. As bumble bees also provide an important ecological service, both bumble bees and freshwater mussels demonstrate how conservation and economic motivations for utilizing population genetics can become intertwined.

In natural populations, genetic analysis can help managers decide whether to combine smaller populations or maintain multiple small, isolated populations (Geist, 2014). The authors of the study wanted to analyze whether the movement of individuals would benefit the metapopulation of freshwater mussels in totality. While larger populations may seem beneficial and a manner of increasing genetic flow, it is necessary to balance inbreeding and outbreeding effects (Geist, 2014). Genetic diversity can be decreased by combining two populations who are quite genetically different from one another (Geist, 2014). This is because the traits of one population can lead to an immediate survival advantage, losing the genetic makeup of the immediately less fit population. Conservation managers must to be able to make informed decisions about how to best support species. This becomes a higher priority when species provide necessary services, such as water filtration in the case of mussels, and pollination, in the case of native bees.

Finally, it is always important to be critical of the results of population genetic studies. Some species of Bombus in North America showed sufficient genetic exchange between populations while being in decline (Cameron et al., 2011). Therefore, the presence of genetic flow does not preclude a population or species from the risk of extinction. Conversely, the
absence of genetic flow and limited genetic diversity does not necessarily result in negative
effects on a species. Population genetics maintains that this decrease in genetic variability due to
bottlenecks then leads to an increased chance of extinction for a population (Maruyama &
Fuerst, 1985). Upon review, this has been shown to not necessary be true. The effect of the
bottleneck depends on the genetic composition of the species, specifically the relationship
between the new allelic composition and associated phenotypes and the adaptive potential of the
population (Bouzat, 2014). Combining physiological and metabolic techniques with molecular
analysis can be used to determine specific associations between genotype and fitness within
certain environments (Geist, 2014). This gives researchers a clearer picture of which populations
are most fit, as opposed to just those that are most genetically diverse. Due to this, it is necessary
to understand that while genetic analysis can predict a specific outcome for a species this should
be carefully considered.

Population genetics has been used to investigate nesting density, foraging range (Knight et
al., 2005), the number of males a queen mates with (Estoup, Scholl, Pouvreau, & Solignac, 1995),
and population structure of several bumble bee species (Ellis & Goulson, 2006; Koch, Looney,
Sheppard, & Strange, 2017). One study estimated nesting density and foraging range by utilizing
the fact that it is easy to recognize individuals from the same colony because all workers are
sisters (Knight et al., 2005). Measuring whether queens were monogamous or polygamous also
relied on the fact that all workers are sisters (Estoup et al., 1995). Estimates of bumble bee
population structure must also take into account the fact that all members of a colony are related
to one another (Ellis & Goulson, 2006; Koch et al., 2017). Therefore, it is necessary to examin
the high relatedness between individuals within colonies that is described in “Bumble
Bee Colony Structure and the Effects on Genetics”

In order to create a more accurate prediction based on the genetic data, a researcher should analyze a species holistically. This includes taking environmental conditions within a habitat and the specific genetics of the species being studied into account. The effects of environmental stress and inbreeding have been shown to decrease population size synergistically (Ellis & Goulson, 2006). Genetic differentiation of a species cannot be accurately analyzed without knowledge of specific ecology, behavior, life history and physiology of a species (Baer & Schmid-Hempel, 1999; Geist, 2014). The effect of a species mobility on genetic flow discussed earlier is not accurate without the consideration of the behavioral ecology of a species. Even species that can travel vast differences still may not have good genetic diversity, due special behavioral connections to local habitat (Karl, Castro, Lopez, Charvet, & Burgess, 2011). Therefore, it is necessary to understand the entire system in order to correctly analyze the genetic analysis of a species.

Primer on Landscape Genetics

Landscape genetics uses the principles of population genetics combined with spatial and environmental data in order to analyze the way the environment and environmental heterogeneity affect the flow of genes across an area (Segelbacher et al., 2010). It has been shown that microsatellite analysis has been strengthened when combined with landscape data (Jehle & Arntzen, 2002). Therefore, its use in studies allows researchers to understand how geographic features affect how populations interact with one another. This is an expansion on the previous ability of population genetics to only examine discrete populations. It also allows us to understand more micro-evolutionary processes that are occurring within a geographic area.
Studies that apply landscape genetics identify geographic barriers to genetic flow, whether naturally occurring or anthropogenically created. This then allows for continued extrapolation; how should we modify our cities to decrease the impact on wildlife, how do we design future cities with these principles in mind?

Landscape genetics can also help you identify which loci within the genome are experiencing selection (Schwartz et al., 2003). This can be used by researchers to understand which traits confer fitness within a particular environment, by testing how genes can affect fitness in the natural systems (Geist, 2014). It also allows the researcher to understand which parts of a genome are losing diversity. If the chosen study site is in an environment that is heavily modified by humans, it could allow a researcher to identify the anthropogenic effect on the environment on the molecular level.

There are four key patterns in landscape genetics: clines, isolation by distance, metapopulation dynamics, and stochastic drivers of diversity (Schwartz et al., 2003). Clines are patterns across continuous gradients. Isolation by distance is a measure of genetic variation over geographic distance. Metapopulations are population of the same species that are separated spatially, but still interact with each other. Stochastic patterns are those not explained by the other three concepts (Schwartz et al., 2003).

An example of landscape genetics in practice is a study investigating bumble bee genetics in the Pacific Northwest. Four sympatric species, species that evolved in the same geographic area, were investigated (Koch et al., 2017). Some of the species had differing ecological niches, specifically limitations for the environments they could survive in (Koch et al., 2017). The results from this study can be used to predict possible effects of climate change on these different species, depending on niche (Koch et al., 2017). As seen in Figure 3, there was less isolation by
distance in species that are able to inhabit broad ranges of elevation compared to those who inhabit limited elevation ranges (Koch et al., 2017). For the species shown in part A and B, *B. sylvicola* and *B. mixtus*, genetic differentiation increases over geographic distance. Compared to the species in part C and D, *B. melanopygus* and *B. flavifrons*, which show little difference in genetic differentiation across geographic distance, *B. sylvicola* and *B. mixtus* are genetically isolated. *B. sylvicola* and *B. mixtus* also have a comparatively limited range of altitudes in which they are found. Therefore, this analysis showed that species whose niche allows them to travel through the environment more easily have better gene flow compared to those niche limits their mobility (Koch et al., 2017). Studies like this use ecological data about species range and niche to understand identified barriers to genetic flow. This includes the geographical barriers that some species may be unable to cross. This is especially important within the context of climate change. If a species is not able to navigate a route that allows them to expand their territory northward, then it will go extinct when its southern range is no longer habitable. This study also postulates that species with more limited niches, in this case, the species with inability to survive at a large range of altitudes, are more susceptible to extinction (Koch et al., 2017). This can help determine which species are more genetically isolated and identify possible reasons isolated for reasons for that isolation. It also identifies important areas and corridors within the geographic landscape that need to be protected, or at least not modified past the point of usability for the species in question.
Figure 3. Genetic Differentiation Between Populations Over Geographic Distance of Four Bumble Bee Species. $F_{ST}(1-F_{ST})$ on the y-axis represents levels of genetic differentiation between populations. As the value of this statistic increases, there is increases differentiation between populations. Taken from (Koch et al., 2017).

Microsatellites

The advent of microsatellite loci was a great advancement within the field of population genetics (Chase et al., 2018). Prior to this, allozymes were commonly used to differentiate genotypes of individuals within a population in order to understand the genetic structure within
these populations (Chase et al., 2018; Dow & Ashley, 2015; Jehle & Arntzen, 2002). Allozymes are the unique chemical structures of a specific enzyme, and reflect the allelic differences necessary to create them (Chase et al., 2018; Dow & Ashley, 2015). One drawback of this method is that allozymes are not known nor easily identified for all species one would like to study. This is especially pertinent in understudied or yet to be identified species. Another drawback of using allozymes is that the analysis is based on the assumptions of both population isolation and equal mating success of all adults within the population (Dow & Ashley, 2015). While it is still necessary for microsatellites to be developed, methods for identifying many loci within the genome have been published and could be replicated for other species (Stolle et al., 2009).

Microsatellite loci are an example of a neutral genetic markers. Microsatellites consistent of flanking sequence and short (2-4) base sequences that are repeated throughout (Balloux & Lugon-Loulin, 2002). Alleles are defined by the number of repeats (Fig 4). When combined with other microsatellites, they can identify individuals, and measure the relatedness between individuals (Chase et al., 2018; Dow & Ashley, 2015). Microsatellites are neutral genetic markers that are not part of genes, and most likely do not affect the behavior, health, or appearance of the individual. Therefore, microsatellite genotypes, for the most part, do not experience selective pressure because they do not affect survival (Sunnucks, 2000). Microsatellites can be used as a good indicator of the degree of a genetic relationship between individuals. Applications of microsatellites include maternity and paternity analysis, and estimating effective population size (Chase et al., 2018). Microsatellites are also codominant, and therefore, both alleles can be distinguished from one another and provide useful information to the researcher (Balloux & Lugon-Loulin, 2002).
One of the main advantages of using microsatellites is the high variability between members of a population (Chase et al., 2018), but this can also be a disadvantage. This high variability is due to the ability of these sites to mutate. There are two accepted mutation models for microsatellite loci: the Stepwise Mutation Model (SMM) and the Infinite Allele Model (IAM) (Balloux & Lugon-Loulin, 2002). SMM assumes there is equal probability of adding or deleting a single unit repeat in either direction (Balloux & Lugon-Loulin, 2002). This model also assumes that alleles of similar repeat number are more closely related than alleles with greater size differences (Balloux & Lugon-Loulin, 2002). IAM assumes that all identical alleles are due to shared ancestral history (Balloux & Lugon-Loulin, 2002). It also states that mutations will never create the same allele twice (Balloux & Lugon-Loulin, 2002). That statistics associated with each model will be detailed in “Statistics for Microsatellite Analysis”.

The applications of microsatellites in population genetics include determining the number of populations and subpopulations of a species within a certain area, source/sink dynamics, and extinction recolonization frequencies (Jehle & Arntzen, 2002). Microsatellites can also be used to identify possible migrant individuals within populations, by looking at comparative level of
heterogeneity or differences compared to the rest of the population (Andersen et al., 2001). Therefore, this can be used to identify populations that are experiencing immigration (Andersen et al., 2001). It is also important to note that social structure and demography have a large influence on how results from studies using microsatellites should be interpreted, as expanded upon at the end of “Population Genetics and Conservation Case Studies”. Microsatellites are also useful for analyzing how genetics change over a geographic area and/or time (Dow & Ashley, 2015). In respect to bumble bees, microsatellites can be used to investigate mating behavior, foraging range, and colony density (Herrmann et al., 2007).

It is important to consider the sensitivity of the genetic markers when designing a study (Chase et al., 2018; Sunnucks, 2000). When a research question is investigating groups that are more distantly related evolutionarily, the marker used must be relatively conserved as the number of differences shown by a more sensitive marker would be so great that they could not easily be analyzed (Dow & Ashley, 2015). Therefore, when comparing different species, it is valuable to choose a nuclear gene that is relatively conserved among organisms (Sunnucks, 2000). In contrast, when investigating a question that compares closely related individuals, a more variable genetic marker, such as microsatellites, is useful as it can show the subtle differences between individuals (Sunnucks, 2000). These differences in sensitivity can be used strategically, as richer analysis can be performed by combining DNA markers such as mitochondrial DNA with microsatellite loci (Karl et al., 2011). As my project was only investigating species within the same genus, microsatellites loci where the appropriate choice.

Another important factor to consider when undertaking a population genetics study is the time scale of sampling methods. If genetic structuring exists between generations, this can affect the results of a study that samples specimens of different generations (Balloux & Lugon-Loulin,
2002). All workers within a bumble bee colony die off the in fall as described in more detail in the next section. As I only collected workers, and my bumble bee collection occurred only during one summer field season, I know that all my specimens are of the same generation. Therefore, this was not a relevant factor within my experimental design, and I would not have had to test for temporal genetic structuring.

Mutations in microsatellite loci persist within a population because they do not affect the survival of the individual, unlike mutations that occur within an actual gene. This aspect of microsatellites makes it difficult to differentiate between alleles that are similar due to relatedness, or those due to chance. Size homoplasy is when two organisms have the same allele for a microsatellite not due to common ancestry, but due to a series of mutations (Balloux & Lugon-Loulin, 2002). The amount of homoplasy for each microsatellite loci depends on mutation rate, the effective population size of the species that the loci corresponds to, and the time divergence between populations (Estoup, Jarne, & Cornuet, 2002). This can complicate analysis, especially because size homoplasy has been documented between two populations that do not share gene flow (Balloux & Lugon-Loulin, 2002). One factor that contributes to size homoplasy is the maximum size constrain for microsatellites (Balloux & Lugon-Loulin, 2002). This limits the possible number of alleles within a population, which increases the probability for two of the same allele to form via mutation. Examples such as this are why it is important to understand dynamics and limitations of microsatellites when using them within a study (Putman & Carbone, 2014). Size homoplasy is not possible under IAM, as this model states that no allele shall be created twice (Balloux & Lugon-Loulin, 2002; Estoup et al., 2002). Therefore, any statistic based off of this mutation model does not account for size homoplasy.
The analysis of individual microsatellites can be strengthened by understanding the mutation models described earlier as well as the specific qualities of a microsatellite that affect mutation. One factor that influences the dynamics of microsatellite mutation is the length of the microsatellite allele (Balloux & Lugon-Loulin, 2002; Schlotterer, 2000). The larger the allele, the more likely that subsequent mutations will lead to a decrease in the number of repeats (Schlotterer, 2000), and longer alleles are more likely to mutate than shorter alleles (Balloux & Lugon-Loulin, 2002). Microsatellite repeats can be di-, tri-, or tetra-nucleotide (Balloux & Lugon-Loulin, 2002). The type of repeat, as well as the base composition of the repeat affect mutation rate (Balloux & Lugon-Loulin, 2002). There are three types of microsatellite repeats: perfect, compound, and interrupted. Perfect microsatellites only contain repeats that are exact copies of one another (Perfect, n.d). Compound microsatellites contain one or more sections with unique repeated sequences (Bull, Pabo, & Freimer, 1999). Interrupted microsatellites contain one repeated sequence, with another sequence, different from the repeated sequence, interspersed within a series of repeats (Oliveira, Pádua, Zucchi, Vencovsky, & Vieira, 2006).

Researchers have various options to investigate the mutation dynamics of specific microsatellite loci, both experimental and theoretical. Microsatellite mutation can be estimated if both the effective population size and the mutation model for the loci and population being investigated (Schlotterer, 2000). This can be difficult as the effective population size is sometimes derived by researchers using microsatellite data itself. Another issue with this method is that it is biased due to the assumptions of the model used to calculate the rate. The dynamics of microsatellite mutation can also be inferred by looking at the distribution of loci within a population (Schlotterer, 2000). This method of estimation relies on the assumption that the
population is at equilibrium (Schlotterer, 2000). Both of these approaches are limited by the models they are based upon.

If there is sufficient data from various generations and familial relationships between individuals is known, a pedigree analysis can be done to see how microsatellites change over time (Schlotterer, 2000). This is difficult because it requires a special set of circumstances, and if data is even available, it is usually limited (Schlotterer, 2000). Another method to understand the microsatellite mutation dynamics is clone a allele of a microsatellite into a vector and connect it to a reporter gene (Schlotterer, 2000). As mutations occur, the reporter gene will turn on and off (Schlotterer, 2000). Microsatellite mutations change the number of repeats, and therefore the length of the sequence, which then shifts the reading frame of the reporter gene, turning it on and off (Schlotterer, 2000). This method is also limited. It cannot be used to investigate loci that have tetra-nucleotide repeats because allele number does not affect the reading frame of the reporter gene. Even with di- and tetra-nucleotide repeats, this method is not sensitive enough to tell the different between a shift in one or two repeats. Figure 5 shows how an additional di- and tetra-nucleotide repeats shifts the reading frame by adding incomplete codon worth of bases is added before the reporter gene. This also holds true when two sets of repeats are added before a reporter gene. The reading frame is only maintained when three repeats are added to the loci. This means that the reporter gene would be turned off for two out of three repeat numbers, decreasing the sensitivity of this method of microsatellite mutation analysis. Also, unless this method is supplemented with other experimental analysis, such as PCR, there is no way to tell if the mutation created a new allele that is shorter or longer than the original allele.
The last two methods described provide allelic data that can be useful to modify current or create new models for microsatellite mutation. This is because both approaches show how the proportion of microsatellite alleles change in a population over time. This documented change can be held up against the predicted results based on the established models. If consistent differences are seen between predicted and actual change over time, models can be modified to account for these consistent differences.

Microsatellites have allowed researchers to effectively answer questions that prior methods were unable to, making it a popular marker for population genetics studies (Chase et al., 2018; Dow & Ashley, 2015; Jehle & Arntzen, 2002). The strength of microsatellites in their high variability and the lack of selection they experience within the genome (Chase et al., 2018). For these reasons and the sensitivity of microsatellites (Sunnucks, 2000), I used this marker when completing my experimental design. One disadvantage of using microsatellites is size homoplasy, when two organisms have the same allele due to mutations instead of common ancestry (Balloux & Lugon-Loulin, 2002). There are two main models for understanding microsatellite mutation dynamics, IAM and SMM, each with their own statistic (Balloux & Lugon-Loulin, 2002). Each model has its strengths and weaknesses and this is why studies usually report both (Balloux & Lugon-Loulin, 2002). One way to either improve current models or tailor models to specific microsatellite loci, is to investigate the mutation dynamics.
experimentally (Schlotterer, 2000), but this requires a time investment and as limited sensitivity. By understanding the limitations of microsatellites as well as the models and statistics used to analyze them, I would be better able to interpret any results from the study I designed.

Bumble Bee Colony Structure and the Effects on Genetics

Bumble bees are within the insect family of Hymenoptera. The behavior of living socially in hives is exhibited in many species of social Hymenopterans. This is also called eusocial behavior, and it requires the cooperation of many individuals to function (Wilson, 1975). Only certain individuals reproduce, and the majority of the population within the hive devotes their lives to foraging, raising young, or other non-reproductive tasks.

Bumble bee colonies follow a yearly cycle (Smith, n.d.). In the spring, queens who mated with males last fall emerge from hibernation (Smith, n.d.). These queens find a place to nest and build up enough energy resources to start laying eggs and caring for the first of their brood (Smith, n.d.). The location a queen chooses to nest is based on preferences that differ between different species of bumble bees (Smith, n.d.). The first of her brood develops into adults who start to build up the colony (Smith, n.d.). This includes physically building the nest and foraging for the queen (Smith, n.d.). These individuals are called workers and are all female (Smith, n.d.). At this stage, the focus of the colony is population growth and colony development (Smith, n.d.).

In late summer and early fall, the queen lays eggs that will become males (Smith, n.d.). After this, the queen may or may not lay eggs that will become new queens (Smith, n.d.). New queens and males mate in late fall, and then old queens, workers, and males all die (Smith, n.d.). The new queen goes into hibernation and starts the cycle over (Smith, n.d.). Tropical bumble
bees are an exception to the annual cycle, as it is not necessary for an individual queen to hibernate to avoid harsh winter weather conditions (Smith, n.d.).

Queens have the ability to control the sex of the individuals that will hatch out of the eggs she lays (Smith, n.d.). When the queen mated the fall before her hibernation, the sperm is stored in her spermatheca (Smith, n.d.). If the queen lets sperm combine with the egg that is being released, then the individual will be a diploid female (Smith, n.d.). This means this individual has two sets of chromosomes. This is the case with both female workers and new queens (Smith, n.d.). If a queen does not release sperm to fertilize the egg, the individual will be a haploid male that only has one set of chromosomes (Smith, n.d.). Species, like bumble bees, whose sex is determined by ploidy are haplo-diploid (Smith, n.d.). Males are only produced right before new queens are about to mate because they do not forage or complete any tasks to help the colony run (Smith, n.d.).

Every individual in a bumble bee colony is related to one another. Since every colony has one queen, she is everyone’s mother. All of the workers, new queens, and males are siblings. The level of relatedness between these siblings depends on how many males their queen mated with last fall. Female individuals in the hive that were created using sperm from the same male are more related than human siblings. This is because both human males and females are diploid. His children have equal chance of getting one two sets of genes (as well as neutral genetic material). The passing down of genes functions in the same way for mothers. Due to this pattern of inheritance and pure chance, most human siblings share about 50% of their genetic material, 25% from their father and 25% from their mother. Since bumble bee males are haploid, they only have one set of genes to pass down. Therefore, every female whose eggs were fertilized by the same male have an identical 50% of genetic material from their father, and 25% from their
mother. Therefore, bumble bee siblings can be as much as 75% related to each other as compared to 50% of other sets of full siblings.

Queens and males are the only ones who sexually reproduce creating a high level of genetic relatedness within the colony (Herrmann et al., 2007). This stands in contrast to species in which every individual has the ability to participate in sexual reproduction. This same mechanism occurs within other hive living Hymenopteran species, such as honey bees. Therefore, genetic structuring of social hymenoptera will be unique in that there is very little genetic variation within hives, creating a smaller effective population size which is measured by the number of colonies (Ellis & Goulson, 2006; Herrmann et al., 2007; Knight et al., 2005). It has also been shown that the effects of a decrease in genetic variability are less documented in social Hymenoptera (Herrmann et al., 2007). These are important factors to keep in mind when examining population dynamics of bumble bees because the same assumptions used for looking at other species may not be applicable.

My Hypothesis

Originally, my experimental plan included sampling from four locations shown in Figure 11. I believed that there would be high genetic similarity between the more Northern sites located in Michigan and between the two more Southern sites located in Wisconsin. I also hypothesized that the sites separated by water would have relatively low genetic similarity. Though bumble bee are able to travel across bodies of water (Osborne et al., 2018), this can be more difficult as it requires constant motion, as there is no place to land or rest. There are also no food resources for individuals to consume and use as energy to continue their trek across the water. If the water is not a significant barrier to gene flow, then there would be more similarity
between the two sites in the middle of the map than between the sites that either of these sites are connected to via land, as the two middle sites are the shortest distance away from one another. These additional sites were not incorporated into the study for the same reason that I was unable to fully analyze the specimens collected from the Appleton area.

As I was unable to create this original experiment, my hypothesis shifted. I now hypothesize that harsher methods of DNA extraction are necessary to extract DNA from bumble bee legs, due to the tough outer cuticle surrounding these cells. Though, once DNA is extracted, all PCR amplification should have no special methodological requirements.

Figure 6. Map of Original Sampling Locations. The black circles on the map encapsulate the field sites in which “Bombus” specimens were to be collected. The site in the lower left hand corner is the Appleton area from which the samples used in this methodology were collected from. The site on the Door Peninsula is from Lawrence’s Northern Campus, Bjorklunden, where another lab group from the biology department was going to collect specimens. The site to the right of Escanaba is where my research advisor collected specimen when on a field excursion with a summer class from the University of Michigan. The site in the upper right corner is the University of Michigan’s Biological Field Station where my research advisor collected specimen with the same summer class. All the “Bombus” specimens were collected during the summer of 2018.
Materials and Methods:

Field Collection and Species Identification:

During the summer of 2017, starting July 17th, I began collecting bumble bees. Bumble bees were collected while foraging on flowers using a net. The bumbles were transferred from the net into a glass jar, and sacrificed with aceton. The samples were transferred to falcon tubes containing a label listing the field site and the date. The specimens were preserved in 95% ethanol.

This method of sampling is useful because there is a large amount of DNA to work with, as well as a specimen that can be examined many times in order to ascertain as well as confirm species identification. Yet, there are some obvious drawbacks including emotional strain, and as an ethical conflict for a researcher whose goal it is to preserve bumble bees. This concern is pertinent especially when the loss of workers can become a strain on a colony (Holehouse et al., 2003). This effect is more pronounced in small colonies due to a larger proportional loss of their workforce, and the fact that smaller colonies are more vulnerable than larger colonies (Holehouse et al., 2003). Therefore, I decided to investigate alternative methods. The paper that I based my experiment off of utilized a cooler to chill to the bumble bees until they were immobilized and then detached a leg from each specimen (Koch et al., 2017). The only drawback to this method is that identification must be done in the field and confirmation can only be done based on photos taken with whatever equipment can be transported out into the field. One study compared two non-lethal sampling methods: hemolymph extraction and tarsal sampling (Holehouse et al., 2003). Hemolymph extraction was shown to create a higher chance of worker mortality, though it did provide a higher quality DNA sample (Holehouse et al., 2003). Tarsal sampling included cutting off 2mm of the terminal tarsus, though not cutting the metatarsus, with
a clean scalpel (Holehouse et al., 2003). This cutting did not significantly affect the time of foraging trips, or pollen or nectar collection (Holehouse et al., 2003). Therefore, were I to repeat this experiment in the future, I would obtain my DNA samples through the non-lethal method of tarsal clippings.

I sampled at all sites for half a person hour, except for the first day of sampling at Lawrence University. The other sites that were sampled at are: City Park, Buboltz Hive, Memorial Park, Riverview Hive, Peabody Park, Riverview Control, Pierce Park, Telulah Park, Bubotlz Field, Heckrodt Wetland Preserve, Purdy Hive, and Thousand Island Nature Preserve. At Buboltz Hive and Field, the sampling was done by myself and one of my lab mates simultaneously for 15 minutes. More information on these sites is given in Table 1.

<table>
<thead>
<tr>
<th>Site</th>
<th>Number of Specimen Collected</th>
<th>Number of Species Collected*</th>
<th>Number of Person Hours</th>
<th>Date of Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>City Park</td>
<td>10*</td>
<td>5</td>
<td>0.5</td>
<td>July 27th, 2017</td>
</tr>
<tr>
<td>Gordon Buboltz Nature Preserve Hive Site</td>
<td>6*</td>
<td>4**</td>
<td>0.5</td>
<td>July 25th, 2017</td>
</tr>
<tr>
<td>Appleton Memorial Park</td>
<td>5</td>
<td>3</td>
<td>0.5</td>
<td>July 18th, 2017</td>
</tr>
<tr>
<td>Riverview Gardens Hive</td>
<td>9</td>
<td>6</td>
<td>0.5</td>
<td>July 28th, 2017</td>
</tr>
<tr>
<td>Peabody Park</td>
<td>4</td>
<td>3</td>
<td>0.5</td>
<td>July 27th, 2017</td>
</tr>
<tr>
<td>Riverview Gardens Control</td>
<td>4</td>
<td>3</td>
<td>0.5</td>
<td>July 28th, 2017</td>
</tr>
<tr>
<td>Pierce Park</td>
<td>2</td>
<td>2</td>
<td>0.5</td>
<td>July 28th, 2017</td>
</tr>
<tr>
<td>Lawrence University</td>
<td>15</td>
<td>6</td>
<td>Unknown, but &gt;0.5</td>
<td>July 17th, 2017</td>
</tr>
<tr>
<td>Telulah Park</td>
<td>4</td>
<td>2</td>
<td>0.5</td>
<td>July 28th, 2017</td>
</tr>
</tbody>
</table>

* Indicates sites in which the one of the specimen collected was destroyed during an effort to take to bumble bees off a pin for an untested idea for a method of DNA extraction. ** Indicates a site at which there is one less species due to the destruction of specimens described before. The specimens from Thousand Island Nature Preserve and Heckrodt Wetland Preserve have not been typed to species yet.
<table>
<thead>
<tr>
<th>Location</th>
<th>Species</th>
<th>Number Collected</th>
<th>Number of Collection Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gordon Buboltz Nature Preserve</td>
<td>B. vagans</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Field Site</td>
<td>B. citrinus</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>B. fervidus</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>B. variabilis</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>B. pensylvanicus</td>
<td>3*</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>B. terricola</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>B. auricomus</td>
<td>4*</td>
<td>3**</td>
</tr>
<tr>
<td></td>
<td>B. grieseocollis</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>B. rufocinctus</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>B. impatiens</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>B. perplexus</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>B. borealis</td>
<td>1*</td>
<td>1**</td>
</tr>
</tbody>
</table>

I dried all specimen and then put them on an insect pin. I pinned a label with the date that the specimen was collected, and a specimen identification label. I gave each specimen an alphanumeric code for subsequent identification. I also put an additional species label on the pin after species identification was done. Species identification was based on examination of the pinned specimen was completed by Israel del Toro with my assistance. I took lateral, dorsal, and facial images of each specimen in order to allow for confirmation or reassignment of species. Images were taken using LAS V4.9™ software which allows images to be taken through the lens of a Leica™ M84 at 0.75 magnification. The lateral and dorsal images were too large to be captured.
in one image and the software’s stitching function was shown not to be an effective method of creating a lateral image of each insect. Instead, I took images by focusing the microscope in the same manner, and using an iPhone camera to take an image through the eye piece. Specimens from Heckrodt Wetland Preserve and Thousand Islands Nature Preserve have not been typed to species yet, but this will be done based solely on the microscope images.

**Appleton Bee Sites**

![Map of Field Sites in the Appleton Area.](image)

**Figure 7. Map of Field Sites in the Appleton Area.**

DNA Extraction:

The first five methods I used to extract DNA from specimen utilized a Quiagen DNeasy blood and tissue kit. For the most part, I followed the protocol included with the kit. The common steps used in all five methods are adding 180ul of ATL buffer with 20ul of proteinase K to a bumble bee leg in a 1.5mL Eppendorf tube. Then I transferred the Eppendorf tube between
“hot” (90-95°C) heat blocks and “warm” heat block (30-55°C). This was done to lyse the cells, and allow proteinase K to break down all of the proteins within the cells. Between each movement, the tube was vortexed. Then I added 200ul of AL buffer and the tube was vortexed. I added 200ul of 100% ethanol, and I vortexed the solution. I pipetted the solution into a DNeasy™ spin column with a 2mL collection tube. I then spun the tube for 1 minute @ 10,000 rpm. I discarded the flow through and the collection tube. I then added 500uL of AW1 buffer to the spin column and spun the column for 1 minute @ 10,000 rpm. I disposed of the flow through and collection tube again. I added 500ul of AW2 and put the spin column in a centrifuge for 3 minutes @ 14,000 rpm. Again, I discarded the flow through and the collection tube. I then added 200uL of AE elution buffer and let the tube incubate on the lab bench for 1 minute before I spun the tube for 1 minute at 10,000 rpm. I repeated the same step. Then I tested each sample for concentration and purity using the Nanodrop™. I followed this protocol, because I had had previous success in DNA extraction using this same kit.

Table 3. Summary of Methods Varied for DNA Extraction.

<table>
<thead>
<tr>
<th>Method of DNA Extraction</th>
<th>Leg Preparation</th>
<th>Volume of Elution Buffer (in uL)</th>
<th>Heat Block Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>Freeze/thaw</td>
<td>200</td>
<td>66°C for one hour, at 90°C for 15 minutes, 37°C for 1 minute, and 90°C for 15 minutes</td>
</tr>
<tr>
<td>2nd</td>
<td>Liquid nitrogen and mortar and pestle</td>
<td>200</td>
<td>55°C for an hour, 95°C for 15 minutes, 37°C heat block for 1 minute, and 95°C heat block for 15 minutes</td>
</tr>
<tr>
<td>3rd</td>
<td>Cracked with scalpel in petri dish</td>
<td>200</td>
<td>55°C for an hour</td>
</tr>
<tr>
<td>4th</td>
<td>Cracked with scalpel in petri dish</td>
<td>25</td>
<td>55°C for an hour</td>
</tr>
<tr>
<td>5th</td>
<td></td>
<td>25</td>
<td>55°C for an hour</td>
</tr>
</tbody>
</table>
My first method of DNA extraction started with me taking the specimen leg out of ethanol. I had stored various legs with specimen, site, and date labels in tubes with 100% ethanol. I put the leg into 180uL of ATL buffer in a 1.5mL tube. I put this tube in the deep freeze for over ten minutes. I added Proteinase K. Then I put the Eppendorf tube in a 66C heat block for one hour, at 90C for 15 minutes, 37C for 1 minute, and 90C for 15 minutes. The results from this extraction was a concentration of 6.4ng/uL with a 260/280 ratio of 2.29. These results showed that if any DNA was extracted, it was at a low concentration and was not pure. This led me to investigating a new method of leg preparation prior to starting the protocol from the Quiagen™ kit.

My second of method of DNA extraction utilized liquid nitrogen to flash freeze the mesothoracic leg in order to completely break open the cell. After I poured liquid nitrogen over the leg, I then ground it with a mortar and pestle. After I transferred the ground leg to 1.5mL tube with a scoopula, I added the ATL buffer and proteinase K and vortexed the solution. I put the samples were then put in a 55C heat block for an hour, 95C heat block for 15 minutes, 37 heat block for 1 minute, and finally a 95C heat block for 15 minutes. I followed the protocol from above, and had an extraction with a concentration of 2.1ng/uL and a 260/280 ratio of 2.34. Though concentrations are less representative of the actual concentration of DNA when the 260/280 ratio is not within a close range of 1.8, seeing that this method resulted in an even lower concentration than the first led me right back to the drawing board.

Before starting my third attempt at DNA extraction, I compared my protocol to one used to extract DNA for museum bumble bee specimens (Strange, Knoblett, & Griswold, 2009). These specimens in this study were all on insect pins, and completely dried out. This made me postulate that the ethanol was have an impact of the lysis buffer, possibly affecting its ability to
break open the cells. Therefore, I pinned a specimen and let it dry for 24 hours before preparing it. After cutting off the mesothoracic leg off with a scalpel that was cleaned with ethanol prior, I used the scalpel to crack open the exoskeleton on a petri dish. This was done to expose the white muscle fibers within the leg. After adding the mixture of proteinase K and ATL buffer, I put the Eppendorf tube in a 55C head block for an hour, then vortexed the solution and proceeded with the protocol from the kit. The change in heat treatment was also influenced by the protocol written by the methodology used to extract DNA from museum specimens, as it was more specific to bumble bees. The resulting concentration of the extraction was 5.4 ng/µL with a 260/280 ratio of 2.56.

Another poor result made me reanalyze the protocol provided by the kit. The previous spring I had taken a molecular biology course with Eric Lewlyn in which I did a DNA extraction during the lab portion of the class. I accessed my online lab notebook from this time, and found a large difference in the amount of AE elution buffer used. The lab protocol written and given to us by Eric Lewlyn said to use 25µL of elution buffer twice as opposed to 200µL twice. The protocol written by Eric Lewlyn resulted in a relatively high concentration of extracted DNA the previous spring. I suspected using 8 times more elution buffer was one of the reasons my concentration kept coming out so low. I took this into consideration and changed this in my protocol for the next two attempts at DNA extraction. Upon discovering this misstep, and discussing it with a professor (Kim Dickson), I decided to see whether or not I could spin down my previous DNA extractions to increase their concentration. Beth DeStasio taught me how to use to Speed Vac® Plus in order to do just that. I had the samples in the machine all day, and noted no significant decrease in liquid volume. I Nanodropped™ one sample, but I did not record
the results due to the poor quality and that I knew it would be impossible to use these extractions in any later lab work.

I began the forth methodological attempt at DNA extraction by breaking off the second leg of the specimen with forceps and scalpel in a petri dish. I then put the broken sections of leg using the forceps into the 1.5 Eppendorf tube. I put the tube was in a 55C heat block for an hour, and then all protocol occurred as listened in the first paragraph of this section except for the switch from 200uL to 25uL of AE elution buffer. This resulted in a concentration of 24.3ng/uL. This was not a large leap in concentration, and therefore, I continued working on my DNA extraction methodology.

The fifth method attempted to extract DNA again used mechanical force to physically to expose cells and break through the tough cuticle. Unlike prior protocol, the scalpel crushed up the sample while the leg was in the 1.5mL Eppendorf tube where the initial reaction would take place. This was done to make sure even the small bits of leg that resulted from the pulverization of the leg ended up in the extraction, possibly increasing the resulting DNA concentration obtained. I held the samples at 56C for 1 hour and then I took the samples out and allowed them to cool to room temperature. After this, I followed the protocol listed in the kit, except for the change made to the volume of AE elution buffer that was referenced above. Four specimen were selected for DNA extraction during this trial run. The resulting reaction was first analyzed via Nanodrop™. After receiving the most positive results throughout this process as of yet, the four samples were loaded into a gel with a Hi-Lo™ ladder.
Figure 8. Gel Electrophoresis Confirmation of DNA Extraction. The first lane of this gel is a Hi-Lo™ marker. The other four lanes are the DNA extractions from my fifth method of DNA extraction. This along gel was run to confirm that presence of DNA along with Nanodropping™. The four bands around 200 bp show that DNA was successfully extracted via this method.

Table 4. Summary of the Results from the Preliminary Methods of DNA Extraction. The 260/280 ratio of for the 4th method of DNA extraction was not saved on a flash drive, nor recorded in my lab notebook. This sample was never Nanodropped™ again to ascertain this value due to the poor quality of the sample. For the 5th method of DNA extraction the concentration and 260/280 ratios are given as average +/- one standard deviation as four specimen had their DNA extracted via this method.

<table>
<thead>
<tr>
<th>Method of DNA Extraction</th>
<th>Replicates</th>
<th>Concentration (ng/uL)</th>
<th>260/280 ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>1</td>
<td>6.4</td>
<td>2.29</td>
</tr>
<tr>
<td>2nd</td>
<td>1</td>
<td>2.1</td>
<td>2.34</td>
</tr>
<tr>
<td>3rd</td>
<td>1</td>
<td>5.4</td>
<td>2.56</td>
</tr>
<tr>
<td>4th</td>
<td>1</td>
<td>24.3</td>
<td>n/a</td>
</tr>
<tr>
<td>5th</td>
<td>4</td>
<td>27.1+/-21.0</td>
<td>2.105+/-0.13</td>
</tr>
</tbody>
</table>

The final method I settled on for DNA extraction did not use a Quiagen DNeasy™ blood and tissue kit. Instead, this was based off of methods used in Jonathon Koch’s laboratory. This method could either be done in individual PCR tubes, 8 strips, or a 96 well plate. I filled a well with 150uL of a 5% Chelex™ and 5uL of proteinase K (solution created by combining 10mg with 1000uL of water) (Koch, 2015). Then I put a leg from each specimen into an individual well and broken using a pair of iris scissors. I washed the scissors with 100% ethanol and wiped down them with a Kimwipe™. Then I moved the samples are then moved into the thermocycler.
I set the thermocycler to 50°C for 60 minutes, 99°C for 15 minutes, 37°C for 1 minute, 99°C for 15 minutes, and then held at 15°C. One of the major advantages of this method was the number of samples that could be processed at once. The DNA extraction kit required many steps and movements between heat blocks that would have made processing all 81 (originally 83 samples) very time consuming. This method had the most consistent 260/280 ratio, and resulted in the highest concentrations (Table 5). I only utilized an 8 strip to extract DNA from 8 samples during the first trial run of the protocol. Once this run showed very promising results, I extracted DNA from the remaining 73 samples in a 96 well plate. 13 of the specimen had their DNA extracted a second time due to some samples not being of sufficient volume to proceed with additional experiments. All Nanodrop™ results are stored on a personal flash drive and have all been sent to a unique project email as a backup.

<table>
<thead>
<tr>
<th>Final DNA Extraction Method</th>
<th>Concentration (ng/μL)</th>
<th>260/280 ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>90.3</td>
<td>1.78</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>990</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Primer Development:

In order to create a PCR product, (described in further detail “PCR”), it is first necessary to have the correct primers. Primers are short single-stranded strings of bases (~18-25bp) that flank the section of interest within the DNA sequence, one binds before the section on the coding strand, and the other binds after the section on the complementary strand. Primers can either be specific to a certain gene or loci, or can intentionally be made general, in order to bind to several loci within the genome. Within PCR, well-designed primers are necessary to amplify a desired sequence.
Primer research began by identifying useful primers for genetic analysis from a Jonathon Koch paper on landscape genetics (Koch et al., 2017). From there, I began to use tools such as NCBI BLAST to try to identify the sequence for these microsatellites. This method was ineffective because these databases only contain genes, and not microsatellite data.

This led me to a literature search for studies that used microsatellites to investigate different aspects of bumble bee ecology. I was eventually able to find four loci that were consistently found within several species of *Bombus*, B10, B11, B124, and B126 (Estoup et al., 1995). At first, I only ordered the forward primers for B10 and B11. I realized how it was impossible to achieve positive results in any of my reactions when I went to order primers for the two other loci. I will go into the development of methodology in “PCR”. The primers I ordered are Value Custom Oligos from Thermosfisher. The use of fluorescent primers later in the experimental processed is documented in the section “A Poor Man’s Approach to Genotyping…” (Schuelke, 2000).

Table 6. Summary of Microsatellite Primers, Core Sequence and Annealing Temperatures.

<table>
<thead>
<tr>
<th>Microsatellite</th>
<th>Primers</th>
<th>Core Sequence</th>
<th>Annealing Temperature (degrees C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B10</td>
<td>Forward: 5’-GTGTAACTTTCTCTCGACAG-3’ Reverse: 5’-GGGAGATGGATATAGATGAG-3’</td>
<td>(CT)<em>{4}TT(CT)</em>{13}</td>
<td>52</td>
</tr>
<tr>
<td>B11</td>
<td>Forward: 5’-GCAACGAAACTCGAAATCG-3’ Reverse: 5’-GTTTCATCCAAGTTTCATCCG-3’</td>
<td>(CT)<em>{5}…(CT)</em>{n}</td>
<td>52</td>
</tr>
<tr>
<td>B124</td>
<td>Forward: 5’-GCAACAGGTCGGGTAGAG-3’ Reverse: 5’-CAGGATAGGGTAGTAACGAG-3’</td>
<td>(CT)<em>{8}TCCTCTTTCCAC(CT)</em>{14}CC TC(GC)...(CCCT)_{8}</td>
<td>57</td>
</tr>
<tr>
<td>B126</td>
<td>Forward:</td>
<td>(CT)<em>{12}GT(CT)</em>{16}</td>
<td>57</td>
</tr>
</tbody>
</table>
PCR

PCR stands for polymerase chain reaction. This is a method for making many copies (in the millions) of a section of DNA. This is done by designing primers as described in the section “Primer Development”. Template DNA is heated to a point where it becomes single-stranded, the temperature is then lowered allowing the primers to anneal, or attach itself to the DNA. From there, the temperature of the reaction is raised to allow for a DNA polymerase to extend or add on bases after the location of primer attachment. At first, the products of this process can be quite large, but eventually products that are only section of DNA between the forward and reverse primer predominate. This process is usually repeated around 30 times, creating many copies of the DNA that can either be sequenced or measured for base pair length.

As said above, the first attempts at PCR were futile since I was only adding a forward primer to each reaction. Despite this, during this time there were many valuable lesson learned along the way. The first thermocycler settings I designed, based off of previous cycler patterns I had seen used in studies were used (except for one exception) throughout until fluorescently tagged primers were added to my protocol (Estoup et al., 1995; Koch et al., 2017). The thermocycler settings started with 95C for 7 minutes and then 30 rounds of 95 C for 30 seconds, 52C for thirty seconds for the B10 or B11 loci or 57C for thirty seconds for B124 or B126 loci, then 72C for thirty seconds. The cycle then ends with a final ten-minute elongation at 72 C and once this cycle ends, the samples are held at 4C.

After my first run through of PCR, I inferred the 8 PCR tubes I had been using may not have been closed correctly or they may not have been placed firmly within the slots. When I
opened the thermocycler, I found most of the tubes open and with very little liquid inside. From then on, I was more careful from then on both when sealing the lid on all tubes, as well as when placing those tubes within the slots. This lesson taught me to remember to follow good lab practices.

I was unable to analyze these samples, but when I did so in the future I used a Thermo Scientific NanoDrop 2000 Spectrophotometer™. A DNA sample is measured by pipetting 1uL onto the platform of the device. From there, light is passed through the sample and the resulting wavelengths inform the concentration and purity of the sample. In order to measure DNA, the corresponding Thermo Scientific NanoDrop 2000™ software must be set to Nucleic Acids.

The next time I ran PCR, I moved the samples from the thermocycler directly to the freezer. When I tried to Nanodrop™ the samples the next day, I had an issue pipetting the samples. At first, I thought the samples were not defrosted thoroughly, but I then realized the contents of the tube had a jelly like consistency. This substance was impossible to Nanodrop™, and I knew this was a bizarre result based on my previous experience doing PCR. Despite this, I thought there could be some correlation between the odd consistency and immediately transferring the samples to the freezer without checking their concentration first. I performed another round of PCR and had the same result. At this point in the process, I had been weighing out 0.001g of bovine serum albumin (BSA) and was adding this solid reagent to reaction in the PCR tube. I could see that the BSA was not dissolving well. I had a theory that this reagent could be causing the jelly like consistency and spoke about this with a professor (Kim Dickson). I was informed that it is quite a difficult process to get solid BSA into solution. I was then given BSA solution from her stock supply. This was a valuable lesson even though BSA did not end up in the final reagent list.
After I started using a BSA solution instead of solid BSA, I still did not see any PCR amplification. My next attempt at troubleshooting was adjusting the annealing temperatures. I thought my PCR was unsuccessful due to a lack of primer binding. The previous annealing temperature I was using (52°C), used in a published paper by Estoup et al., 1995. Therefore, I was unsure if this method would fix my issues with PCR, but I was willing to try anything due to my frustration. I tried to amplify both B10 and B11 creating temperature gradient for the annealing temperature on the thermocycler from 51°C to 55°C. I placed 10 individual PCR tubes that were placed throughout this temperature gradient. I recorded the location of each sample within the gradient. If there was any successful PCR, I would have used the corresponding annealing temperature in all subsequent reactions. I later ran these samples through a 2% agarose gel I mixed along with a Hi-Lo and a 1 kb ladder. All PCR reactions were mixed with gel red in order to be visualized on the gel. Due to the lack of amplification, there was no band in any of these lanes on the gel (Figure 8).
After attempting PCR several times, running gels that yielded similar results and consulting with a professor (Kim Dickson), I hypothesized that the issue with my PCR may have been degraded dNTPs. I had also already received some information from Jonathon Koch’s lab, which performed a similar experiment on a larger scale. This included a recipe for a dNTP mixture that the lab mixed themselves as opposed to buying a premixed solution. Therefore, I bought individual nucleotides, so I could create a fresh mixture for my experiments periodically. This method is also economical, as a large volume of dNTPs can be mixed using the individual nucleotides. The dNTP mixture is created by combining 360uL of milliQ water, 10uL of 100mM dCTP, 10uL of 100mM dATP, 10uL of 100mM dGTP, and 10uL of 100mM dTTP. I still created

---

**Figure 9.** Attempted PCR Conformation via Gel Electrophoresis. The top well on the top of the image contained a 1 kb marker and the bottom well contained a Hi-Lo marker. The other 10 lanes in between the two markers contained 10 PCR reactions and gel red. As no DNA resulted from these reactions.
my own mixture of dNTPs even after I figured out that this was not the main issue I was having with PCR.

As described in “Primer Development”, when I was ordering the primers for the two other microsatellite loci that I planned on analyzing, I realized that I had made an error the last time I was ordering primers. I saw that both the forward and reverse primer were listed in the paper I was referencing (Estoup et al., 1995). I had previously only ordered the forward primer for the B10 and B11 loci. Therefore, even if I had a perfect mix of reagents and perfect annealing temperature, PCR would still be impossible because forward primers by themselves are unable to create a usable PCR product.

Once both the forward and reverse primers were ordered, I attempted PCR again. I only attempted on 8 samples (4 from Telulah Park and 4 from Heckrodt Wetland Preserve) on an 8 strip of PCR tubes. I got poor concentration read outs as well as 260/280 ratios. I postulated this was because I used milliQ water as my blank instead of GoTaq™ 5X flexibuffer, which made up the majority of my reaction. This lead me to realize that I was using too much flexibuffer (5.85uL in a total volume of 10uL) in my PCR reaction. I adjusted all of my reaction volumes and settled on a final reaction mixture shown in Table 7. I stopped using BSA within my reactions because I ran out of my stock one night and found that I was able to successfully amplify my microsatellites without BSA in the reaction.

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Original Volumes (in uL)</th>
<th>Final Volumes (in uL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reverse Primer</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Forward Primer</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>MgCl₂ (25mM)</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>MilliQ Water</td>
<td>4</td>
<td>4.8</td>
</tr>
<tr>
<td>dNTPs</td>
<td>0.6</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Table 7. Reagents for PCR. The original volumes of reagents refer to the volumes I used when I first successfully performed PCR. The final volumes refer to those I used after I ran out of BSA, and which includes the PCR that I performed to send in for fragment length analysis.
Due to the volume of reactions I was performing, I started to make a master mix of reagents for each loci. I would multiply the volume of each reagent, only excluding the template DNA, by the same number of possible reactions. I would record the volumes I included in the mixture, so I could be able to analyze the volumes later and check my math if the reaction did not go well. After all the reagents were combined in a 1.5 Eppendorf tube, I would vortex the tube, and then spin the tube for 30s @ 5,000rpm. I would pipette 9uL of this solution into each PCR tube, and then 1uL of extracted DNA. This allowed me to save time, and also kept me from getting tired and making mistakes.

When I first tried to amplify one microsatellite for all 81 samples, I used a 96 well plate. I did not Nanodrop™ the samples because I could tell that the reagents were not concentrated at the bottom of the tube when I took the samples out the thermocycler. I realized that my method of tapping the plate down against the lab bench was not sufficient to pool all of the reagents at the bottom of the well. I then investigated the tools at my disposal to spin down a 96 well plate. I tried one centrifuge meant to hold 96 well plates, but I found that it did not spin fast enough to pool my reagents on the bottom of each well. I then realized that all my reactions would have to take place in 8 strip PCR wells. This was less convenient, but still feasible.

I then labeled all 8 strips with a color that represented the microsatellite loci that was being amplified and a number at the front of each strip that was used as a reference for which specimen’s DNA was used as template DNA. The same template DNA was put in the same position within the 8 strip PCR wells for all microsatellite loci and recorded. During my first bulk PCR, I attempted to multiplex the loci that shared the same annealing temperatures by putting two sets of primers in each reaction (B10 with B11, and B124 and B126). I realized that

<table>
<thead>
<tr>
<th>GoTaq™</th>
<th>0.5</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>GoTaq™ flexibuffer</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>BSA</td>
<td>0.8</td>
<td>n/a</td>
</tr>
</tbody>
</table>
since there was no way to differentiate the two PCR products, and I had to redo my PCR reactions.

I then redid all my PCR reactions for each microsatellite loci individually. I prepared this samples to be analyzed with each loci individually pipetted in their own 96-well plate. MilliQ water (9uL) was combined with 1uL of each PCR product. I sent these plates to the Yale DNA Analysis Facility on Science Hall. I recorded the position of each specimen within the 96 well plate in an excel document which I printed out and placed within my lab notebook. I received a call later that week from the manager of the facility. There were no results from my fragment analysis because none of my PCR products included any fluorescent tag. I did not understand that this fluorescent tag was necessary for the type of fragment analysis done at this facility. A capillary genetic analyzer detects how long it takes between the beginning of capillary electrophoresis, and the recognition of a fluorescent molecule (LifeTechnologiesCorp, 2015). Longer DNA fragments take longer to travel through a capillary than smaller fragments during electrophoresis, and this property can be used to ascertain the length of a fragment (LifeTechnologiesCorp, 2015).

The manager at the facility (Carol Mariani) who is also a university professor whose lab is involved with many experiments that include microsatellite analysis, advised me on a method of adding fluorescence to samples in an inexpensive manner. The method of adding fluorescent primers to a sample is described in the section titled “A Poor Man’s Approach to Genotyping” (Schuelke, 2000). This protocol as described in ““A Poor Man’s Approach to Genotyping…” (Schuelke, 2000)” was followed through for each microsatellite loci separately and loci were combined as described earlier in this section on a 96 well plate. I sent these samples to the facility, and received an email that my results came in. I was unsure of how to interpret results,
so I called the professor who I had spoken with before. The professor opened up my results in the proper software and found that fluorescence was recorded at the same point within every sample. This meant that the 3730xl-Capillary Genetic Analysis machine recognized that there was fluorescence within the sample, but that all of the fragments were the same length. The most probable cause of this uniform hump was lack of PCR product created with a fluorescent tag. My Nanodrop™ results made me quite sure that PCR had occurred. Therefore, my next step was to try to identify the issue with fluorescent primer binding. I went into the lab and examined my primer order form. I found that my M13 tagged forward primer and M13 tagged fluorescent primer with the exact same version of the M13 sequence instead of two complementary M13 sequences. I traced this error back to an email I received, in which I was given an example of the primers I should order. I put in this primer sequence into the order form exactly without considering how these primers sequences needed to be in order to build the product I wanted. Therefore, it was impossible for the fluorescently tagged primers to bind.

<table>
<thead>
<tr>
<th></th>
<th>Concentration (ng/µL)</th>
<th>260/280 ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>531</td>
<td>1.75</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>182</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Gel Electrophoresis

Gel Electrophoresis uses the movement of DNA through a polyacrylamide gel to demonstrate the presence of DNA and measure length of a DNA fragment, as well as the shape of DNA (ex. linear or coiled). I only used this method to measure of fragment length. It utilizes the DNA’s negative charge. By pipetting DNA combined with a fluorescent dye into well in a gel, and then running an electrical charge through the gel, the DNA migrates toward the positive electrode. Smaller fragments are better able to travel better through the matrix of the gel, and
therefore migrate faster. When combined with a marker that produces bands of known base pair sizes, the fluorescent band of a DNA segment can easily be assigned a length in base pairs.

I used gel electrophoresis to confirm the presence of DNA after DNA extraction. When I began this project, I had thought that I would also be able to analyze the length of my PCR products using gel electrophoresis. When I discussed the best way to accomplish this with a professor (Kim Dickson), I was advised to mix my own gels as opposed to using the premixed gels the Biology stockroom has on hand. This would have allowed me to use a higher percent agarose mix in order to get a crisper result. I mixed the gel by combining 4uL gel red, 0.8g of agarose, and 40mL of 1xTAE in a 50mL Falcon™ tube. I then screwed on cap gently, in order to allow for some release of gas. Then I filled a beaker around ¾ full and the tube was placed in the water. I first microwaved this solution for 30 second pulses until the agarose began to boil. When the solution boiled, I took the tube out of the water and swirled and inverted in order to better get the agarose into solution, and prevent a gritty gel. I microwaved the gel for 10 second pulses. Once I boiled the gel 3 or more times, the gel was allowed to cool for a short period. I then poured the gel is into the casting tray with the well combs in position. As described above, I found that gel electrophoresis was not a useful method of separating the different alleles. This is because it is not possible to distinguish the difference between alleles that vary by under 20 bases pairs. This is why I sent my samples to the Yale DNA Analysis Facility on Science Hill as described in the section on PCR.

“A Poor Man’s Approach to Genotyping…” (Schuelke, 2000):

After sending in the samples that could not be read due to their lack of fluorescence, I was advised on to use a comparatively inexpensive method to fluorescently tag my
microsatellites. Fluorescently tagged primers cost around 16 times more than unlabeled primers used in this project. Instead of buying a set of fluorescent primers for each of the 4 loci I was using in my study, I added an M13 tag to the beginning of each forward primer. The M13 tag is a known sequence used in molecular biology (Schuelke, 2000). The complementary sequence to M13 is ordered with a fluorescent pigment on the end (Schuelke, 2000). Two of these primers were ordered, and they were identical except one had a 5’ HEX tag and the other a 5’ 6-FAM tag. In the reaction, the volume of the primer with the fluorescent tag is twice as much as the volume of the M13 tagged forward primer for each locus. In theory, during the first rounds of PCR, the M13 tagged forward primer was used create product (Schuelke, 2000). Product that includes the microsatellite and with the additional M13 sequence at the beginning of the fragment builds up within the reaction (Schuelke, 2000). This allows the fluorescently tagged primer to anneal and start creating product, especially because it exists at a double the concentration than the other forward primer within the reaction (Schuelke, 2000). Table 7 shows that 1μL of the forward and reverse primer were previously added to each reaction. With this addition to the protocol, the volume of primers added to the reaction are: 0.4μL reverse primer, 0.4μL complementary M13 fluorescently tagged primer, and 0.2μL of forward primer with additional M13 sequence. This is depicted visually in Figure 9.
Nested PCR reactions, such as the one described above, allow researchers use creative methods to build the molecular products, such as fluorescently tagged microsatellites or genes with particular mutations, in an inexpensive manner. This has many benefits. It allows molecular research to be done at a variety of institutions as opposed to only those with a large enough budget to afford these products. This increases the combined body of work and allows the field to advance at a faster pace. It also gives individuals who are new to the field the ability to investigate questions that have specific molecular requirements without having to invest, and therefore risk, large amounts of lab budgets.
Figure 11. Concept Map of My Experimental Plan. Steps in blue are those been completed in the experimental process, and steps in pink are those that have not been completed. The steps in pink will be expanded upon in the discussion and conclusion. “Species Identification” is connected to “Run Statistical Analysis” because while this step can be done sooner, it is only necessary to do this before analysis on microsatellite data.

Results

Species Diversity and Distribution

I worked 63 specimens typed to species. While there were no species was collected at all Appleton area locations and most species were collected at multiple sampling locations. The one exception was B. borealis, which was only found at the Gordon Buboltz Nature Preserve Field location. B. greiseocollis was collected at 7 out of the 10 different field sites, making it the most
widespread according to my limited sample size. 9 out of the 12 species were only found at 3 or less locations (Figure 11).

![Stacked Bar Chart of Locations where Species were Collected](image)

*Figure 12. Stacked Bar Chart of Locations where Species were Collected. Count refers to the number of specimen collected for each species. This data is also shown in Table 2. All species have the same genus ("Bombus"), so are only listed by their species name. Lawrence University was shortened to LawrenceU.*

No sampling location had only one species that was collected. Lawrence University and Riverview Hive had the highest number of species collected at each site, with six species found at both sites (Figure 12). Though, it is important to note that the most land area cover and time was spent collecting specimens occurred on Lawrence University’s campus. Therefore, this species diversity shown at this site may solely be indicative of that increased sampling effort. In general, as the number of specimens collected increased at a location, the number of species collected also increased.
DNA Extraction

The protocol used in my first attempts at DNA extraction were influenced by procedures I had done before in classroom lab settings. Specifically, I was extracting DNA from model organisms such as C. elegans and E. coli. There is a large body of knowledge and developed procedures for studying these species. Both of organisms also have cell walls that are easily broken into. Therefore kits, such as the Quiagen DNeasy Blood and Tissue Kit™, and the included standardized protocols can easily extract DNA in high concentrations with good purity.

In contrast, insects have tough exoskeletons that can be difficult to break into. Even when an insect’s leg is macerated to reveal muscle tissue, these kits to not utilize harsh enough methods to break chitin. There are procedures that can be performed on a DNA sample before putting it
through a kit’s extraction protocol, but these are time intensive, such as putting a sample through several freeze thaw cycles in order to break open cell walls. Even if using the Quiagen™ DNA extraction kit was effective, this method would have been quite time consuming as it requires the repetition of several step process for every individual specimen.

The protocol for DNA extraction that was included in the final methodology is more useful for many reasons. As mentioned before, it took less time. It also produced more consistent results. I believe that this occurred because the Chelex™ solution was harsh enough to break through insect cuticle. The standard deviation of the average concentration (Table 5) is quite high (standard deviation 990, average of 90.2), but that is due to a few low concentrations within the average. What impressed me most about this method of DNA extraction was 260/280 ratio (average 1.78, standard deviation 0.43). If DNA is available in a relatively pure state, even small concentrations can be effectively PCR amplified. In my final run of PCR, the average concentration was 531ng/uL (standard deviation of 182) with a 260/280 ratio of 1.75 (standard deviation of 0.05). These results assured me that even though the fragment analysis did not turn up any results, there was PCR product.

Establishing Final Protocol

When I first began this project, I knew that this project would require a high level of independence. As there is not professor at Lawrence University whose research focuses on the subject of molecular ecology, there were no established protocol, or previous student’s lab notebook that I could use to inform my experimental and methodological design. I was lucky enough to receive a paper written by Jonathon Koch, “Patterns of population genetic structure and diversity across bumble bee communities in the Pacific Northwest”, from one of my
mentor’s (Relena Ribbons). The study featured in this paper utilized methodology that was easily scaled down to my project. As I said before, I even received extra material on methods from Koch (Koch, 2015). From this, I was able to create a base experimental plan in order to answer the question I described in “My Hypothesis”.

Troubleshooting both PCR and DNA extraction taught me a large amount about both processes. In the case of DNA extraction, I found out more about the chemical and mechanical methods necessary to break open cells and nuclei. This experience also taught me how protocol can be modified based on the specific challenges of a situation, such as breaking open stubborn cuticle. Troubleshooting PCR required a different set of experimental skills. Details are quite important, especially in molecular laboratory work. Perfecting the reagents and their volumes within my PCR reaction was at times a slow and frustrating process. Even when other individuals looked over my reagent list, some of my errors were missed due to assumptions based on wording, as was the case during the period in which I was only using forward primers when performing PCR. My next struggle was figuring out how to properly multiplex samples in order to save money on fragment analysis. I put two sets of untagged primers that shared the same annealing temperature in a reaction, but then realized that it would be impossible to differentiate these fragments from one another. Yet, my next step in my process I made a similar mistake. Though I amplified all loci independently of one another, I combined unmarked PCR products with one another. I did not realize that even if these samples had not been combined, a capillary genetic analyzer needs fluorescence in other to read fragment length. I had not been thinking about the requirements of the technology used to analyze fragment length. The next error that I committed was in primer design. Instead of ordering complementary M13 sequences, I maintained the same sequence on my forward primers and my fluorescently tagged primers,
making it impossible for them to bind. This taught me to always think about what I’m trying to create when ordering supplies, especially primers.

Both troubleshooting processes taught me how to be more deliberate as a scientist, and as well as push past my frustrations in order to achieve the goals I set forth. I can say with certainty that with the protocol I have now developed, that I would be able to complete my experiment. This would require following “The Final Protocol”. The next steps would be to convert the files into a readable format and then use a software to score the alleles as described in “Microsatellite Length Analysis”. From there, data analysis can be performed (statistics described “Statistics for Microsatellite Analysis”). These results would then be overlaid on a geographic map and interpreted (as described in “Overlaying Results on a Geographic Map”).

The Final Protocol

Specimen Collection:

- Collect bumble bees using a net.
- Chill the specimen until it becomes immobile.
- Take images of specimen and either type to species immediately or do so later based on these images.
- Cut off 2mm of the metatarsus from the mesothoracic leg with a clean scalpel and store this in 100% ethanol with a specimen tag including collection data, collection sites, species identification, and specimen number.

DNA Extraction:

- Note: This can take place either in a single PCR tube, 8 strip, or 96-well plate.
• Pipette 150uL of %5 Chelex™ directly after vortexing, in order to make sure the Chelex™ is in solution, into a PCR well.

• Pipette 5uL of proteinase K into the well. (Proteinase K solution is made by combining 10mg with 1000uL of milliQ water)

• Add the 2mm metatarsus clipping to the PCR well and broken along the side of the well using forceps that are cleaned before and after using 100% ethanol.

• Flick or shake PCR well, in order to mix the reagents, and then tap it against the table to pool the reagents at the bottom of the well.

• Place the PCR well into a thermocycler with the following settings: 50C for 60 minutes, 99C for 15 minutes, 37C for 1 minute, 99C for 15 minutes, and held at 15C.

• Nanodrop™ this solution to measure the concentration and purity.

PCR

• Create a master mix for the number of PCR reactions you plan to set up. Creating a mix for more than the number of reactions you plan on setting up is advisable in order make sure you do not run out of master mix. The reagents and the volumes that should be multiplied to create the master mix for one locus are:
  o 0.85uL of GoTaq Flexibuffer™
  o 0.5uL of GoTaq™
  o 0.6uL of dNTPs (10mM)
  o 4.8uL of milliQ water
  o 0.5uL of MgCl₂ (25mM)
  o 0.4uL reverse primer
  o 0.4uL M13 primer with 5’ fluorescent tag
0.2uL forward primer with an M13 tag on the 5’ end

- Note: When ordering the M13 primer with the 5’ fluorescent tag and the forward primer with an M13 tag on the 5’ end, make sure the two M13 sequences used are complementary to one another.

- Vortex this solution and then spin for 30 seconds at 5000 rpm.

- Pipette 9uL of this mix to each PCR well.

- Add 1uL of template DNA to each well, documenting beforehand which template DNA is going into which well.

- Spin the wells down before placing them into the thermocycler.

- Thermocycler setting:
  1. 95C for 7 minutes
  2. 95C for 30 seconds
  3. 52C (for B10 and B11) or 57C (for B124 and B126) for 45 seconds
  4. 72C for 45 seconds
  5. Repeat steps (2-4) 30 times
  6. 94C for 30 seconds
  7. 53C for 45 seconds
  8. 72C for 45 seconds
  9. Repeat steps (6-8) 8 times
  10. 72C for 10 minutes
  11. Hold at 4C

- Nanodrop™ each reaction to measure purity and concentration.

Preparing Samples for Fragment Analysis
• Pipette 9μL of milliQ water into a well of a 96-well plate.

• Pipette 1μL of the PCR product into the well, denoting the location of each product within the 96-well plate.

• Firmly secure strip caps on the plate.

• Wrap the plate in parafilm to assure the caps stay on during transport.

• Put wrapped plates in an envelope with padding and the fragment analysis order form received after submitting the order on GeneSifter®

(https://yale.genesifter.net/gsle/mainPage)

• Address the envelope to:

  DNA Analysis Facility on Science Hill
  Attn: Carol Mariani
  170 Whitney Ave
  ESC Room 150
  New Haven, CT 06511
  Phone: 203-432-7394

Discussion/Conclusion

Microsatellite Length Analysis

Alleles are can be scored using various software programs designed to do so, such as GeneMapper™ (Yale, n.d.). The files received from the Yale DNA Analysis Facility on the Hill need to be converted before they can be read by allele scoring programs. Gm Convert™ is one program that can be used for file conversion (Yale, n.d.). Many more options for fragment analysis can be found in the fragment analysis tab on the website for Yale’s DNA facility’s website. Allele scoring must be done before any statistical analysis can occur.
Statistics for Microsatellite Analysis

Neither SMM or IAM are perfect representations of microsatellite mutation dynamics. Both have their strengths and weakness, and each have a specific statistic based on the assumptions of these models (Balloux & Lugon-Loulin, 2002). $F$ statistics are based on IAM. $F_{ST}$ is a measurement of inbreeding that examines the correlation between the subpopulation an organism lives in and their genotype (Balloux & Lugon-Loulin, 2002). One advantage of $F$ statistics is its sensitivity to mutation (Balloux & Lugon-Loulin, 2002). There is also established interpretations of scores within certain ranges (Balloux & Lugon-Loulin, 2002). The disadvantage of this statistic is that it does not account for size homoplasy (Balloux & Lugon-Loulin, 2002).

$R$ statistics are based on SMM. $R_{ST}$ is a comparable to $F_{ST}$ operating under SMM (Balloux & Lugon-Loulin, 2002). $R$ statistics better represent microsatellite dynamics, especially in a structured populations and when analyzing loci which mutate in a stepwise fashion as described by SMM (Balloux & Lugon-Loulin, 2002). This statistic is limited by its assumption that there is an infinite possible number of alleles (Balloux & Lugon-Loulin, 2002). This is an inaccurate assumption, as microsatellites have shown to have to not mutate to have more than a certain number of repeats that is unique to each microsatellite loci (Balloux & Lugon-Loulin, 2002). Therefore, there is a maximum number of repeats an allele can reach, limiting the number of possible alleles. Studies will report both $R$ and $F$ statistics because each provides valuable information and it allows the reader and researcher to be critical of their limitations (Balloux & Lugon-Loulin, 2002).
Table 9. Comparison of the Different Statistical Equations based on Different Microsatellite Mutation Models. 'V_p' represents variance in allele frequency among subpopulations. 'p' represents mean allele frequency among subpopulations. 'S' represents the average squared allele size within populations. 'S_w' represents the average sum of squares of the difference in allele size within each subpopulation. 'S_r' represent the difference in allele size between populations. Equations taken from (Balloux & Lugon-Loulin, 2002).

<table>
<thead>
<tr>
<th>Model</th>
<th>IAM</th>
<th>SMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equation for Measuring Inbreeding Correlation</td>
<td>$F_{ST} = \frac{V_p}{p(1 - p)}$</td>
<td>$R_{ST} = \frac{(S - S_w)}{S}$</td>
</tr>
</tbody>
</table>

Overlaying Results on a Geographic Map

After statistical analysis has been done on the data, it is helpful to overlay it on a geographic map. An example of this Figure 14, which represents each population as a circle, whose area corresponds to the size of the circle (Koch et al., 2017). Each color within the circle represents a different genotype (Koch et al., 2017). For B. sylvicola it can be seen that while populations that are located close to one another have relatively similar genetic makeup, those far away from each other are quite different. For B. mixtus, none of the populations have similar genetic makeups. This stands in contrast to B. melanopygus and B. flavifrons, which have quite populations with similar genetic makeups across the entire map. These realizations could have also been made after looking at graphs such as Figure 3, but overlaying this on a geographic map slows the researcher to better hypothesize the mechanisms for genetic isolation. Specifically, in this case, the study site was quite mountainous, and therefore the researcher thought that elevation had a strong effect(Koch et al., 2017), and placed this data on a topographic map. In general, overlaying data figures on a map gives whoever is looking at this figure an opportunity to point of the features in the landscape that may be effecting how individuals move and
reproduce within a landscape, and therefore how genes flow across that landscape. This is how I formed the hypothesis for my experiment as described in “My Original Hypothesis”. By thinking about how geography affects how organisms reproduce across a landscape, it allows for a more dynamic understanding of genetic flow and that is why the field of landscape genetics is so valuable.

![Genetic Composition of Bumble Bee Species](image)

**Figure 14.** Genetic composition of four different bumble bee species in the Pacific Northwest. The different colors represent different haplotypes. Species A and B have more limited niches, and are only able to exist at specific altitudes. Species C and D have broader niches and can exist in a larger range of altitudes. A and C show greater differentiation between populations, indicating lesser gene flow, compared to C and D which have relatively stable gene composition across populations. Taken from (Koch et al., 2017). This data is also represented in Figure 3 of this paper.
Future Directions

Though it is useful to identify species and populations at risk of extinction, this is only the first step. It is important to communicate this information to those not only in the research field, but to the individuals who are primarily involved in conservation. The individuals who are out in the field implementing conservation methods need to know species or locations to apply conservation techniques. It is also necessary to translate this information into relevant statements to share with those policy makers who decide how much and where funding is allocated and how land can be used. A simple land use change, such as a decrease in the frequency of mowing grass and pesticide use in the land underneath powerline easements, has shown to increase native bee diversity (Russell, Ikerd, & Droege, 2005). Actions such as this create a refuge out of an area that is not even usable for most other purposes. This management practice also has the added benefit of costing less than its less environmentally predecessor, making policy mandating this action easily defended. By demonstrating the importance of implementing practices in order to help save struggling pollinator population, policy could be passed easier or enacted faster.

Another study demonstrated that bumble bee diversity within a city was shown to increase specifically with floral diversity. This study also showed that it didn’t matter how “landlocked” (surrounded by concrete) areas with flowers were, bumble bees were able to find them (Marietta et al., 2016). Urban areas can easily create refugia and corridors for bumble bees. By establishing urban gardens throughout a city that include native flowering plants, efforts can be made to conserve bumble bees while creating an aesthetically pleasing landscape. As it has been demonstrated that urban areas have high native bee diversity, conservation efforts can be
low effort, high reward situations (Hall et al., 2017). It has been confirmed that bumble bees colonies respond positively to the beneficial aspects of a habitat, with little negative response to negative aspects of a habitat (Herrmann et al., 2007). The combined ease of creating urban refugia (Marietta et al., 2016) due to the low level of ecological needs that bees have, but are not served by the urban environment combined with the conservation resources available within cities (Hall et al., 2017) make urban areas valuable locations for managers to concentration their conservation efforts.

Population and landscape geneticists need to communicate their results with the public, policy makers, and restoration ecologists. By contributing to a greater understanding of how bee function in urban areas, the creation of bee friendly spaces will be easier (Hall et al., 2017). This is especially important as our world becomes more urbanized. If this is done, large strides can be made to save bumble bees and other native bee species. When conservation integrates ecological data with genetic analysis, it allows researchers and other shareholders to better direct their efforts and resources in order to be the most efficient (Geist, 2014).

Broader Implications

Gene flow has been shown to be important for bumble bee species, as colony size and parasite resistance increases with genetic diversity (Baer & Schmid-Hempel, 1999; Herrmann et al., 2007). By documenting gene flow within a species, it is easier to identify genetically isolated populations, which should be managed independently for conservation purposes (Geist, 2014; Pearse et al., 2006). Yet, defining conservation units and measuring colony health are not the only reasons to use population genetics to understand bumble bee and pollinator ecology.
Genetic analysis has given researchers new methods for investigating ecological questions concerning bumble bees or native insect pollinators in general. When studying a small flying organism, observing behavior is not always feasible (Knight et al., 2005). Additionally, bumble bees nest in the ground and these nests can be difficult to identify (Knight et al., 2005). Therefore, the use of microsatellites to identify the number of males that queens of different species mate typically mate with (Estoup et al., 1995), or to calculate nesting densities (Knight et al., 2005) shows how genetics can be used answer ecological questions in an efficient and effective manner. Molecular ecology opens doors for researchers, by allowing them to use some as small as a molecules to understand species ecology.
References


LifetechnologiesCorp. (2015, July 14). Retrieved May 01, 2018, from https://www.youtube.com/watch?v=RzKFR-n5zs


https://doi.org/10.1007/s13205-017-1030-0

https://doi.org/10.1111/j.1365-294X.2006.02869.x

Perfect Repeats. (n.d.). Retrieved from
https://www.bioinformatics.org/microsatellite/wiki/Main/PerfectRepeats

https://doi.org/10.1016/j.tree.2010.01.007


https://doi.org/10.1016/j.biocon.2005.01.022


https://doi.org/10.1016/S0169-5347(03)00008-9


Whiteley, A. (2017). Population genetics of wild and managed pollinators: implications for crop pollination and the genetic integrity of wild bees. Conservation Genetics, 0(0), 0. https://doi.org/10.1007/s10592-017-0955-5


