10-27-2017 10:30 AM

Assessing Landscape Effects on Genetics and Dispersal of the Rocky Mountain Apollo Butterfly Parnassius smintheus using a Resistance Mapping Approach

Ning Chen
The University of Western Ontario

Supervisor
Dr. Nusha Keyghobadi
The University of Western Ontario

Graduate Program in Biology
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science

Recommended Citation

Chen, Ning, "Assessing Landscape Effects on Genetics and Dispersal of the Rocky Mountain Apollo Butterfly Parnassius smintheus using a Resistance Mapping Approach" (2017). *Electronic Thesis and Dissertation Repository*. 5058.
https://ir.lib.uwo.ca/etd/5058

Abstract

Landscape variables that best explain genetic differentiation may not also best explain dispersal patterns, but many studies use genetic differentiation as a proxy for dispersal. I tested the effects of landscape on both genetic differentiation and dispersal in parallel, to explore whether landscape effects on genetic differentiation between populations and landscape effects on dispersal would be comparable in such contexts. I used circuit theory (Circuitscape) and least cost transect analysis to evaluate the effects of landscape on both movement and genetic differentiation of the butterfly, *Parnassius smintheus*, in the Jumpingpound Ridge study system. Circuit theory and least cost transect analyses did not identify the same best predictors to explain genetic differentiation and dispersal data. Circuit theory produced more accurate results with higher precision. Genetic differentiation should not be used as a sole proxy for dispersal in studies of landscape effects, but should be supplemented by more direct measures of dispersal.

Keywords

Genetic differentiation, dispersal, landscape, butterfly, circuit theory, least cost transect, resistance surface.

Co-Authorship Statement

All work presented in this thesis was completed under the supervision of Dr. Nusha Keyghobadi at the University of Western Ontario. Fieldwork data were collected in collaboration with Dr. Stephen Matter from the University of Cincinnati and Dr. Jens Roland from the University of Alberta. Map and resistance surface creation and data analyses were completed by Helen Chen. This thesis has been written by Helen Chen.

Acknowledgments

First and foremost, I would like to express my heartfelt gratitude to Dr. Nusha Keyghobadi for her supervision and guidance in the past two years. Without her immense knowledge and continuous support, this project could not have happened. Dr. Nusha Keyghobadi is not only an inspiration and role model as a researcher and a teacher, but also as a person I would strive to emulate as I look to the future.

I would like to thank my fellow lab mates in the Keyghobadi Lab group (past and present): Benoit Talbot, Maryam Jangjoo, and Melissa Lucas, for their continued academic and emotional support. I would like to thank Melissa Lucas for her assistance in the statistical analysis and least cost transect portion of my project.

I would also like to express my gratitude to my advisory committee: Dr. Marc-Andre Lachance and Dr. Adam Yates, for their perceptive insights and vast sea of knowledge. Specifically, I would like to thank Dr. Marc-Andre Lachance for his empathy and his philosophy of life.

I would like to thank the Biogeoscience Institute of the University of Calgary as well as the numerous field work teams for their countless hours of butterfly catching and data collection. Without such a complete set of data this project could not have been completed. Specifically, I would like to thank Dr. Stephen F. Matter for his guidance and endless knowledge in the field.

I would like to thank Christine Homuth for her expertise in geography and specifically ArcGIS, without whom this project would have never proceeded past the initial map creation phase.

Lastly, I would like to thank my family for their unwavering support in all that I do and all that I will ever pursue.

Table of Contents

List of Tables

List of Figures

List of Appendices

1 Introduction

1.1 Habitat loss and fragmentation

Loss of habitat, as well as fragmentation of remaining habitat, have become more prevalent in recent times due to human activity (Fahrig, 2003). This has significant detrimental impacts on species diversity and the persistence of natural populations (Zuberogoitia et al., 2013). Shrinking habitats reduce the amount of resources available to the species that depend on it. This can result in more competition for the limited resources, and therefore decrease species survival (Keyghobadi, 2007). Roads, residential sprawl, farmland, and logging practices often develop in ways that break up large, contiguous habitats into smaller, disconnected patches. This affects community composition as it increases the amount of edge habitat, which tends to have less species diversity and richness, and decreases interior habitat area (Theobald et al., 2011). If patches are close enough with no definitive barrier in between, mobile species might still be able to move between them, though incurring higher mortality risks due to increased vulnerability to predation or vehicle strikes (Jr et al., 2016). However, barriers and distances among habitat patches can be so great that populations in different patches are effectively separated from each other, with no further movement and therefore no gene flow between them (Fahrig, 2003). This results in significant genetic effects on the populations (Caplins et al., 2014).

1.2 Genetic effects of habitat loss and fragmentation

The genetic effects of habitat loss and fragmentation are due to both the reduced population sizes and restriction of movement between population patches. Fragmented landscapes have smaller total area and more limited resources, and so can support a smaller number of individuals compared to a non-fragmented landscape. Smaller populations are more prone to the effects of genetic drift (Ouborg et al., 2006). Isolated patches will not have gene flow from neighbouring patches, and so diversity within patches is reduced, and populations occupying patches will gradually become more genetically distant from each other (Jangjoo et al., 2016). Loss of genetic diversity in turn can hinder the ability of the population to adapt to future changes in its environment and potentially cause a decrease in fitness through inbreeding (Kutschera et al., 2016). Studies of the genetic effects of habitat loss and fragmentation are paramount in conservation efforts. It is important to take into account not just the patches of populations, as landscape elements such as elevation and land cover type, as well as the spatial configuration of the populations, will affect the ability of species to disperse between patches (Keyghobadi, 2007). Landscape genetics provides an important way of studying this problem.

1.3 Landscape genetics

Combining landscape ecology and population genetics, the field of landscape genetics presents an interdisciplinary way of analyzing how species of interest interact with their landscape (Manel and Holderegger, 2013). This field of study seeks to consider the quality, configuration, and connectedness of the landscape as possible explanatory variables for gene flow, population structure, and genetic differences between populations (Epps and Keyghobadi, 2015). Sampling of individuals or populations provides genetic data, while surveying at various spatial scales (i.e. from satellite imaging to ground-truthing) provides information about the landscape. Landscape genetics approaches incorporate a measure of the landscape in a way that ideally allows the effects of the landscape on dispersal to be quantified and compared to genetic data (Keller et al., 2013).

1.3.1 Methods in landscape genetics

One of the simplest methods used in landscape genetics is the transect approach, which uses landscape components along a straight line between two sites as explanatory variables to explain genetic differentiation between sampled populations occupying those sites. This method assumes that dispersal is limited only to straight-line movements (Emaresi et al., 2011). Another method is the least cost path (LCP) approach, which is calculated using a raster map (a map composed of cells each able to be characterized by distinct properties). The cumulative cost of the path takes into account the cost of each raster cell crossed by the path and the least cost path is that which is the least costly or resistant between the two sites. The landscape components of the least costly path are used as the explanatory variable in the study (Koen et al., 2012). The different landscape components on the raster map are assigned resistance values based on how they hinder movement, and so a disadvantage of the least cost path method is that it is only as accurate as the researchers' prior knowledge of the ability of a species to move through differing landscape elements such as land cover type and elevation (Adriaensen et al., 2003). These two methods can be combined into the least cost transect analysis (LCTA), in which one first calculates the least cost path before a buffer is added around this path. Adding a buffer gives the path a defined area, which is then treated as a transect. The proportion of each landcover type in the transect area is considered an explanatory variable that explains genetic distances between sites. Least cost transect analysis is also used to rank landcover types as to whether or not they are conducive to movement (Van Strien et al., 2012). Yet another method estimates ecological connectivity between pairs of sites based upon circuit theory. This method involves mapping the intervening area between sites as well as the surrounding landscape to create a raster map of the land cover surface, just as for least cost path analyses. Different land cover types can be assigned different resistance values depending on how difficult they are for the species of interest to move through, creating a resistance surface. Unlike least cost path analysis, however, the circuit theory method of modelling (commonly called Circuitscape due to the software most commonly used to conduct the analysis) considers all possible pathways between each pair of sites across the landscape as, akin to the flow of electricity across a circuit board, it considers potential movement of the organism of interest across the surface as a whole (McRae et al., 2008). Using Circuitscape involves assigning a starting and ending point on the resistance surface, and one final resistance value, called a 'resistance distance', is calculated for that pair of points. This resistance distance changes as the resistance value of individual land cover types change, but this final calculated value does not give any information on the distinct effects of individual land cover types. A model created in Circuitscape is very flexible in its ability to incorporate data of many different scales, landscape features, varying resistances, and dispersal distances (McRae and Beier, 2007). With constantly improving tools available for modelling the landscape and its components, including barriers, edges, and corridors conducive to movement, landscape genetics will continue to become more versatile and functional.

1.3.2 Genetic markers in landscape genetics

The use of molecular markers in landscape genetics has also been increasing as new methods of tracking gene flow and genetic distances are developed. Depending on the timeframe of interest, markers that track evolution at different rates can be selected (Epps and Keyghobadi, 2015). Genetic markers with higher rates of mutations, such as microsatellites and single nucleotide polymorphisms (SNPs), are used to track more recent changes (past tens of generations), while genetic markers with lower rates of mutations, such as subunit ribosomal RNA genes (Woese and Fox, 1977) (past thousands of generations), are used to track changes from further in the past (Wang, 2011). Note that various other factors besides timeframe should be considered when choosing appropriate molecular markers, as they have other distinctive characteristics. For example, microsatellites and panels of genome-wide SNPs represent multiple nuclear loci, while specific genes each represent a single marker (Epps and Keyghobadi, 2015). The use of the correct markers in landscape genetics is paramount as the data are commonly used as a representation of the response variable.

1.3.3 Genetic data as a proxy for movement in landscape genetics

Landscape genetics often focuses indirectly on the effect that landscape has on the movement and dispersal of species, as there are many instances where it would not be practical or feasible to collect enough direct movement data for a study (Levin et al., 2009). Collecting genetic data is often easier than direct methods of collecting movement data, such as in mark-recapture studies. For example, genetic data can be obtained from all the populations in one sampling period, whereas mark-recapture studies require returning to the site multiple times to track the marked specimens, and is therefore not always feasible (Chan and Karczmarski, 2017).

Some landscape genetics studies explicitly use genetic distance data between population sites as an indirect means of measuring dispersal and movement (Spear et al., 2010). For example, Emaresi et al. (2011) surveyed the genetic variation between 19 populations of alpine newt (*Mesotriton alpestris*) at 7 microsatellite loci to determine the effect of habitat fragmentation on movement between populations. This landscape study found that even though *M. alpestris* are readily able to disperse and occur in large population sizes, all of which contribute to increased gene flow, this alpine newt is susceptible to the detrimental effects of landscape fragmentation due to habitat destruction. In another example, Epps et al. (2007) surveyed 26 populations of desert bighorn sheep (*Ovis canadensis nelsoni*) using 14 microsatellite loci to represent dispersal as a part of their least-cost modelling approach. Because insufficient direct data on dispersal of the animals was available, genetic data, and related spatial information, were used to model their movement on the landscape. Their study found that bighorn sheep movements determined in this manner were consistent with known dispersal routes and barriers. More often, using genetic data as a proxy for movement data is implicit. This substitution of genetic differentiation for actual dispersal is based on the assumption that landscape elements will exert an effect on the dispersal of the species, which will then affect the gene flow between populations of that species, which will ultimately lead to genetic differentiation between the populations. This may not always be the case. For example, dispersal only leads to gene flow if the dispersal is followed by successful reproduction at the new location (Levin et al., 2009). As well, gene flow might not lead to perceivable genetic differentiation if the new alleles are lost through genetic drift (Keyghobadi, 2007).

1.4 Study objective

Landscape variables that best explain genetic differentiation may not also best explain dispersal patterns, and yet it is not uncommon for studies to use genetic differentiation as a proxy for dispersal. For my project, I want to test the robustness of this claim and determine the validity and reliability of substituting genetic differentiation for dispersal data in instances where the latter might be difficult to collect.

1.4.1 Study species

My study species is the Rocky Mountain Apollo butterfly, *Parnassius smintheus*, which inhabits high altitude mountainous regions above approximately 2000m. *Parnassius smintheus* is abundant in alpine meadows all throughout the eastern portion of the Rocky Mountains of Canada and the United States (Keyghobadi et al., 1999), where the host plants of the larval stage, the stonecrop (*Sedum lanceolatum* and *Sedum rosaceae)*, grow (Roslin et al., 2008). Adult butterflies feed on nectar of a variety of plants. Males of this species are whiter while females have a darker, peppered wing. These butterflies are univoltine; adults fly and mate from mid-July to the end of August and beginning of September (Keyghobadi et al., 1999). Butterflies of this species usually fly close to the ground, and will land if it becomes too windy. Roland et al. (2000) found that the average dispersal distance, as determined by mark-recapture, is approximately 150m. The females lay eggs on the ground near the host plants, and the eggs overwinter under the snow before hatching the following summer. Population sizes of *P. smintheus* can fluctuate from year to year, as the survival rate of the eggs depend heavily on the winter temperatures and the amount of snow cover (Roland and Matter, 2016). *P. smintheus* has been consistently studied over the past two decades as a model species for population ecology and population genetics (DeChaine et al., 2004; Fownes and Roland, 2002; Ross et al., 2005).

1.4.2 Study location

Parnassius smintheus is an ideal study species due to its abundance, and therefore large available sample size, and the patch quality of its habitat meadows. For these reasons, it has been extensively studied over the last twenty years. The study area is Jumpingpound Ridge in Kananaskis, Alberta, located in the foothills of the Canadian Rocky Mountains just west of Calgary. This landscape consists predominantly of forest and is interspersed by open meadow areas at higher altitudes along the ridgetops. Fifteen grassy meadows (hereafter referred to as ‘sites’) have been identified along the ridgetop on Jumpingpound Ridge and the adjacent Cox Hill (Figure 2.1), where *P. smintheus* and its host plants are found in abundance. The rest of the landscape consists of mainly forests of various coniferous trees with areas of rocky outcroppings and barren, gravelly ground (Keyghobadi et al., 1999). Most meadows are surrounded on all sides by forest, but a few meadow pairs have no forest between them. This resulted from dividing a larger, contiguous meadow into smaller meadows, through the creation of imaginary boundary lines for the purpose of grouping butterfly captures. Butterflies typically fly through meadows along the ridgetop when dispersing, and are rarely recorded flying straight over forested valleys (Roland et al., 2000). As the Kananaskis region represents a multi-use area, some clearings in the forest at lower elevations are the result of logging or other human land use. The total area of interest is approximately 5km by 5km, with the tallest peak (meadow M) at approximately 2.2km above sea level.

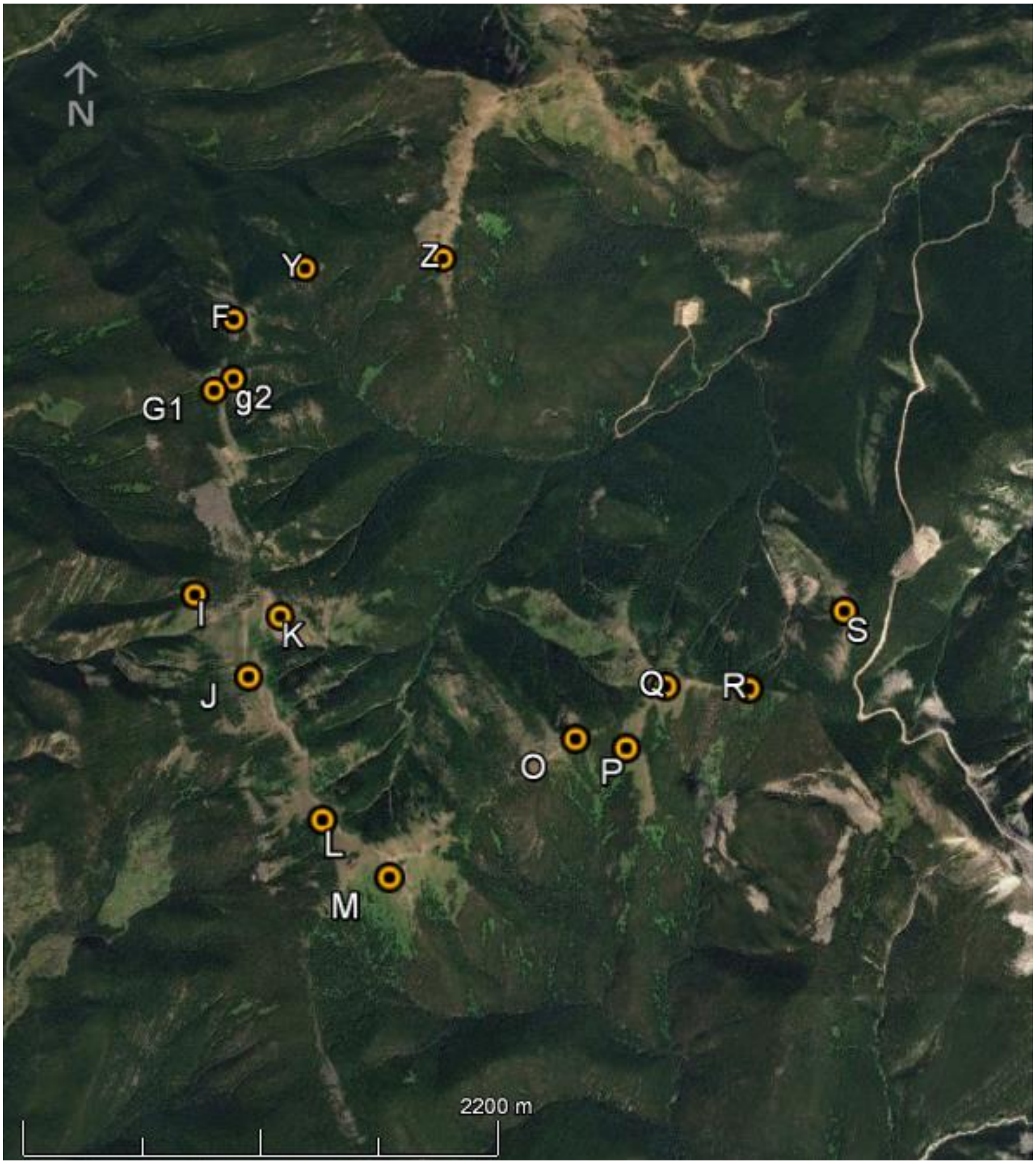

Figure 1.1 Jumpingpound Ridge in Kananaskis, Alberta. The 15 study sites where mark-recapture of *Parnassius smintheus* are conducted each year are identified by a unique letter. Site Z is located on the adjacent Cox Hill. The extent of this study area is approximately 5km by 5km. Aerial image from Google Earth Pro (Image © Google Earth 2016 DigitalGlobe).

1.4.3 Past studies on *Parnassius smintheus* landscape ecology and movement

Roland et al. (2000) looked at the effects of landscape and population size at 17 meadows of Jumpingpound Ridge and 3 meadows of the neighbouring Lusk Ridge. Mark-recapture of *P. smintheus* was conducted to estimate population size, as well as the amount of movement between meadows. Aerial photographs were used to determine meadow and landscape characteristics such as land cover type and elevation. That study found that while open meadows are easy for butterflies to move through, forested areas are twice as resistant to their movements. Butterflies also tend to stay at and seek out sites with large numbers of butterflies.

Matter et al. (2004) looked at the dispersal and survival rate of *P. smintheus* at 21 sites on Jumpingpound Ridge and Lusk Ridge. They found that migration between patches was infrequent, and so a model was used to provide estimates of dispersal. Butterflies in isolated populations had a slightly higher rate of mortality when dispersing, as determined by mark-recapture. Further, the authors found that intervening forest reduced the dispersal distance more than intervening meadows, which means that dispersal from a population depends on population size, land cover type, and distance between populations.

Ross et al. (2005) examined the behaviour of *P. smintheus* with respect to meadow/forest edges on Jumpingpound Ridge. By tracking the movements of male butterflies that had been released in either forest or meadow, they were able to characterize the species' behaviour within each land cover type, as well as at the edges where forest and meadows meet. On Jumpingpound Ridge such edges are quite abrupt and distinct. Male butterflies released in a meadow tended to turn away from forest edges. As well, butterflies flew less frequently and for shorter distances in forests compared to meadow. Ross et al. (2005) also observed that *P. smintheus* flew more readily at higher light intensities, and that light intensity was significantly higher in meadows as compared to forest. Light intensity tended to decrease in forest with increasing distance from the edge. Ross et al. (2005) therefore hypothesized that *P. smintheus* avoids flying in forests largely because of the lower light intensity there. As ectotherms, butterflies require heat from the sun to warm their flight muscles in order to initiate flight, and typically are not able to begin flying during overcast and cooler periods. In addition to land cover, population density may also affect movement of *P. smintheus*. Roland et al. (2000) studied the effect of population size in habitat patches on dispersal of *P. smintheus*. Using mark-recapture methods, they determined that butterflies in large populations were likely to stay in those populations, while butterflies in small populations were more likely to move away from their own populations and towards larger populations. This behaviour is possibly due to increased chances of successful mating in locations with more conspecific butterflies.

1.4.4 Past studies on *Parnassius smintheus* landscape genetics

Keyghobadi et al. (1999) looked at the relationship between the landscape and the genetic structure of butterfly populations. Mark-recapture of the butterflies from all 17 sites provided dispersal information and wing clippings were taken from some marked butterflies for microsatellite analysis at four loci for genetic differentiation. The number of butterfly recaptures between each pair of sites was predicted using a generalized linear model, to obtain non-zero pairwise movement estimates for all site pairs (including those more distant sites between which no actual recaptures were recorded). Topographic maps and aerial photographs were used to determine geographic distances between site pairs. Distances were not measured as straight lines, but rather along the ridge top, following observed patterns of butterfly movement. Furthermore, the total distance between each pair of sites was divided into two components: the distance that was over forest and the distance that was over open meadow. Using Mantel and partial Mantel tests, the study found that genetic distance between sites was negatively correlated with predicted movement, and that the distance through forest between sites was a stronger predictor of genetic distance than distance through meadow. This suggests that simple geographic distance does not explain genetic structure by itself, but the land cover type the butterflies move through is also very important.

Keyghobadi et al. (2005b) examined the effect of habitat connectivity on *P. smintheus* at a larger spatial scale, using a very different approach, to understand the effects of the landscape on genetic differentiation. The study area of approximately 40km x 135km included 27 sites across an extent of the Canadian Rocky Mountains. These sites were all high-altitude, non-forested areas, and were divided into three regions: East Kananaskis, West Kananaskis, and Banff. The East Kananaskis sites were at the foothills of the mountains, which are a hilly landscape of predominantly forest. West Kananaskis and Banff were at higher elevations and less forested. Patches of habitat for *P. smintheus* were bigger and more interconnected in the high-altitude, non-forested regions. Topographic maps were used to determine distances between sites, and landscape metrics were used to describe the patch quality and connectivity of the landscape within each region as they affect *P. smintheus* movement. At each region, a rectangular quadrat was placed around all the sites with a buffer of 2.5km on all four sides to encompass all dispersal. Patch density of high-altitude, non-forested areas within the quadrat was quantified. Tissue samples were collected from butterflies at all sites and typed at seven microsatellite loci in order to compare patterns of genetic differentiation among the three regions. The landscape metrics showed that in East Kananaskis, habitat for the butterfly is more fragmented compared to Banff and West Kananaskis. The proportion of area covered by high-altitude, non-forested areas was also much lower in East Kananaskis. This reduced landscape connectivity between sites in East Kananaskis was, in turn, associated with greater genetic differentiation, lower rates of gene flow, and lower genetic diversity.

Past studies on the landscape genetics of *P. smintheus* used relatively simple approaches to quantifying landscape structure, for example measuring the distance over forest and over open meadow between patches along ridge tops. My project builds upon these earlier studies by using more sophisticated approaches that integrate more information about the landscape.

1.5 Study overview

For my project, I used two relatively recent and complex landscape genetic approaches, specifically the Circuitscape approach and the least cost transect approach, to evaluate the effects of landscape on both movement and genetic differentiation of *P. smintheus* in the Jumpingpound Ridge study system.

I hypothesized that landscape effects on genetic differentiation between populations and landscape effects on dispersal would be comparable and parallel. I predicted that as I compared resistance surfaces of differing cell resistance values as my explanatory variables, while using genetic distance and dispersal as my response variables, the surface that best explained the data would be the same for both response variables (genetic distance and predicted dispersal). Furthermore, I predicted that the ranking of resistance surfaces, from best to worst, should also be similar for both response variables. If this turns out to be the case, it will be strong evidence that 1) genetic differentiation is a reliable substitute in cases where dispersal data cannot be obtained, and 2) the resistance values of the one best resistance surface (i.e., landscape model), as selected by looking at both dispersal data and genetic differentiation, is likely to reflect the real-life resistance values of the land cover types.

This analysis is important because finding that genetic differentiation between populations behaves similarly to dispersal information in landscape genetic analyses gives justification and support to past landscape genetics studies that use this substitution. However, if this analysis finds that genetic differentiation data between populations behaves differently from dispersal data, and that genetic differentiation is explained by different factors than dispersal, then I can recommend that future landscape genetics studies be more cautious of using genetic differentiation as a proxy for limited movement data, and consider using other means of measuring dispersal as well.

2 Methods

2.1 Genetic data and distances

I used genotypes at seven microsatellite loci (Ps50, Ps81, Ps85, Ps76, Ps163, Ps165, Ps262, see Appendix A, (Keyghobadi et al., 1999, 2002)) to determine the genetic distance between populations of *P. smintheus* occupying the different sites (Keyghobadi et al. 1999, 2005a). Microsatellite markers are highly variable and are considered a rapidly evolving, contemporary marker that reflects comparatively recent population genetic changes, on the scale of years to decades for a species that has one generation per year (Wang, 2011).

Small tissue samples from the lower edge of one hindwing were taken from some of the captured butterflies in 1995 and 1996 (Keyghobadi et al., 1999). This method of wing tissue sampling has no effect on a butterfly's survival or dispersal between patches (Koscinski et al., 2011; Roland et al., 2000). Whole butterflies were also collected the last day each site was visited in each year. Keyghobadi et al. (1999, 2005a) extracted DNA, amplified the microsatellite loci using PCR, sized fragments, and determined genotypes on an automated sequencer using Genemapper and Genotyper software (Applied Biosystems). A summary of genetic diversity metrics for each site is provided in Appendix B.

I analyzed the microsatellite genotypes using FreeNA (Chapuis, 2007) to account for the fact that many *P. smintheus* individuals have null alleles that do not amplify by PCR because of mutations in the primer binding regions that flank the microsatellite sequence (Carlsson, 2008). The presence of null alleles at these loci was inferred because of a higher than expected number of homozygotes in the data set, based on Hardy-Weinberg proportions, that cannot be explained by any other processes (Keyghobadi et al., 1999). Because null alleles are alleles at microsatellite loci that do not amplify by PCR due to mutations at the primer binding regions, heterozygotes can appear as homozygotes. For each locus, FreeNA uses the Estimation-Maximization algorithm to estimate simultaneously the frequency of the non-amplifying null allele and all other alleles, and to calculate unbiased estimates of genetic differentiation and distance (Chapuis, 2007). As a genetic distance measure, I used pairwise F_{ST}, which in these populations displays a stronger correlation with geographic distance than alternative genetic distances such as Nei's standard distance or the Cavalli-Sforza chord distance (Caplins et al., 2014). F_{ST} is a fixation index that reflects the loss of heterozygosity resulting from lack of random mating between populations, and provides a measure of genetic differentiation between the site pairs.

2.2 Dispersal data

I used estimates of dispersal between the Jumpingpound Ridge sites that were originally derived by Caplins et al. (2014). These estimates were based on data collected in 1995 and 1996 by Roland et al. (2000) through mark-recapture of *P. smintheus*. *Parnassius smintheus* adults were caught with butterfly nets during their flight season in July and August. Individuals were each marked with a unique three-letter code on their ventral hind wing using a black, permanent ink pen, and the location of the capture was noted before release. Recaptures of marked butterflies provides information on dispersal and distance travelled. However, between most pairs of sites we do not see any butterflies moving in a given year. As a result, direct observations of the number of movements between sites would yield values of zero for most pairs of sites. In order to have a non-zero estimate of the number of butterflies moving between each pair of populations, I used estimates of movement derived by Caplins et al. (2014) from an analysis of the mark-recapture data using the Virtual Migration Model (VMM; Hanski et al., 2000). VMM uses the mark-recapture data and a set of biological and statistical assumptions (Hanski et al., 2000) to obtain maximum likelihood estimates of the number of butterflies moving between all pairs of sites (Hanski et al., 2000). All such estimates are greater than zero, as movement between two distant sites can have a low probability, but is not impossible.

Movement estimates from the VMM are asymmetrical in that the estimated number of individuals moving from site A to B is not necessarily the same as the number moving from B to A. The estimate accounts for different rates of movement in each direction between two sites. However, my analyses required symmetric estimates of dispersal between each pair of sites. Therefore, for each pair of sites, I used the sum of the estimated number of individuals moving in the two directions; this essentially yields an estimate of the total ‘flow’ of individuals between each pair of sites. I then summed these total ‘flow’ values for the two years of data (1995 and 1996), and finally took the natural logarithm of the resulting number. I used the natural logarithm to normalize the dispersal values, as there were a few pairs of sites with very large estimates of movement between them while most pairs had very small estimates of movement.

2.3 Geographic distance

Geographic distance between each pair of sites was measured as the straight line, Euclidean distances between the site centroids using Google Earth Pro. With 15 sites, there were 105 site pairs.

2.4 Landscape analysis

In landscape genetics resistance surfaces are often employed as a way to represent the landscape and to study the movement of species across that landscape (Spear et al., 2010). A resistance surface uses a raster (i.e., pixelated) representation of the landscape and assigns different resistance values to the cells associated with different land cover types. The resistance values represent hypotheses of how difficult each land cover type is for the organism of interest to move through (higher resistance value = more difficult to move through). For my project, I created a landcover map and subsequent resistance surface of Jumpingpound Ridge, to analyze the landscape connectivity between the butterfly populations and the resistance of the landscape.

I exported a high quality aerial image of Jumpingpound Ridge from Google Earth Pro (Image © Google Earth 2016 DigitalGlobe) at its maximum resolution of 4800 by 2360 pixels. While the aerial image was taken years after the genetic and dispersal data was collected, distinct features such as tree lines and high elevation areas along the ridgetop remained unchanged. I stitched multiple exported images together using Adobe Photoshop to encompass the whole study area, plus a buffer of just over one kilometer around all sides so as to not cut off possible movement pathways (Koen et al., 2012). One kilometer is approximately the same distance as the longest recorded butterfly dispersal distance at Jumpingpound Ridge (Roland et al., 2000). The final stitched image was 4762 pixels in width and 5987 pixels in height with 72 dots per inch (dpi) (Figure 2.1). I rasterized this image in ArcGIS 10.3 at a resolution of 2m x 2m per cell, and georeferenced it following the NAD 1983 UTM Zone 11N projected coordinate system.

Figure 2.1 Aerial image of Jumpingpound Ridge, Cox Hill, and surrounding areas in Kananaskis, Alberta. This image is taken from Google Earth Pro (Image © Google Earth 2016 DigitalGlobe) and was used for land cover classification.

I then created a land cover map based on the rasterized and georeferenced aerial images, using ArcGIS's "Supervised Image Classification" tool. I created training samples where I designated areas on the image as either forest, open meadow, or barren rock (including gravel and scree), based on my own familiarity with the study area. ArcGIS uses these samples to create a signature file and is then able to classify every raster cell as one of these three land cover types based on colour bands. The result was a landscape map comprised mostly of forest, with meadow mainly across the ridgetops, and scatterings of barren rock. At this stage of the map, isolated pixels are more prone to classification inaccuracies (Erdey-Heydorn, 2008). I cleaned up the map post-classification using the tools "Majority Filter," "Boundary Clean," "Region Group," "Set Null," and "Nibble" to prevent the occurrence of single isolated pixels. This resulted in a land cover classification map that is more accurate and improved upon the "Supervised Image Classification" tool's ability (Figure 2.2).

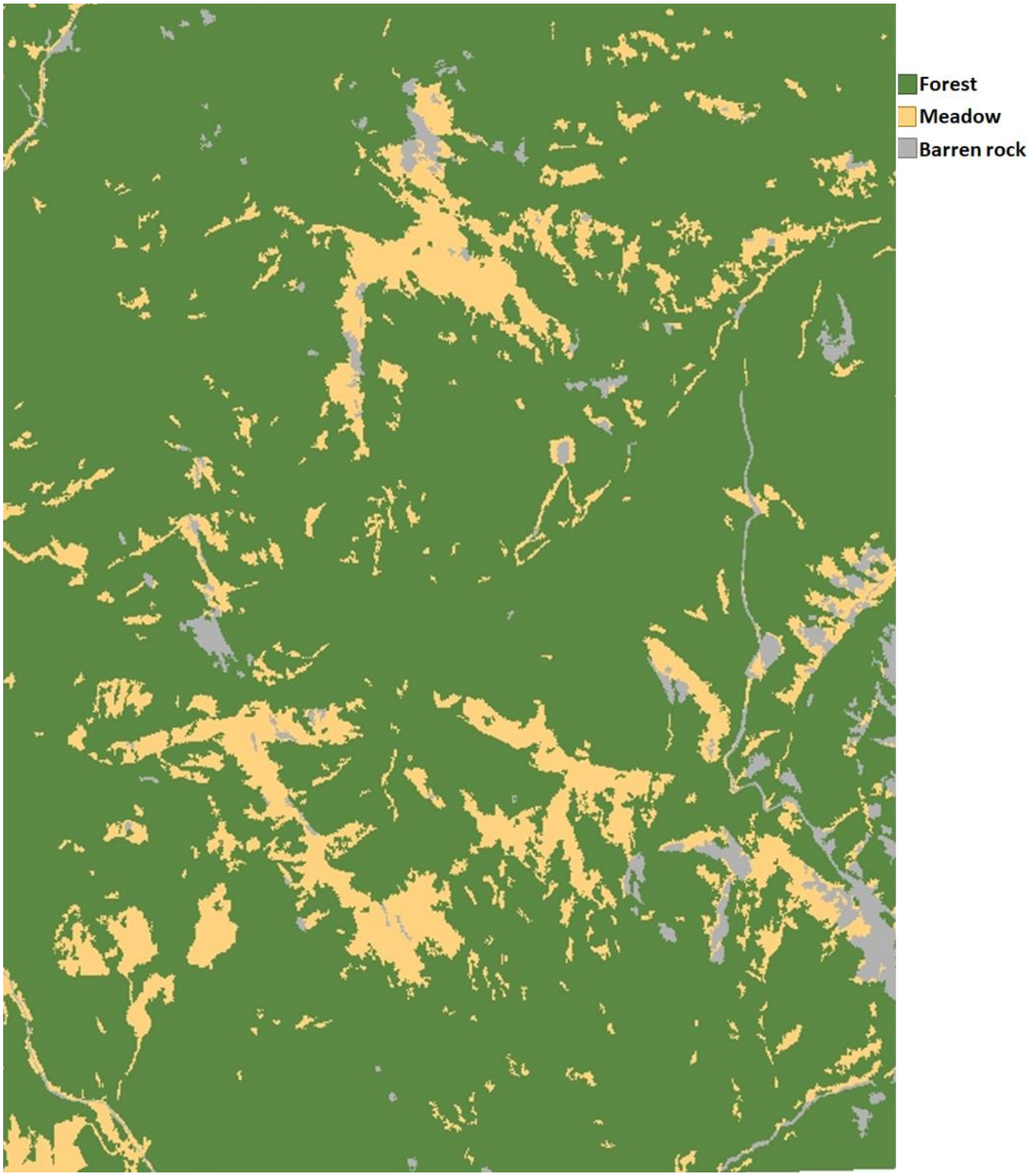

Figure 2.2 The aerial image (Figure 2.1) is rasterized and all cells are classified into one of three land cover types: forest (green), meadow (beige), and barren rock (grey) with the “Supervised Image Classification” tool from ArcGIS. The cleaned-up image is shown here.

The image classification process of ArcGIS was not able to differentiate high elevation butterfly meadows from low elevation clearings, as they look similar in colour on the aerial image. To address this issue, I included a Digital Elevation Model (DMTI Spatial Inc., 2003) (Figure 2.3) and used the elevation data to differentiate areas initially classified as meadow into the two aforementioned categories. The cut-off point of 1920m in elevation was chosen to differentiate meadows that could potentially be inhabited by butterflies (elevations of 1920m or higher) from low elevation non-habitat clearings (elevations of below 1920m). I had two requirements when deciding the elevation cut-off. First, the cut-off had to allow all sampled meadows where we know, through mark-recapture, that butterflies are found, to be classified as high elevation meadow. Site Y had the lowest elevation of the study meadows at approximately 1950m. The second requirement was that the elevation cut-off point had to allow the scattering of small meadow patches along the ridge top between the main meadows to be classified as habitat meadows, as butterflies have been captured there and are known to use these areas to disperse between meadows. At a cut-off of 1920m these two conditions are comfortably met. Incorporating the elevation data resulted in a final complete land cover map with four categories: forest, barren, high elevation meadow, and low elevation clearing. Note that while this elevation cut-off was the main method of differentiation between high and low elevation non-forested vegetation, it was manually adjusted and fine-tuned in certain areas of the map further to reflect knowledge of actual butterfly locations acquired through years of mark-recapture fieldwork (Figure 2.4).

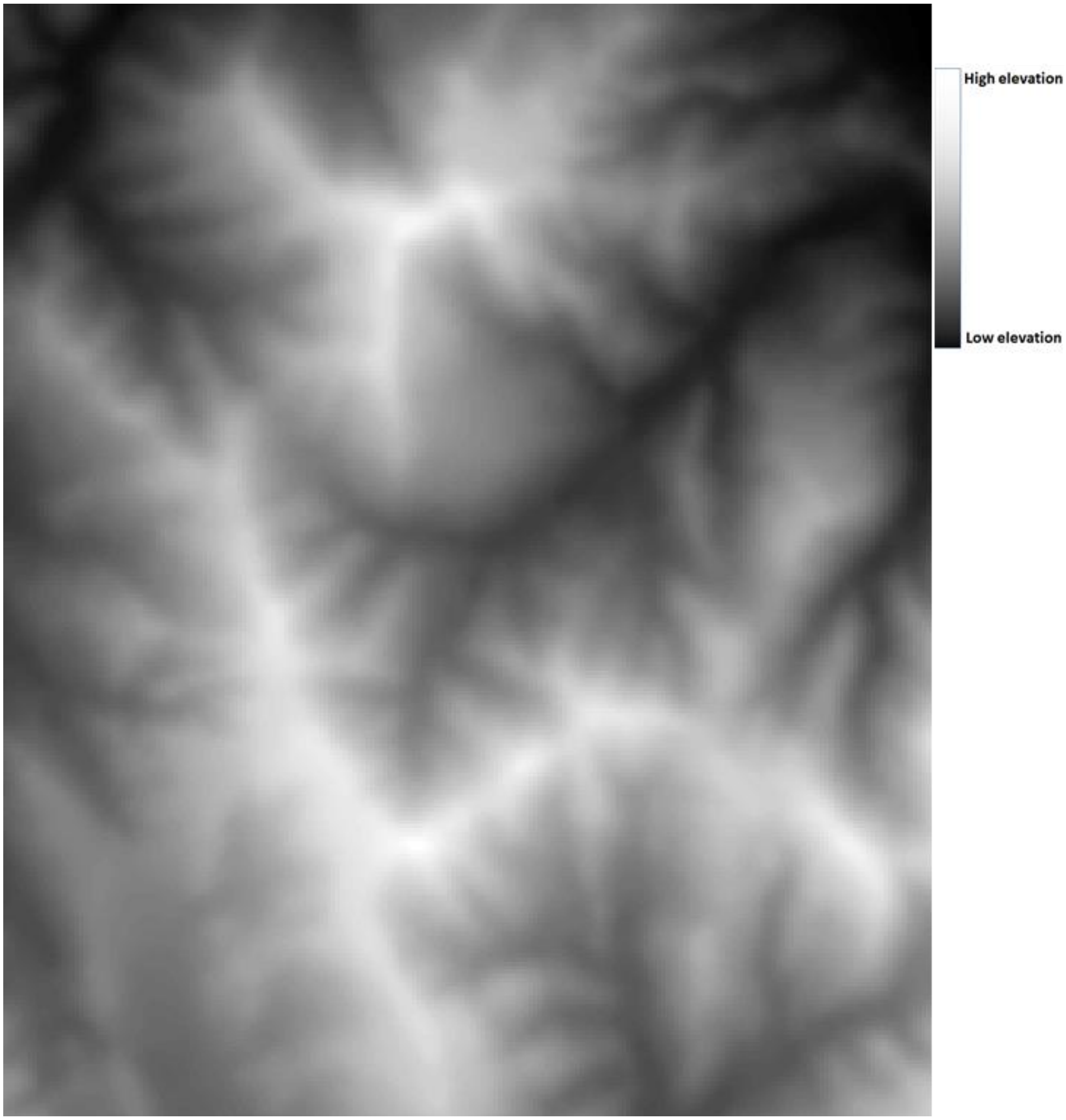

Figure 2.3 Digital elevation model of Jumpingpound Ridge, Cox Hill, and surrounding areas in Kananaskis, Alberta (DMTI Spatial Inc., 2003). Each cell has an elevation value, displayed here as a gradient from low (black) to high elevation (white).

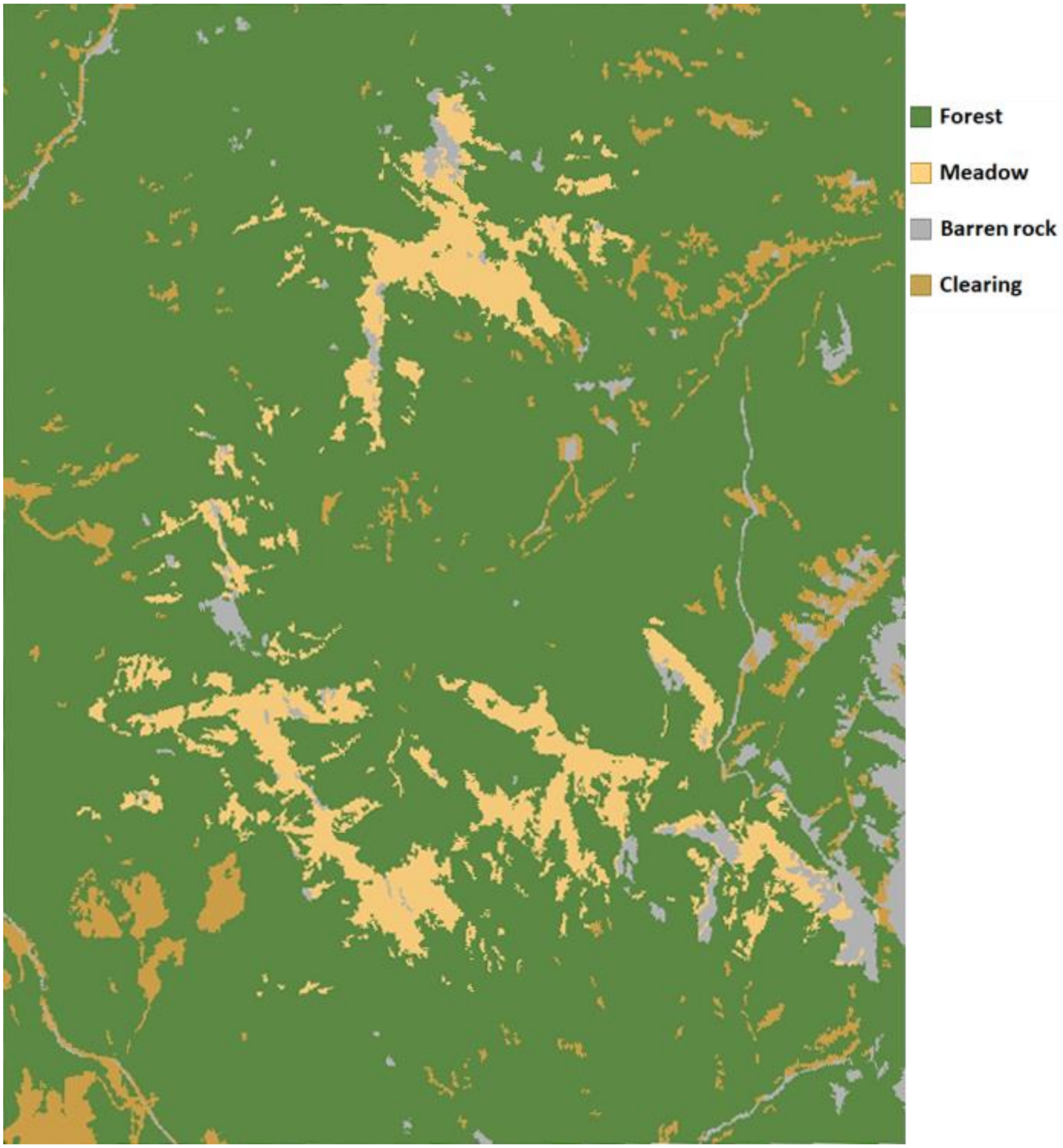

Figure 2.4 Combined result of the classified land cover (Figure 2.2) and the digital elevation model (Figure 2.3). Previous 'meadow' land cover is now classified into either high elevation meadow (beige) or low elevation clearing (brown) at the elevation cut off of 1920m. Only high elevation meadows are butterfly habitat. This image has been cleaned up so that isolated single pixels do not occur. The cut off elevation was adjusted in certain areas of the map to more accurately reflect knowledge of butterfly locations. This is the final map used in the creation of resistance surfaces.

2.4.1 Resistance values

Resistance values were assigned to different land cover types. These values represent hypotheses of how resistant the different land covers are to *P. smintheus* movement across the landscape. The resistance values are relative to each other and so have no units. As high elevation meadows containing host plants are ideal habitat for *P. smintheus*, I assigned an unchanging resistance value of 1 to them. Low elevation meadows are theoretically never visited by *P. smintheus*, and so I designated them as “no data” cells that have no nodes and cannot be travelled through. In order to narrow down the remainder of the possible combinations of resistance value sets, I first conducted a sensitivity test to determine the impact of varying resistance values for barren rock, relative to meadow. With meadow assigned resistance 1 and forest assigned (somewhat arbitrarily) resistance 3, the resistance value of barren rock was varied from 1 to 16 (values of 1, 2, 4, 8, and 16 were tested). For each resistance value of barren rock, a new resistance surface was created (meadow and forest resistances were unchanging across all these resistance surfaces). For each of these resistances surfaces I generated a resistance distance between all pairs of sites using Circuitscape (see below for description). I then made scatterplots of the resistance distances from the resistance surfaces (two resistance distances at a time). If varying the resistance value of barren rock does not have a large impact, then I expect these scatterplots to show straight lines. The results were all linear graphs, which tells me that varying the resistance value of barren rock does not have much of an impact on any final resistance surfaces I create. This is reasonable because there are very few barren rock cells on the resistance surface. Ecologically we know that barren rock does not hinder butterfly movement nearly as much as forest, and that many habitat meadows where butterflies are still present include small rocky areas within them. As a result, I subsequently tested resistance value sets where the resistance for both high elevation meadow and barren rock were fixed at a value of 1. For forest cell resistance, values of 1, 2, 3, 4, 10, 50, and 99 were used in the Circuitscape analysis, while values of 2, 3, 5, 10, 50, and 99 were used in the least cost transect analysis (see below for description). Roland et al. (2000) have estimated that forest is most likely between 2-3 times more resistant than meadow to butterfly movement. The exceptionally large resistance values of 50 and 99 were included to test if they give significantly different results from the more reasonable values, and to see if the effect of forest resistance plateaus at a certain point.

The resistance surfaces derived from each of the seven sets of resistance values were then compared to both the genetic distance and the estimated movement of *P. smintheus*, separately, using two commonly used LG approaches: Circuitscape and least cost transect analysis.

2.5 Circuitscape

Circuitscape (McRae and Nürnberger, 2006) is an approach, and an open source program, that is able to estimate connectivity and resistance across my land cover map, using the principles of electric circuit theory. This method treats the map as though it were a circuit board. The raster cells are considered nodes for conductivity, and adjacent cells are connected by resistors that differ in their strength depending on the resistance value assigned to the corresponding land cover type (Figure 2.5). Circuitscape calculations require that each of the 15 sites be assigned one focal node. For a given pair of sites, total conductance across the entire surface is then calculated between their focal nodes (McRae and Beier, 2007). The inverse of this conductance can then be used as measure of the resistance of the entire landscape to flow (either direct flow of individuals, or gene flow) between the two sites. This inverse of conductance is referred to as a ‘resistance distance’. I used the center point of all butterfly captures for each site in 1995 and 1996 as the location of the site’s focal node (where movement from each site begins and ends) in Circuitscape (Appendix C). For meadows I and K, these center points initially occurred on forest and barren rock, respectively, and so were moved slightly to the nearest neighbouring meadow cell. For each resistance surface (i.e., each set of resistance values) separately, I calculated the resistance for all 105 pairs of sites using the Circuitscape ArcGIS toolbox, which allowed me to call Circuitscape directly while working in ArcMap (McRae and Beier, 2007).

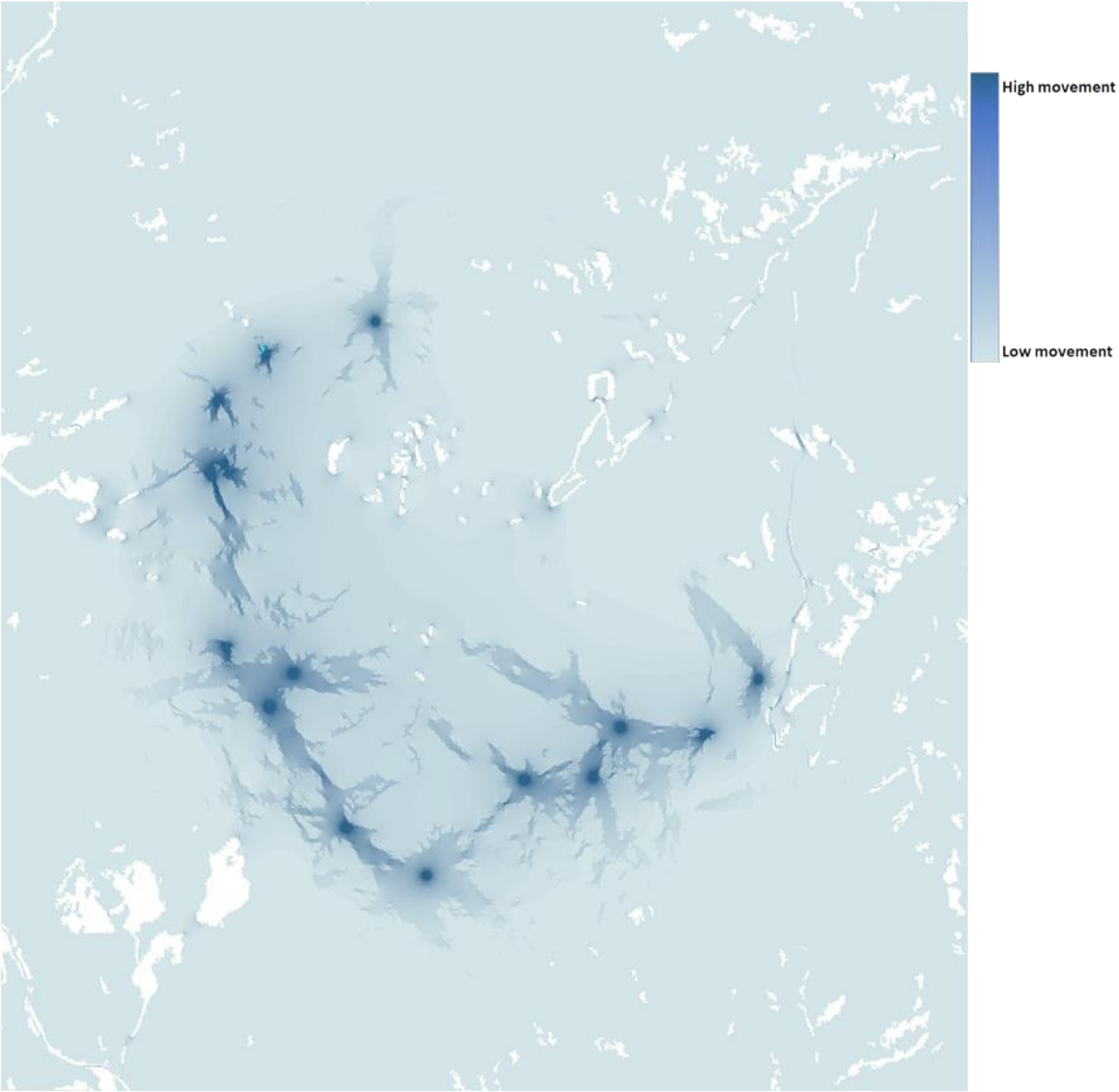

Figure 2.5 An example of a Circuitscape (McRae and Beier, 2007) resistance surface (created based on circuit theory) where forest land cover is twice as resistant as meadow and barren rock land cover. Darker areas represent corridors of high movement. Because forest is more resistant, the majority of movement occurs within meadow and barren rock land cover.

2.6 Least cost transect analysis

I also conducted least cost transect analysis (LCTA), a novel method of landscape genetic modelling (Van Strien et al. 2012). Instead of looking at the flow or resistance between a pair of sites across the entire resistance surface, LCTA first identifies and uses the least cost path between the two sites. The least cost path is the single pathway between two sites that has the least total resistance to movement; that is, it is the path with the lowest cumulative resistance of all the cells the pathway passes through. The least cost path is not necessarily the shortest path between two sites because in certain instances it would be “cheaper”, resistance-wise, to take a longer detour path if the straight-line path passes through many cells of high resistance. Depending on the assigned resistance values for the different land cover types, the least cost path can vary from a straight line if the resistance values of all land cover types are 1, to bigger and bigger detours that seek out low resistance cells as the resistance of immediate cells between the two sites increases. For the purpose of creating accurate least cost paths, I changed low elevation meadows from cells with ‘no data’ to having a resistance value the same as forest, so as not to bias artificially the route of the least cost path. However, as these low elevation meadows were mostly near the edge of the map and away from most least cost pathways, this change did not have any effect on the final analyses.

Once the least cost path is established, buffers of varying widths are added around the least cost path to create ‘least cost transects’ (Van Strien et al., 2012). I used buffer widths of 4m, 20m, 100m, 200m, 400m, 800m, and 1200m. The proportion of each land cover type (in this case, forest, meadow, barren rock) in this transect is then calculated by dividing the number of cells of that land cover type in the transect by the total number of cells in the transect. These proportions are estimated for all pairs of sites and are then used as predictors in statistical models where the response is either genetic distance (the usual landscape genetics approach) or movement between each pair of sites. Full models included proportions of all three land cover types as additive predictors, and all possible reduced models with one or two of the land cover types as predictors were also tested.

2.7 Statistical analysis

Because the pairwise distances among a set of objects (whether based on genetic, dispersal, or landscape data) are not independent of each other, traditional methods of regression and correlation analysis are not appropriate. Historically, landscape genetics studies have used the Mantel Test (Mantel, 1963) and the Partial Mantel Test (Smouse et al., 1986) to test the significance of correlations among pairwise measures of geographic distance, genetic distance, and some measure of the landscape resistance. These methods, particularly the Partial Mantel Test, have been criticized as they cannot accommodate the intrinsic autocorrelation in pairwise comparisons (Yang and Williams, 2004), and have both low power and high rates of type I error (Guillot and Rousset, 2013). Spatial autocorrelation is often implicit in landscape ecology and landscape genetic studies and so it has been suggested the Mantel Test and the Partial Mantel should be avoided for this purpose (Guillot and Rousset, 2013).

The maximum-likelihood population effects model (MLPE) has been proposed as an alternative method to deal with pairwise data. Since each observation is based on information from two sites, this is a mixed model that includes a random effect that accounts for the pairwise correlation structure of the data (Clarke et al., 2002). That is to say, in the context of my project and other landscape genetic studies, the random effect is the effect of having distance measures from pairs of sites. I used MLPE to select the best set of explanatory variables that explain genetic distances or movement, separately, among *P. smintheus* populations.

2.7.1 Circuitscape statistical analysis

I created my models in the “nlme” (Clarke et al., 2002) package in R (R Core Team, 2017). Using Generalized Least Square (GLS), I modeled genetic distance (F_{ST}) as a function of Circuitscape distance, with forest cell resistance set to values of 1, 2, 3, 4, 10, 50, and 99. I fit the models using maximum likelihood (ML). ML can be used to compare models with different fixed factors and the same random factor. I use the R package “corMLPE” (Pope, 2014) to allow me to account for the correlation structure of site pairs, as a random factor for the GLS model.

I created one statistical null model and two biological null models. The statistical null model is a GLS model with the genetic distance data as the response variable and with the random effect only. My first biological null model had straight-line Euclidean distances between site pairs as the fixed predictor. The predictor in my second biological null model was resistance distance derived from a surface in which the resistance of all land cover types is set to 1 (that is to say, all land cover types have the same effect on movement and gene flow). These latter two models do not take into account land cover resistance, but represent different ways of accounting for only geographic distance between sites, and so they are considered biological null models. Note that these two models should theoretically give the same result because straight-line distance should be identical to a resistance distance if all land cover types are set to a cell resistance of 1; however, due to the way Circuitscape analyzes paths through the square cells of the resistance surface, the resulting values are slightly different. Indeed, I found that straight-line geographic distance was strongly correlated with Circuitscape resistance derived from the uniform resistance surface ($r = 0.973$, $P < 0.001$)

I compared all nine models (two null models plus seven models each derived from a surface with a different resistance value for forest) using the corrected Akaike information criterion (AICc) to determine which model best explains the genetic distance between sites (Akaike, 1974). I then recalculated the best model using restricted maximum likelihood (REML) instead of ML to obtain unbiased estimates of coefficients. I repeated the whole analysis, from nlme model creation to model selection and coefficient estimation, with movement data between site pairs instead of genetic distance data as the response variable.

2.7.2 Least cost transect statistical analysis

The statistical analysis of LCT was also conducted using Generalized Least Square (GLS) models fit using (ML), and using corMLPE to account for the pairwise structure of the data via a random effect in the model. I conducted separate analyses with each of genetic distance and movement as the response variables.

I analyzed 49 groups of models using the LCT analysis, by having all possible combinations of the seven transect widths (4m, 20m, 100m, 200m, 400m, 800m, 1200m) as well as seven different forest cell resistance values (1, 2, 3, 5, 10, 50, 99) that were used to establish the initial least cost paths. In each of these 49 groups, there was a full model with the proportion of all three land cover types in the least cost transects as the additive predictors, as well as all possible reduced models (six total) with proportions of one or two of the landcover types as predictors. This resulted in seven models in each of the 49 groups. The predictor variables of proportion forest, meadow, and barren rock were centered, separately, for each model. I again used AICc to compare the resulting 343 (7 x 7 x 7) models.

Following Van Strien et al. (2012), I first determined which combination of transect width and resistance cell value for forest (used to determine the initial least cost paths) yielded the best model, based on comparison of AICc values of the full models. Among the seven models derived from that particular combination of transect width and forest resistance value, I then determined the set of predictors (i.e. full or reduced model) that best explained the response variable. I then re-fit this best model using REML to obtain unbiased estimates of regression coefficient.

3 Results

3.1 Resistance surface analysis with Circuitscape

For both the genetic distance and dispersal response variables, the statistical null model performed the worst by far, having the highest AICc value (and therefore the largest ΔAICc) (Table 3.1, Table 3.2). The model with forest resistance of 99 also performed poorly.

When considering genetic distance response variable (pairwise F_{ST}), the model with the lowest AICc was the one in which resistance distance was calculated with forest resistance set to 10 (i.e., in which forest is hypothesized to be ten times more resistant to gene flow than meadow and barren rock land cover; Table 3.1, Figure 3.1). However, the models with forest resistance of 1, 2, 3, 4, and 50, as well as the biological null model, were within 2 AIC of this model. Therefore, all of these models were equally well supported.

Table 3.1 Circuitscape analysis: comparison of maximum likelihood population effects (MLPE) models explaining genetic differentiation (F_{ST}) between populations of the butterfly, *Parnassius smintheus*. All models included a random effect that accounts for the pairwise correlation structure of the data. The statistical null model included only the random effect. The predictor variable in the biological null was straight-line geographic distance. The predictor in all other models was resistance distance calculated from a resistance surface using Circuitscape. The resistance value assigned to forest land cover varied among the resistance surfaces, while meadow and barren rock land cover were consistently assigned a resistance value of 1. The corrected Akaike information criterion (AICc) and the difference in AICc from the top model (ΔAICc) are presented for each model. Models are listed in rank order of AICc.

Model	ΔAICc	AICc
Forest cell resistance 10	0	-780.8097
Forest cell resistance 4	0.2798	-780.5299
Forest cell resistance 3	0.4146	-780.3951
Forest cell resistance 2	0.6543	-780.1554
Forest cell resistance 50	0.8842	-779.9255
Forest cell resistance 1	1.4221	-779.3876
Biological Null	1.432	-779.3777
Forest cell resistance 99	2.0531	-778.7566
Statistical Null	20.6904	-760.1193

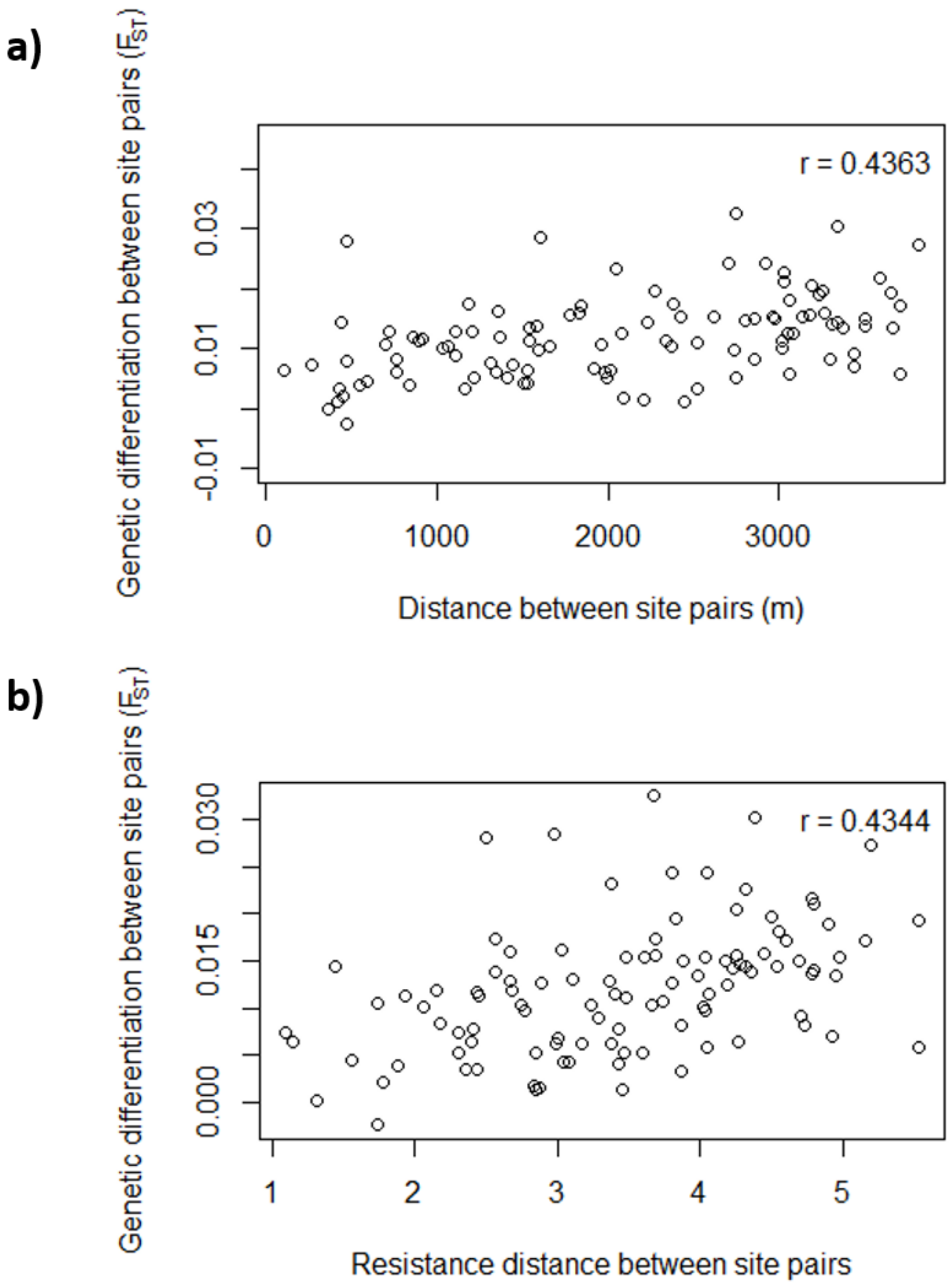

Figure 3.1 Effects of geographic distance and landscape resistance on genetic differentiation in the butterfly, *Parnassius smintheus*. a) Correlation between straight-line distance between sites and genetic differentiation between sites. b) Correlation between the resistance distance between sites, derived from Circuitscape analysis of the resistance surface of the model with lowest AICc (in which forest has a resistance of 10), and genetic differentiation between sites.

When considering the dispersal response variable, the model with lowest AICc was the one in which resistance distance was calculated with forest resistance set to four (i.e., in which forest is hypothesized to be four times more resistant to gene flow than meadow and barren rock land cover; Table 3.2, Figure 3.2). The second, third, and fourth best models involved forest resistance set to 10, 3, and 2, respectively. However, these latter models all had ΔAICc values less than 2, and were therefore not appreciably better than the top model. All other models had ΔAICc greater than two. It is interesting to note that the biological null model of isolation-by-distance and the model involving a forest resistance of one differed substantially in their explanatory power (AICc values were more than 2 apart) (Table 3.2).

Table 3.2 Circuitscape analysis: comparison of maximum likelihood population effects (MLPE) models explaining estimated dispersal between populations of the butterfly, *Parnassius smintheus*. All models included a random effect that accounts for the pairwise correlation structure of the data. The statistical null model included only the random effect. The predictor variable in the biological null model was straight-line geographic distance. The predictor in all other models was resistance distance calculated from a resistance surface using Circuitscape. The resistance value assigned to forest land cover varied among the resistance surfaces, while meadow and barren rock land cover were consistently assigned a resistance value of 1. The corrected Akaike information criterion (AICc) and the difference in AICc from the top model (ΔAICc) are presented for each model. Models are listed in rank order of AICc.

Model	Δ AICc	AICc
Forest cell resistance 4	0	-143.5361
Forest cell resistance 10	0.2975	-143.2386
Forest cell resistance 3	0.3463	-143.1898
Forest cell resistance 2	1.0193	-142.5168
Forest cell resistance 1	2.6232	-140.9129
Biological Null	6.2972	-137.2389
Forest cell resistance 50	9.2656	-134.2705
Forest cell resistance 99	15.7186	-127.8175
Statistical Null	76.31821	-67.21789

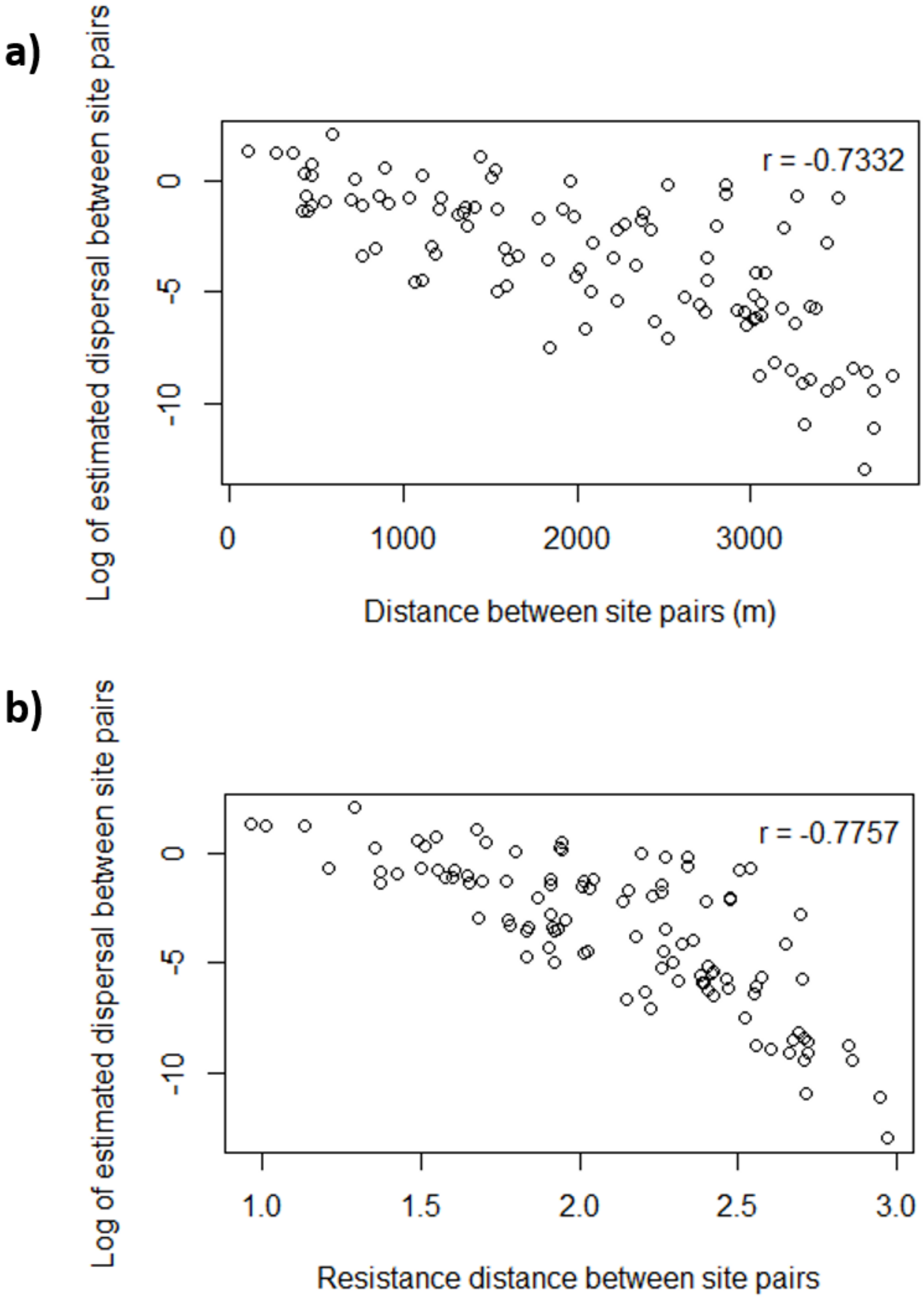

Figure 3.2 Effects of geographic distance and landscape resistance on estimated dispersal in the butterfly, *Parnassius smintheus*. a) Correlation between straight-line distance between sites and estimated dispersal between sites. b) Correlation between the resistance distance between sites, derived from Circuitscape analysis of the resistance surface of the model with lowest AICc (in which forest has a resistance of 4), and estimated dispersal between sites.

Using restricted maximum likelihood (REML) to evaluate the model with lowest AICc for the genetic distance response variable (the model with resistance distances calculated from a resistance surface in which forest resistance is set to 10), the estimate of the coefficient for the effect of resistance distance was 0.0029 (SE = 0.0006). Using REML to evaluate the model with lowest AICc for the dispersal response variable (the model with resistance distance calculated from a resistance surface in which forest resistance is set to 4) the estimate of the coefficient for the effect of resistance distance was -5.3642 (SE = 0.34).

3.2 Resistance surface analysis with least cost transect

In the least cost transect analysis I also tested seven different resistance values for forest (1, 2, 3, 5, 10, 50, 99), as well as seven different buffer widths. For each of these 49 combinations, there were seven models that included different land cover combinations as predictor variables, for a total of 343 models. These models were compared separately to the genetic distance and dispersal response variables (Table 3.3, Table 3.4).

For the genetic distance response variable, the model with lowest AICc included only the proportion of barren rock in the least-cost transect as the predictor. The least-cost transects were obtained by using a resistance surface with forest resistance of 99 and a buffer width of 400m (AICc = -773.753; Table 3.3, Figure 3.3, Figure 3.4, Figure 3.5). Using REML to evaluate this model, the estimate of the coefficient for the effect of barren rock was 0.0082 (SE = 0.02). Additional models that were equally well supported (ΔAICc<2) also all had barren rock as the predictor and were obtained using a resistance surface with forest resistance of 99, and had buffer widths ranging from 4-200m.

Table 3.3 Least cost transect analysis: comparison of maximum likelihood population effects (MLPE) models explaining genetic differentiation (F_{ST}) between populations of the butterfly, *Parnassius smintheus*. The models are sorted by ΔAICc values. 'Resistance of forest' is the resistance value assigned to forest in the surface that was used to generate the least cost paths. 'Transect width' is the buffer applied to the least cost path to create the least cost transect. 'Model predictors' refer to the combination of predictors (proportion of meadow, barren rock, or forest land cover in the least cost transect) that were included in the model; 'Full' means that all three predictors were included. The corrected Akaike information criterion (AICc) and the difference in AICc from the top model (ΔAICc) are presented for each model.

Transect width (m)	Resistance of forest	Model predictors	AICc	ΔAICc
400	99	Barren rock	-773.7531704	0.00
4	99	Barren rock	-773.3124666	0.44
20	99	Barren rock	-773.2502485	0.50
200	99	Barren rock	-772.0599796	1.69
100	99	Barren rock	-771.9382379	1.81
4	99	Full	-771.2008074	2.55
800	99	Barren rock	-770.6326855	3.12
400	99	Full	-770.6009269	3.15
20	99	Full	-769.8550316	3.90
1200	1	Full	-769.6837173	4.07
100	99	Full	-769.0200657	4.73
200	99	Full	-768.9171334	4.84
800	1	Full	-768.796252	4.96
1200	99	Barren rock	-764.8249207	8.93

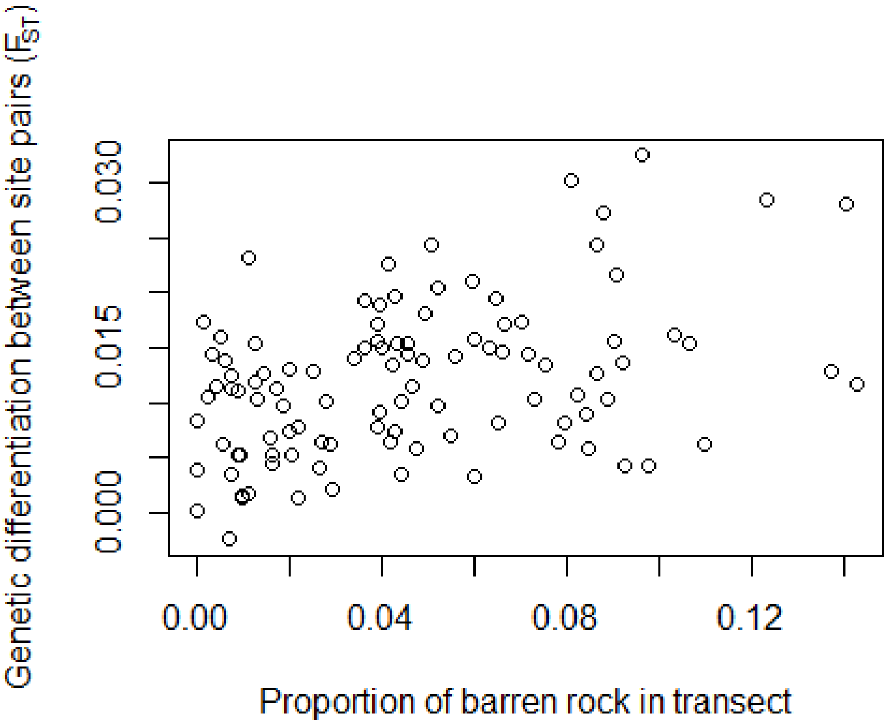

Figure 3.3 Relationship between the proportion of barren rock (best predictor) in a 400m wide least-cost transect and the genetic differentiation between sites. The least cost transects followed least cost paths derived from a resistance surface in which meadow and barren rock land covers had a resistance value of 1, and forest land cover had a resistance value of 99. This was the model with the lowest AICc in the genetic differentiation least cost transect analysis.

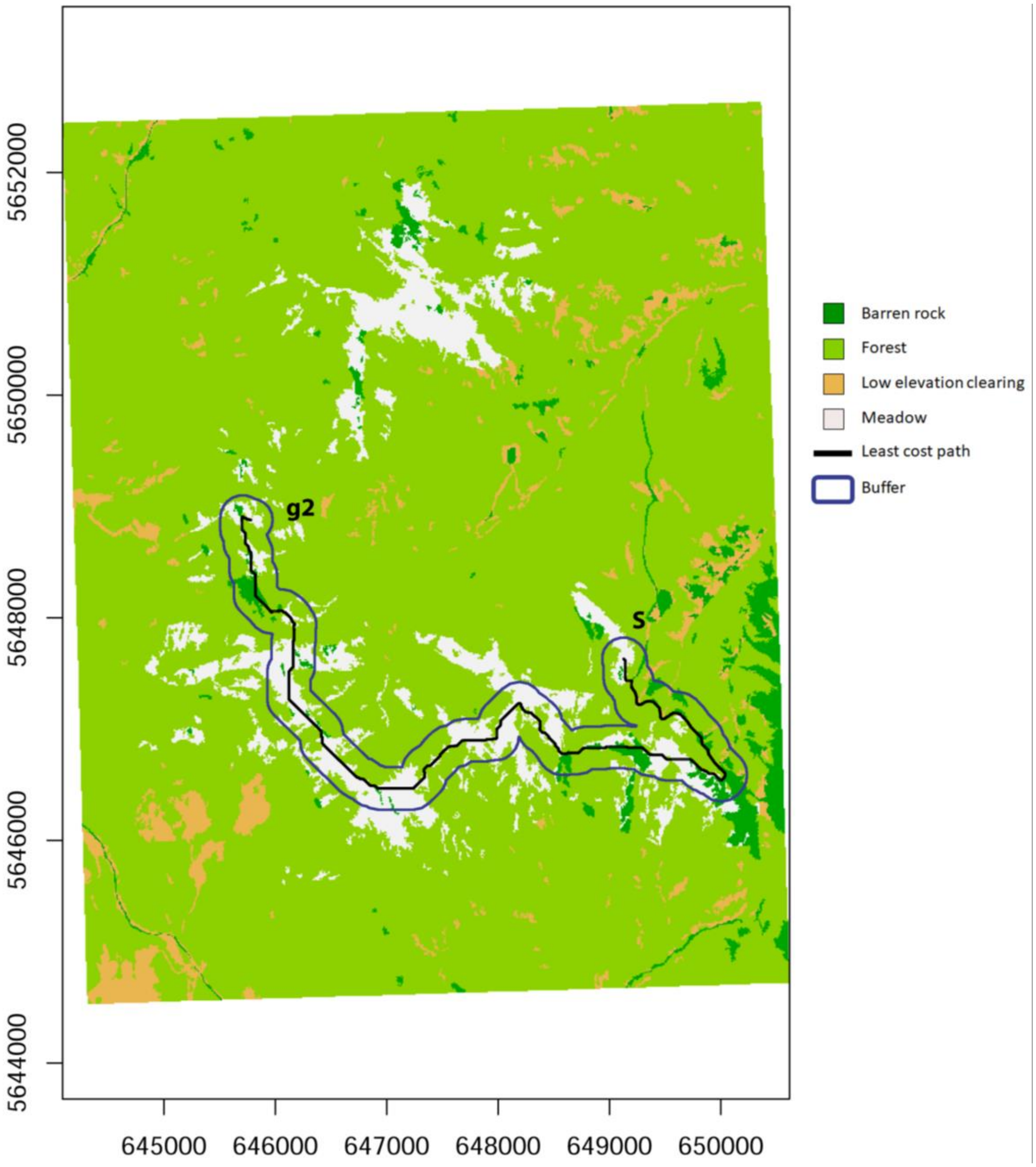

Figure 3.4 Example least cost transect from a resistance surface with forest resistance of 99 and a transect buffer of 400m. The transect between sites g2 and S is shown. The proportion of barren rock in such transects was the best predictor of genetic differentiation (F_{ST}) between sites. The least cost path (black line) is the path of least total resistance that *Parnassius smintheus* would travel if forest land cover is 99 times more resistant than meadow and barren rock. Blue lines encompass the transect around this least cost path. This map displays UTM coordinates (eastings and northings) and is in UTM zone 11.

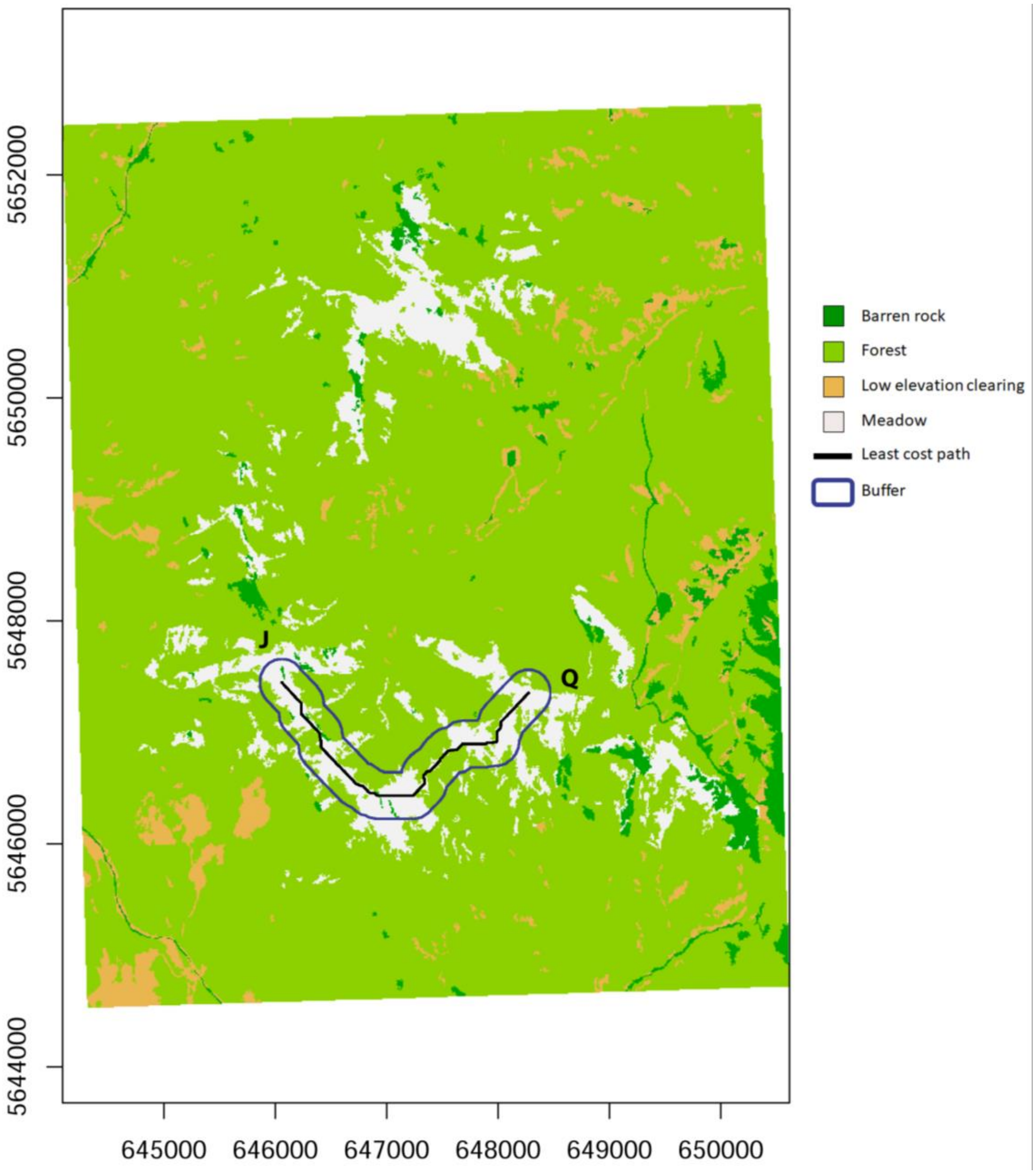

Figure 3.5 Example least cost transect from a resistance surface with forest resistance of 99 and a transect buffer of 400m. The transect between sites J and Q is shown. The proportion of barren rock in such transects was the best predictor of genetic differentiation (F_{ST}) between sites. The least cost path (black line) is the path of least total resistance that *Parnassius smintheus* would travel if forest land cover is 99 times more resistant than meadow and barren rock. Blue lines encompass the transect around this least cost path. This map displays UTM coordinates (eastings and northings) and is in UTM zone 11.

For the estimated dispersal response variable, the model with lowest AIC was the full model (including the proportion of meadow, barren rock, and forest in the least-cost transect as predictors), and was based on transects obtained by using a resistance surface with forest resistance set to 1 and a buffer width of 100m (AICc = 435.616; Table 3.4, Figure 3.6, Figure 3.7). Using REML to evaluate this model, the estimated coefficients for the effects of the predictors were: 106.98 for forest (SE = 27.54), 114.04 for barren rock (SE = 27.56), and 118.24 for meadow (SE = 26.99). Additionally, there was a second model that was equally well supported (ΔAICc<2) and also had the full model as the predictor and was also obtained using a resistance surface with forest resistance of 1, and had a buffer width of 200m.

Table 3.4 Least cost transect analysis: comparison of maximum likelihood population effects (MLPE) models explaining estimated dispersal between populations of the butterfly, *Parnassius smintheus*. The best full model and the best reduced model for each transect width are shown. 'Resistance of forest' is the resistance value assigned to forest in the surface that was used to generate the least cost paths. 'Transect width' is the buffer applied to the least cost path to create the least cost transect. 'Model predictors' refer to the combination of predictors (proportion of meadow, barren rock, or forest land cover in the least cost transect) that were included in the model; 'Full' means that all three predictors were included. The corrected Akaike information criterion (AICc) and the difference in AICc from the top model (ΔAICc) are presented for each model.

Transect width (m)	Resistance of forest	Model predictors	AICc	Δ AICc
200	1	Full	435.610122	0.00
400	1	Full	436.3274735	0.72
100	1	Full	440.7016591	5.09
1200	1	Meadow & Barren Rock	440.9148251	5.30
1200	1	Full	441.6391906	6.03
400	1	Meadow & Barren Rock	443.6985641	8.09
800	1	Meadow & Barren Rock	447.2351396	11.63
200	1	Meadow & Barren Rock	447.7571256	12.15
800	1	Full	449.286683	13.68
20	1	Full	450.0983835	14.49
100	1	Meadow	452.6934986	17.08
4	1	Full	453.7383441	18.13
20	1	Meadow & Barren Rock	461.0940229	25.48
4	1	Meadow & Forest	462.6728983	27.06

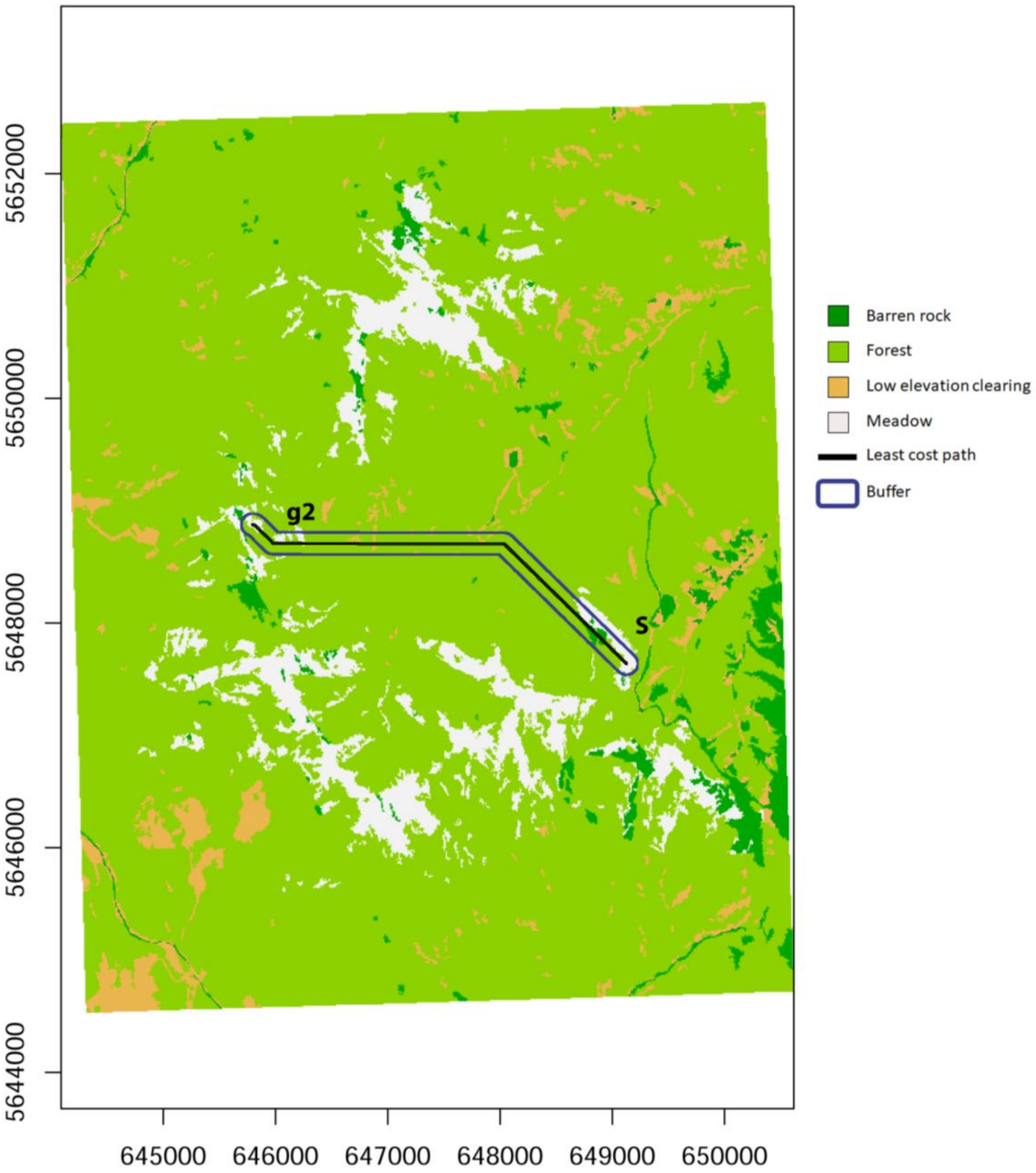

Figure 3.6 Example least cost transect from a resistance surface with forest resistance of 1 and a transect buffer width of 200m. the transect between sites g2 and S is shown. The proportion of forest, meadow, and barren rock in such transects were the best predictors of estimated dispersal between sites. The least cost path (black line) is the path of least total resistance that *Parnassius smintheus* would travel if forest land cover has the same resistance as meadow and barren rock. Blue lines encompass the transect around this least cost path. This map displays UTM coordinates (eastings and northings) and is in UTM zone 11.

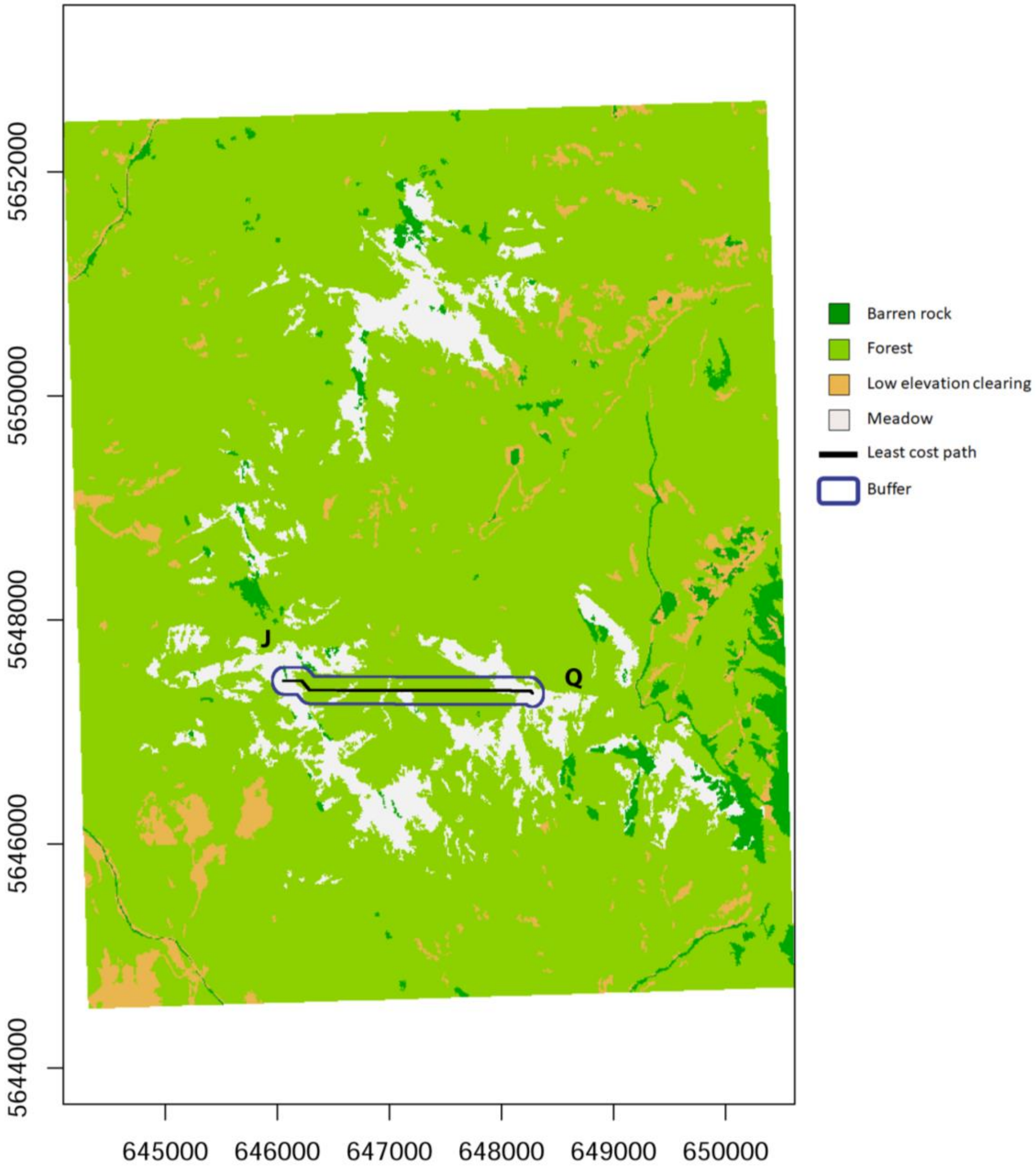

Figure 3.7 Example least cost transect from a resistance surface with forest resistance of 1 and a transect buffer width of 200m. The transect between sites J and Q is shown. The proportion of forest, meadow, and barren rock in such transects were the best predictor of estimated dispersal between sites. The least cost path (black line) is the path of least total resistance that *Parnassius smintheus* would travel if forest land cover has the same resistance as meadow and barren rock. Blue lines encompass the transect around this least cost path. This map displays UTM coordinates (eastings and northings) and is in UTM zone 11.

4 Discussion

4.1 Project objective

This project is unique in using different landscape genetics approaches to examine, in parallel, the effects of landscape structure on both dispersal and genetic data. With this, I test the hypothesis that landscape effects on genetic differentiation between populations, and landscape effects on dispersal, are comparable and parallel. Furthermore, I hope to provide more empirical insight into how two commonly used methods of characterizing landscape effects on gene flow (Circuitscape and least cost transect) differ from each other. The results of my study are relevant to understanding both the ecology and evolution of *Parnassius smintheus*, and more broadly to the field of landscape genetics studies in general.

4.2 Landscape effects on genetic differentiation versus dispersal

A key finding of my study was that genetic differentiation and dispersal were best explained by different sets of models (in this case the best model set refers to all models within two AICc of the model with the lowest AICc), regardless of the method of analysis used. Using Circuitscape, when genetic differentiation was the response variable, the model with lowest AIC had resistance distance derived from a surface with forest resistance set to ten times that of meadow and barren rock. Also, most other forest resistance values, with the exception of 99, explained genetic differentiation equally well. When dispersal was the response variable, the model with lowest AIC had resistance distance derived from a surface with forest resistance set to four times that of meadow and barren rock. Also, only three other models, with forests resistance ranging from two to ten, explained the dispersal data equally well. Therefore, using Circuitscape, genetic and dispersal data were explained by different, albeit overlapping, model sets. In least cost transect analysis (LCTA), genetic differentiation and dispersal were also best explained by different, non-overlapping, sets of models with different predictors. A difference in the results for genetic versus dispersal data is not entirely surprising. While dispersal is an important determinant of gene flow, and hence genetic differentiation, the two are not identical (Eriksen et al., 2014). Genetic differentiation and dispersal data may differ for several reasons. First, not all dispersal necessarily results in gene flow and an effect on genetic differentiation. Only dispersal that is accompanied by reproduction will affect genetic differentiation (Levin et al., 2009). Second, our dispersal data were based on movement of the butterflies tracked through mark-recapture for the specific years that the fieldwork was conducted. In contrast, the genetic differentiation data are the result of dispersal and reproduction over many past generations of *P. smintheus* that have led up to observed patterns at the time of sample collection (Keyghobadi et al., 2005b). Because of this, genetic differentiation paints more of a picture across time and cumulative generations, whereas dispersal data are more akin to a snapshot at the time of collection (Orsini et al., 2013). Finally, unlike dispersal data, genetic differentiation data will also be influenced by genetic drift, which is in turn a function of effective population size (Hoeck et al., 2010).

Overall, my results did not strongly support the hypothesis that inferred effects of the landscape on genetic differentiation and on dispersal would be highly similar. This occurred despite the fact that genetic differentiation between sites and estimated dispersal between sites are positively correlated in this system (Caplins et al., 2014; Keyghobadi et al., 1999). However, my ability to select among competing models was limited, particularly for the genetic response variable. Therefore, I cannot confidently identify a single best model for each of the response variables.

An important result was that genetic data seemed to provide less precise estimates of landscape effects than did the dispersal data. While Circuitscape analyses indicated that genetic differentiation was best explained by a resistance surface in which forest resistance was ten-fold greater than meadow resistance, it is important to note that all models, other than the statistical null model and the model with forest resistance of 99, were within 2 AICc of the best model. That is to say, all of these models were comparable in their ability to explain genetic distances. This suggests that while the Circuitscape method appears to detect an effect of intervening forest cover on genetic distance, it is unable in this case to differentiate among a broad range of forest resistance values. In this case, one can only conclude that the genetic differentiation data suggest a forest resistance between 1 and 50 times that of meadow and barren rock. This is a large range and would not be useful in most situations for making any kinds of management or conservation decisions. In contrast, the best model explaining dispersal was derived from a surface in which forest has a resistance of 4 although surfaces with forest resistance of 2, 3, and 10 were within 2 AIC of this best model. This suggests that forest is approximately 2 to 10 times more resistant to butterfly dispersal than meadow and barren rock, providing a much narrower range of likely forest resistance values than for the genetic data. The inability to discriminate among a broad range of forest resistance values when inferring effects of forest on genetic differentiation might be due to the fact that the relationship between genetic data and the landscape is inherently highly variable. The genetic data are based on a relatively small number of genetic markers, represent effects of movement and reproduction over many generations, and are also affected by genetic drift and effective population size. My study suggests that extrapolation of results based on genetic data to inferences about dispersal and movement should be made with caution, and that ideally both types of data should be used in conjunction.

4.3 Comparison to previous work on *Parnassius smintheus*

Previous studies on *P. smintheus* have consistently identified forested land cover as a barrier to both dispersal and gene flow (Keyghobadi, 2007). Compared to meadow land cover, past studies based on analysis of mark-recapture data have estimated that forest is from two (Roland et al., 2000) to 2.9 times (Matter et al., 2004) as resistant to movement compared to open meadow. I also found evidence in my analyses that forested land cover was more resistant to both dispersal and gene flow than open meadow. The estimated effects of forest on genetic differentiation and dispersal that I detected using Circuitscape were more consistent with previous studies than the effects inferred using LCTA (Matter et al., 2004; Roland et al., 2000).

Circuitscape analyses indicated that genetic differentiation was best explained by a resistance surface in which forest resistance ranged from 1 to 50. Using restricted maximum likelihood (REML) to evaluate the model with the lowest AICc (in which forest resistance was ten-fold greater than meadow resistance), the estimate of the coefficient for the effect of resistance distance was 0.0029 (SE = 0.0006), which suggests that as the amount of forest land cover increases, the genetic differentiation between sites also increases, consistent with previous genetic studies in this system (Caplins et al., 2014; Keyghobadi et al., 1999).

When dispersal was the response variable, Circuitscape analysis identified both a narrower range of optimal forest resistance values (2 to 10 times more resistant than meadow and barren rock), and an effect of forest that was quite concordant with previous estimates in this system (previous estimates suggest forest is 2 to 2.9 times more resistant to dispersal than meadow; Matter et al., 2004; Roland et al., 2000). In this case, the model with lowest AICc explaining dispersal was derived from a surface in which forest has a resistance of 4. Using REML to evaluate this model, the estimate of the coefficient for the effect of resistance distance was -5.3642 (SE = 0.34). As this coefficient is negative, it suggests that as forest land cover increases, the amount of dispersal decreases, which is consistent with forest hindering butterfly movement.

Using LCTA however, barren rock was inferred to be the best predictor of genetic differentiation. This result is surprising because there is comparatively little barren rock land cover on this landscape. However, patches of barren rock are located in corridors between sites that are expected to have high amounts of butterfly movement. For example, the large patch of barren rock between sites G1 and K would need to be traversed by any butterflies travelling from the northern patches of Z, Y, F, g2, and G1 to the southern patches of I, K, J, L, and M. This may explain why barren rock, the land cover type with the least coverage, is identified by LCTA as an important land cover in this system. Using REML to evaluate this best model, the estimate of the coefficient for the effect of barren rock was 0.0082 (SE = 0.02). There is a positive relationship between the amount of barren rock between sites and genetic differentiation, giving support to this possible explanation.

LCTA identified the full model, which includes all the predictor variables of meadow, forest, and barren rock, as the best model explaining the dispersal data. Using REML to evaluate this best model, the estimated coefficients for the effects of the predictors were: 106.98 for forest (SE = 27.54), 114.04 for barren rock (SE = 27.56), and 118.24 for meadow (SE = 26.99). These positive coefficients suggest that dispersal increases as the proportion of each of the three land covers increases. It makes sense for the full model to be the best predictor, as it takes into account all land cover types, which will all exert effects on the dispersal of *P. smintheus* as either conductors (meadow and barren rock) or resistors (forest). It also makes sense for a higher proportion of meadow, and perhaps the barren rock land covers, to promote dispersal between sites. However, a positive effect of forest on dispersal is not concordant with previous work in this system.

In the LCTA, I also found that the inferred optimal buffer widths ranged from 4m - 400m and 200m - 400m for genetic differentiation and dispersal, respectively. A broader range of inferred optimal buffer widths for genetic data reflects the greater imprecision of inferences derived from genetic data versus dispersal data, as discussed in section 4.2. Interestingly, the optimal buffer width for the dispersal data in the LCTA is very close to the mean distances individuals moved, as reported by Roland et al. (2000) for the same dataset (131.9 m and 131.6 m for males and females respectively in 1995; 162.4 m and 118.0 m for males and females respectively in 1996). These LCTA results support prior information about the spatial scale at which the butterfly is responding to the landscape using a new approach. Knowing what spatial scale is relevant to the species one is interested in provides information about what spatial lens should be used to study the species, and, particularly in terms of conservation, gives a clearer idea of the size of corridors that would be needed to support movement through the landscape.

4.4 Circuitscape versus least cost transect analyses

Various methods have been used in landscape genetics to estimate the effects of intervening landscape on genetic differentiation between sites. A common method is the linear transect (sometimes referred to as a vector) approach, where metrics such as land cover type, precipitation, or elevation are evaluated along linear transects connecting pairs of sites, and the effects of these pairwise environmental and landscape variables on genetic distances are evaluated (Murphy et al., 2010). Another method that was very commonly used in the earlier days of the field, but has now largely been replaced by the Circuitscape approach, is the least cost path approach. Using a resistance surface of the landscape that represents hypotheses about the conductance (permeability) of different landscape elements, the least cost paths between pairs of sites were determined. Either the total length of these paths (Spear et al., 2005), or the total accumulated cost of moving along the paths (Cushman et al., 2006) were then used as predictors of genetic distance.

The least cost transect method (Van Strien et al., 2012) builds upon the two more conventional methods of linear transects and least cost path. Instead of imposing a straight corridor between pairs of sites, LCTA allows the corridor to further take into account hypothesized effects of the landscape by following the least cost path. This additional complexity is meant to allow LCTA to improve upon these more conventional methods of studying land cover, movement, and gene flow across the landscape, and in certain situations has indeed been shown to perform better (Van Strien et al., 2012). A further advantage of LCTA, which is shared with linear transect analysis, is that by testing different buffer widths, it is able to suggest the spatial scale at which the landscape affects the species of interest (Murphy et al., 2010).

In my study however, I found that the LCTA method performed quite poorly at explaining both genetic differentiation and estimated dispersal. My LCTA results were not concordant with the Circuitscape results or with previous estimates of landscape effects derived using other approaches in this system (Matter et al., 2004; Roland et al., 2000). For example, LCTA suggested that barren rock and not forest was the best predictor of genetic differentiation. Also, LCTA suggested that forest had a positive effect on dispersal, which is contradictory to all previous work in this system, as well as to my Circuitscape results. LCTA also initially identified a resistance surface in which forest had a resistance of one (i.e., not different from meadow and barren rock) as the optimal resistance surface to describe estimated dispersal between sites. This resulted in least cost transects (Figure 3.6, Figure 3.7) that did not follow the ridge tops along which *P. smintheus* is expected to move (Roland et al., 2000). Overall, many of my LCTA results were unexpected and did not make sense.

There are a number of reasons why LCTA may have performed poorly in my study. First, the landscape in which I was working was relatively simple in containing very few different land cover types. I distinguished only three land cover types (forest, meadow, barren rock) at the elevations that would be relevant to *P. smintheus.* Furthermore, one of the land covers, barren rock, was not very abundant. As a result, the proportion the different land covers in the least cost transects were highly correlated (r > 0.9). In particular, the proportions of meadow and forest were highly negatively correlated, such that as one increased the other decreased. This collinearity among these proportion variables is problematic, and would make it difficult to tease apart the effects of the land covers using a transect-based approach. Indeed, LCTA has been used successfully in more complex landscapes, where there are more distinct land cover types whose proportions within the least cost transects are not correlated (Van Strien et al., 2012). Second, my study sites were arranged essentially along a single dimension by virtue of being located on the spine of a ridge (i.e., sites are in a line following the ridge-top). This means that the least cost transects for many site pairs were similar and partially overlap. The unidimensional distribution of the sites could also affect my ability to tease apart the effects of the different land cover types as predictor variables. Overall, LCTA may have been more sensitive to idiosyncrasies in the configuration of the landscape specific to my study. It is likely that LCTA will perform better in studies that have a larger number of distinct and uncorrelated land cover variables, as well as a more even arrangement of sites across two dimensions on the landscape.

Circuitscape, based on circuit theory, represents an extension of least cost path analysis and was introduced as a way of quantifying landscape effects, particularly for conservation purposes, in a way that can account for “multiple pathways linking populations,” which previous methods were unable to do (McRae and Beier, 2007). Circuit theory, as a tool for modelling connectivity, has been shown to predict gene flow patterns accurately across different scales and distances (McRae et al., 2008). Uniquely, Circuitscape is able to take into account the whole landscape as all cells of the resistance surface have an impact on the final resistance distance calculations. If the landscape outside the range of the least cost transect also contributes to gene flow or dispersal between sites, then Circuitscape analysis may provide more meaningful results than LCTA. In particular, LCTA is based on initially determining a single least cost path between each pair of sites. Therefore, LCTA may still suffer to some degree from ignoring much of the landscape.

Furthermore, Circuitscape does not suffer from a certain degree of circularity that is present in LCTA. In LCTA, the least cost path is created given a particular resistance surface, which already represents a specific hypothesis about the effects of each land cover type. The predictor variables in the analysis are then the proportion of different land cover types in transects following these paths. By assigning different resistance values to the land covers initially, and then using the proportions of those same land covers in the resulting transect as predictors, circularity is arguably introduced into the LCTA. In my analyses for example, in most resistance surfaces that I tested forest was assigned a higher resistance value compared to meadow and barren rock. The resulting least cost paths in these surfaces would avoid forest cells and travel through meadow and barren rock cells when possible. All least cost transects would then be likely to contain a maximal amount of meadow and barren rock. This may have contributed to barren rock being the best predictor of genetic differentiation in my LCTA. These potential limitations of LCTA support the conclusion that, for my particular study at least, Circuitscape is a more appropriate approach for estimating landscape effects on both genetic differentiation and dispersal.

4.5 Implications for conservation

In alpine habitats, the effects of climate change, as well as fire suppression practices, tend to cause tree lines to move to increasingly higher elevations (Ee et al., 2015). This will potentially cause forests to encroach into alpine meadows, causing meadow size reduction as well as isolation (Roland et al., 2000). Such an encroachment will further reduce the area of habitat available to *P. smintheus* through the edge avoidance effect (Ross et al., 2005). In broad terms, the edge effect occurs when species prefer one habitat type over another, and so will avoid the boundaries where the preferred habitat transitions to another land cover type (Ries and Debinski, 2001). In *P. smintheus,* this behaviour causes the butterflies to fly away from forest edges and towards the center of meadows. Small meadows (less than 50 m wide) are prone to this effect throughout and may become unsuitable butterfly habitat as a result of further shrinkage (Ross et al., 2005).

These changes will invariably result in a decrease of resources and habitat available for *P. smintheus*, and so will have the potential to reduce their population sizes and habitat patch connectivity in the future. My results, particularly from Circuitscape analysis, support previous studies that show that intervening forests can reduce dispersal and increase genetic differentiation among populations of *P. smintheus*. Therefore, my results reinforce the predicted isolating effects of encroaching tree lines on *P. smintheus* (Keyghobadi et al., 1999; Roland et al., 2000; Ross et al., 2005). Forest encroachment not only affects *P. smintheus*, but also other species that depend on alpine meadows (Illerbrun and Roland, 2011).

With global trends of climate change and habitat fragmentation and loss resulting in loss of biodiversity, landscape genetics is an essential tool that can provide in-depth analysis at relevant scales and actionable results for the purposes of conservation (Epps and Keyghobadi, 2015). Understanding the effects of individual land cover types on dispersal and gene flow is paramount in understanding and quantifying species responses to a changing landscape (Manel and Holderegger, 2013). With limited resources and time, knowing which habitat patch, landscape corridor, or population should be prioritized to maximize conservation efforts will make a big difference (Dennis et al., 2013). I have shown through my project that, in order to get an accurate and precise representation of how the dispersal patterns of the species of interest are affected by the landscape, genetic differentiation between population sites should ideally not be the only source of species ‘movement data’. I suggest that by combining genetic data with direct dispersal data, researchers can create improved models of landscape connectivity and dispersal of the species of interest.

4.6 Next steps

The LCTA may not have performed well in my study system because of the occurrence of few land cover types and the arrangement of sites along a single ridge top. Future studies wishing to continue this line of inquiry about the landscape effects on genetic differentiation as well as dispersal could address similar questions at locations with different landscape composition and a more even, two-dimensional configuration of sites. Indeed, replicating my study across several other landscapes would be a powerful way of assessing the performance of the different approaches I used. It would also be interesting to extend my analyses to other species, as the approaches I used should be applicable to any species with distinct populations occurring in landscapes with discrete land cover types. It would also be valuable to conduct similar analyses in species with smaller population sizes and with less interconnected populations, as the results would then be more applicable to endangered species in need of conservation.

4.7 Conclusions

It is common for landscape genetics studies to use genetic differentiation, either implicitly or explicitly, as a proxy for dispersal. But landscape variables that best explain genetic differentiation may not also best explain dispersal patterns. I tested the effects of the landscape on genetic differentiation and dispersal in parallel, to explore the extent to which using genetic differentiation can provide insights into landscape effects on dispersal, using both circuit theory (Circuitscape) and LCTA. The results of my findings did not support my hypothesis that landscape effects on genetic differentiation between populations and landscape effects on dispersal would be highly similar. In particular, the precision of estimates derived using genetic data may be much lower than that of estimates derived using dispersal data directly. My study suggests that in studies that aim to determine landscape effects on animal movement and dispersal, genetic differentiation should ideally not be used as a simple proxy for dispersal, but when possible should be used in conjunction with other, more direct measures of dispersal.

References

Adriaensen, F., Chardon, J.P., De Blust, G., Swinnen, E., Villalba, S., Gulinck, H., and Matthysen, E. (2003). The application of “least-cost” modelling as a functional landscape model. Landsc. Urban Plan. *64*, 233–247.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Trans. Autom. Control *19*, 716–723.

Caplins, S.A., Gilbert, K.J., Ciotir, C., Roland, J., Matter, S.F., and Keyghobadi, N. (2014). Landscape structure and the genetic effects of a population collapse. Proc. Biol. Sci. *281*, 20141798.

Carlsson, J. (2008). Effects of microsatellite null alleles on assignment testing. J. Hered. *99*, 616–623.

Chan, S.C.Y., and Karczmarski, L. (2017). Indo-Pacific humpback dolphins (*Sousa chinensis*) in Hong Kong: Modelling demographic parameters with mark-recapture techniques. PLOS ONE *12*, e0174029.

Chapuis, M.-P. (2007). Microsatellite null alleles and estimation of population differentiation. Mol. Biol. Evol. *24*, 621–631.

Clarke, R.T., Rothery, P., and Raybould, A.F. (2002). Confidence limits for regression relationships between distance matrices: estimating gene flow with distance. J. Agric. Biol. Environ. Stat. *7*, 361–372.

Cushman, S.A., McKelvey, K.S., Hayden, J., Schwartz, M.K., Helberg, A.E.M.E., and Losos, E.J.B. (2006). Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. Am. Nat. *168*, 486–499.

DeChaine, E.G., Martin, A.P., and Porter, A. (2004). Historic cycles of fragmentation and expansion in *parnassius smintheus* (papilionidae) inferred using mitochondrial dna. Evolution *58*, 113–127.

Dennis, R.L.H., Dapporto, L., Dover, J.W., and Shreeve, T.G. (2013). Corridors and barriers in biodiversity conservation: a novel resource-based habitat perspective for butterflies. Biodivers. Conserv. *22*, 2709–2734.

DMTI Spatial Inc. (2003). DMTI Digital Elevation Model (DEM). [d30m082j15] (Markham, Ontario).

Ee, J. van, Westbrook, M., Yadav, S., Roland, J., and Matter, S.F. (2015). Predicting the effects of forest encroachment for *Sedum lanceolatum* and *Rhodiola integrifolia* at a subalpine ecotone. Alp. Bot. *125*, 125–135.

Emaresi, G., Pellet, J., Dubey, S., Hirzel, A.H., and Fumagalli, L. (2011). Landscape genetics of the Alpine newt (*Mesotriton alpestris*) inferred from a strip-based approach. Conserv. Genet. *12*, 41–50.

Epps, C.W., and Keyghobadi, N. (2015). Landscape genetics in a changing world: disentangling historical and contemporary influences and inferring change. Mol. Ecol. *24*, 6021–6040.

Eriksen, R.L., Hierro, J.L., Eren, Ö., Andonian, K., Török, K., Becerra, P.I., Montesinos, D., Khetsuriani, L., Diaconu, A., and Kesseli, R. (2014). Dispersal pathways and genetic differentiation among worldwide populations of the invasive weed *Centaurea solstitialis* L. (Asteraceae). PLoS ONE *9*, e114786.

Fahrig, L. (2003). Effects of habitat fragmentation on biodiversity. Annu. Rev. Ecol. Evol. Syst. *34*, 487–515.

Fownes, S., and Roland, J. (2002). Effects of meadow suitability on female behaviour in the alpine butterfly *Parnassius smintheus*. Ecol. Entomol. *27*, 457–466.

Guillot, G., and Rousset, F. (2013). Dismantling the Mantel tests. Methods Ecol. Evol. *4*, 336–344.

Hanski, I., Alho, J., and Moilanen, A. (2000). Estimating the parameters of survival and migration of individuals in metapopulations. Ecology *81*, 239–251.

Hoeck, P.E.A., Bollmer, J.L., Parker, P.G., and Keller, L.F. (2010). Differentiation with drift: a spatio-temporal genetic analysis of Galápagos mockingbird populations (Mimus spp.). Philos. Trans. Biol. Sci. *365*, 1127–1138.

Illerbrun, K., and Roland, J. (2011). Treeline proximity alters an alpine plant–herbivore interaction. Oecologia *166*, 151–159.

Jangjoo, M., Matter, S.F., Roland, J., and Keyghobadi, N. (2016). Connectivity rescues genetic diversity after a demographic bottleneck in a butterfly population network. Proc. Natl. Acad. Sci. *113*, 10914–10919.

Jr, L.C., Stanton, J.C., Lima, F., Uezu, A., Perilli, M.L., and Akçakaya, H.R. (2016). Implications of Fine-Grained Habitat Fragmentation and Road Mortality for Jaguar Conservation in the Atlantic Forest, Brazil. PLoS One San Franc. *11*.

Keller, D., Holderegger, R., and Strien, M.J. (2013). Spatial scale affects landscape genetic analysis of a wetland grasshopper. Mol. Ecol. *22*, 2467–2482.

Keyghobadi, N. (2007). The genetic implications of habitat fragmentation for animals. Can. J. Zool. *85*, 1049–1064.

Keyghobadi, N., Roland, J., and Strobeck, C. (1999). Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilionidae). Mol. Ecol. *8*, 1481–1495.

Keyghobadi, N., Roland, J., and Strobeck, C. (2002). Isolation of novel microsatellite loci in the Rocky Mountain apollo butterfly, *Parnassius smintheus*. Hereditas *136*, 247–250.

Keyghobadi, N., Roland, J., Matter, S.F., and Strobeck, C. (2005a). Among- and within-patch components of genetic diversity respond at different rates to habitat fragmentation: an empirical demonstration. Proc. R. Soc. Lond. B Biol. Sci. *272*, 553–560.

Keyghobadi, N., Roland, J., and Strobeck, C. (2005b). Genetic differentiation and gene flow among populations of the alpine butterfly, *Parnassius smintheus*, vary with landscape connectivity. Mol. Ecol. *14*, 1897–1909.

Koen, E.L., Bowman, J., and Walpole, A.A. (2012). The effect of cost surface parameterization on landscape resistance estimates. Mol. Ecol. Resour. *12*, 686–696.

Koscinski, D., Crawford, L.A., Keller, H.A., and Keyghobadi, N. (2011). Effects of different methods of non-lethal tissue sampling on butterflies. Ecol. Entomol. *36*, 301–308.

Kutschera, V.E., Frosch, C., Janke, A., Skírnisson, K., Bidon, T., Lecomte, N., Fain, S.R., Eiken, H.G., Hagen, S.B., Arnason, U., et al. (2016). High genetic variability of vagrant polar bears illustrates importance of population connectivity in fragmented sea ice habitats. Anim. Conserv. *19*, 337–349.

Levin, S.A., Carpenter, S.R., Godfray, H.C.J., Kinzig, A.P., Loreau, M., Losos, J.B., Walker, B., and Wilcove, D.S. (2009). The Princeton Guide to Ecology (Princeton University Press).

Manel, S., and Holderegger, R. (2013). Ten years of landscape genetics. Trends Ecol. Evol. *28*, 614–621.

Mantel, N. (1963). Chi-Square tests with one degree of freedom; extensions of the Mantel- Haenszel procedure. J. Am. Stat. Assoc. *58*, 690–700.

Matter, S.F., Roland, J., Moilanen, A., and Hanski, I. (2004). Migration and survival of *Parnassius smintheus*: detecting effects of habitat for individual butterflies. Ecol. Appl. *14*, 1526–1534.

McRae, B.H., and Beier, P. (2007). Circuit theory predicts gene flow in plant and animal populations. Proc. Natl. Acad. Sci. U. S. A. *104*, 19885–19890.

McRae, B.H., and Nürnberger, B. (2006). Isolation by resistance. Evolution *60*, 1551–1561.

McRae, B.H., Dickson, B.G., Keitt, T.H., and Shah, V.B. (2008). Using circuit theory to model connectivity in ecology, evolution, and conservation. Ecology *89*, 2712–2724.

Murphy, M.A., Evans, J.S., and Storfer, A. (2010). Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics. Ecology *91*, 252–261.

Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., and Meester, L. (2013). Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Mol. Ecol. *22*, 5983–5999.

Ouborg, N.J., Vergeer, P., and Mix, C. (2006). The rough edges of the conservation genetics paradigm for plants. J. Ecol. *94*, 1233–1248.

Pope, N. (2014). corMLPE: A correlation structure for symmetric relational data.

R Core Team (2017). R: A language and environment for statistical computing (Vienna, Austria: R Foundation for Statistical Computing).

Ries, L., and Debinski, D.M. (2001). Butterfly responses to habitat edges in the highly fragmented prairies of central Iowa. J. Anim. Ecol. *70*, 840–852.

Roland, J., and Matter, S.F. (2016). Pivotal effect of early-winter temperatures and snowfall on population growth of alpine *Parnassius smintheus* butterflies. Ecol. Monogr. *86*, 412–428.

Roland, J., Keyghobadi, N., and Fownes, S. (2000). Alpine Parnassius butterfly dispersal: effects of landscape and population size. Ecology *81*, 1642–1653.

Roslin, T., Syrjälä, H., Roland, J., Harrison, P.J., Fownes, S., and Matter, S.F. (2008). Caterpillars on the run – induced defences create spatial patterns in host plant damage. Ecography *31*, 335–347.

Ross, J.A., Matter, S.F., and Roland, J. (2005). Edge avoidance and movement of the butterfly *Parnassius smintheus* in matrix and non-matrix habitat. Landsc. Ecol. *20*, 127–135.

Smouse, P.E., Long, J.C., and Sokal, R.R. (1986). Multiple regression and correlation extensions of the mantel test of matrix correspondence. Syst. Biol. *35*, 627–632.

Spear, S.F., Peterson, C.R., Matocq, M.D., and Storfer, A. (2005). Landscape genetics of the blotched tiger salamander (*Ambystoma tigrinum melanostictum*). Mol. Ecol. *14*, 2553–2564.

Spear, S.F., Balkenhol, N., Fortin, M.-J., McRae, B.H., and Scribner, K. (2010). Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. Mol. Ecol. *19*, 3576–3591.

Theobald, D.M., Crooks, K.R., and Norman, J.B. (2011). Assessing effects of land use on landscape connectivity: loss and fragmentation of western U.S. forests. Ecol. Appl. *21*, 2445–2458.

Van Strien, M.J., Keller, D., and Holderegger, R. (2012). A new analytical approach to landscape genetic modelling: least-cost transect analysis and linear mixed models. Mol. Ecol. *21*, 4010–4023.

Wang, I.J. (2011). Choosing appropriate genetic markers and analytical methods for testing landscape genetic hypotheses. Mol. Ecol. *20*, 2480–2482.

Woese, C.R., and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proc. Natl. Acad. Sci. *74*, 5088–5090.

Yang, R.-C., and Williams, R. (2004). A likelihood-based approach to estimating and testing for isolation by distance. Evolution *58*, 1839–1845.

Zuberogoitia, I., Zalewska, H., Zabala, J., and Zalewski, A. (2013). The impact of river fragmentation on the population persistence of native and alien mink: an ecological trap for the endangered European mink. Biodivers. Conserv. *22*, 169–186.

Appendices

Appendix A The seven microsatellite loci used for genotyping *Parnassius smintheus*. Expected heterozygosity and observed heterozygosity were averaged over all 15 sites. This table is reproduced based on information from Table 1 in Keyghobadi et al. (1999) and Table 1 from Keyghobadi et al. (2002).

Locus Name	No. of alleles	Size range (bp)	Mean H_e	Mean H_o	Primers
Ps50	8	98 - 126	0.762	0.562	**FAM** GATCACCGAGAAAGAGAAAG TTTTTGCGTCTGTTACATAA
Ps81	7	122 - 133	0.582	0.206	**TET** AAATGGAGCAATTATACCTA GTTGCCCCGTTGAGTGAAAG
Ps85	8	118 -135	0.553	0.318	**HEX** CACGCTCTGGCACTATCTACC TGCGCAGATAGGGCTGAC
Ps76	29	260 - 364	0.86	0.399	**FAM** GGCAAATACCCTCCCTA GTAACGCTCAGTAAATCTGC
Ps163	9	283 - 136	0.221	0.234	**TET** CATTACCGAAACACGCACTT GTTTGCCAGGTCACGTTTAGGA
Ps165	36	180 - 250+	0.878	0.472	**HEX** CATGCGTAAATGTTGTAA CTAAACTAGGCGACGAAC
Ps262	33	71 - 175	0.911	0.83	**TET** TTTGGTGTGTGCAAATGAAA TGCGACTGGATGGGATT

Appendix B Genetic diversity of the 15 sites at Jumpingpound Ridge. A total of 474 *Parnassius smintheus* individuals were sampled. 'Mean no. alleles' refers to the number of observed alleles, including a null allele, and is averaged over seven microsatellite loci. Avg. H_e is the unbiased estimate of expected heterozygosity averaged over loci. This table is reproduced based on information from Table 1 in Keyghobadi et al. (2005).

Site	Sample size	Mean no. alleles	Avg. H_e
F	41	10.9	0.743
G1	40	11.3	0.698
g2	40	10.6	0.754
I	21	8.7	0.766
J	31	10.1	0.732
K	40	10.9	0.764
L	40	10	0.715
M	38	10.9	0.724
O	12	6.7	0.674
P	39	11	0.751
Q	40	10.6	0.727
R	24	9.3	0.734
S	14	6.9	0.643
Y	13	7.4	0.712
Z	41	10.1	0.722

Appendix C The location of center points of butterfly capture in 1995 and 1996 in each of the 15 sites at Jumpingpound Ridge in UTM coordinates. These center points were used as the location of the site's focal node (where movement from each site begins and ends) in Circuitscape. The study site is in UTM zone 11.

Site	Eastings	Northings
F	645710	5649352
G1	645678	5648879
g2	645770	5648926
I	645797	5647823
J	646060	5647425
K	646204	5647643
L	646554	5646678
M	646063	5646382
O	647666	5646994
P	648105	5647016
Q	648269	5647334
R	648808	5647312
S	649131	5647648
Y	646209	5649652
Z	646693	5649860

Appendix D Least cost transect analysis: comparison of maximum likelihood population effects (MLPE) models explaining genetic differentiation (F_{ST}) between populations of the butterfly, *Parnassius smintheus*. All models are shown. 'Resistance of forest' is the resistance value assigned to forest in the surface that was used to generate the least cost paths. 'Transect width' is the buffer applied to the least cost path to create the least cost transect. 'Model predictors' refer to the combination of predictors (proportion of meadow, barren rock, or forest land cover in the least cost transect) that were included in the model. The corrected Akaike information criterion (AICc) are presented for each model.

		Model predictors						
Resistance of forest	**Transect width (m)**	**Meadow**	**Forest**	**Barren rock**	**Meadow & forest**	**Meadow & Barren rock**	**Forest & Barren Rock**	**Meadow, forest, & Barren rock**
1	4	-763.99	-764.23	-758.33	-762.04	-762.29	-762.03	-760.70
2	4	-764.76	-760.09	-762.77	-762.92	-762.95	-763.01	-761.55
3	4	-761.76	-758.16	-764.46	-762.27	-762.29	-762.29	-760.06
4	4	-759.22	-757.97	-760.79	-758.42	-758.61	-758.59	-757.70
10	4	-759.12	-757.99	-760.78	-758.35	-758.67	-758.62	-766.92
50	4	-764.31	-757.97	-767.77	-765.75	-765.69	-765.68	-763.59
99	4	-764.86	-758.83	-773.31	-772.07	-771.67	-771.75	-771.20
1	20	-764.03	-764.44	-758.60	-762.24	-762.53	-762.28	-760.94
2	20	-766.27	-761.47	-763.20	-764.41	-764.47	-764.57	-763.90
3	20	-762.09	-758.47	-764.77	-762.57	-762.69	-762.71	-761.15
4	20	-759.63	-758.02	-761.19	-758.77	-758.99	-759.00	-758.37
10	20	-759.30	-757.96	-761.12	-758.57	-758.93	-758.91	-766.62
50	20	-764.21	-758.06	-767.86	-765.88	-765.91	-765.91	-763.67
99	20	-764.87	-758.00	-773.25	-771.47	-771.16	-771.19	-769.86
1	100	-765.86	-766.06	-758.72	-763.89	-764.21	-763.86	-764.41
2	100	-768.60	-765.56	-762.22	-766.60	-766.61	-766.61	-764.36
3	100	-764.55	-761.45	-763.94	-763.72	-763.80	-763.83	-761.77
4	100	-760.80	-759.10	-762.13	-759.70	-760.22	-760.34	-761.82
10	100	-760.09	-758.55	-762.32	-759.47	-760.17	-760.23	-766.63
50	100	-764.56	-759.90	-767.60	-766.30	-766.30	-766.30	-764.05
99	100	-765.11	-759.40	-771.94	-770.24	-769.91	-769.88	-769.02
1	200	-765.95	-765.83	-758.62	-763.78	-764.02	-763.63	-764.91

2	200	-765.97	-763.46	-761.41	-764.53	-764.41	-764.28	-762.77
3	200	-762.97	-760.69	-763.39	-762.66	-762.57	-762.53	-760.46
4	200	-759.39	-758.27	-763.00	-759.61	-760.82	-760.80	-762.85
10	200	-758.98	-758.07	-763.19	-759.70	-761.17	-761.06	-767.12
50	200	-762.21	-758.90	-767.59	-765.55	-765.57	-765.57	-763.32
99	200	-762.57	-758.55	-772.06	-770.26	-769.87	-769.88	-768.92
1	400	-765.35	-764.39	-757.96	-763.65	-763.20	-762.69	-764.51
2	400	-764.22	-761.98	-762.07	-764.11	-763.58	-763.35	-764.34
3	400	-761.70	-759.81	-764.78	-763.55	-762.94	-762.88	-762.33
4	400	-758.44	-757.96	-764.80	-761.30	-763.70	-763.45	-764.21
10	400	-758.25	-757.97	-765.02	-761.68	-764.82	-764.38	-768.73
50	400	-760.18	-758.10	-768.97	-766.92	-766.90	-766.90	-764.68
99	400	-760.28	-757.97	-773.75	-772.85	-772.69	-772.72	-770.60
1	800	-767.45	-765.72	-757.97	-767.42	-765.74	-764.78	-768.80
2	800	-762.77	-760.93	-761.82	-764.08	-763.66	-763.51	-761.97
3	800	-760.18	-759.04	-763.01	-761.80	-761.32	-761.26	-759.67
4	800	-758.01	-758.01	-763.65	-760.11	-762.23	-762.09	-761.11
10	800	-757.96	-758.12	-764.05	-760.46	-763.53	-763.24	-763.58
50	800	-758.73	-757.97	-767.14	-765.64	-765.03	-765.06	-763.55
99	800	-758.59	-758.18	-770.63	-768.58	-769.02	-769.00	-766.84
1	1200	-764.38	-762.71	-757.97	-764.16	-762.30	-761.09	-769.68
2	1200	-759.50	-758.66	-759.04	-759.08	-758.35	-758.12	-757.79
3	1200	-758.19	-757.97	-759.21	-757.81	-757.09	-757.04	-756.38
4	1200	-758.10	-758.38	-759.08	-756.97	-757.43	-757.37	-755.37
10	1200	-758.23	-758.63	-759.29	-757.31	-758.15	-758.00	-756.45
50	1200	-758.02	-758.37	-761.90	-760.77	-759.73	-759.79	-760.52
99	1200	-758.01	-758.83	-764.82	-762.80	-762.83	-762.83	-760.58

Appendix E Least cost transect analysis: comparison of maximum likelihood population effects (MLPE) models explaining estimated dispersal between populations of the butterfly, *Parnassius smintheus*. All models are shown. ‘Resistance of forest’ is the resistance value assigned to forest in the surface that was used to generate the least cost paths. ‘Transect width’ is the buffer applied to the least cost path to create the least cost transect. ‘Model predictors’ refer to the combination of predictors (proportion of meadow, barren rock, or forest land cover in the least cost transect) that were included in the model. The corrected Akaike information criterion (AICc) are presented for each model.

		Model predictors						
Resistance of forest	**Transect width (m)**	**Meadow**	**Forest**	**Barren rock**	**Meadow & forest**	**Meadow & Barren rock**	**Forest & Barren Rock**	**Meadow, forest, & Barren rock**
1	4	464.10	464.85	529.68	465.09	462.67	466.79	453.74
2	4	517.29	515.92	530.09	515.93	515.52	516.16	505.53
3	4	527.79	529.21	529.76	529.99	529.96	530.15	522.38
4	4	531.18	523.40	528.90	522.80	524.43	523.29	512.76
10	4	530.89	515.46	527.82	514.15	514.31	514.18	516.34
50	4	529.64	517.92	530.14	520.11	519.46	520.09	517.97
99	4	528.94	524.51	531.29	526.41	525.66	526.50	516.89
1	20	461.98	463.63	529.47	463.29	461.09	465.35	450.10
2	20	512.22	511.02	529.59	510.73	510.26	511.02	499.95
3	20	526.72	527.84	529.89	528.86	528.78	529.04	522.98
4	20	531.29	528.69	529.31	528.03	529.08	528.53	520.57
10	20	531.14	525.05	528.41	523.57	523.82	523.66	525.36
50	20	530.32	525.28	530.47	527.15	526.82	527.09	527.63
99	20	529.67	528.64	531.26	530.73	530.37	530.79	522.36
1	100	452.69	455.15	528.94	454.31	452.96	456.23	440.70
2	100	502.08	497.58	530.38	499.41	498.29	499.62	474.98
3	100	519.65	518.51	530.43	520.51	520.11	520.62	514.49
4	100	530.75	531.26	530.07	531.66	532.27	532.25	526.95
10	100	531.12	531.13	529.05	530.27	530.64	530.55	531.32
50	100	531.28	530.81	530.33	531.82	531.70	531.74	533.38
99	100	531.21	531.19	531.29	533.40	533.35	533.39	522.58
1	200	448.21	449.45	526.75	449.62	447.76	451.29	435.61

2	200	497.22	494.96	530.96	496.53	494.87	497.02	473.32
3	200	516.57	516.34	529.97	518.13	517.60	518.40	512.22
4	200	531.22	531.07	529.56	529.66	531.15	530.84	526.20
10	200	531.28	530.34	528.70	527.80	528.57	528.34	529.05
50	200	531.04	529.84	530.03	530.15	529.93	530.00	531.16
99	200	530.84	530.55	531.26	532.69	532.48	532.61	521.26
1	400	447.05	445.45	525.15	447.07	443.70	447.61	436.33
2	400	498.55	497.61	530.88	499.34	498.27	499.73	489.58
3	400	520.02	520.38	529.29	522.17	521.90	522.46	518.46
4	400	530.23	527.01	529.70	520.51	524.73	523.24	518.64
10	400	528.52	521.79	528.77	511.31	512.87	512.00	513.53
50	400	526.21	519.30	530.03	518.00	516.31	517.20	509.80
99	400	525.72	522.40	531.28	524.14	522.66	523.71	506.76
1	800	456.29	447.84	512.90	449.48	447.24	448.40	449.29
2	800	500.90	497.04	530.31	498.93	499.64	499.08	500.75
3	800	525.62	524.25	531.20	525.52	526.10	525.83	527.32
4	800	524.76	522.43	531.29	524.12	524.95	524.35	526.12
10	800	513.94	505.01	531.09	505.02	504.54	504.49	506.64
50	800	506.77	500.17	531.21	502.37	499.53	502.24	492.50
99	800	508.58	509.49	530.60	509.39	508.00	510.50	496.90
1	1200	474.43	452.86	491.06	448.67	440.91	443.36	441.64
2	1200	519.03	511.22	513.68	504.74	503.11	503.26	505.36
3	1200	530.79	529.09	518.70	520.49	519.99	519.93	522.12
4	1200	516.89	525.95	518.29	513.69	509.68	508.23	509.04
10	1200	497.23	510.14	521.13	498.83	497.17	494.77	496.22
50	1200	488.65	504.94	528.72	490.83	490.52	494.17	488.48
99	1200	489.55	515.30	519.75	490.49	490.56	490.88	492.63

Curriculum Vitae

Helen Chen

Education

M.Sc. Biology, Sept. 2015 – Present
University of Western Ontario, London, ON

BES. Honours Environment and Resource Studies, Co-op, Apr. 2014
University of Waterloo, Waterloo, ON

Work Experience

Teaching Assistant, *University of Western Ontario*, London, ON, 2015 – 2017

Field Research Assistant, *University of Waterloo,* Waterloo, ON, 2013 – 2014

Awards

Biology Travel Award, University of Western Ontario, 2017

Conference Presentations

Chen, H., Matter, S.F., Roland, J., and Keyghobadi, N. (2017) The differing effects of landscape on the genetic differentiation and dispersal of the Rocky Mountain apollo butterfly *Parnassius smintheus*. Entomological Society of Ontario 154th Annual Conference. Guelph, ON, Canada. [Poster]

Chen, H., Matter, S.F., Roland, J., and Keyghobadi, N. (2017) A resistance mapping approach to assess landscape effects on genetics and dispersal of the Rocky Mountain Apollo butterfly. Canadian Society for Ecology and Evolution. Victoria, BC, Canada. [Oral]

Chen, H., Matter, S.F., Roland, J., and Keyghobadi, N. (2016) Assessing landscape effects on genetics and dispersal of the Rocky Mountain Apollo Butterfly *Parnassius smintheus* using a resistance mapping approach. GIS Day 2016. London, ON, Canada. [Poster]

Chen, H., Matter, S.F., Roland, J., and Keyghobadi, N. (2016) Assessing landscape effects on genetics and dispersal of the Rocky Mountain Apollo Butterfly *Parnassius* *smintheus* using a resistance mapping approach. Biology Graduate Research Forum. London, ON, Canada. [Poster]

Outreach and Service

Animal Foster, Animal Love Foundation, Animal Outreach, 2015 - 2017

Butterfly Outreach, Meadowlily Nature Preserve, 2016 – 2017

Conference Session Moderator, Canadian Society for Ecology & Evolution, May 9/2017

Greenhouse Seed Collection and Planting, Thames Talbot Land Trust, 2016-2017

Site Restoration and Transplanting, Thames Talbot Land Trust, June 14/2016